ACS SyntheticBiology

pubs.acs.org/synthbio

Research Article

# A High-Quality Genome-Scale Model for *Rhodococcus opacus* Metabolism

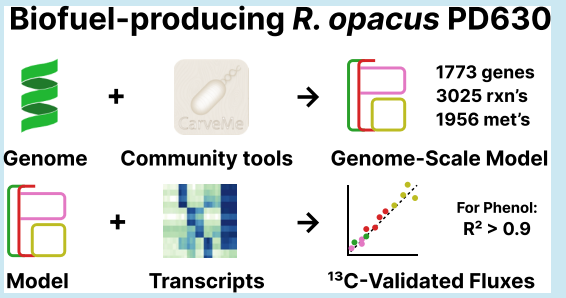
Garrett W. Roell, Christina Schenk, Winston E. Anthony, Rhiannon R. Carr, Aditya Ponukumati, Joonhoon Kim, Elena Akhmatskaya, Marcus Foston, Gautam Dantas, Tae Seok Moon,* Yinjie J. Tang,* and Hector García Martín*

Cite This: *ACS Synth. Biol.* 2023, 12, 1632−1644

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** *Rhodococcus opacus* is a bacterium that has a high tolerance to aromatic compounds and can produce significant amounts of triacylglycerol (TAG). Here, we present iGR1773, the first genome-scale model (GSM) of *R. opacus* PD630 metabolism based on its genomic sequence and associated data. The model includes 1773 genes, 3025 reactions, and 1956 metabolites, was developed in a reproducible manner using CarveMe, and was evaluated through Metabolic Model tests (MEMOTE). We combine the model with two Constraint-Based Reconstruction and Analysis (COBRA) methods that use transcriptomics data to predict growth rates and fluxes: E-Flux2 and SPOT (Simplified Pearson Correlation with Transcriptomic data). Growth rates are best predicted by E-Flux2. Flux profiles are more accurately predicted by E-Flux2 than flux balance analysis (FBA) and parsimonious FBA (pFBA), when compared to 44 central carbon fluxes measured by $^{13}$C-Metabolic Flux Analysis ($^{13}$C-MFA). Under glucose-fed conditions, E-Flux2 presents an $R^2$ value of 0.54, while predictions based on pFBA had an inferior $R^2$ of 0.28. We attribute this improved performance to the extra activity information provided by the transcriptomics data. For phenol-fed metabolism, in which the substrate first enters the TCA cycle, E-Flux2's flux predictions display a high $R^2$ of 0.96 while pFBA showed an $R^2$ of 0.93. We also show that glucose metabolism and phenol metabolism function with similar relative ATP maintenance costs. These findings demonstrate that iGR1773 can help the metabolic engineering community predict aromatic substrate utilization patterns and perform computational strain design.




**KEYWORDS:** *ATP maintenance, genome-scale models, omics data, $^{13}$C-metabolic flux analysis, predictive biology*

Downloaded via UNIV OF HAWAII on September 15, 2023 at 02:58:05 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

## 1. INTRODUCTION

*Rhodococcus opacus* PD630 (hereafter, *R. opacus*) is a Gram-positive aerobic bacterium known for its pronounced ability to produce triacylglycerol, a biofuel precursor, from aromatic monomers.[1,2] *R. opacus* can be used as a 'biological funnel' to convert heterogeneous mixtures of aromatic compounds from the thermal or catalytic deconstruction of lignin into lipid-based biofuels.[3] Its natural tolerance toward the aromatic compounds from lignin deconstruction is partially attributed to a high-flux $\beta$-ketoadipate pathway that facilitates aromatic catabolism. The $\beta$-ketoadipate pathway converts aromatic compounds into acetyl-CoA and succinyl-CoA,[4] both of which enter central metabolism via the TCA cycle. High TCA cycle flux produces large amounts of ATP and NADH, and as a result, *R. opacus* can synthesize highly reduced products.[2,5]

Previous work on *R. opacus* has identified aromatic tolerance and utilization mechanisms based on transcript profile changes that do not cause large amounts of flux rewiring and that are not dependent on many genetic mutations. The transcriptome and fluxome of the wild type were examined when grown with sugars and model lignin monomers (i.e., aromatics) for a base strain as well as for adaptively evolved mutants.[5−7] A key finding is that the adaptive mutants could achieve optical densities ($OD_{600}$) up to 1900% higher than the wild-type strain when grown on high concentrations of aromatics, despite a limited number of mutations (∼12 single nucleotide polymorphisms on average) and limited flux rewiring.[5,7] The mutants, however, show big differences in their transcriptomic profiles when compared to the wild-type strain, which may account for their abilities to tolerate and utilize higher concentrations of aromatics.[6,7] In addition, the molecule-level mechanisms for aromatic substrate utilization and regulation have been elucidated.[8] Despite these advances in understanding the metabolism and gene regulation in *R. opacus*, a predictive genome-scale model derived from its genome has yet to be developed.

**Received:** November 16, 2022
**Published:** May 15, 2023

ACS Publications

© 2023 The Authors. Published by American Chemical Society

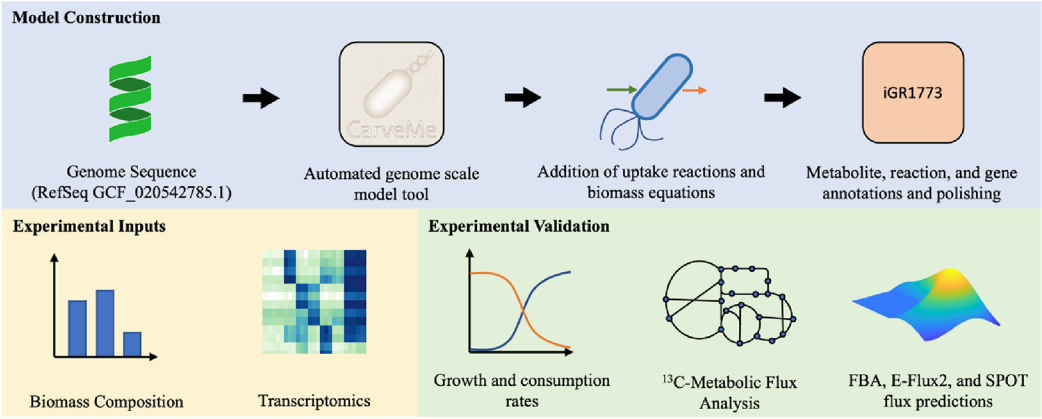



**Figure 1.** Reconstruction details and model validation. A draft version of the model was created through CarveMe, which was then augmented with relevant uptake and biomass reactions and then manually curated to yield the iGR1773 *R. opacus* GSM.

Genome-scale models (GSMs) are comprehensive mathematical summaries of the reactions encoded in an organism's genome. For example, flux balance analysis (FBA) uses GSM to optimize metabolic fluxes through mass balance constraints under the assumption that these fluxes maximize biomass production (i.e., produce the maximum growth rate).[9] The FBA method has been successful when modeling fast-growing lab-adapted species, but it is less accurate for organisms with slower growth rates.[10] Using data reflecting the internal state of the cell (e.g., omics data) is expected to improve the accuracy of flux predictions. In contrast to input and output flux measurements, omics data are not as straightforward to integrate.[11] A variety of Constraint-Based Reconstruction and Analysis (COBRA) methods that integrate omics data have been developed including iMAT,[12,13] GIMME,[14,15] E-Flux,[16] E-Flux2 and SPOT,[17] tFBA,[18] GX-FBA,[19] FCG,[20] and CORDA.[21] Such methods may be used to leverage high throughput transcriptomics data to improve model predictions.[22−25] There is, however, no 'best' method to guarantee the most accurate predictions under all circumstances, so care must be taken to identify differences, benefits, and drawbacks of each prediction method in order to apply the method that is most suited to a particular system.[26]

Here, we present and validate iGR1773, the first GSM for *R. opacus* derived from its genome, providing a comprehensive description of its internal metabolism and a valuable tool to integrate omics data into metabolic flux predictions. iGR1773 consists of 3025 reactions and 1956 metabolites obtained from annotating its completed genome,[27] adding the corresponding metabolic reactions, and testing the predictions derived by it. Although previous publications have reported a GSM for *R. opacus* PD630,[28,29] this model did not use an annotation of the *R. opacus* PD630 genome. This model repurposed the *Rhodococcus jostii* GSM[30] by doing some minor modifications including setting fluxes to polyhydroxyalkanoates (PHA), polyhydroxyvalerates (PHV), glycogen, and acetate to zero and adjusting the TAG reaction to reflect the fatty acid composition of *R. opacus* PD630.[28] Notably, iGR1773 was validated in three different ways: via the Metabolic Model Test (MEMOTE) suite,[31] by checking growth rate predictions, and by comparing flux predictions to $^{13}$C-metabolic flux analysis ($^{13}$C-MFA) results. Flux and growth rate predictions from the model were obtained through several COBRA methods, including parsimonious FBA (pFBA) and two methods that integrate transcriptomic data: E-Flux2 and SPOT. Briefly, E-Flux2 uses transcript measurements as upper and lower bounds for flux values, and SPOT finds the maximum correlation between transcript levels and reaction rates. These methods were chosen because the solutions they produce are non-degenerate, and they have been validated by previous studies.[17] We found that of the three COBRA methods, E-Flux2 provided the best predictions for growth rates and central carbon fluxes, providing, with iGR1773, an accurate predictive method for future *R. opacus* studies.

## 2. RESULTS AND DISCUSSION

iGR1773 was created through CarveMe[32] and manually curated by refining the reversibility of two reactions based on thermodynamics and adding transport reactions needed for ATP synthesis. MEMOTE[31] was used to ascertain the quality of the reconstruction, testing on par with state-of-the-art models. We tested iGR1773's predictive capabilities in two different ways: by comparing quantitative predictions of growth rates with experimentally measured growth rates and by comparing flux predictions with $^{13}$C-MFA measurements. Growth rate and flux predictions were obtained through three different methods: pFBA, EFlux-2, and SPOT. FBA works by providing the fluxes that maximize biomass production whereas pFBA adds an extra step, in which the sum of squared fluxes is minimized while the biomass production flux is held at its maximum. EFlux-2 and SPOT work differently: they do not assume maximum biomass production but constrain fluxes based on transcriptomic measurements. E-Flux2 determines fluxes by solving a transcript-adjusted FBA problem, and SPOT constrains fluxes by maximizing the correlation between fluxes and transcripts. Additionally, $^{13}$C-MFA and pFBA were used to determine that phenol and glucose metabolisms operate at roughly the same maintenance cost (i.e., similar amounts of ATP are lost to non-growth purposes per mmol of substrate consumed).

