Predicting synthetic rescues in metabolic networks

A. E. Motter, N. Gulbahce, E. Almaas, A.-L. Barabási

SUPPLEMENTARY INFORMATION

In silico models

The *E. coli* K-12 MG1655 reconstructed metabolic network considered in the paper consists of $m = 537$ metabolites and $n = 728$ biochemical reactions [1]. This *E. coli* model has been tested extensively in the experimental validation of FBA and MOMA predictions [2] (see below for comparison with other models). The vector $\boldsymbol{\nu}$ includes all internal and transport metabolic fluxes bounded according to Eq. (1) as in Ref. [1]. The internal fluxes are unrestricted: $(\alpha_j, \beta_j) = (0, \infty)$ for irreversible reactions and $(\alpha_j, \beta_j) = (-\infty, \infty)$ for reversible reactions. Transport fluxes for metabolites that leave the cell are unrestricted in the outward direction. Uptake fluxes for carbon dioxide, potassium, sulfate and inorganic phosphate are unrestricted: $(\alpha_j, \beta_j) = (0, \infty)$. Transport fluxes for other substrates are restricted to zero ($\alpha_j = \beta_j = 0$) when the substrate is not available and to $(\alpha_j, \beta_j) = (0, \nu_j^{\max})$ when the substrate is available in the medium. The limited substrates in the minimal uptake basis used in the paper are a carbon source (10 mmol/g DW-h), oxygen (20 mmol/g DW-h), and ammonia (100 mmol/g DW-h). The rich medium contains limited amount of acetate, α-ketoglutarate, arabinose, glucose, lactate, malate, mannose, pyruvate, succinate, and sucrose. The ATP maintenance flux, ATPM, is set to $(\alpha_j, \beta_j) = (7.6, 7.6)$, and $(\alpha_j, \beta_j) = (0, 0)$ for the reactions associated with genes *edd*, *aldH*, *deoC*, and *poxB* [2].

The *H. pylori* 26695 [3] and *S. cerevisiae* S288C [4] models used in our comparison across organisms have $(m, n) = (485, 554)$ and $(646, 1266)$, respectively, for the numbers of metabolites and reactions. The minimal medium for *H. pylori* consists of 20 mmol/g DW-h of L-alanine, D-alanine, L-arginine, L-histidine, L-isoleucine, L-leucine, L-methionine, L-valine, glucose, Iron (II and III), phosphate, sulfate, pimelate, and thiamine, 5 mmol/g DW-h of oxygen, and unlimited access to water and protons. The minimal medium for *S. cerevisiae* includes unlimited access to water, protons, phosphate, carbon dioxide, potassium, and sulfate, and limited access to a carbon source (10 mmol/g DW-h), oxygen (20 mmol/g DW-h), and ammonia (100 mmol/g DW-h). The rich medium contains limited amount of acetate, ethanol, galactose, glucose, glycerol, sorbitol, and xylose. In addition, similar to Ref. [4],

the reactions ACOAH and GLUSx are constrained to zero, and the ATPM flux is set to $(\alpha_j, \beta_j) = (1, 1)$. Under anaerobic conditions, in addition to setting oxygen uptake to zero, ergosterol and zymosterol are made available to allow growth.

Biomass flux and growth rate

For exponential growth, the biomass production is governed by equation $\frac{1}{N}\frac{dN}{dt} = \kappa$, where κ is the growth rate and $\frac{1}{N}\frac{dN}{dt}$ is proportional to the biomass flux G. Therefore, the normalized growth rate $\bar{\kappa} = \kappa/\kappa^{wt}$ equals the normalized biomass flux $\bar{G} = G/G^{wt}$ used throughout the paper.

Gene abbreviations

We have specified *E. coli* gene complexes either by their full names, e.g., gene complex *sucAB*, or represented them by the first gene in the complex, e.g., *cyoA* for complex *cyoABCD*. For *E. coli pfk*-deficient mutants, we delete both *pfkA* and *pfkB* genes, and for *S. cerevisiae* we delete both *pfk1* and *pfk2* genes.

