Horizontal transfer of a pathway for coumarate catabolism unexpectedly inhibits purine nucleotide biosynthesis

Dan M. Close[1,#,†], Connor J. Cooper[2,#], Xingyou Wang[3], Payal Chirania[2,4], Madhulika Gupta[1], John R. Ossyra[2], Richard J. Giannone[4,5,6], Nancy Engle[1,5,6], Timothy J. Tschaplinski[1,5,6], Jeremy C. Smith[1,7], Lizbeth Hedstrom[8], Jerry M. Parks[1,2], and Joshua K. Michener[1,5,6,*]

(1) Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
(2) Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN, 37996, USA
(3) Graduate Program in Chemistry, Brandeis University, 415 South Street, Waltham MA 02454, USA
(4) Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37830, USA
(5) BioEnergy Science Center
(6) Center for Bioenergy Innovation
(7) Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee Knoxville, Knoxville, Tennessee 37996, USA
(8) Departments of Biology and Chemistry, Brandeis University, 415 South Street, Waltham MA 02454, USA
(#) These authors contributed equally to this work
(†) Current address: 490 BioTech, Inc. 2450 E.J. Chapman Drive, Knoxville, Tennessee 37996, USA

(*) Correspondence should be sent to JKM (865-576-7957, michenerjk@ornl.gov)

Keywords: biosystem design, pathway optimization, molecular dynamics, IMPDH, lignin catabolism

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States

Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

Summary:

A microbe's ecological niche and biotechnological utility are determined by its specific set of co-evolved metabolic pathways. The acquisition of new pathways, through horizontal gene transfer or genetic engineering, can have unpredictable consequences. Here we show that two different pathways for coumarate catabolism failed to function when initially transferred into *Escherichia coli*. Using laboratory evolution, we elucidated the factors limiting activity of the newly acquired pathways and the modifications required to overcome these limitations. Both pathways required host mutations to enable effective growth with coumarate, but the necessary mutations differed. In one case, a pathway intermediate inhibited purine nucleotide biosynthesis, and this inhibition was relieved by single amino acid replacements in IMP dehydrogenase. A strain that natively contains this coumarate catabolism pathway, *Acinetobacter baumannii*, is resistant to inhibition by the relevant intermediate, suggesting that natural pathway transfers have faced and overcome similar challenges. Molecular dynamics simulation of the wild type and a representative single-residue mutant provide insight into the structural and dynamic changes that relieve inhibition. These results demonstrate how deleterious interactions can limit pathway transfer, that these interactions can be traced to specific molecular interactions between host and pathway, and how evolution or engineering can alleviate these limitations.

55 **Introduction:**

56 Microbes can use a wide variety of compounds as carbon and energy sources. Expanding the
57 breadth of compounds that a strain can catabolize can allow access to new environmental niches
58 or enable engineered microbes to use new feedstocks. Correspondingly, catabolic pathways are
59 frequently transferred between strains, either in nature through horizontal gene transfer (HGT) or
60 in the laboratory through metabolic engineering (Pál *et al.*, 2005; Nielsen and Keasling, 2016).
61 However, newly-acquired pathways often fail to function effectively in their new host (Porse *et*
62 *al.*, 2018). In these cases, productive use of a new pathway may require post-transfer refinement
63 to optimize expression and minimize deleterious interactions (Michener *et al.*, 2014; Clark *et al.*,
64 2015). The pathway activity immediately following transfer may be very different from the
65 potential activity after optimization, complicating predictions about engineering or HGT.

66 We have explored this issue using pathways for catabolism of lignin-derived aromatic
67 compounds, since these pathways are widespread in nature (Bugg *et al.*, 2011), are often
68 transferred by HGT (Chain *et al.*, 2006), have biotechnological applications (Ragauskas *et al.*,
69 2014), and involve challenging biochemistry (Masai *et al.*, 2007). We previously constructed
70 strains of *E. coli* that grow with the model lignin-derived compounds protocatechuate (PCA) and
71 4-hydroxybenzoate (4-HB) as sole sources of carbon and energy using the 3,4-cleavage pathway
72 for protocatechuate catabolism from *Pseudomonas putida* and a 4-hydroxybenzoate 3-
73 monooxygenase from *P. putida* or *Paenibacillus* sp. JJ-1b (Clarkson *et al.*, 2017; Standaert *et al.*,
74 2018). Introduction of the relevant catabolic pathways was not sufficient to enable rapid growth
75 with either carbon source. We then used experimental evolution to select for strains with
76 improved growth. By resequencing the evolved variants and reconstructing mutations in the
77 parental strains, we identified causal mutations that improved function of the heterologous
78 pathway.

79 In this work, we extended those pathways to allow growth with a model phenylpropanoid,
80 coumarate. There are two known oxidative routes for coumarate catabolism, differing in their
81 specific reaction chemistry and resulting intermediates (Figure 1). These pathways are
82 exemplified by the *hca* pathway from *Acinetobacter baylyi* ADP1 (Parke and Ornston, 2003) and
83 the *cou* pathway from *Rhodococcus jostii* (Otani *et al.*, 2014). Both pathways begin by
84 conjugating the phenylpropanoid substrate to coenzyme A. The *hca* pathway then uses a retro-
85 aldol reaction to produce an intermediate benzaldehyde derivative, while the *cou* pathway uses a

86 hydrolytic retro-Claisen reaction to produce the benzoate derivative directly. Since these two
87 phenylpropanoid pathways use different biochemistry and intermediates, their interactions with
88 the host may also differ substantially (Kim and Copley, 2012). Identifying the likeliest pairing of
89 host and pathway, either for engineering or HGT, will depend on understanding the specific
90 challenges imposed by each potential pathway and the mechanisms to overcome these challenges
91 available to the host.

92 Using a combination of engineering and evolution, we constructed and optimized both
93 representative pathways for phenylpropanoid catabolism in *E. coli*. We show that pathway
94 activity is initially limited due to pathway-specific molecular interactions that can readily be
95 alleviated through point mutations to the host. Similar compensatory mechanisms are present in a
96 strain that natively contains the appropriate pathway. Molecular dynamics simulations of the
97 wild-type and mutant enzymes demonstrate how subtle modifications to the enzyme distant from
98 the active site can relieve inhibition while preserving catalysis. Identifying and alleviating the
99 specific molecular interactions between an engineered metabolic pathway and its heterologous
100 host will aid our efforts to rapidly engineer metabolic capabilities.

101

102 **Results**

103 Combining engineering and evolution enabled coumarate catabolism

104 We designed and synthesized two constructs containing genes for phenylpropanoid import
105 and degradation, each of which converts coumarate into 4-hydroxybenzoate (Figure 1 and Figure
106 S1). Each pathway was introduced into *E. coli* strains, JME38 and JME50, that had previously
107 been engineered to grow with 4-HB using *pobA* and *praI*, respectively (Standaert *et al.*, 2018).
108 None of the engineered strains acquired the immediate ability to grow with coumarate as the sole
109 source of carbon and energy (Figure S2).

110 To understand the factors preventing pathway function, we used experimental evolution to
111 select for strains with the ability to catabolize coumarate. Three replicate cultures of each
112 engineered strain were propagated in minimal medium containing 1 g/L coumarate (~6.1 mM).
113 After 300 generations, individual mutants were isolated from each population and characterized
114 for growth with protocatechuate (PCA), 4-HB, coumarate, and caffeate. Representative isolates
115 were chosen for each replicate population for further characterization. All isolates could grow

with PCA and coumarate, though growth with caffeate and 4-HB varied between replicates (Figure S2).

Genome resequencing and reconstruction identified causal mutations

The genomes of the selected isolates were resequenced to identify new mutations (Table S1 and Table S4). Several of the mutations have previously been described for their effects on catabolism of 4-HB, such as synonymous mutations to the gene encoding the 4-hydroxybenzoate monooxygenase *pobA* (Standaert *et al.*, 2018). Among the strains with the *hca* pathway, five of the six isolates had additional mutations to the native gene *guaB*, encoding inosine monophosphate (IMP) dehydrogenase (IMPDH), and to the intergenic region between *hcaB* and *hcaC* in the engineered pathway. The exception was JME96, which had a mutation to *rpoS*, encoding the RNA polymerase sigma factor σ^{38}, that is expected to be highly pleiotropic (Saxer *et al.*, 2014).

In the strains with the *cou* pathway, the acquired mutations were less consistent across replicates, with several mutations to genes that are expected to be pleiotropic. However, parallel mutations were observed in JME106 and JME109, with mutations to both *couL* and *nadR*. The mutations to *couL*, which encodes the CoA ligase, were coding mutations, L192R and S134Y. NadR is involved in both regulation and catalysis for NAD salvage (Kurnasov *et al.*, 2002). One of the mutations to *nadR* led to a frameshift that precisely removed the C-terminal ribosylnicotinamide kinase (RNK) domain, which converts N-ribosylnicotinamide into β-nicotinamide mononucleotide during NAD salvage (Kurnasov *et al.*, 2002). Similarly, the second *nadR* mutation also occurred in the RNK domain. The physiological consequences of these mutations are unclear.