**2.1. Model Attributes and Refinement of Draft Reconstruction.** iGR1773 was generated from a recent genome annotation[27] and the genome-to-GSM tool CarveMe[32] (Figure 1). The draft model produced by CarveMe was accurate but required manual changes: two reactions needed to have their flux bounds adjusted to match known thermodynamic patterns. In the draft model, the succinate dehydrogenase reaction (EC 1.3.5.1; succinate + FAD ↔ fumarate + $FADH_2$) allowed flux only in the reverse direction. Based on $^{13}$C data demonstrating a complete TCA cycle in the forward direction,[5] this reaction was allowed to have forward and reverse flux. Additionally, the draft


https://doi.org/10.1021/acssynbio.2c00618
ACS Synth. Biol. 2023, 12, 1632−1644

model contained a thermodynamically infeasible cycle that allowed the model to produce unrealistic amounts of ATP. This flaw was traced to two versions of 3-hydroxyadipyl-CoA dehydrogenase (EC 1.1.1.35): one version of the reaction was 3-oxoacyl-CoA + NADH + $H^+$ ↔ 3-hydroxyacyl-CoA + $NAD^+$ and the other version was 3-hydroxyacyl-CoA → 3-oxoacyl-CoA + $H_2$. When combined, this reaction pair has the net effect of converting NADH and $H^+$ into $H_2$ and $NAD^+$. The resultant $H_2$ could then be used to pump $H^+$ into the periplasm by a hydrogenase reaction (EC 1.12.5.1; $H_2$ + $2H^+_{cytosolic}$ + menaquinone → $2H^+_{perisplasm}$ + menaquinol), with subsequent periplasmic $H^+$ used to drive ATP synthase to produce an unrealistic quantity of ATP. The reaction of 3-hydroxyacyl-CoA → 3-oxoacyl-CoA + $H_2$ was blocked to prevent this loop from generating ATP. Four reactions were added to the draft model to allow hydrogen ions travel to the periplasm to drive ATP synthase flux. These reactions included cytochrome b6/f complex periplasm, active co2 transporter facilitator (periplasm), cytochrome c oxidase, and cytochrome oxidase bd. These reactions allow reduced energy-carrying molecules, like plastoquinol and ferrocytochrome, to participate in moving hydrogen ions to the periplasm. After these manual changes, the finalized model contained 3025 reactions and 1956 metabolites (Table 1).

**Table 1. iGR1773 Model Statistics**

| | |
|---|---|
| Genes | |
| total genes | 1773 |
| Reactions | |
| total reactions | 3025 |
| transport reactions | 824 |
| purely metabolic reactions | 1862 |
| Metabolites | |
| total metabolites | 1956 |
| Model Properties | |
| metabolic coverage | 1.71 |
| degrees of freedom | 847 |
| compartments | 3 |

**2.2. Model Evaluation through MEMOTE.** The *R. opacus* GSM was evaluated with MEMOTE,[31] producing a score commensurate with the best in the field. MEMOTE addresses the problem of assessing the quality of GSMs, given their complexity (GSMs often include thousands of metabolites and reactions that are assigned to subcellular locations). Adequate model quality tests are critical because mass balance or stoichiometric errors can render erroneous model predictions. The annotated and curated model was determined to have 100% stoichiometric consistency, 100% mass balance, and 100% metabolite connectivity. The annotation scores consist of 79% for metabolites, 77% for reactions, 33% for genes, and 100% for SBO (systems biology ontology). MEMOTE scores are designed to reflect the average completeness of annotations across databases since there are multiple databases for genome-scale model data (e.g., BiGG and KEGG). For each category (e.g., metabolites, reactions, and genes), a score is calculated for each database as a percentage of the category members that contain an annotation corresponding to that database. The overall MEMOTE score for the category is calculated by averaging the database-specific annotation scores. The overall score for the model was 91%. As a reference, a recent *E. coli* GSM, iML1515, has an overall MEMOTE score of 91%.[33]

**2.3. Experimental Calculation of Growth Parameters.** *R. opacus* grown in glucose showed a significantly higher substrate uptake rate ($P < 0.001$, two-tailed Student's *t* test) and yield than when it was grown in phenol ($P < 0.001$, two-tailed Student's *t* test). Paired sets of time course growth and consumption curves were used to determine the growth rate, yield coefficient, and substrate uptake rate of wild type *R. opacus* when grown on phenol or glucose, and for an adapted mutant strain, PVHG6, when grown on phenol. The fitted parameters (Table 2) were confirmed by plotting fitted growth and consumption curves against measured data (Figures S1−S3). The higher uptake rate and yield contributed to the faster growth rate of *R. opacus* in glucose than in phenol. The aromatic adapted strain, PVHG6, had a faster growth rate in phenol than the wild-type strain ($P = 0.002$, two-tailed Student's *t* test). The mutant was developed through ∼30 passages of *R. opacus* grown on a mixture of aromatic substrates including phenol. This process selected for mutations that increased growth rate, so the observed difference between WT and PVHG6 was expected. While the mutant's growth rate in phenol was higher than that of WT, the biomass yield showed no difference between the two strains ($P = 0.09$, two-tailed Student's *t* test).

**Table 2. Fitted Growth Parameters for Wild-Type (WT) and Aromatic-Adapted (PVHG6) Strains[a]**

| | growth rate | yield coefficient | substrate uptake rate |
|---|---|---|---|
| WT phenol | 0.065 ± 0.001 | 0.048 ± 0.005 | 1.4 ± 0.2 |
| PVHG6 phenol | 0.080 ± 0.003 | 0.040 ± 0.002 | 2.0 ± 0.1 |
| WT glucose | 0.260 ± 0.005 | 0.073 ± 0.004 | 3.6 ± 0.2 |

[a]Growth rate has units of $h^{-1}$, yield coefficient has units of g biomass/mmol substrate, and substrate uptake rate has units of mmol substrate/g biomass/h. All values are averages and standard deviations of three biological replicates.

**2.4. Growth Rate Predictions.** iGR1773 predicted growth rates in an acceptable, but by no means perfect, manner (Figure 2). The method that provided the most accurate predictions was E-Flux2, with SPOT generating the least accurate predictions. pFBA produced predictions that were somewhat less accurate than those provided by E-Flux2. The fact that enzyme constraints increase the accuracy of growth rate predictions over unbounded pFBA is consistent with recent reports from *Saccharomyces cerevisiae* genome-scale modeling.[34,35] Growth rates under phenol were lower and better-predicted than those under glucose.

E-Flux2 made the most accurate growth rate predictions, while the other methods either displayed larger errors (pFBA) or completely failed (SPOT) (Figure 2). It is not surprising to see SPOT predict null growth rates since it is based on maximizing the correlation between fluxes and transcripts and not maximizing growth. pFBA and E-Flux2 both typically predict faster growth rates than those that have been measured experimentally. pFBA is expected to overestimate growth rates by aiming to predict the maximum theoretical growth rate. We would expect that the actual growth rate would be less than the theoretical maximum due to other factors. For example, soil bacteria such as *R. opacus* need to consume many carbon sources, and maintaining this ability imposes a cost on the growth rate for any one carbon source. Additionally, pFBA seeks out the most efficient use of carbon resources and does not


https://doi.org/10.1021/acssynbio.2c00618
ACS Synth. Biol. 2023, 12, 1632−1644

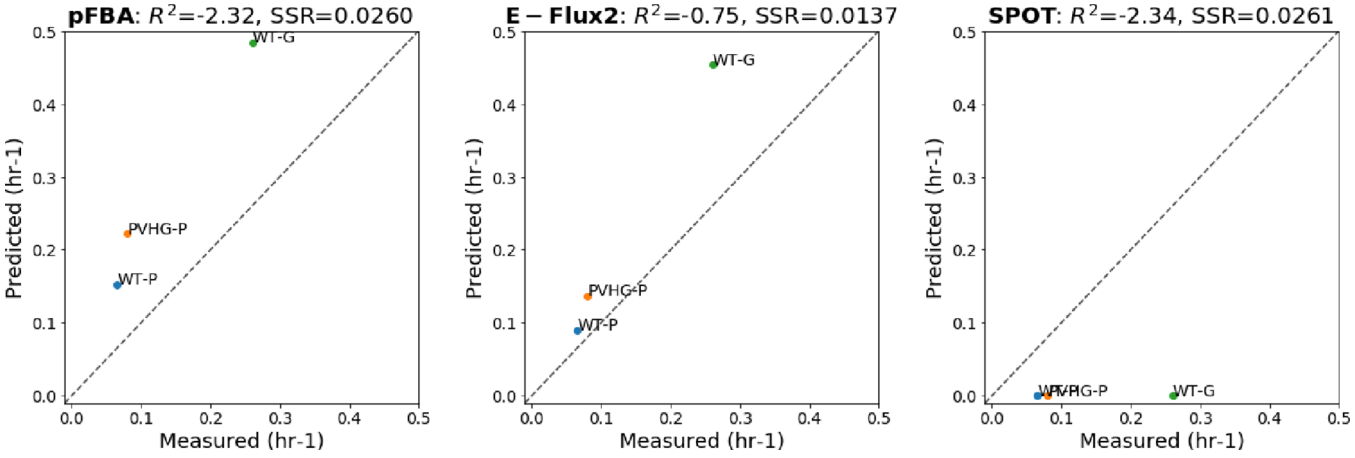



**Figure 2.** Growth rate predictions. Growth rate predictions are acceptable, but not perfect. Comparison of observed growth rates and model predicted growth rates for wild type consuming glucose (WT-G), wild type consuming phenol (WT-P), and aromatic-adapted strain consuming phenol (PVHG-P). SPOT completely fails. The points represent growth rates with units ($h^{-1}$). SSR = sum of squared residuals.

factor in competing interests, including the cost to make the enzymes. Since enzyme cost is not included in pFBA calculations, pathways with high carbon efficiency are preferred even though these pathways may have low *in vivo* flux due to the overall resource cost in producing the corresponding enzymes.[36]

Growth rates under phenol were lower, and better predicted, than growth rates under glucose. Typically, carbon sources that are consumed through the TCA cycle (e.g., acetate, succinate, and fumarate) result in lower growth rates than for growth on sugars since TCA cycle metabolites are generally more oxidized than sugars. Additionally, when TCA cycle metabolites are used as sole carbon sources, gluconeogenesis is required to produce amino acid precursors. Unlike glycolysis, which produces energy molecules, gluconeogenesis consumes ATP and NADH. Furthermore, phenol is a toxic substance, which imposes an additional metabolic burden via stress response.