Sensitivity analysis

We tested numerous variants of the gene deletions described in the paper. The overall conclusion is that the findings are quite robust and do not rely on fine tuning of parameters or environmental conditions. In particular, we observe that different combinations of individual rescue deletions can be used to buffer a given deleterious mutation and that a given set of deletions can often buffer different deleterious mutations. For example, any of the sets of rescue deletions corresponding to Table S I are capable of restoring more than 45% of the biomass production after the deletion of any of the genes *pfk*, *fbaA*, and *tpiA*. This result is partially related to the fact that these genes are involved in closely connected reactions of the central metabolism. More important, we observe that no individual gene is essential to the design of a strongly rescuing gene deletion. For example, in Figure 3 essentially the same results are obtained when any given gene is excluded in advance from the combination of individual gene deletions.

We also performed a systematic study of the robustness of our predictions when the metabolic state of the wild-type organism is perturbed. Specifically, we investigated how the biomass fluxes vary before (G^1_{MOMA}) and after (G^2_{MOMA}) rescue deletions when the wild-type fluxes are displaced from an optimal point $\boldsymbol{\nu}^{wt}_{FBA}$ predicted by FBA. As an example, consider *tpiA*-deficient *E. coli* mutants rescued with forty targeted gene deletions as shown

in Figure 3(b). With the rescue deletions, the predicted biomass flux of the organisms increase from $\bar{G}^1_{MOMA} = 0$ to $\bar{G}^2_{MOMA} = 0.48$. As shown in Figure S1 for perturbations of up to 5% of the wild-type biomass flux G^{wt}_{FBA}, after the rescue deletions, the histogram of perturbed biomass fluxes is sharply peaked around 0.48; before the rescues, the histogram is very sharply peaked around zero, indicating that the original mutants are nonviable.

In addition, we investigated the effects of alternate optimal solutions of the wild type on the MOMA predicted growth. It has been previously shown that different optima can yield different MOMA predicted growths for a given strain [5, 6]. To sample the alternate optima, we used a mixed integer linear programming formulation [7] introduced and discussed in Refs. [5, 8, 9]. When the optimal solution is degenerate, the space of optimal solutions has dimension larger than zero and includes infinitely many points. Using *tpiA*-deficient *E. coli* mutants and the 10-gene rescue set *aceA*, *gadA*, *gadB*, *gpt*, *gsk*, *lpdA*, *pat*, *putA*, *tynA* and *xapB* as in Figure 3(b), we sampled 1000 of these optimal solutions and calculated the MOMA growth before and after recovery (see Figure S2). For all the sampled alternate wild-type solutions, the rescued mutants are viable, and their MOMA predicted growth rates are more than 9% of the wild-type growth rate.

Robustness across *in silico* models

We also tested the robustness of the predictions across different *in silico* models. In particular, for *E. coli* K-12 MG1655, we compared the predictions of the model *i*JE660a [1] used in the paper with the model *i*JR904 [10], which has 304 more unique reactions. We considered the model *i*JE660a both when the Entner-Doudoroff pathway is turned on and when it is turned off. In addition, genes *aldH*, *deoC*, and *poxB* are also made available when the Entner-Doudoroff pathway is turned on. Figure S3 shows a significant agreement between the predictions drawn from these models. Twelve out of the fifteen gene-deficient mutants that are eligible for improvement in these models consistently support a recovery of more than 15% of the wild-type biomass flux for all three variants.

Supplementary data on double gene deletions

In Figure S4 we show the gene deletion results for all gene pairs of *E. coli*. We plot the probability distribution of $\Delta\bar{G} = \bar{G}_{ij} - \min\{\bar{G}_i, \bar{G}_j\}$, where $\bar{G}_{ij}$ is the MOMA predicted growth rate after the concurrent deletion of genes i and j, and $\bar{G}_i$ and $\bar{G}_j$ are the MOMA predicted growth rates after the single deletion of gene i and gene j, respectively. Synthetic

lethality is defined as $\bar{G}_{ij} < 0.01$ for $\bar{G}_i > 0.01$ and $\bar{G}_j > 0.01$. A gene pair is considered to be synthetically recoverable if $\Delta\bar{G} > 0.01$. In Figure S5, we also show the corresponding gene deletion results for all gene pairs of *S. cerevisiae*.