To test the causality of the identified mutations, we reconstructed representative mutations in the engineered parental strains. We assumed that parallel mutations to a given gene produced similar effects, and therefore only tested one representative mutation (*e.g.* D243G in *guaB*). Two mutations, to *pobA* and *hcaABCK,* were necessary for growth with coumarate in JME64, while a third mutation to *guaB* significantly increased growth (Figure 2A). Similarly, mutations to *pobA*, *couLHTMNO*, and *nadR* were all required for growth with coumarate using the *cou* pathway in JME65 (Figure 2B).

146 Synonymous mutations to *pobA* mutation have previously been shown to increase expression
147 of PobA by destabilizing secondary structures in the mRNA (Standaert *et al.*, 2018). To
148 understand the effect of the intergenic mutations upstream of *hcaC*, we measured protein
149 expression levels in the engineered strains. As expected, the mutation to *pobA* increased
150 expression of PobA by approximately 9-fold, while the *hcaC* mutation increased expression of
151 both HcaB and HcaC by roughly 2-fold (Figure S3). The intergenic mutation before *hcaC* is
152 predicted to increase the translation rates by approximately 10-fold through modulation of the
153 strength of the ribosome binding site (Espah Borujeni *et al.*, 2014).
154 Parallelism of mutations within replicates of a pathway, but divergence between pathways,
155 strongly suggests that the mutations are specific to a particular pathway. To test this hypothesis,
156 we replaced the *hca* pathway in JME131 with either the wild-type or evolved *cou* pathways.
157 Neither strain was able to grow with coumarate as the sole source of carbon and energy.
158
159 Inhibitory cross-talk between engineered and native pathways limits function
160 A mutation to *guaB* was necessary for growth with coumarate using the *hca* pathway.
161 IMPDH, encoded by *guaB*, converts inosine monophosphate (IMP) to xanthosine
162 monophosphate (XMP) with the reduction of NAD^+ during guanine nucleotide biosynthesis
163 (Hedstrom, 2009). Five independent amino acid replacements in IMPDH were identified: A48V,
164 D243G, G330D, L364Q, and P482L. IMPDH uses different conformations to catalyze each step
165 of the catalytic cycle: an open conformation for hydride transfer that produces a covalent
166 intermediate with the catalytic C305 (E-XMP*) and a closed conformation for hydrolysis of the
167 E-XMP* (Figure S4) . We generated homology models of wild-type *E. coli* IMPDH in both the
168 closed and open conformations. The mutations are distant from each other and from the active
169 site, with no obvious effect on catalysis (Figure 3).
170 To understand the consequences of these mutations, we measured metabolite levels in the
171 parent and engineered strains during growth with coumarate. Consistent with our genetic
172 analysis above, we chose to focus on the D243G mutation. Compared to the D243G *guaB* mutant
173 (JME131), the strain with wild-type *guaB* (JME129) showed higher levels of inosine nucleotides
174 (Figure 4A). We hypothesized that growth with coumarate led to inhibition of IMPDH and
175 accumulation of IMP, and that this inhibition was relieved in the *guaB* mutants. To determine
176 whether inhibition of nucleotide biosynthesis limited growth with coumarate, we supplemented

177 the growth medium with guanosine. Addition of guanosine increased growth with coumarate in a
178 strain with the wild-type IMPDH, but not the mutant (Figure 4B).
179 Mutations to *guaB* improved growth with the *hca* pathway but not with the *cou* pathway. The
180 *hca* pathway produces an intermediate, 4-hydroxybenzaldehyde, that is not present in the *cou*
181 pathway (Figure 1). To test whether this intermediate was responsible for the inhibition of
182 IMPDH, we grew strains containing WT and mutant IMPDH in varying concentra tions of 4-
183 hydroxybenzaldehyde (Figure S5). Both strains were inhibited by high concentrations of 4-
184 hydroxybenzaldehyde, but the mutation to *guaB* decreased inhibition. In strain JME129, addition
185 of guanosine provided no advantage during growth with glucose but relieved inhibition by 4-
186 hydroxybenzaldehyde (Figure S6).
187 Next, we purified WT and mutant IMPDH and measured inhibition *in vitro* with 4-
188 hydroxybenzaldehyde. This compound is a weak inhibitor of WT *Ec*IMPDH, with a $K_{i,app}$ of 320
189 ± 20 μM (Figure 4C). Introduction of the D243G replacement had little effect on catalytic
190 activity (Table 1) but increased the $K_{i,app}$ to 1250 ± 50 μM, indicating a substantial reduction of
191 inhibition in the mutant. The *hca* pathway that we used came from *A. baylyi* ADP1, and we
192 hypothesized that the native IMPDH of this strain would have faced similar selective pressures to
193 minimize inhibition by 4-hydroxybenzaldehyde. As a surrogate, we tested the IMPDH of *A.*
194 *baumannii*, since this strain contains a homologous *hca* pathway (83-94% amino acid identity)
195 and IMPDH (91% amino acid identity). As predicted, the *A. baumannii* IMPDH has a $K_{i,app}$ of
196 720 ± 30 μM, substantially higher than that of wild-type *Ec*IMPDH. Further kinetic
197 characterization of these IMPDH orthologs is summarized in Table 1.
198
199 <u>Subtle changes in enzyme dynamics relieve inhibition while maintaining catalysis</u>
200 To gain insight into possible mechanisms of inhibition by 4-hydroxybenzaldehyde, we
201 measured enzyme activity at varying inhibitor concentrations. 4-Hydroxybenzaldehyde is an
202 uncompetitive inhibitor with respect to IMP and a noncompetitive (mixed) inhibitor with respect
203 to NAD^+ for both wild type and D243G *Ec*IMPDH (Figure S7). Similar patterns of inhibition
204 have been observed for compounds that bind in the NAD^+ site (Makowska-Grzyska *et al.*, 2015).
205 In addition, we computationally docked 4-hydroxybenzaldehyde to a model of wild-type IMPDH
206 in the open conformation of the apoenzyme as well as to IMP-bound and IMP/NAD^+-bound
207 states. In both the apoenzyme and IMP-bound models, the majority of the top poses of 4-

208 hydroxybenzaldehyde were found to occupy the NAD^+ binding site, approximately 20 Å from
209 D243 (Figure S8 and Table S6). Therefore, to relieve inhibition by 4-hydroxybenzaldehyde, the
210 D243G substitution would need to perturb the structure or dynamics of the distant active site.
211 Understanding the mechanism of this perturbation required additional analysis. The
212 hydrolysis of E-XMP* is the slow step in the IMPDH reaction, so E-XMP* is the predominant
213 enzyme complex (Hedstrom, 2009). Therefore, a decrease in the affinity of 4-
214 hydroxybenzaldehyde for E-XMP* can account for resistance to inhibition. Hydrolysis of E-
215 XMP* requires a conformational change wherein a mobile protein flap folds into the cofactor
216 binding site. We assessed the effect of the D243G mutation on the active site by performing
217 molecular dynamics (MD) simulations of wild-type and D243G mutant IMPDH in the covalently
218 bound E-XMP* state using the closed conformation model, since the flap is disordered in crystal
219 structures of the open conformation. In the simulations of the wild-type *E. coli* IMPDH, D243
220 forms stable hydrogen bonds with the side chains of K87 and R219 and also with the backbone
221 of V220 (Figure 5A). In the absence of this hydrogen bonding network in the mutant (Figure
222 5B), G243 adopts two different conformations, one that resembles the wild type in which G243
223 is close to but not interacting with K87, R219, and V220, and another in which G243 is
224 positioned farther away from these residues when a new hydrogen bond is formed with Q272
225 (Figure 5C and Figure S9). However, it was not obvious how these local changes around the
226 mutation site propagate to the active site, which is located on the opposite side of the β-barrel.
227 To identify changes in protein dynamics resulting from the D243G mutation, we calculated
228 root-mean-squared fluctuations (RMSFs) for chain A in both the wild-type and mutant. In both
229 systems, high RMSFs were observed over the entire flap region (Figure S10). Upon inspection of
230 specific interactions of flap residues, we found that the mutation leads to changes in hydrogen-
231 bonding interactions with other residues on chain A or the adjacent chain D (Figure S11 and
232 Figure S12). These interactions resulted in reorientation of a loop on the flap that could alter
233 inhibitor binding. Details of specific hydrogen bonding interaction changes are described in the
234 Supporting Information. Despite these changes in flap conformations and dynamics, the catalytic
235 dyad remains in close proximity to the covalent intermediate, poised for catalysis (Figure S13).
236 RMSF analysis also revealed that helix α2 (residues 76-89) and helix α8 (residues 230-241)
237 fluctuate more in the mutant than in the wild type (Figure S10). Helix α8 is downstream of the
238 mutation site (Figure 6A). Therefore, the higher fluctuations of this helix and the adjacent helix