A possible explanation for why the growth rate predictions are better for phenol than for glucose is that there is only one catabolic pathway for phenol while there are multiple options for glucose. Specifically, phenol degradation into TCA cycle metabolites has low degrees of freedom. Conversely, there are multiple pathways for glucose catabolism, including glycolysis (EMP), Entner−Doudoroff (ED), and pentose phosphate pathways. These pathways can be flexibly regulated and are underdetermined by pFBA.

**2.5. Comparison of Model Predictions and $^{13}$C-MFA Fluxes.** When compared to fluxes measured by $^{13}$C-MFA, the flux predictions from the COBRA methods were more accurate for phenol metabolism than for glucose metabolism. Among the COBRA methods we tried, EFlux-2 provided the best predictions, whereas SPOT provided the worst predictions for the phenol uptake case but the second best for glucose. pFBA provided the same results as FBA, which were very good for phenol but not very accurate for glucose. The comparison of predicted fluxes with $^{13}$C-MFA flux measurements is the most rigorous test of GSM and COBRA methods since $^{13}$C-MFA measurements are the gold standard for quantifying intracellular reaction rates,[37] and they provide detailed information about central metabolism instead of aggregated measurements (e.g., just growth rate). $^{13}$C-MFA, however, is an expensive procedure to carry out.[38] Thus, it typically provides fewer conditions for comparison than grow/no grow tests or growth rates. However, the reduction in conditions is more than compensated for by the increased metabolic resolution.

*2.5.1. Comparison of Phenol Flux Predictions and $^{13}$C-MFA Fluxes.* $^{13}$C-MFA of phenol metabolism was obtained from a previous publication.[5] The glucose $^{13}$C-MFA data was obtained following the same procedure as discussed in that publication. The transcriptomics data and growth curves for phenol came from Henson et al.[7] The glucose growth curves and consumption data are new in this work, and they were generated from cultures grown under the same conditions as the Henson et al. data (except for the carbon source). For comparisons with $^{13}$C-MFA data, the carbon source uptake rates for pFBA, E-Flux2, and SPOT were normalized to 100 units (instead of the experimentally determined mmol substrate/g biomass/h), in accordance with $^{13}$C-MFA convention.

For the phenol case, intracellular fluxes were accurately predicted by the COBRA methods (Figures 3 and 4). Fluxes predicted by E-Flux2 were very close to the fluxes measured through $^{13}$C-MFA ($R^2$ = 0.96 without considering ATP maintenance). pFBA predicted fluxes that were slightly less accurate than those predicted by E-Flux2 ($R^2$ = 0.93). Though minor compared to the other methods, the largest divergences between E-flux2 predictions and $^{13}$C-MFA measurements were found in anaplerotic reactions and transport reactions. For pFBA, the trend continued with the largest divergences coming from anaplerotic reactions and transport reactions. The prediction errors for anaplerotic reactions may be a result of their underdetermined nature in $^{13}$C-MFA due to reactions with matching labeling patterns. pFBA and E-Flux2 both under-predicted the flux of $CO_2$ out of the cell, which is a direct consequence of growth rate overprediction. Since these methods assume complete carbon efficiency to maximize biomass, it is expected that they would underestimate the amount of carbon lost as $CO_2$.

For phenol, SPOT's predictions were the least accurate ($R^2$ = 0.66). Despite the decent $R^2$ value, a closer analysis of SPOT's predictions, guided by biochemical knowledge, shows that it generates an unrealistic metabolic flux profile. Most notably, SPOT critically underestimates TCA cycle fluxes, especially with respect to isocitrate dehydrogenase, $\alpha$-ketoglutarate dehydrogenase, and succinyl-CoA synthetase. When phenol uptake was normalized to 100 units, each of these reactions had errors over 100 units. Particularly noteworthy are isocitrate dehydrogenase


https://doi.org/10.1021/acssynbio.2c00618
ACS Synth. Biol. 2023, 12, 1632−1644

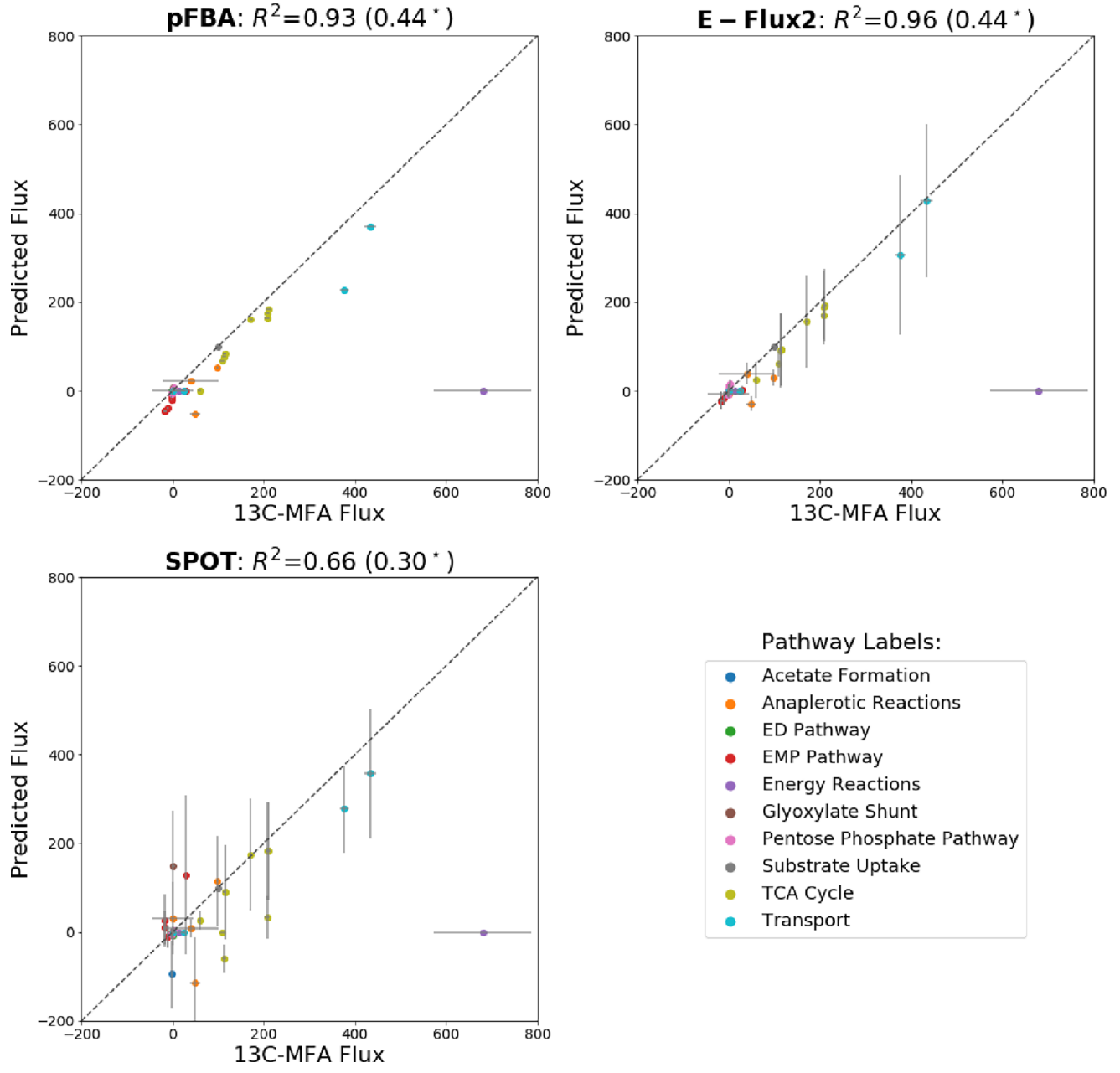



**Figure 3.** Flux predictions for phenol metabolism. Predictions are accurate for the three COBRA methods. The *y* axis represents the predicted flux by each of the COBRA methods (pFBA, E-Flux2, and SPOT) and the *x* axis represents the flux measured via $^{13}$C-MFA. The fluxes are normalized to the carbon source uptake (units are mmol reaction/100 mmol phenol uptake). The first $R^2$ value does not include ATP maintenance reaction and the $R^2$ value in parentheses includes the ATP maintenance reaction. The *x* axis error bars are 90% confidence intervals as determined via $^{13}$C-MFA, and if applicable, the *y* axis error bars are standard deviations of flux predictions made from three biological replicates of transcriptomics data.

and $\alpha$-ketoglutarate dehydrogenase, which were predicted to have negative and zero flux, respectively (Table S1). To compensate for the underpredictions of the TCA cycle reactions, the flux through the glyoxylate shunt was over-predicted. SPOT predicted the flux of isocitrate lyase to be ~150 units, while the $^{13}$C-MFA determined its flux to be only 0.4 (Figure 4). This discrepancy casts doubt on the viability of SPOT as a widely applicable standalone method for predicting fluxes from transcript data.