239 α2 in the mutant are likely due to the loss of the hydrogen bonding network formed by D243
240 with K87, R219, and V220 as well as the formation of a new hydrogen bond between G243 and
241 Q272 in the mutant. The N-terminal end of helix α2 and the C-terminal end of helix α8 are
242 located near the NAD^+ binding site, which is also the predicted binding site for 4-
243 hydroxybenzaldehyde based on docking to the open conformation model (Figure S8). Therefore,
244 these perturbations to helices α2 and α8 represent another mechanism for the D243G mutation to
245 affect inhibitor binding.
246 To understand the mechanism by which helix dynamics could affect inhibitor binding, we
247 combined these MD results with our previous docking studies. In six of the top 10 docking
248 poses, the phenolic hydrogen of the inhibitor forms hydrogen-bonding interactions with either
249 the side chain or backbone of D248. The carbonyl oxygen of the inhibitor also interacts with the
250 side chain of S250 (Figure 6B). D248 and S250 are located on the β-sheet (β11) downstream of
251 the mutation site and are in close proximity to helices α2 and α8 as well as the loop on the flap
252 that showed different interactions in the wild-type and mutant simulations. Thus, changes in the
253 structure and dynamics of these regions around the NAD^+ binding site likely disrupt inhibitor
254 access and binding.
255
256 **Discussion**
257 In this work, we have recapitulated the process of HGT and demonstrated the necessity for
258 host adaptations to accommodate the *hca* pathway in both *E. coli* and *A. baumannii*. We
259 identified a novel interaction between the newly introduced pathway and the endogenous
260 metabolism, as well as the physiological and biochemical consequences of this interaction.
261 Finally, we demonstrated how single point mutations to an essential host protein alter its
262 conformational dynamics to prevent binding of the novel inhibitor while still preserving
263 catalysis.
264 Highly similar *hca* pathways are present in various beta- and gamma-proteobacteria. Further
265 HGT of this pathway would require either a host with an IMPDH homolog that is resistant to
266 inhibition by 4-hydroxybenzaldehyde, or post-transfer selection for mutations that relieve
267 inhibition. Understanding these types of limitations on HGT, and the mechanisms by which
268 organisms evolve to avoid them, will aid in our ability to predict and manipulate horizontal gene
269 transfer (Michener *et al.*, 2014; Clark *et al.*, 2015).

270 In combination, our results suggest that introduction of the *hca* pathway allowed only limited
271 growth with coumarate because accumulation of 4-hydroxybenzaldehyde inhibited the native *E.*
272 *coli* IMPDH. This inhibitory cross-talk results in nucleotide starvation and impairs growth and
273 phenylpropanoid catabolism. Mutations to *guaB* prevent inhibition by 4-hydroxybenzaldehyde
274 and allow growth with coumarate. There is no *a priori* reason to expect that a pathway for
275 degradation of an aromatic compound would interact with a native pathway for nucleotide
276 biosynthesis. Phenolic amides such as feruloyl amide have been shown to inhibit a different step
277 in nucleotide biosynthesis (Pisithkul *et al.*, 2015), but neither the substrate nor products of
278 coumarate degradation are toxic at the relevant concentrations (Figure S5, and Clarkson et al.,
279 2017; Standaert et al., 2018). These types of inhibitory cross-talk are likely to be common with
280 heterologous engineered metabolic pathways, though they are rarely identified and alleviated
281 (Kizer *et al.*, 2008; Kim and Copley, 2012; Michener *et al.*, 2012).
282 In particular, inhibition of microbial growth by aldehydes is commonly observed, though the
283 mechanisms of toxicity can rarely be traced to a specific interaction (Mills *et al.*, 2009; Clarkson
284 *et al.*, 2014; Yi *et al.*, 2015). Enzymatic pathways have frequently evolved to limit the release of
285 free aldehydes, for example through enzymatic channeling (Huang *et al.*, 2001). It is unclear
286 whether channeling between HcaA and HcaB limits the release of free aldehydes in either the
287 native or heterologous hosts. Mutations that increase tolerance to free aldehydes generally do so
288 either by increasing export of the toxic compound or by performing redox chemistry to remove
289 the aldehyde functionality (Mukhopadhyay, 2015). In this work, we have shown an example of
290 aldehyde toxicity that acts through a single protein and can be relieved by point mutations to the
291 associated gene. For the D243G mutant, biochemical assays revealed mixed inhibition that was
292 relieved through mutation. Other examples of nonspecific toxicity may prove to be similarly
293 specific when characterized fully.
294 MD simulations provided insight into how a single amino acid substitution distant from the
295 active site could relieve inhibition while maintaining catalysis. The mutation is located at the N-
296 terminal end of β11, which is near the NAD^+ binding site where the inhibitor was predicted to
297 bind based on docking calculations. The simulations showed local changes in the hydrogen
298 bonding networks at the mutation site, which led to changes in the dynamics of the catalytic flap
299 and helices α2 and α8 near the inhibitor binding site. In addition, the catalytic dyad showed only
300 minor perturbations and remained poised for catalysis.

301 Across the replicate populations, many mutations were highly pleiotropic, including large
302 insertions and deletions flanked by insertion sequences as well as mutations to core
303 transcriptional machinery such as *rho* and *rpoB*. Duplications frequently spanned the insertion
304 sites for engineered operons, suggesting that expression of the heterologous genes was limiting.
305 By comparing across replicates, we were able to identify a set of point mutations that allowed
306 growth with coumarate as the sole source of carbon and energy. However, a reconstructed strain
307 containing these mutations does not grow as quickly with coumarate as the evolved isolates,
308 suggesting that some of the remaining mutations provided additional fitness benefits (Figure 2
309 and Figure S2).
310 The two 4-HB monooxygenases, *praI* and *pobA*, are 60% identical at the nucleotide level,
311 and the associated enzymes have 54% amino acid identity. We previously demonstrated that
312 these enzymes required different optimization solutions to enable growth with 4-HB (Standaert
313 *et al.*, 2018). In contrast, in this experiment, the evolutionary solutions were very similar. Even
314 in the optimized strain, JME131, the growth rate with coumarate was lower than the growth rate
315 with PCA. We hypothesize that the conversion of coumarate into 4-HB is the rate-limiting step,
316 and that the conversion of 4-HB into PCA by either enzyme was sufficient under these
317 circumstances.
318 Multiple mutations were identified in the heterologous *cou* and *hca* pathways. In the *hca*
319 pathway, these mutations served to increase expression of the pathway, either through pathway
320 duplication or by intergenic mutations that affected translation, specifically increasing expression
321 of the CoA ligase HcaC. The *cou* pathway mutations were coding mutations to a single gene, the
322 *couL* that encodes a CoA ligase, and decrease expression of that enzyme (SI Appendix, Figure
323 S3). These differential evolutionary responses could arise from different initial expression levels
324 of the two CoA ligases, for example due to the placement of *couL* at the beginning of an operon
325 and *hcaC* at the end. Further biochemical analysis will be required to precisely identify the
326 consequences of these mutations.
327 We have described the use of experimental evolution to identify and alleviate deleterious
328 interactions between engineered metabolic pathways for coumarate catabolism and native
329 pathways for nucleotide biosynthesis and cofactor salvage. Many engineered pathways place a
330 substantial burden on the production host, yet understanding and accommodating these
331 interactions remains challenging. Evolution can simplify this optimization process by directly

332 selecting for mutations that eliminate the inhibition. As we did with *guaB*, researchers can then
333 work backwards from the evolutionary solutions to understand the factors that were initially
334 limiting productivity and the biochemical solutions to overcome those problems. By solving
335 more problems of this sort, we will develop design rules for future forward engineering of
336 metabolic pathways and better predictions of the likelihood of pathway transfer by HGT.
337

338 **Experimental Procedures**

339 Strains and chemicals

340 Unless otherwise noted, all chemicals were purchased from Sigma-Aldrich (St. Louis, MO)
341 or Fisher Scientific (Fairlawn, NJ) and were molecular grade. All oligonucleotides were ordered
342 from IDT (Coralville, IA). *E. coli* strains were routinely cultivated at 37 °C in LB containing the
343 necessary antibiotics (50 mg/L kanamycin or 50 mg/L spectinomycin). Growth assays with
344 aromatic substrates were performed in M9 salts medium containing 300 mg/L thiamine and 1
345 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). PCA and 4-HB were dissolved in water at 5
346 g/L, filter sterilized, and added at a final concentration of 1 g/L. Coumarate and caffeate were
347 dissolved in DMSO at 100 g/L and added at a final concentration of 1 g/L. The addition of 1%
348 DMSO did not affect growth. The pH of the substrates was not controlled, as PCA oxidation
349 occurred more rapidly at neutral pH.
350

351 Plasmid construction

352 Plasmids pJM219 and pJM223, containing the *cou* and *hca* expression constructs, were
353 synthesized by the Joint Genome Institute. As described previously, the pathway design used
354 synthetic promoters, terminators, and custom ribosome binding sites (Salis *et al.*, 2009; Chen *et*
355 *al.*, 2013; Kosuri *et al.*, 2013; Espah Borujeni *et al.*, 2014). Plasmids expressing sgRNA for
356 chromosomal modifications were constructed as described previously, using an inverse PCR to
357 linearize the expression vector followed by assembly with synthesized oligonucleotides
358 (Clarkson *et al.*, 2017). Plasmid pJM303, expressing the D243G mutant of the *E. coli* IMPDH,
359 was constructed by amplifying the mutant *guaB* allele from JME89 and cloning it into pMCSG7
360 under the control of a T7 promoter. Plasmids and primers used in this work are described in
361 Tables S2 and S3.
362

363 Strain construction

364 Genome modifications were performed as described previously, using the lambda-red
365 recombineering system in combination with Cas9-mediated selection (Jiang *et al.*, 2015;
366 Clarkson *et al.*, 2017). Integration cassettes were amplified from synthesized plasmids or the
367 chromosomal DNA of mutant strains, as needed.

368

369 Experimental evolution

370 Parental strains were streaked to single colonies. Three colonies from each strain were grown
371 to saturation in LB + 1 mM IPTG, then diluted 128-fold into M9 + 1 mM IPTG + 1 g/L
372 coumarate + 50 mg/L PCA and grown at 37 ºC. When the cultures reached saturation, typically
373 after two days during the initial stages, they were diluted 128-fold into fresh medium. As the
374 growth became more robust, the PCA concentration was decreased. Cultures derived from
375 JME65 and JME67 required the addition of PCA for 150 generations, and in some cases took
376 two days to reach saturation even after 300 generations. One replicate culture, JME65-C, became
377 contaminated with a different coumarate-degrading strain. This contamination was not
378 discovered until resequencing, and consequently the culture was not restarted.

379 After 300 generations, the evolved cultures were streaked to single colonies. Six isolates
380 from each replicate culture were tested for growth with PCA, 4-HB, coumarate, and caffeate.
381 Representative isolates were selected and validated, followed by genome resequencing.

382

383 Genome resequencing

384 Genomic DNA was isolated using a Blood and Tissue kit (Qiagen, Valencia, CA), according
385 to the manufacturer's directions. The DNA was then sequenced by the Joint Genome Institute on
386 a MiSeq (Illumina, San Diego, CA) to approximately 75x coverage.

387

388 Growth rate measurements

389 Growth rates were measured as described previously (Clarkson *et al.*, 2017). Briefly, cultures
390 were grown overnight to saturation in M9 + 1 mM IPTG + 2 g/L glucose. They were then diluted
391 100-fold into fresh M9 + IPTG containing the appropriate carbon source and grown as triplicate
392 100 µL cultures in a Bioscreen C plate reader (Oy Growth Curves Ab Ltd, Helsinki, Finland).

393 Growth rates were calculated using CurveFitter software based on readings of optical density at
394 600 nm (Delaney *et al.*, 2013).
395
396 Proteomic measurements
397 Engineered *E. coli* strains were grown to saturation in 5 mL cultures of M9 + 2 g/L glucose +
398 1 mM IPTG. They were then diluted 100-fold into triplicate 5 mL of the same medium and
399 grown to mid-log phase. The cells were separated by centrifugation, washed twice with water,
400 and frozen in LN_2 for later analysis.
401 Processing for LC-MS/MS analysis was performed as previously described (Clarkson *et al.*,
402 2017). Briefly, crude protein lysates were obtained by bead beating cells in sodium deoxycholate
403 (SDC) lysis buffer (4% SDC, 100 mM ammonium bicarbonate, pH 8.0). Cleared protein lysates
404 were then adjusted to 10 mM dithiothreitol and incubated at 95 °C for 10 min to denature and
405 reduce proteins. Cysteines were alkylated/blocked with 30 mM iodoacetamide and 250 µg
406 transferred to a 10-kDa MWCO spin filter (Vivaspin 500, Sartorius) for *in situ* clean-up and
407 digestion with sequencing-grade trypsin (G-Biosciences). The tryptic peptide solution was then
408 spin-filtered through the MWCO membrane, adjusted to 1% formic acid to precipitate residual
409 SDC, and SDC precipitate removed from the peptide solution with water-saturated ethyl acetate
410 extraction. Peptide samples were then concentrated via SpeedVac (Thermo Fisher) and
411 quantified by BCA assay (Pierce) prior to LC-MS/MS analysis.
412 Peptide samples were analyzed by automated 2D LC-MS/MS analysis using a Vanquish
413 UHPLC plumbed directly in-line with a Q Exactive Plus mass spectrometer (Thermo Scientific)
414 outfitted with a triphasic MudPIT back column (RP-SCX-RP) coupled to an in-house pulled
415 nanospray emitter packed with 30 cm of 5 µm Kinetex C18 RP resin (Phenomenex). For each
416 sample, 5 µg of peptides were loaded, desalted, separated and analyzed across two successive
417 salt cuts of ammonium acetate (50 mM and 500 mM), each followed by 105 min organic
418 gradient, as previously detailed (Clarkson *et al.*, 2017). Eluting peptides were measured and
419 sequenced by data-dependent acquisition on the Q Exactive MS.
420 MS/MS spectra were searched against the *E. coli* K-12 proteome concatenated with
421 exogenous Pca, Hca, and Cou pathway proteins, common protein contaminants, and decoy
422 sequences using MyriMatch v.2.2 (Tabb *et al.*, 2007). Peptide spectrum matches (PSM) were
423 required to be fully tryptic with any number of missed cleavages; a static modification of

424 57.0214 Da on cysteine (carbamidomethylated) and a dynamic modification of 15.9949 Da on
425 methionine (oxidized) residues. PSMs were filtered using IDPicker v.3.0 (Ma *et al.*, 2009) with
426 an experiment-wide false-discovery rate controlled at < 1% at the peptide-level. Peptide
427 intensities were assessed by chromatographic area-under-the-curve and unique peptide intensities
428 summed to estimate protein-level abundance. Protein abundance distributions were then
429 normalized across samples and missing values imputed to simulate the MS instrument's limit of
430 detection. Significant differences in protein abundance were assessed by pairwise T-test.
431
432 Metabolite measurements
433 Strains JME129 and JME131, picked from single colonies on freshly-streaked plates, were
434 grown in M9 + 2 g/L glucose + 1 mM IPTG overnight at 37 °C. Cultures were diluted 100-fold
435 into fresh M9 + 1 g/L coumarate + 1 mM IPTG and regrown to mid-log phase, then quickly
436 concentrated to 100 μL by centrifugation and plated on sterile 50 mm mixed cellulose ester 0.2
437 μm Whatman filters placed on agar plates containing M9 + 1 g/L coumarate + 1 mM IPTG
438 (Rabinowitz and Kimball, 2007). The plates were incubated for a further 1 h at 37 °C before
439 analysis.
440 Metabolites were quickly extracted by scraping cells off the permeable membrane and
441 immediately placing them in 100 μL of pre-chilled extraction solvent (40% acetonitrile, 40%
442 methanol, 20% H_2O, 0.1 M formic acid; stored at -20°C) (Rabinowitz and Kimball, 2007). The
443 extraction was then placed at -20°C for 15 min, cells/debris pelleted at 21,000 x g for 5 min at
444 4°C, and supernatants transferred to cold autosampler vials. Five microliters of each extract was
445 then analyzed by high-resolution LC-MS/MS using a Vanquish UHPLC plumbed directly in-line
446 with a Q Exactive Plus mass spectrometer (Thermo Scientific) outfitted with an in-house pulled
447 nanospray emitter, as previously described (Cecil *et al.*, 2018). Samples (n=4 per strain;
448 biological replicates) were analyzed across 3 separate LC-MS injects, with technical replicates
449 (n=3) and blank runs spread across the entire sampling campaign time of 20 h. Precursor
450 abundances were derived for specific nucleotide phosphates (AxP, GxP, IxP) via Skyline
451 (MacLean *et al.*, 2010) at 5 ppm accuracy and normalized to IPTG, which is taken up by cells
452 but not metabolized. Student's T test was performed to compare the normalized nucleotide
453 phosphate pools (merged signals from tri-, di-, and monophosphate versions) across samples to
454 identify differences between the two *E. coli* strains.