E-Flux2 and SPOT were also applied to phenol metabolism in the PVHG6 strain. Since pFBA does not take transcript measurements into account, its predictions are the same for the wild type and mutant strains. Overall, the transcript profiles of the two strains on phenol were very similar,[7] so it was expected that the mutant strain flux predictions from EFlux-2 and SPOT would be similar to the wild type predictions. Indeed, EFlux-2 makes accurate flux predictions for phenol metabolism in the mutant strain (wild type EFlux-2, $R^2$ = 0.96; mutant EFlux-2, $R^2$ = 0.95; Figure S4). Interestingly, despite similar transcriptomics measurements, SPOT's predictions of fluxes in the mutant strain are different from the wild type (Table S2) (wild type SPOT, $R^2$ = 0.66; mutant SPOT, $R^2$ = 0.39 Figure S4). The greater difference of SPOT's predictions between the strains compared to E-Flux2 demonstrates that E-Flux2 is more robust to small changes in transcript values than SPOT. As in the


https://doi.org/10.1021/acssynbio.2c00618
ACS Synth. Biol. 2023, 12, 1632−1644

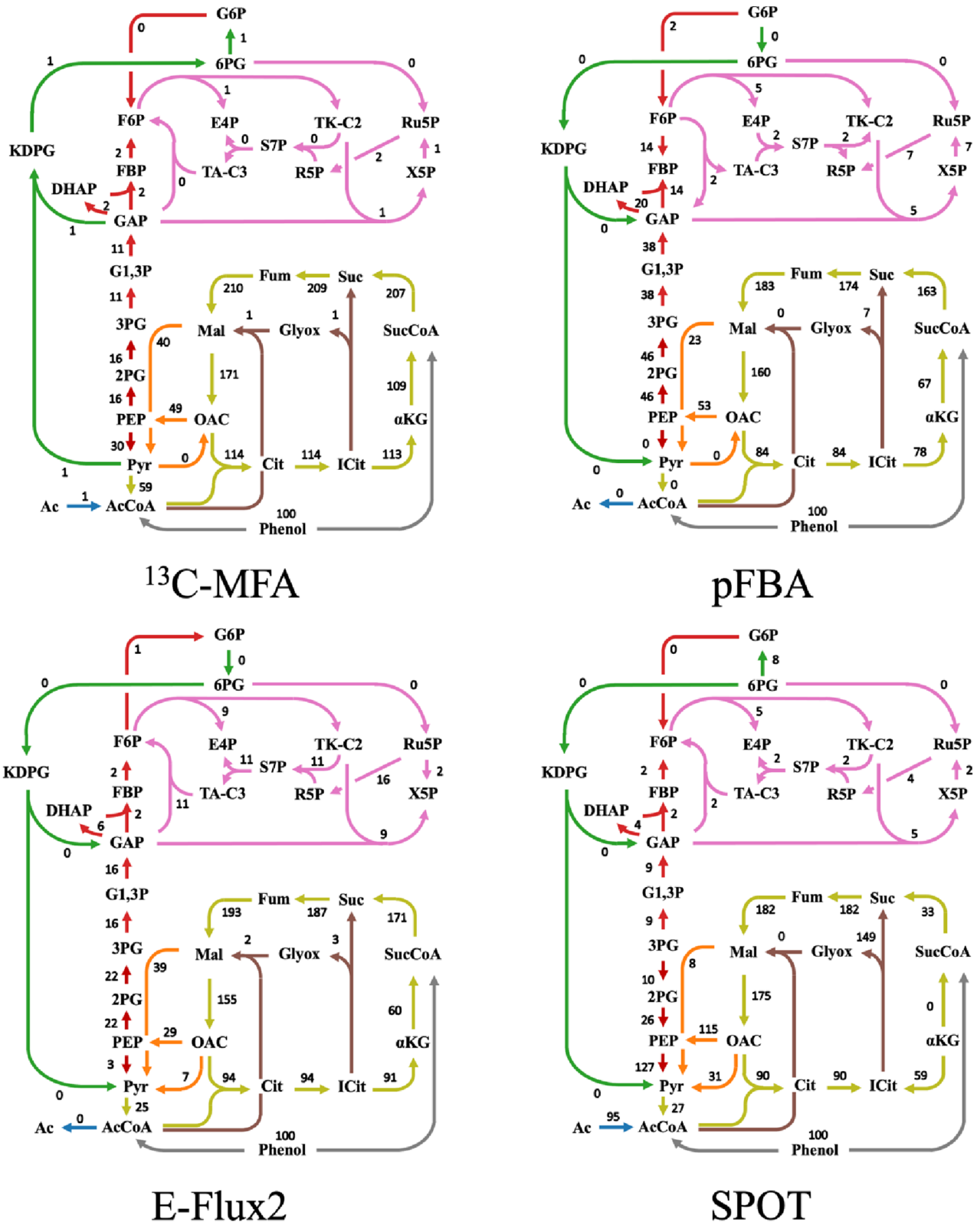



**Figure 4.** Phenol flux maps. Flux map predictions when phenol is the sole carbon source. The flux values are relative flux distributions based on 100 mmol of phenol consumed by the cell to generate 100 mmol of influx toward both acetyl-CoA and succinyl-CoA. A mapping of abbreviations to metabolite names is given in Table S8.

wild type's phenol condition, the largest errors in SPOT's mutant predictions occurred in the TCA cycle (Figure S5).

*2.5.2. Comparison of Glucose Flux Predictions and $^{13}$C-MFA Fluxes.* In the case of glucose, each of the three predictive methods show limitations (Figure 5). As observed with the phenol condition, E-Flux2 had the best predictions, though in this case, its predictions only fit moderately well ($R^2$ = 0.63). SPOT's predictions had the second best fit for glucose ($R^2$ =



https://doi.org/10.1021/acssynbio.2c00618
ACS Synth. Biol. 2023, 12, 1632−1644

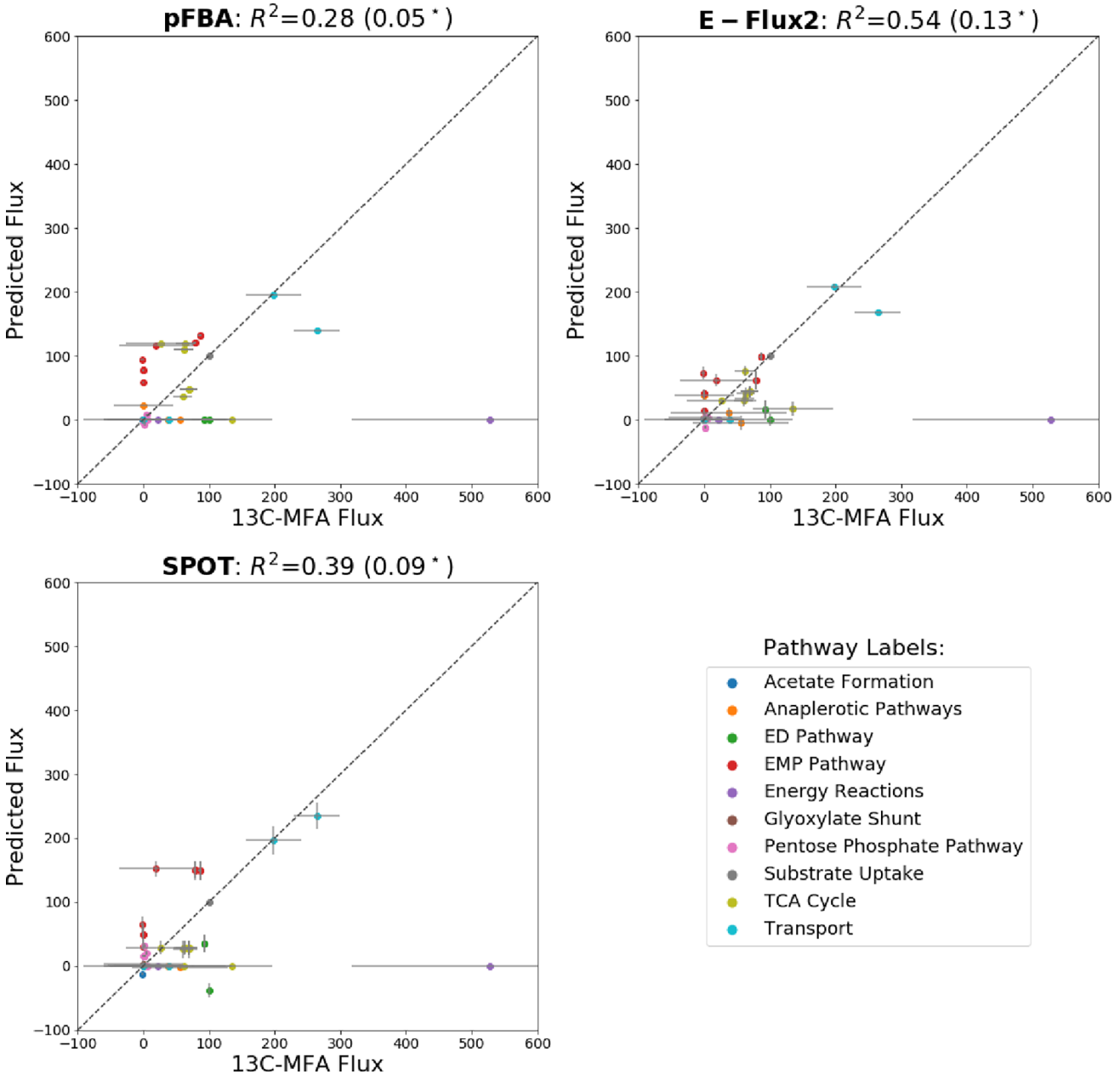



**Figure 5.** Glucose metabolism flux predictions. Glucose metabolism flux predictions are much less accurate for the COBRA methods considered here. Comparison of $^{13}$C-MFA fluxes with model-predicted fluxes for glucose metabolism in the wild type strain. Horizontal and vertical axes and error bars are as described in Figure 3. In the same way, the $R^2$ value in parentheses is the $R^2$ value when ATP maintenance is included in the calculation.