455

456 IMPDH expression and purification

457 NAD^+ was purchased from Roche, IMP and EDTA were purchased from Fisher, MOPS was
458 purchased from Sigma, DTT and IPTG were purchased from GoldBio. EcIMPDH/WT and
459 AbIMPDH were purified as previously described (Makowska-Grzyska *et al.*, 2015). pJM303,
460 expressing EcIMPDH/D243G was transformed into BL21(*ΔguaB*) cells that lack endogenous
461 EcIMPDH (MacPherson *et al.*, 2010). An overnight culture (5 mL) was diluted into 1 L of fresh
462 LB broth containing 100 µg/mL ampicillin and grown at 37 ºC. Once the culture reached an
463 OD600 of 0.6-0.8, IPTG was added to a final concentration of 0.25 mM to induce expression of
464 IMPDH. After 13 h at 30 ºC, the cells were collected by centrifugation. All the operations below
465 were performed at 4 ºC. The pellet was resuspended in 50 mL phosphate buffer (pH = 8.0)
466 containing 1 mM dithiothreitol (DTT) and sonicated. The debris was removed by centrifugation
467 at 10,000 g at 4 ºC for 1 h.

468 The enzyme in the supernatant was purified by nickel affinity chromatography. The Ni-NTA
469 resin equilibrated with water and phosphate buffer (pH = 8.0). Lysate was loaded onto the 10 mL
470 column with 5 mL resin and washed with 50 mL phosphate buffer (pH = 8.0) then 50 mL
471 phosphate buffer containing 25 mM imidazole. Enzyme was eluted in 25 ml phosphate buffer
472 with 250 mM imidazole. The fractions with IMPDH activity were identified by enzyme activity
473 assays, combined and dialyzed in 25 mM HEPES (pH = 8.0), 1mM DTT and 1 mM EDTA.
474 Protein concentration was determined by Bio-Rad Bradford assay using IgG as a standard. The
475 assay over-estimates the concentration of IMPDH by a factor of 2.6, and protein concentration
476 was adjusted accordingly (Wang *et al.*, 1996).

477

478 IMPDH enzyme assays

479 The IMPDH reaction was monitored by measuring the rate of NADH production on a
480 Shimadzu UV-1800 Spectrometer at λ = 340 nm. MOPS buffer (pH = 7.0) was used to reduce
481 the background absorbance of 4HB. The assay buffer was composed by 20 mM MOPS (pH
482 =7.0), 100 mM KCl, 1 mM EDTA and 1 mM DTT. The final volume of each cuvette was 1 mL.

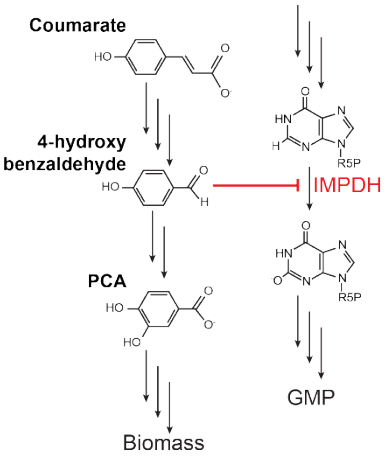
483 Kinetic parameters with respect to NAD^+ were determined by measuring the initial velocity
484 for varying concentrations of NAD^+ at a fixed saturating concentration of IMP (1.2 mM) and 50
485 nM of enzyme. Kinetic parameters with respect to IMP were determined by measuring the initial

486 velocity for varying concentrations of IMP at a fixed saturating concentration of NAD^+ (2.5 mM)
487 and 50 nM of enzyme. Initial velocities were plotted against substrate concentrations and the
488 data were fit using SigmaPlot. The values of $K_{i,app}$ were determined by measuring the initial
489 velocities for the reaction of IMPDH (20 nM) in the presence of varied concentrations (0-300
490 μM) of 4-hydroxybenzaldehyde at 12 μM of IMP and 500 μM of NAD^+. The inhibition by 4-
491 hydroxybenzaldehyde under physiological concentrations of IMP (270 μM) and NAD^+ (2500
492 μM) were determined as well (Bennett *et al.*, 2009; Park *et al.*, 2016).
493
494 Inhibitor Characterization
495 The inhibition mechanism was characterized by varying the concentration of 4HB from 0 to 200
496 μM at fixed concentration of substrates (see following) and enzyme (50 nM). The values of $K_{i,app}$
497 with respect to IMP were determined using a fixed NAD+ concentration (1 mM; $K_m(NAD^+)$ =
498 210 μM for EcIMPDH/WT; $K_m(NAD^+)$ = 210 μM for *Ec*IMPDH/D243G) and the values of $K_{i,app}$
499 with respect to NAD^+ were determined by using a saturating IMP concentration (300 μM;
500 K_m(IMP) = 21 μM for *Ec*IMPDH/WT; K_m(IMP) = 7.9 μM for *Ec*IMPDH/D243G). The
501 inhibition mechanism was determined by fitting to the equations for uncompetitive (Eq.1) and
502 noncompetitive (Eq.2) inhibition using SigmaPlot.
503 Equation for uncompetitive inhibition:
504 $$v = V_{max}[S]/ \{K_m + [S] (1 + [I]/K_{ii})\} \quad \text{(Eq. 1)}$$
505 Equation for mixed noncompetitive inhibition:
506 $$v = V_{max}[S]/ \{(K_m(1+[I]/K_{is}) + [S] (1 + [I]/K_{ii})\} \quad \text{(Eq. 2)}$$
507
508 Homology modeling
509 To generate models of IMPDH from *E. coli* (accession number P0ADG7) with the
510 corresponding cofactors and substrates, HHpred (Zimmermann *et al.*, 2017) was used to search
511 the Protein Data Bank for suitable structural templates. Five templates were chosen on the basis
512 of their similarity to the query sequence, inclusion of cofactors and substrates, or both (Table
513 S5). The sequences were aligned with MAFFT (L-INS-i) (Katoh and Standley, 2013) (Figure
514 S15) and homology models were generated using RosettaCM (Song *et al.*, 2013) with fragment
515 files obtained from the Robetta web server (Kim *et al.*, 2004; Gront *et al.*, 2011). The top-scoring
516 model was used for docking of IMP, NAD^+, and 4-hydroxybenzaldehyde.

517 The flap containing the catalytic dyad (R401 and Y402) is not resolved in most X-ray crystal
518 structures of IMPDH but is present in the structure of the phosphate-bound "apoenzyme" from
519 *Bacillus anthracis* (PDB entry 3TSB) (Makowska-Grzyska *et al.*, 2015). Thus, we used this
520 structure as a template to generate homology models of *E. coli* IMPDH in the closed
521 conformation and the top scoring model was selected to generate a model of the C305-XMP*
522 covalent intermediate for molecular dynamics simulations. Sequences were aligned with MAFFT
523 (L-INS-i) (Katoh and Standley, 2013) (Figure S16).
524
525 Ligand docking
526 Structure files in mol2 format for IMP (ZINC04228242), NAD^+ (ZINC08214766), and 4-
527 hydroxybenzaldehyde (ZINC00156709) were obtained from http://zinc.docking.org (Irwin *et al.*,
528 2012). RosettaLigand (Meiler and Baker, 2006) was used to dock 4-hydroxybenzaldehyde into
529 the active site of the IMPDH model following a previously described protocol (Combs *et al.*,
530 2013). Top binding poses were ranked on the basis of their '*interface_delta*' score in Rosetta
531 energy units. Additional details of the homology modeling and ligand docking are provided in
532 the Supporting Information.
533
534 Molecular dynamics simulations
535 Initial coordinates for XMP were extracted from the crystal structure of *B. anthracis*
536 IMPDH complexed with XMP (PDB ID 3TSD), and the covalently bound C305-XMP* complex
537 was generated for chain A using the Molefacture plugin in VMD (Humphrey *et al.*, 1996). The
538 force field toolkit (ffTK) plugin in VMD was used to generate CHARMM-compatible force field
539 parameters for C305-XMP*. Gaussian09 was used to perform geometry optimizations, compute
540 Hessian matrices, and calculate water interaction energies of the C305-XMP* fragment. NAMD
541 2.11 was used for charge, bond, angle, and dihedral optimization (Phillips *et al.*, 2005).
542 The CHARMM36 force field (Best *et al.*, 2012) and TIP3P water model (Jorgensen *et al.*,
543 1983) were used to describe the protein and solvent, respectively. Each system (wild-type and
544 mutant) was solvated in a periodic box of 168 Å × 168 Å × 104 Å and 0.15 M KCl ions were
545 added using CHARMM-GUI (Jo *et al.*, 2008), resulting in a system of ~274,000 atoms. All-atom
546 molecular dynamics (MD) simulations were performed using the OpenMM 7.0 package
547 (Eastman *et al.*, 2017) with GPU acceleration using CUDA 7.2. Ten thousand steps of energy

548 minimization were performed to eliminate clashes, followed by equilibration in the NPT
549 ensemble at 310 K. Temperature was maintained using the Langevin thermostat with a damping
550 coefficient of 1 ps^{-1}. To enable a 5-fs time step, which was used in all simulations, all bond
551 lengths were constrained to their equilibrium distances and the masses of hydrogens were
552 repartitioned to the parent heavy atoms (Eastman *et al.*, 2017). Five separate runs of 100 ns each
553 were performed for both the wild type and mutant IMPDH and the last 50 ns of each run was
554 used for analysis.
555
556 **Data Availability**
557 The data that support the findings of this study are available from the corresponding
558 author upon reasonable request. Genome sequences of JME64-67 are available in GenBank
559 through BioProject PRJNA559875.
560
561 **Acknowledgments**
562 Genome resequencing and analysis was performed by Christa Pennacchio, Natasha Brown,
563 Anna Lipzen, and Wendy Schackwitz at the Joint Genome Institute. DNA synthesis was
564 performed by Jan-Fang Cheng, Samuel Deutsch, and Miranda Harmon-Smith at the Joint
565 Genome Institute. The work conducted by the U.S. Department of Energy Joint Genome
566 Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the
567 U.S. Department of Energy under Contract No. DE-AC02-05CH11231. SJC was supported by
568 NIH/NIGMS-IMSD Grant No. R25GM086761. This material is based upon work supported by
569 the National Science Foundation Graduate Research Fellowship under Grant No. (2017219379).
570 This work was supported by the BioEnergy Science Center and The Center for Bioenergy
571 Innovation, both U.S. Department of Energy Research Centers supported by the Office of
572 Biological and Environmental Research in the DOE Office of Science; the National Institutes of
573 Health (GM054403 to LH); and the ORNL Laboratory Directed Research and Development
574 program (#8949 to JMP). This work also used resources of the Compute and Data Environment
575 for Science (CADES) and the Oak Ridge Leadership Computing Facility (OLCF) at ORNL. Oak
576 Ridge National Laboratory (ORNL) is managed by UT-Battelle, LLC, for the DOE under
577 Contract No. DE-AC05-00OR22725.
578

579 **Graphical Abstract:**



580

581

582 **Abbreviated Summary:**

583 A pathway for catabolism of coumarate, a biotechnologically relevant lignin-derived aromatic
584 compound, was engineered into *Escherichia coli*. Accumulation of an intermediate, 4-
585 hydroxybenzaldehyde, was shown to inhibit a key enzyme in purine nucleotide biosynthesis,
586 IMP dehydrogenase (IMPDH). Inhibition could be relieved through point mutations to IMPDH
587 that altered enzyme dynamics without disrupting the catalytic center.

588

589 **References:**

590 Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Dien, S.J. Van, and Rabinowitz, J.D.
591 (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in
592 Escherichia coli. *Nat Chem Biol* **5**: 593–599 http://www.nature.com/articles/nchembio.186.
593 Accessed February 20, 2018.

594 Best, R.B., Zhu, X., Shim, J., Lopes, P.E.M., Mittal, J., Feig, M., and MacKerell, A.D.
595 (2012) Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting
596 Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J Chem*
597 *Theory Comput* **8**: 3257–3273 http://pubs.acs.org/doi/10.1021/ct300400x. Accessed May 23,
598 2019.

599 Bugg, T.D.H., Ahmad, M., Hardiman, E.M., and Rahmanpour, R. (2011) Pathways for
600 degradation of lignin in bacteria and fungi. *Nat Prod Rep* **28**: 1883–1896
601 http://xlink.rsc.org/?DOI=c1np00042j. Accessed July 20, 2016.

602 Cecil, J.H., Garcia, D.C., Giannone, R.J., and Michener, J.K. (2018) Rapid, parallel

603 identification of pathways for catabolism of lignin-derived aromatic compounds in
604 Novosphingobium aromaticivorans. *Appl Environ Microbiol* AEM.01185-18
605 http://www.ncbi.nlm.nih.gov/pubmed/30217841. Accessed October 8, 2018.
606 Chain, P.S.G., Denef, V.J., Konstantinidis, K.T., Vergez, L.M., Agulló, L., Reyes, V.L., *et*
607 *al.* (2006) Burkholderia xenovorans LB400 harbors a multi-replicon, 9.73-Mbp genome shaped
608 for versatility. *Proc Natl Acad Sci U S A* **103**: 15280–7
609 http://www.ncbi.nlm.nih.gov/pubmed/17030797. Accessed July 12, 2018.
610 Chen, Y.-J., Liu, P., Nielsen, A.A.K., Brophy, J.A.N., Clancy, K., Peterson, T., and Voigt,
611 C.A. (2013) Characterization of 582 natural and synthetic terminators and quantification of their
612 design constraints. *Nat Methods* **10**: 659–664
613 http://www.nature.com/doifinder/10.1038/nmeth.2515. Accessed November 9, 2016.
614 Clark, I.C., Melnyk, R.A., Youngblut, M.D., Carlson, H.K., Iavarone, A.T., and Coates, J.D.
615 (2015) Synthetic and evolutionary construction of a chlorate-reducing *Shewanella oneidensis*
616 MR-1. *MBio* **6**: e00282-15 http://mbio.asm.org/lookup/doi/10.1128/mBio.00282-15. Accessed
617 July 20, 2016.
618 Clarkson, S.M., Hamilton-Brehm, S.D., Giannone, R.J., Engle, N.L., Tschaplinski, T.J.,
619 Hettich, R.L., and Elkins, J.G. (2014) A comparative multidimensional LC-MS proteomic
620 analysis reveals mechanisms for furan aldehyde detoxification in Thermoanaerobacter
621 pseudethanolicus 39E. *Biotechnol Biofuels* **7**: 165 https://doi.org/10.1186/s13068-014-0165-z.
622 Clarkson, S.M., Kridelbaugh, D.M., Elkins, J.G., Guss, A.M., and Michener, J. (2017)
623 Construction and optimization of a heterologous pathway for protocatechuate catabolism in
624 Escherichia coli enables rapid bioconversion of model lignin monomers. *Appl Environ Microbiol*
625 **83**: e01313-17 http://www.biorxiv.org/content/early/2017/01/06/094805. Accessed June 28,
626 2017.
627 Combs, S.A., DeLuca, S.L., DeLuca, S.H., Lemmon, G.H., Nannemann, D.P., Nguyen,
628 E.D., *et al.* (2013) Small-molecule ligand docking into comparative models with Rosetta. *Nat*
629 *Protoc* **8**: 1277–1298 http://www.nature.com/doifinder/10.1038/nprot.2013.074. Accessed
630 March 12, 2018.
631 Delaney, N.F., Kaczmarek, M.E., Ward, L.M., Swanson, P.K., Lee, M.-C., and Marx, C.J.
632 (2013) Development of an optimized medium, strain and high-throughput culturing methods for
633 Methylobacterium extorquens. *PLoS One* **8**: e62957

634 http://dx.plos.org/10.1371/journal.pone.0062957. Accessed July 15, 2016.
635 Eastman, P., Swails, J., Chodera, J.D., McGibbon, R.T., Zhao, Y., Beauchamp, K.A., *et al.*
636 (2017) OpenMM 7: Rapid development of high performance algorithms for molecular dynamics.
637 *PLOS Comput Biol* **13**: e1005659 https://dx.plos.org/10.1371/journal.pcbi.1005659. Accessed
638 May 23, 2019.
639 Espah Borujeni, A., Channarasappa, A.S., and Salis, H.M. (2014) Translation rate is
640 controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding
641 at upstream standby sites. *Nucleic Acids Res* **42**: 2646–59
642 http://www.ncbi.nlm.nih.gov/pubmed/24234441. Accessed July 15, 2016.
643 Gront, D., Kulp, D.W., Vernon, R.M., Strauss, C.E.M., and Baker, D. (2011) Generalized
644 Fragment Picking in Rosetta: Design, Protocols and Applications. *PLoS One* **6**: e23294
645 http://dx.plos.org/10.1371/journal.pone.0023294. Accessed March 12, 2018.
646 Hedstrom, L. (2009) IMP dehydrogenase: structure, mechanism, and inhibition. *Chem Rev*
647 **109**: 2903–28 http://www.ncbi.nlm.nih.gov/pubmed/19480389. Accessed March 28, 2017.
648 Huang, X., Holden, H.M., and Raushel, F.M. (2001) Channeling of Substrates and
649 Intermediates in Enzyme-Catalyzed Reactions. *Annu Rev Biochem* **70**: 149–180.
650 Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J Mol*
651 *Graph* **14**: 33–8, 27–8 http://www.ncbi.nlm.nih.gov/pubmed/8744570. Accessed May 23, 2019.
652 Irwin, J.J., Sterling, T., Mysinger, M.M., Bolstad, E.S., and Coleman, R.G. (2012) ZINC: A
653 Free Tool to Discover Chemistry for Biology. *J Chem Inf Model* **52**: 1757–1768
654 http://pubs.acs.org/doi/10.1021/ci3001277. Accessed March 12, 2018.
655 Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., and Yang, S. (2015) Multigene editing in
656 the Escherichia coli genome via the CRISPR-Cas9 system. *Appl Environ Microbiol* **81**: 2506–
657 2514 http://aem.asm.org/lookup/doi/10.1128/AEM.04023-14. Accessed July 15, 2016.
658 Jo, S., Kim, T., Iyer, V.G., and Im, W. (2008) CHARMM-GUI: A web-based graphical user
659 interface for CHARMM. *J Comput Chem* **29**: 1859–1865
660 http://doi.wiley.com/10.1002/jcc.20945. Accessed May 23, 2019.
661 Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983)
662 Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**: 926–935
663 http://aip.scitation.org/doi/10.1063/1.445869. Accessed May 23, 2019.
664 Katoh, K., and Standley, D.M. (2013) MAFFT Multiple Sequence Alignment Software