0.45), and pFBA's predictions were largely inaccurate ($R^2$ = 0.28) (Figure 5). One major difference between the three methods occurred in the predictions for the glucose uptake pathways. Two of these pathways, the EMP pathway and the ED pathway, share the enzymes that connect glyceraldehyde-3-phosphate to pyruvate but differ in their initial enzymes. Between the two, *R. opacus* shows a strong preference for the ED pathway, with approximately 95% of glucose consumed via this pathway despite a complete EMP pathway also being present.[39] While the two run essentially in parallel, this stark disparity is nonetheless unexpected, as the EMP pathway produces an extra molecule of ATP per molecule of glucose metabolized.[36] Potentially, the enzyme efficiency of the ED pathway explains this preference. Predictably, while $^{13}$C-MFA determined that 93% of glucose was consumed through the ED pathway, pFBA predicted that the ED pathway would have zero flux because creating extra ATP helps facilitate reactions including the biomass production reaction. Interestingly, the methods that incorporate transcriptomics into the genome-scale model recapitulate some ED flux. E-Flux2 and SPOT predict 21% and 38% of glucose consumption to occur via the ED pathway, respectively (Figure 6). These non-zero ED flux values contribute to the increased accuracy of the transcriptomics-based methods over FBA-based methods.

Similar to the predictions made for phenol growth conditions, pFBA predicted TCA cycle fluxes of glucose metabolism with less accuracy than E-Flux2 and SPOT. pFBA overestimated the fluxes of $\alpha$-ketoglutarate dehydrogenase, succinate dehydrogenase, fumarase, and malate dehydrogenase (Table S3). All of these enzymes, except fumarase, produce reducing equivalents


https://doi.org/10.1021/acssynbio.2c00618
ACS Synth. Biol. 2023, 12, 1632−1644

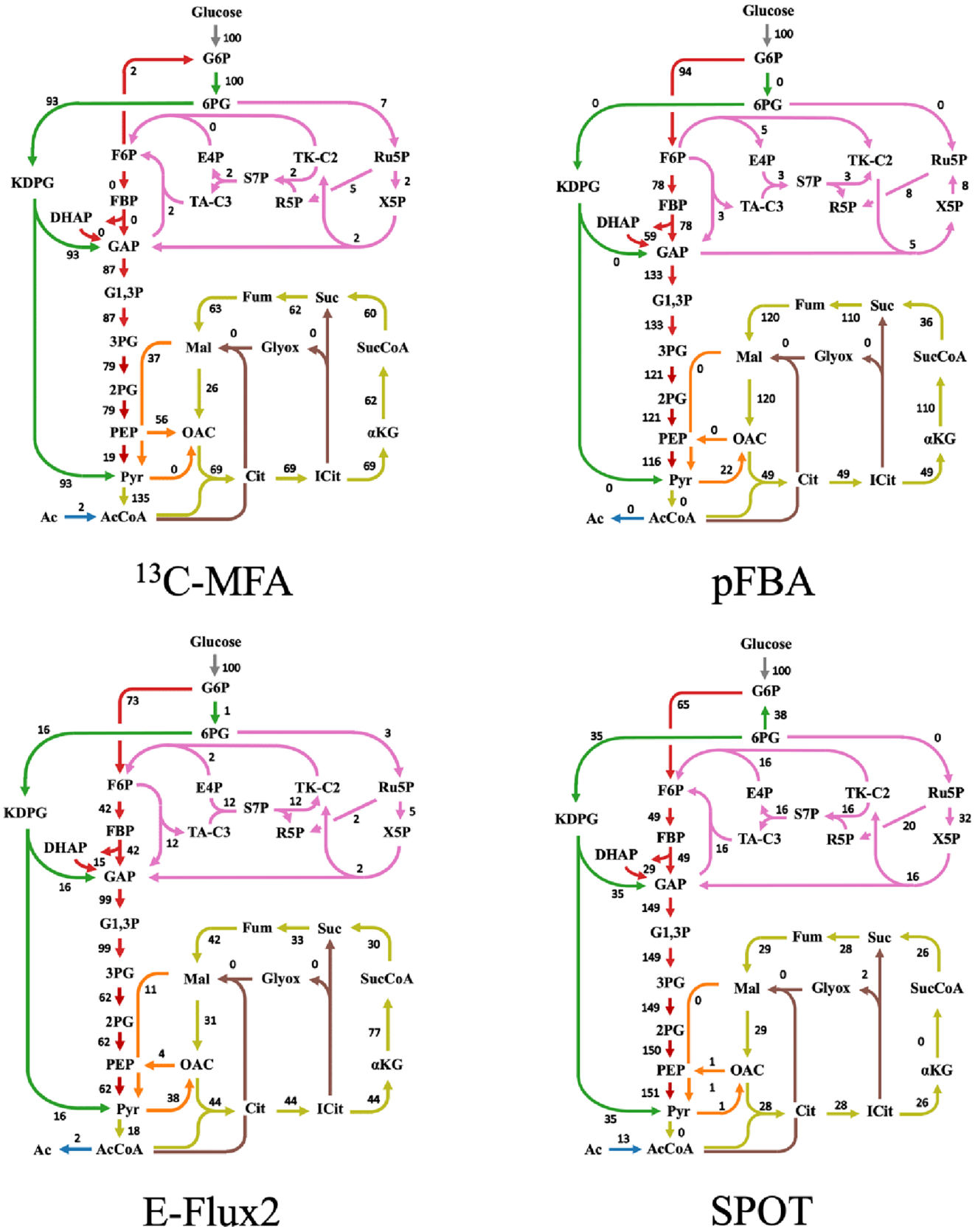



**Figure 6.** Glucose flux maps. Flux map predictions when glucose is the sole carbon source. The flux values are relative flux distributions based on 100 mmol of phenol consumed by the cell to generate 100 mmol of influx toward both acetyl-CoA and succinyl-CoA.

in the form of NADH or $FADH_2$. FBA and pFBA's overprediction of these TCA cycle reactions results in additional energy molecules and carbon losses.

**2.6. ATP Maintenance Flux Upper Bound Estimates.** Multiple methods for determining the non-growth associated ATP maintenance flux (NGAM) show that glucose metabolism



https://doi.org/10.1021/acssynbio.2c00618
ACS Synth. Biol. 2023, 12, 1632−1644

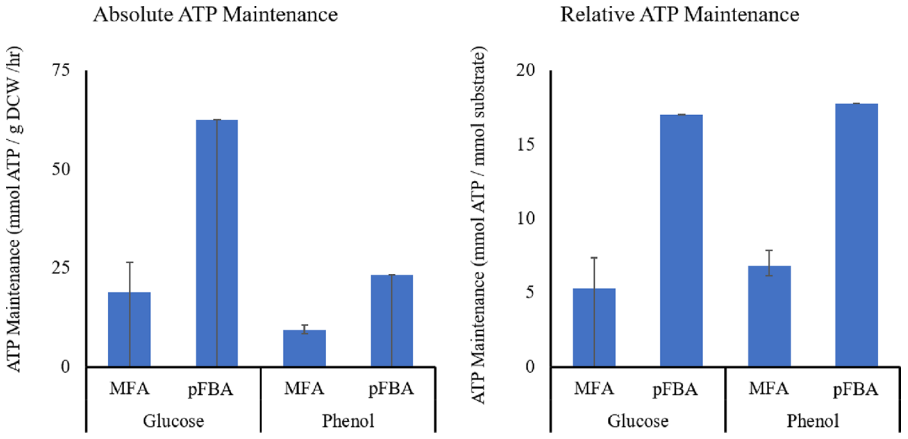



**Figure 7.** ATP maintenance flux as determined by metabolic flux analysis (MFA) and flux balance analysis (pFBA). Absolute ATP maintenance is the mmol of ATP used by 1 g of dry cell weight per hour, and relative ATP maintenance is the mmol of ATP used per mmol of either glucose or phenol consumed.

and phenol metabolism function with similar efficiency (i.e., relative ATP used for maintenance). NGAM is the amount of ATP generated in a metabolic model that is not consumed by the reactions in the model. It is thought that this excess ATP is used for cellular "housekeeping" tasks such as maintaining ionic gradients and producing enzymes via transcription and translation.[40] A cell is considered to be operating at higher efficiency when its ATP maintenance flux is low as less ATP is "lost" to non-growth purposes.

The GSM calculated non-growth associated ATP maintenance flux via FBA. When ATP maintenance loss is high, less biomass can be produced because ATP (growth associated) is a reactant in the biomass equation, and ATP is a required cofactor for many reactions that produce biomass precursors. The flux configuration with the maximum growth rate has zero ATP maintenance flux, and the flux configuration with the maximum ATP maintenance flux has zero biomass production (Figure S6). The true ATP maintenance loss can be estimated by mapping the experimental growth rate onto the ATP maintenance flux vs growth rate curve. This method gives the same result as fixing the growth rate and then calculating the maximum ATP maintenance flux (using fixed growth associated ATP maintenance). Using this method, the model predicts that the non-growth associated ATP maintenance flux was 23.4 mmol ATP per gram dry cell weight per hour when consuming phenol and 63.0 mmol ATP per gram dry cell weight per hour when consuming glucose (Figure 7). In $^{13}$C-MFA, ATP maintenance flux is a fitted variable constrained by amino acid labeling patterns. The $^{13}$C-MFA ATP maintenance flux was 9.2 mmol ATP per gram dry cell weight per hour when consuming phenol and 18.9 mmol ATP per gram dry cell weight per hour when consuming glucose (Figure 7).

The ATP maintenance flux calculated using FBA is roughly three times greater than the value determined by $^{13}$C-MFA (Figure 7), a discrepancy that can be traced to FBA's fundamental assumption that cells are optimized to maximize biomass production. As described above, FBA was used to estimate the ATP maintenance flux by fixing the model's growth rate to the experimental growth rate and then maximizing the amount of ATP maintenance flux. As a result, the ATP maintenance value predicted by FBA represents the upper bound of possible ATP maintenance values in the same way that FBA's growth rate predictions represent the theoretical maximum growth rate. Interestingly, while glucose had a higher absolute ATP maintenance flux per hour than phenol, when the data was normalized per mmol of substrate uptake, this difference was largely eliminated. This indicates that per mole of substrate, both conditions use roughly the same amount of ATP for non-growth activities despite the difference in uptake rates.