665 Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**: 772–780
666 https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mst010. Accessed March 12,
667 2018.
668 Kim, D.E., Chivian, D., and Baker, D. (2004) Protein structure prediction and analysis using
669 the Robetta server. *Nucleic Acids Res* **32**: W526–W531 https://academic.oup.com/nar/article-
670 lookup/doi/10.1093/nar/gkh468. Accessed March 12, 2018.
671 Kim, J., and Copley, S.D. (2012) Inhibitory cross-talk upon introduction of a new metabolic
672 pathway into an existing metabolic network. *Proc Natl Acad Sci* **109**: E2856–E2864
673 http://www.ncbi.nlm.nih.gov/pubmed/22984162. Accessed August 10, 2017.
674 Kizer, L., Pitera, D.J., Pfleger, B.F., and Keasling, J.D. (2008) Application of Functional
675 Genomics to Pathway Optimization for Increased Isoprenoid Production. *Appl Environ*
676 *Microbiol* **74**: 3229–3241 http://aem.asm.org/content/74/10/3229.abstract.
677 Kosuri, S., Goodman, D.B., Cambray, G., Mutalik, V.K., Gao, Y., Arkin, A.P., *et al.* (2013)
678 Composability of regulatory sequences controlling transcription and translation in *Escherichia*
679 *coli*. *Proc Natl Acad Sci U S A* **110**: 14024–9 http://www.ncbi.nlm.nih.gov/pubmed/23924614.
680 Accessed November 9, 2016.
681 Kurnasov, O. V, Polanuyer, B.M., Ananta, S., Sloutsky, R., Tam, A., Gerdes, S.Y., and
682 Osterman, A.L. (2002) Ribosylnicotinamide kinase domain of NadR protein: identification and
683 implications in NAD biosynthesis. *J Bacteriol* **184**: 6906–17
684 http://www.ncbi.nlm.nih.gov/pubmed/12446641. Accessed August 25, 2017.
685 Ma, Z.-Q., Dasari, S., Chambers, M.C., Litton, M.D., Sobecki, S.M., Zimmerman, L.J., *et*
686 *al.* (2009) IDPicker 2.0: Improved protein assembly with high discrimination peptide
687 identification filtering. *J Proteome Res* **8**: 3872–3881
688 http://www.ncbi.nlm.nih.gov/pubmed/19522537. Accessed April 12, 2017.
689 MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., *et al.*
690 (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics
691 experiments. *Bioinformatics* **26**: 966–968 https://academic.oup.com/bioinformatics/article-
692 lookup/doi/10.1093/bioinformatics/btq054. Accessed October 8, 2018.
693 MacPherson, I.S., Kirubakaran, S., Gorla, S.K., Riera, T. V., D'Aquino, J.A., Zhang, M., *et*
694 *al.* (2010) The Structural Basis of Cryptosporidium-Specific IMP Dehydrogenase Inhibitor
695 Selectivity. *J Am Chem Soc* **132**: 1230–1231 http://pubs.acs.org/doi/abs/10.1021/ja909947a.

696 Accessed February 20, 2018.
697 Makowska-Grzyska, M., Kim, Y., Maltseva, N., Osipiuk, J., Gu, M., Zhang, M., *et al.*
698 (2015) A Novel Cofactor-binding Mode in Bacterial IMP Dehydrogenases Explains Inhibitor
699 Selectivity. *J Biol Chem* **290**: 5893–5911 http://www.ncbi.nlm.nih.gov/pubmed/25572472.
700 Accessed February 20, 2018.
701 Masai, E., Katayama, Y., and Fukuda, M. (2007) Genetic and biochemical investigations on
702 bacterial catabolic pathways for lignin-derived aromatic compounds. *Biosci Biotechnol Biochem*
703 **71**: 1–15 http://www.tandfonline.com/doi/full/10.1271/bbb.60437. Accessed November 2, 2016.
704 Meiler, J., and Baker, D. (2006) ROSETTALIGAND: Protein-small molecule docking with
705 full side-chain flexibility. *Proteins Struct Funct Bioinforma* **65**: 538–548
706 http://doi.wiley.com/10.1002/prot.21086. Accessed March 12, 2018.
707 Michener, J.K., Camargo Neves, A.A., Vuilleumier, S., Bringel, F., and Marx, C.J. (2014)
708 Effective use of a horizontally-transferred pathway for dichloromethane catabolism requires
709 post-transfer refinement. *Elife* **3**
710 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4271186&tool=pmcentrez&renderty
711 pe=abstract. Accessed July 14, 2016.
712 Michener, J.K., Nielsen, J., and Smolke, C.D. (2012) Identification and treatment of heme
713 depletion attributed to overexpression of a lineage of evolved P450 monooxygenases. *Proc Natl*
714 *Acad Sci* **109**: 19504–19509 http://www.pnas.org/cgi/doi/10.1073/pnas.1212287109. Accessed
715 July 14, 2016.
716 Mills, T.Y., Sandoval, N.R., and Gill, R.T. (2009) Cellulosic hydrolysate toxicity and
717 tolerance mechanisms in Escherichia coli. *Biotechnol Biofuels* **2**: 26
718 https://doi.org/10.1186/1754-6834-2-26.
719 Mukhopadhyay, A. (2015) Tolerance engineering in bacteria for the production of advanced
720 biofuels and chemicals. *Trends Microbiol* **23**: 498–508
721 http://www.sciencedirect.com/science/article/pii/S0966842X1500102X.
722 Nielsen, J., and Keasling, J.D. (2016) Engineering Cellular Metabolism. *Cell* **164**: 1185–
723 1197 http://www.sciencedirect.com/science/article/pii/S0092867416300708. Accessed August
724 10, 2017.
725 Otani, H., Lee, Y.-E., Casabon, I., and Eltis, L.D. (2014) Characterization of p-
726 hydroxycinnamate catabolism in a soil Actinobacterium. *J Bacteriol* **196**: 4293–303