## 3. CONCLUSIONS

In this article, we present a GSM for *R. opacus* PD630: iGR1773. This model provides a tool for predicting this organism's metabolism and can help fulfill its potential as a platform for converting lignin derivatives into liquid fuels and chemicals. iGR1773 was validated with the Metabolic Model Test (MEMOTE) suite,[31] by checking growth rate predictions, and through comparisons of flux predictions via COBRA methods to $^{13}$C-MFA measurements. The COBRA method that provided the most accurate predictions was E-Flux2 followed by pFBA and SPOT. In general, the COBRA methods were more accurate for phenol than for glucose. Additionally, the model was used to demonstrate that *R. opacus*' metabolic network operates with similar efficiency when consuming phenol or glucose. We expect this GSM to be a stepping-stone toward building progressively more predictive models of *R. opacus* metabolism that will guide future metabolic engineering efforts.

## 4. MATERIALS AND METHODS

**4.1. Strains and Data.** The data used in this manuscript originated either in previous publications[5,7] or are newly reported in this work (Table 3). The experiments in this work used *Rhodococcus opacus* PD630 (DSMZ 44193) as the wild-

**Table 3. Sources of the Experimental Data Used in This Paper**

| | phenol (wild type and PVHG6) | glucose (wild type) |
|---|---|---|
| transcript data | Henson et al. (2018)[7] | Henson et al. (2018)[7] |
| growth curves | Henson et al. (2018)[7] | first published in this paper |
| substrate consumption curves | Henson et al. (2018)[7] | first published in this paper |
| $^{13}$C-metabolic flux analysis | Roell et al. (2019)[5] | first published in this paper |
| biomass composition | first published in this paper | first published in this paper |


https://doi.org/10.1021/acssynbio.2c00618
ACS Synth. Biol. 2023, 12, 1632−1644

type strain and a *Rhodococcus opacus* PD630 mutant strain PVHG6, which had previously been adaptively evolved on a mixture of phenol, vanillate, guaiacol, 4-hydroxybenzoate, and guaiacol.[7] All data was generated from fermentation experiments wherein *R. opacus* was cultured in minimal media B with either phenol or glucose as the sole carbon source and 1 g/L ammonium sulfate as the nitrogen source.[41] The transcript data used in this analysis comes from a previous publication[7] stored in the National Center for Bioinformatics Sequence Read Archive in bioproject PRJNA431604, and the data was reprocessed to count per million (CPM) normalization. The growth curve data for phenol conditions, $OD_{600}$, and substrate consumption data, were from a previous report,[7] while the glucose data was generated in this work. The $^{13}$C-MFA data for phenol was previously reported,[5] and the glucose data was obtained using the same procedure as described therein. The biomass composition data for both phenol and glucose was obtained using a custom spectrophotometry method described in Section 4.4.

**4.2. Draft Model Reconstruction and Gap Filling.** The initial version of the GSM for *R. opacus* was made using CarveMe, an automated tool developed to produce GSMs.[32] For this reconstruction, the following versions were used: CarveMe 1.5.1, Diamond 0.9.14, and CPLEX 12.10.0.0. CarveMe follows a top-down approach where a universal model and genome sequence are the only required inputs to construct a model in a fast and reproducible manner. The GSM was based on a recent genomic sequence of the *Rhodococcus opacus* PD630 (Refseq ID: GCF_020542785.1).[27] The initial model was made using the command line command "carve r_opacus_bologna.faa -u grampos -o r_opacus_bologna_raw.xml". After the model generation, this initial draft model was also gap-filled to ensure growth on M9 and LB media using the command "gapfill r_opacus_bologna.xml -m M9,LB -o r_opacus_bologna_gapfilled.xml".

**4.3. Addition of Uptake Reactions (Notebook A).** As generated by CarveMe, the GSM did not contain uptake reactions for all the carbon sources *R. opacus* can metabolize, so these reactions were added in notebook A. This initial model contained all the reactions needed for the model to consume several carbon sources including glucose, 4-hydroxybenzoate, vanillate, and benzoate. To account for growth with phenol, the metabolites for extracellular and intracellular phenol were added as well as the reactions for phenol exchange (adding phenol to the medium), phenol transport (phenol entering the cell), and phenol monooxygenase (phenol + NADH + $O_2$ + $H^+$ → catechol + $NAD^+$ + $H_2O$; EC 1.14.13.244). For growth with guaiacol, intracellular and extracellular guaiacol were added, and so were reactions for exchange, transport, and guaiacol-*o*-demethylase (guaiacol + NADPH + $O_2$ → catechol + formaldehyde + $NADP^+$ + $H_2O$; EC 1.14.14.-). Additionally, an intracellular metabolite for triacylglycerol (TAG) and reactions for its production from 1,2-diacyl-*sn*-glycerol and palmitoyl-CoA and transport out of the cell were added to the model. The bounds of two reactions, catalyzed by 3-hydroxyadipyl-CoA dehydrogenase and succinate dehydrogenase, were adjusted to avoid thermodynamically infeasible cycles. This notebook also contains tests to ensure that the model can explain the growth in glucose, phenol, vanillate, 4-hydroxybenzoate, guaiacol, and benzoate. In addition to these aromatic carbon sources, *R. opacus* PD630 has also been shown to be able to consume mannitol, ribose, xylose, lactose, and maltose as sole carbon sources according to the BacDive page for DSMZ 44193. The model from CarveMe was able to consume all these carbon sources without the need for manual edits.

**4.4. Addition of Custom Biomass Reactions (Notebook B).** The biomass composition of *R. opacus* when grown with various substrates was quantified in terms of carbohydrate, lipid, and protein fractions. Carbohydrates were measured using a hydrolysis procedure. Lipid extraction, purification, and measurement were conducted using the Bligh and Dyer method.[42] Proteins were measured with an L-8800 AAA Hitachi High-Speed Amino Acid Analyzer. These measurements are summarized in Table S4.

The biomass composition data and previously reported amino acid data[5] were used to make customized biomass equations for the *R. opacus* GSM when grown in glucose or phenol. These biomass equations were based on the *Bacillus subtilis* biomass equation that comes by default with CarveMe for Gram-positive bacteria.[43] In the customized *R. opacus* biomass equations, the coefficients for precursors that are not amino acids, lipids, or carbohydrates (e.g., energy molecules and salts) are the same as they are in the *B. subtilis* biomass equation. The coefficients of lipid and carbohydrate precursors were scaled proportionally to the measured amount of lipids or carbohydrates in *R. opacus*. The amino acid coefficients were calculated using the measured milligrams of amino acids per gram of biomass and the measured mole percentage of each amino acid. Table S5 contains a comparison of the biomass equations for *R. opacus* with phenol, *R. opacus* with glucose, and *B. subtilis*.

**4.5. Addition of Metabolite, Reaction, and Gene Annotations (Notebook C).** The reconstruction from CarveMe included detailed metabolite and reaction annotations. The only metabolites in the *R. opacus* model that were not included in the BiGG Universal model were guaiacol and triacylglycerol.[9] All but 25 of the reactions in the *R. opacus* model were found in the Universal model, so these reactions were left unannotated. The model's gene IDs are the NCBI non-redundant protein accession numbers (with the prefix 'WP_') from the NCBI database (Refseq ID: GCF_020542785.1).[44] The proper system biology ontology (SBO[45]) numbers were also added to all metabolites, reactions, and genes. Further, since the annotations in the Universal model are the Python type, List, they were converted into dictionaries with keys to match MEMOTE's requirements.

**4.6. Experimental Determination of Growth Rate and Substrate Uptake Rate (Notebook D).** Experimental growth rates were calculated by first collecting time-course $OD_{600}$ data from fermentations with 5 mM phenol or glucose as the carbon source and 1 g/L ammonium sulfate as the nitrogen source. The growth rate was calculated using the slope of the log-transformed OD vs time regression since the growth in the exponential phase follows the equation $X(t) = X_0 e^{\mu t}$, where $X(t)$ represents the OD at time $t$, $X_0$ is the initial OD, $\mu$ is the growth rate in $h^{-1}$, and $t$ is the time in hours. The yield coefficient (g biomass/mmol substrate) was determined using the slope of the line made when plotting the amount of substrate consumed vs the amount of biomass produced. The substrate consumption rate (mmol substrate/g biomass/h) was calculated by dividing the growth rate ($h^{-1}$) by the yield coefficient (g biomass/mmol substrate). For each of the three conditions (wild-type phenol, wild-type glucose, and PVHG6 phenol), there were three biological replicates of growth and consumption data. The growth parameters were calculated individually for each trial and then averaged for each condition (Table 2).


https://doi.org/10.1021/acssynbio.2c00618
ACS Synth. Biol. 2023, 12, 1632−1644

**4.7. Growth Rate Simulations.** The *R. opacus* GSM was used to make growth rate predictions. While GSMs are stoichiometric models without a time component, when the input and output reactions are properly scaled, these models can be used to predict growth rates.[46] The model was calibrated to simulate the behavior of 1 g of dry cell weight for 1 h. The substrate uptake rate was set to the amount of substrate, in mmol, that 1 g of biomass would consume in 1 h, and the biomass formation reaction was set up so that its flux would equal the amount of biomass in grams produced in 1 h. Growth rate ($\mu$) is defined according to the equation $dX/dt = \mu X$, where $dX/dt$ is the rate of change of biomass and $X$ is the biomass concentration. Translating to the GSM, $dX/dt$ is equal to the biomass flux, and since the model was scaled for 1 g of biomass ($X = 1$), the biomass flux is equal to the growth rate.

**4.8. Comparison with $^{13}$C-MFA.** Another approach for validating the GSM is to compare its flux predictions with fluxes determined using $^{13}$C-MFA. Since the $^{13}$C-MFA metabolic network contains ~70 reactions and the iGR1773 GSM contains ~2300 reactions, reactions from the two cannot be directly compared. A mapping of reactions from the GSM to the $^{13}$C-MFA reactions was made to compare genome-scale flux predictions and $^{13}$C-MFA measurements (Table S6). Some reactions in the $^{13}$C-MFA model involve multiple reactions in the GSM. This can happen when two reactions occur in series or when they occur in parallel. An example of reactions in series is the conversion of 3-phosphoglycerate to phosphoenolpyruvate. In the GSM, 3-phosphoglycerate is converted to 2-phosphoglycerate and then to phosphoenolpyruvate, while in the $^{13}$C-MFA, 3-phosphoglycerate is directly converted to phosphoenolpyruvate. The minimum flux value of reactions in series was compared to $^{13}$C-MFA flux. Additionally, some reactions in the $^{13}$C-MFA have multiple reactions that act in parallel in the GSM. An example is malate dehydrogenase. In the $^{13}$C-MFA, there is only a single isozyme (that produces NADH), while in the GSM, there are isozymes that produce NADH, menaquinone, and ubiquinone. The sum of fluxes of parallel reactions was compared to $^{13}$C-MFA flux. The quality of GSM fit was determined by calculating the $R^2$ (coefficient of determination) between the GSM model fluxes and the $^{13}$C-MFA fluxes with and without the ATP maintenance flux.[47]

**4.9. Methods to Predict Fluxes from Transcripts.** E-Flux2 predicts fluxes from transcripts by solving an FBA problem where the upper and lower bounds for each reaction have been modified according to the absolute expression for the corresponding gene.[17] The underlying idea is that, given a limited translational efficiency and enzyme accumulated over the time, the mRNA level can be considered as an approximate upper bound on the maximum amount of metabolic enzyme available and hence as a bound on reaction rates. Briefly, after a suitable flux bound normalization, the upper bound for each flux with transcript information is substituted by the absolute expression for the corresponding gene (for a positive upper bound, zero otherwise). If the reaction is reversible, the lower bound is substituted by the negative value of the absolute expression for the corresponding gene (if lower bound is negative, zero otherwise). An FBA problem is solved using these bounds and, as a last step, which differentiates E-Flux2 form its previous version of E-Flux,[16] the norm of the resulting flux is minimized. This ensures a single solution, unlike E-Flux. SPOT, instead of optimizing growth, maximizes the correlation between fluxes and the measured transcript profile, as determined through the Pearson correlation coefficient.[17] The assumption is that enzymatic transcript concentrations and fluxes tend to be as proportional to each other as allowed by stoichiometric constraints and enzyme presence. SPOT transforms the problem into an equivalent semi-definite programming problem that can be solved efficiently (eq 8 in ref 17), which is the version we use here.

**4.10. Summary of Jupyter Notebooks in This Publication.** Table S7 contains the list of the Jupyter notebooks used for creating the figures in this paper.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

The data used in this project is publicly available at https://github.com/LBLQMM/RhodococcusGSM.

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.2c00618.

Table S1: wild-type phenol 13C-MFA vs genome-scale model predicted fluxes; Table S2: PVHG6 phenol $^{13}$C-MFA vs genome-scale model predicted fluxes; Table S3: wild-type glucose 13C-MFA vs genome-scale model predicted fluxes; Table S4: biomass composition of *B. subtilis* and *R. opacus* with glucose and *R. opacus* with phenol; Table S5: *R. opacus* PD630 biomass reactants for phenol and glucose conditions; Table S6: mapping of reactions from $^{13}$C-MFA to genome-scale model reactions; Table S7: description of the notebooks used to generate and test iGR1773; Table S8: mapping of abbreviations to metabolite names; Figure S1: WT phenol growth and consumption data; Figure S2: PVHG6 phenol growth and consumption data; Figure S3: WT glucose growth and consumption data; Figure S4: phenol metabolism flux predictions based on PVHG6 strain transcripts; Figure S5: phenol flux maps based on PVHG6 transcripts; Figure S6: FBA prediction of ATP maintenance vs growth rate (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Authors

**Tae Seok Moon** − *Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States*; Email: hgmartin@lbl.gov

**Yinjie J. Tang** − *Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States*; Email: yinjie.tang@wustl.edu

**Hector García Martín** − *Biological Systems and Engineering Division, Lawrence Berkeley National Lab, Berkeley, California 94720, United States; BCAM - Basque Center for Applied Mathematics, Bilbao 48009, Spain; DOE Agile BioFoundry, Emeryville, California 94608, United States; DOE Joint BioEnergy Institute, Emeryville, California 94608, United States;* orcid.org/0000-0002-4556-9685; Email: tsmoon7@gmail.com

### Authors

**Garrett W. Roell** − *Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States*

**Christina Schenk** − *BCAM - Basque Center for Applied Mathematics, Bilbao 48009, Spain; Biological Systems and*

*Engineering Division, Lawrence Berkeley National Lab, Berkeley, California 94720, United States;* orcid.org/0000-0002-7817-6757

**Winston E. Anthony** − *The Edison Family Center for Genome Sciences and Systems Biology, Washington University in St. Louis School of Medicine, St. Louis, Missouri 63110, United States; Department of Pathology and Immunology, Washington University in St. Louis School of Medicine, St. Louis, Missouri 63108, United States*

**Rhiannon R. Carr** − *Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States*

**Aditya Ponukumati** − *Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States*

**Joonhoon Kim** − *DOE Agile BioFoundry, Emeryville, California 94608, United States; DOE Joint BioEnergy Institute, Emeryville, California 94608, United States*

**Elena Akhmatskaya** − *BCAM - Basque Center for Applied Mathematics, Bilbao 48009, Spain; Biological Systems and Engineering Division, Lawrence Berkeley National Lab, Berkeley, California 94720, United States; IKERBASQUE, Basque Foundation for Science, Bilbao 48009, Spain*

**Marcus Foston** − *Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, United States;* orcid.org/0000-0002-4227-0362

**Gautam Dantas** − *The Edison Family Center for Genome Sciences and Systems Biology, Washington University in St. Louis School of Medicine, St. Louis, Missouri 63110, United States; Department of Pathology and Immunology and Department of Molecular Microbiology, Washington University in St. Louis School of Medicine, St. Louis, Missouri 63108, United States; Department of Biomedical Engineering, Washington University in St. Louis, St Louis, Missouri 63130, United States; Department of Pediatrics, Washington University School of Medicine in St Louis, St Louis, Missouri 63110, United States;* orcid.org/0000-0003-0455-8370

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.2c00618

### Author Contributions

T.S.M., Y.J.T., and H.G.M. conceived the project. G.W.R., C.S., and H.G.M. conceived the methods of GSM construction and application of omics data. W.E.A., R.R.C., and A.P. prepared data for modeling. All authors wrote and proofread the paper.

### Notes

The authors declare no competing financial interest.

The code for this project and the model itself are open source. They are available at https://github.com/LBLQMM/RhodococcusGSM.

## ■ ACKNOWLEDGMENTS

This project was supported by the US DOE (DE-SC0018324) and the Office of Science Graduate Student Research (SCGSR) Program. C.S. and E.A. were also supported by the Basque Government through the BERC 2018−2021 program and by the Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2017-0718. G.R. received funding from Defense Advanced Research Projects Agency B-SURE program (HR001122S0010) for this research. The views, opinions and/or findings expressed should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

## ■ REFERENCES

(1) Gani, A.; Naruse, I. Effect of cellulose and lignin content on pyrolysis and combustion characteristics for several types of biomass. *Renewable Energy* **2007**, *32*, 649−661.

(2) Alvarez, H. M.; Mayer, F.; Fabritius, D.; Steinbüchel, A. Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630. *Arch. Microbiol.* **1996**, *165*, 377−386.

(3) Chatterjee, A.; DeLorenzo, D. M.; Carr, R.; Moon, T. S. Bioconversion of renewable feedstocks by Rhodococcus opacus. *Curr. Opin. Biotechnol.* **2020**, *64*, 10−16.

(4) Fuchs, G.; Boll, M.; Heider, J. Microbial degradation of aromatic compounds — from one strategy to four. *Nat. Rev. Microbiol.* **2011**, *9*, 803−816.

(5) Roell, G. W.; Carr, R. R.; Campbell, T.; Shang, Z.; Henson, W. R.; Czajka, J. J.; Martín, H. G.; Zhang, F.; Foston, M.; Dantas, G.; Moon, T. S.; Tang, Y. J. A concerted systems biology analysis of phenol metabolism in Rhodococcus opacus PD630. *Metab. Eng.* **2019**, *55*, 120−130.

(6) Yoneda, A.; Henson, W. R.; Goldner, N. K.; Park, K. J.; Forsberg, K. J.; Kim, S. J.; Pesesky, M. W.; Foston, M.; Dantas, G.; Moon, T. S. Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630. *Nucleic Acids Res.* **2016**, *44*, 2240−2254.

(7) Henson, W. R.; Campbell, T.; DeLorenzo, D. M.; Gao, Y.; Berla, B.; Kim, S. J.; Foston, M.; Moon, T. S.; Dantas, G. Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus. *Metab. Eng.* **2018**, *49*, 69−83.

(8) Diao, J.; Carr, R.; Moon, T. S. Deciphering the transcriptional regulation of the catabolism of lignin-derived aromatics in Rhodococcus opacus PD630. *Commun. Biol.* **2022**, *5*, 1109.

(9) King, Z. A.; Lu, J.; Dräger, A.; Miller, P.; Federowicz, S.; Lerman, J. A.; Ebrahim, A.; Palsson, B. O.; Lewis, N. E. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **2016**, *44*, D515−D522.

(10) Feist, A. M.; Palsson, B. O. The biomass objective function. *Curr. Opin. Microbiol.* **2010**, *13*, 344−349.

(11) Dahal, S.; Yurkovich, J. T.; Xu, H.; Palsson, B. O.; Yang, L. Synthesizing Systems Biology Knowledge from Omics Using Genome-Scale Models. *Proteomics* **2020**, *20*, 1900282.

(12) Zur, H.; Ruppin, E.; Shlomi, T. iMAT: an integrative metabolic analysis tool. *Bioinformatics* **2010**, *26*, 3140−3142.