727 http://www.ncbi.nlm.nih.gov/pubmed/25266382. Accessed May 19, 2017.
728 Pál, C., Papp, B., and Lercher, M.J. (2005) Adaptive evolution of bacterial metabolic
729 networks by horizontal gene transfer. *Nat Genet* **37**: 1372–1375
730 http://www.nature.com/articles/ng1686. Accessed March 16, 2018.
731 Park, J.O., Rubin, S.A., Xu, Y.-F., Amador-Noguez, D., Fan, J., Shlomi, T., and Rabinowitz,
732 J.D. (2016) Metabolite concentrations, fluxes and free energies imply efficient enzyme usage.
733 *Nat Chem Biol* **12**: 482–489 http://www.nature.com/articles/nchembio.2077. Accessed February
734 20, 2018.
735 Parke, D., and Ornston, L.N. (2003) Hydroxycinnamate (hca) Catabolic Genes from
736 Acinetobacter sp. Strain ADP1 Are Repressed by HcaR and Are Induced by Hydroxycinnamoyl-
737 Coenzyme A Thioesters. *Appl Environ Microbiol* **69**: 5398–5409
738 http://aem.asm.org/content/69/9/5398.abstract.
739 Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., *et al.* (2005)
740 Scalable molecular dynamics with NAMD. *J Comput Chem* **26**: 1781–1802
741 http://doi.wiley.com/10.1002/jcc.20289. Accessed May 23, 2019.
742 Pisithkul, T., Jacobson, T.B., O'Brien, T.J., Stevenson, D.M., and Amador-Noguez, D.
743 (2015) Phenolic Amides Are Potent Inhibitors of De Novo Nucleotide Biosynthesis. *Appl*
744 *Environ Microbiol* **81**: 5761–72 http://www.ncbi.nlm.nih.gov/pubmed/26070680. Accessed
745 August 4, 2017.
746 Porse, A., Schou, T.S., Munck, C., Ellabaan, M.M.H., and Sommer, M.O.A. (2018)
747 Biochemical mechanisms determine the functional compatibility of heterologous genes. *Nat*
748 *Commun* **9**: 522 http://www.nature.com/articles/s41467-018-02944-3. Accessed March 13, 2018.
749 Rabinowitz, J.D., and Kimball, E.H. (2007) Acidic Acetonitrile for Cellular Metabolome
750 Extraction from Escherichia coli. https://pubs.acs.org/doi/full/10.1021/ac070470c. Accessed
751 September 24, 2018.
752 Ragauskas, A.J., Beckham, G.T., Biddy, M.J., Chandra, R., Chen, F., Davis, M.F., *et al.*
753 (2014) Lignin Valorization: Improving Lignin Processing in the Biorefinery. *Science (80-)* **344**:
754 709-+ http://www.ncbi.nlm.nih.gov/pubmed/24833396. Accessed July 19, 2016.
755 Salis, H.M., Mirsky, E.A., and Voigt, C.A. (2009) Automated design of synthetic ribosome
756 binding sites to control protein expression. *Nat Biotechnol* **27**: 946–950
757 http://www.nature.com/doifinder/10.1038/nbt.1568. Accessed July 15, 2016.

758 Saxer, G., Krepps, M.D., Merkley, E.D., Ansong, C., Deatherage Kaiser, B.L., Valovska,
759 M.-T., *et al.* (2014) Mutations in Global Regulators Lead to Metabolic Selection during
760 Adaptation to Complex Environments. *PLoS Genet* **10**: e1004872
761 http://dx.plos.org/10.1371/journal.pgen.1004872. Accessed August 25, 2017.
762 Song, Y., DiMaio, F., Wang, R.Y.-R., Kim, D., Miles, C., Brunette, T., *et al.* (2013) High-
763 Resolution Comparative Modeling with RosettaCM. *Structure* **21**: 1735–1742
764 https://www.sciencedirect.com/science/article/pii/S0969212613002979?via%3Dihub. Accessed
765 March 12, 2018.
766 Standaert, R.F., Giannone, R.J., and Michener, J.K. (2018) Identification of parallel and
767 divergent optimization solutions for homologous metabolic enzymes. *Metab Eng Commun* **6**:
768 56–62 https://www.sciencedirect.com/science/article/pii/S2214030118300063. Accessed May 4,
769 2018.
770 Tabb, D.L., Fernando, C.G., and Chambers, M.C. (2007) MyriMatch: highly accurate
771 tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J Proteome*
772 *Res* **6**: 654–61 http://www.ncbi.nlm.nih.gov/pubmed/17269722. Accessed April 12, 2017.
773 Wang, W., Papov, V. V., Minakawa, N., Matsuda, A., Biemann, K., and Hedstrom, L.
774 (1996) Inactivation of Inosine 5‘-Monophosphate Dehydrogenase by the Antiviral Agent 5-
775 Ethynyl-1-β- d -Ribofuranosylimidazole-4-Carboxamide 5‘-Monophosphate †. *Biochemistry* **35**:
776 95–101 http://pubs.acs.org/doi/abs/10.1021/bi951499q. Accessed February 20, 2018.
777 Yi, X., Gu, H., Gao, Q., Liu, Z.L., and Bao, J. (2015) Transcriptome analysis of
778 Zymomonas mobilis ZM4 reveals mechanisms of tolerance and detoxification of phenolic
779 aldehyde inhibitors from lignocellulose pretreatment. *Biotechnol Biofuels* **8**: 153
780 http://www.ncbi.nlm.nih.gov/pubmed/26396591. Accessed July 27, 2018.
781 Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., *et al.* (2017) A
782 Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core.
783 *J Mol Biol* https://www.sciencedirect.com/science/article/pii/S0022283617305879. Accessed
784 March 12, 2018.
785

786 **Tables:**

787 **Table 1**: Kinetic Parameters for *Ec*IMPDH/WT, *Ec*IMPDH/D243G and *Ab*IMPDH. The values
788 are the average and range of two independent experiments. a. UC inhibition versus IMP. b. NC
789 inhibition versus NAD^+. n.d., not determined.

	*Ec*IMPDH/WT	*Ec*IMPDH/D243G	*Ab*IMPDH
k_{cat} (s^{-1})	0.97 ± 0.05	0.50 ± 0.04	0.86 ± 0.02
K_m(IMP) (μM)	21 ± 3	7.9 ± 1	16 ± 3
K_m(NAD) (μM)	210 ± 20	210 ± 40	208 ± 17
K_{ii}(NAD) (μM)	4300 ± 530	3500 ± 690	10600 ± 1100
$K_{i,app}$(4HB) (μM)	320 ± 30	1250 ± 50	720 ± 30
IMP K_{ii} (μM) [a]	930 ± 70	570 ± 30	n.d.
NAD^+ K_{is} (μM) [b]	490 ± 100	450 ± 90	n.d.
K_{ii} (μM) [b]	1200 ± 100	450 ± 90	n.d.

790

791 **Figure Legends:**

792 **Figure 1:** Two routes, exemplified by the *hcaABC* pathway from *A. baylyi* ADP1 and the
793 *couLMNO* pathway from *R. jostii*, convert the phenylpropanoid coumarate to 4-HB, which is
794 then oxidized to PCA. For simplicity, cofactors and the resulting acetyl-CoA are not shown.

795

796 **Figure 2:** Reconstruction identifies causal mutations. (A) For the *hca* pathway, a triple mutant,
797 containing mutations to *pobA* (synonymous, T3T), *guaB* (D243G), and *hcaC* (intergenic
798 mutation, G→A 47 bp upstream of start codon), and the three double mutants were grown in
799 minimal medium with 1 g/L of the indicated substrate as the sole source of carbon and energy.
800 (B) As in 3A, except using the *cou* pathway and mutations to *pobA* (T3T), *nadR* (Δ326-410), and
801 *couL* (S134Y). Error bars show one standard deviation, calculated from three biological
802 replicates. Specific mutations are described in Tables S1 and S4.

803

804 **Figure 3:** Beneficial mutations to IMPDH are distant from the active site. (A) Structural model
805 of *E. coli* IMPDH colored by chain. (B) Chain A of IMPDH in the closed conformation,
806 highlighting the loop containing C305-XMP* and the active site flap containing the R401-Y402
807 catalytic dyad.

808

809 **Figure 4:** Accumulation of 4-hydroxybenzaldehyde inhibits IMPDH. (A) Metabolomics of
810 purine nucleobases and derivatives. Different phosphorylation states could not be distinguished
811 and are reported as the total nucleotide pool. Each of the four biological replicates is shown as a
812 separate measurement, normalized to the mean for JME129. The solid bar indicates the median,
813 as a guide for the eye. N.S.: Not significant; **: $p < 0.01$. (B) Strains with wild-type and mutant
814 versions of IMPDH were grown in medium containing 1 g/L coumarate with and without the
815 addition of 5 mg/L guanosine (~18 μM). No growth was observed with guanosine alone. Error
816 bars show the standard deviation, calculated from three biological replicates. (C) Enzyme
817 variants were purified and assayed *in vitro* for inhibition by 4-hydroxybenzaldehyde. Curves
818 show a model fit, using the calculated inhibition constants. Error bars show the standard
819 deviation, calculated from three biological replicates.

820

821 **Figure 5:** A beneficial mutation to IMPDH affects enzyme structural dynamics. Hydrogen bond
822 network around the D243G mutation site for (A) wild-type and (B) mutant IMPDH from MD
823 simulations. (C) Heavy atom distance distributions from five independent simulations of each
824 closed conformation model are shown.

825

826 **Figure 6:** (A) Selected snapshots of the flap from MD simulations of wild-type and mutant
827 IMPDH in the closed conformation. Cα atoms of key residues whose interactions differ between
828 wild type and mutant simulations (D50, R386, S399, S405, D410, E418) are shown as spheres
829 and labeled. A selected docking pose is shown for 4-hydroxybenzadehyde in the IMP-bound
830 open conformation after superposition with the closed conformation snapshots. (B) Selected
831 docking pose showing 4-hydroxybenzaldehyde interactions with residues D248 and S250. Flap
832 residues were omitted as they were not resolved in the templates used to model the open
833 conformation.
834