(13) Shlomi, T.; Cabili, M. N.; Herrgård, M. J.; Palsson, B. Ø.; Ruppin, E. Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.* **2008**, *26*, 1003−1010.

(14) Becker, S. A.; Palsson, B. O. Context-Specific Metabolic Networks Are Consistent with Experiments. *PLoS Comput. Biol.* **2008**, *4*, No. e1000082.

(15) Schmidt, B. J.; Ebrahim, A.; Metz, T. O.; Adkins, J. N.; Palsson, B. Ø.; Hyduke, D. R. GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics* **2013**, *29*, 2900−2908.

(16) Colijn, C.; Brandes, A.; Zucker, J.; Lun, D. S.; Weiner, B.; Farhat, M. R.; Cheng, T.-Y.; Moody, D. B.; Murray, M.; Galagan, J. E. Interpreting Expression Data with Metabolic Flux Models: Predicting Mycobacterium tuberculosis Mycolic Acid Production. *PLoS Comput. Biol.* **2009**, *5*, No. e1000489.

(17) Kim, M. K.; Lane, A.; Kelley, J. J.; Lun, D. S. E-Flux2 and SPOT: Validated Methods for Inferring Intracellular Metabolic Flux Distributions from Transcriptomic Data. *PLoS One* **2016**, *11*, No. e0157101.

(18) van Berlo, R. J. P.; de Ridder, D.; Daran, J.-M.; Daran-Lapujade, P. A. S.; Teusink, B.; Reinders, M. J. T. Predicting Metabolic Fluxes Using Gene Expression Differences As Constraints. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2011**, *8*, 206−216.

(19) Mahadevan, R.; Schilling, C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **2003**, *5*, 264−276.
(20) Kim, H. U.; Kim, W. J.; Lee, S. Y. Flux-coupled genes and their use in metabolic flux analysis. *Biotechnol. J.* **2013**, *8*, 1035−1042.
(21) Schultz, A.; Qutub, A. A. Reconstruction of Tissue-Specific Metabolic Networks Using CORDA. *PLoS Comput. Biol.* **2016**, *12*, No. e1004808.
(22) Fondi, M.; Liò, P. Multi -omics and metabolic modelling pipelines: Challenges and tools for systems microbiology. *Microbiol. Res.* **2015**, *171*, 52−64.
(23) Hyduke, D. R.; Lewis, N. E.; Palsson, B. Ø. Analysis of omics data with genome-scale models of metabolism. *Mol. BioSyst.* **2013**, *9*, 167−174.
(24) Reed, J. L. Shrinking the Metabolic Solution Space Using Experimental Datasets. *PLoS Comput. Biol.* **2012**, *8*, No. e1002662.
(25) Blazier, A. S.; Papin, J. A. Integration of expression data in genome-scale metabolic network reconstructions. *Front. Physiol.* **2012**, *3*, 299.
(26) Machado, D.; Herrgård, M. Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism. *PLoS Comput. Biol.* **2014**, *10*, No. e1003580.
(27) Firrincieli, A.; Grigoriev, B.; Dostálová, H.; Cappelletti, M. The Complete Genome Sequence and Structure of the Oleaginous Rhodococcus opacus Strain PD630 Through Nanopore Technology. *Front. Bioeng. Biotechnol.* **2022**, *9*, No. 810571.
(28) Sundararaghavan, A.; Mukherjee, A.; Sahoo, S.; Suraishkumar, G. K. Mechanism of the oxidative stress-mediated increase in lipid accumulation by the bacterium, R. opacus PD630: Experimental analysis and genome-scale metabolic modeling. *Biotechnol. Bioeng.* **2020**, *117*, 1779−1788.
(29) Kim, H. M.; Chae, T. U.; Choi, S. Y.; Kim, W. J.; Lee, S. Y. Engineering of an oleaginous bacterium for the production of fatty acids and fuels. *Nat. Chem. Biol.* **2019**, *15*, 721−729.
(30) Tajparast, M.; Frigon, D. Genome-scale metabolic model of Rhodococcus jostii RHA1 (iMT1174) to study the accumulation of storage compounds during nitrogen-limited condition. *BMC Syst. Biol.* **2015**, *9*, 43.
(31) Lieven, C.; Beber, M. E.; Olivier, B. G.; Bergmann, F. T.; Ataman, M.; Babaei, P.; Bartell, J. A.; Blank, L. M.; Chauhan, S.; Correia, K.; Diener, C.; Dräger, A.; Ebert, B. E.; Edirisinghe, J. N.; Faria, J. P.; Feist, A. M.; Fengos, G.; Fleming, R. M. T.; García-Jiménez, B.; Hatzimanikatis, V.; van Helvoirt, W.; Henry, C. S.; Hermjakob, H.; Herrgård, M. J.; Kaafarani, A.; Kim, H. U.; King, Z.; Klamt, S.; Klipp, E.; Koehorst, J. J.; König, M.; Lakshmanan, M.; Lee, D.-Y.; Lee, S. Y.; Lee, S.; Lewis, N. E.; Liu, F.; Ma, H.; Machado, D.; Mahadevan, R.; Maia, P.; Mardinoglu, A.; Medlock, G. L.; Monk, J. M.; Nielsen, J.; Nielsen, L. K.; Nogales, J.; Nookaew, I.; Palsson, B. O.; Papin, J. A.; Patil, K. R.; Poolman, M.; Price, N. D.; Resendis-Antonio, O.; Richelle, A.; Rocha, I.; Sánchez, B. J.; Schaap, P. J.; Malik Sheriff, R. S.; Shoaie, S.; Sonnenschein, N.; Teusink, B.; Vilaça, P.; Vik, J. O.; Wodke, J. A. H.; Xavier, J. C.; Yuan, Q.; Zakhartsev, M.; Zhang, C. MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **2020**, *38*, 272−276.
(32) Machado, D.; Andrejev, S.; Tramontano, M.; Patil, K. R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **2018**, *46*, 7542−7553.
(33) Monk, J. M.; Lloyd, C. J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; Feist, A. M.; Palsson, B. O. iML1515, a knowledgebase that computes Escherichia coli traits. *Nat. Biotechnol.* **2017**, *35*, 904−908.
(34) Lu, H.; Li, F.; Sánchez, B. J.; Zhu, Z.; Li, G.; Domenzain, I.; Marcišauskas, S.; Anton, P. M.; Lappa, D.; Lieven, C.; Beber, M. E.; Sonnenschein, N.; Kerkhoven, E. J.; Nielsen, J. A consensus S. cerevisiae metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat. Commun.* **2019**, *10*, 3586.
(35) Sánchez, B. J.; Zhang, C.; Nilsson, A.; Lahtvee, P.-J.; Kerkhoven, E. J.; Nielsen, J. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol. Syst. Biol.* **2017**, *13*, 935.
(36) Flamholz, A.; Noor, E.; Bar-Even, A.; Liebermeister, W.; Milo, R. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 10039−10044.
(37) Tang, Y. J.; Martin, H. G.; Myers, S.; Rodriguez, S.; Baidoo, E. E. K.; Keasling, J. D. Advances in analysis of microbial metabolic fluxes via 13C isotopic labeling. *Mass Spectrom. Rev.* **2009**, *28*, 362−375.
(38) Nießer, J.; Müller, M. F.; Kappelmann, J.; Wiechert, W.; Noack, S. Hot isopropanol quenching procedure for automated microtiter plate scale 13C-labeling experiments. *Microb. Cell Fact.* **2022**, *21*, 78.
(39) Hollinshead, W. D.; Henson, W. R.; Abernathy, M.; Moon, T. S.; Tang, Y. J. Rapid metabolic analysis of Rhodococcus opacus PD630 via parallel 13C-metabolite fingerprinting. *Biotechnol. Bioeng.* **2016**, *113*, 91−100.
(40) Varma, A.; Palsson, B. O. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Appl. Environ. Microbiol.* **1994**, *60*, 3724−3731.
(41) DeLorenzo, D. M.; Henson, W. R.; Moon, T. S. Development of Chemical and Metabolite Sensors for Rhodococcus opacus PD630. *ACS Synth. Biol.* **2017**, *6*, 1973−1978.
(42) Bligh, E. G.; Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **1959**, *37*, 911−917.
(43) Oh, Y.-K.; Palsson, B. O.; Park, S. M.; Schilling, C. H.; Mahadevan, R. Genome-scale Reconstruction of Metabolic Network in Bacillus subtilis Based on High-throughput Phenotyping and Gene Essentiality Data*. *J. Biol. Chem.* **2007**, *282*, 28791−28799.
(44) Kitts, P. A.; Church, D. M.; Thibaud-Nissen, F.; Choi, J.; Hem, V.; Sapojnikov, V.; Smith, R. G.; Tatusova, T.; Xiang, C.; Zherikov, A.; DiCuccio, M.; Murphy, T. D.; Pruitt, K. D.; Kimchi, A. Assembly: a resource for assembled genomes at NCBI. *Nucleic Acids Res.* **2016**, *44*, D73−D80.
(45) Courtot, M.; Juty, N.; Knüpfer, C.; Waltemath, D.; Zhukova, A.; Dräger, A.; Dumontier, M.; Finney, A.; Golebiewski, M.; Hastings, J.; Hoops, S.; Keating, S.; Kell, D. B.; Kerrien, S.; Lawson, J.; Lister, A.; Lu, J.; Machne, R.; Mendes, P.; Pocock, M.; Rodriguez, N.; Villeger, A.; Wilkinson, D. J.; Wimalaratne, S.; Laibe, C.; Hucka, M.; Le Novère, N. Controlled vocabularies and semantics in systems biology. *Mol. Syst. Biol.* **2011**, *7*, 543.
(46) Österlund, T.; Nookaew, I.; Bordel, S.; Nielsen, J. Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Syst. Biol.* **2013**, *7*, 36.
(47) Castro, A. R.; Rocha, I.; Alves, M. M.; Pereira, M. A. Rhodococcus opacus B4: a promising bacterium for production of biofuels and biobased chemicals. *AMB Express* **2016**, *6*, 35.


https://doi.org/10.1021/acssynbio.2c00618
*ACS Synth. Biol.* 2023, 12, 1632−1644