In Supplementary Information 4 we provide up to 6 choices of single rescue gene for each gene-deficient *E. coli* mutant for which the biomass production rate can be increased by at least 1%. The rescue gene choices are ranked from high to low according to the observed increase in biomass production. The data is provided in separate sheets for the following media: acetate (ac), α-ketoglutarate (akg), arabinose (ara), glucose (glc), lactate (lac), malate (mal), mannose (man), pyruvate (pyr), rich, succinate (succ), and sucrose (suc). In Supplementary Information 5 we provide the same information for each gene-deficient *S. cerevisiae* mutant. The data for *S. cerevisiae* is provided for the following media: acetate (ac), ethanol (eth), galactose (gal), glucose (glc), glucose anaerobic (glc^{a}), glycerol (gly), rich, sorbitol (sor), and xylose (xyl).

Supplementary data on rescue deletions

In Tables SII and SIII we list, respectively, the minimum rescue deletion set necessary to increase the biomass production rate by at least 3% of the wild-type rate for each eligible gene-deficient mutant of *E. coli* and *S. cerevisiae* in minimal glucose media.

In Supplementary Information 2 and Supplementary Information 3 we list the full rescue deletion set to increase the biomass production to the maximum recovery rate (also indicated in the files) for each eligible gene-deficient *E. coli* and *S. cerevisiae* mutant, respectively.

Supplementary analysis of experimental data

We further analyze the experimental data [11] corresponding to Figure 5(a) within the scope of our approach. Our approach is based on the assumption that the deletion of a reaction whose flux is much larger (smaller) than the optimal-state flux will tend to improve (not improve) growth. Using as input the experimentally measured growth rate and uptake oxygen and carbon source rates combined with flux variability analysis [6], we have determined the average flux values of the reactions catalyzed by the deleted genes. Comparing them with the optimal values under the corresponding conditions, our predictions are in agreement with the experimental results for 20 out of the 22 mutants for which growth was measured in Ref. [11]. The results of this analysis are shown in Table SIV.

[1] J. S. Edwards and B. O. Palsson, *The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities*, Proc. Natl. Acad. Sci. U.S.A. **97**, 5528-5533 (2000).

[2] D. Segrè, D. Vitkup, and G. M. Church, *Analysis of optimality in natural and perturbed metabolic networks*, Proc. Natl. Acad. Sci. U.S.A. **99**, 15112-15117 (2002).

[3] I. Thiele, T. D. Vo, N. D. Price, and B. O. Palsson, *An expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): An in silico genome-scale characterization of single and double deletion mutants*, J. Bacteriol. **187**, 5818-5830 (2005).

[4] N. C. Duarte, M. J. Herrgard, and B. O. Palsson, *Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model*, Genome Res. **14**, 1298-1309 (2004).

[5] J. L. Reed and B. O. Palsson, *Genome-scale in silico models of E. coli have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states*, Genome Res., **14**, 1797-1805 (2004).

[6] R. Mahadevan and C. H. Schilling, *The effects of alternate optimal solutions in constraint-based genome-scale metabolic models*, Metabolic Eng. **5**, 264-276 (2003).

[7] Code provided by Dong-Hee Kim, Northwestern University (2007).

[8] S. Lee, C. Phalakornkule, M. M Domach, and I. E. Grossmann, *Recursive MILP model for finding all the alternate optima in LP models for metabolic networks*, Comput. Chem. Eng. **24**, 711-716 (2000).

[9] B. O. Palsson, *Systems Biology* (Cambridge University Press, 2006).

[10] J. L. Reed, T. D. Vo, C. H. Schilling, and B. O. Palsson, *An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR)*, Genome Biol. **4**, R54.1-R54.12 (2003).

[11] S. Fong and B. O. Palsson, *Metabolic gene deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes*, Nat. Genet. **36**, 1056-1058 (2004). Uptake rates obtained through private communication with the authors.

Table SI: The biomass production before and after rescue deletions in *E. coli* mutants in glucose minimal medium. The overlines indicate that the biomass fluxes are normalized by the wild-type flux $G^{wt}_{FBA} = 0.908$ mmol/g DW-h.

mutant	$\bar{G}^1_{FBA}$	$\bar{G}^1_{MOMA}$	$\bar{G}^2_{MOMA}$
pfk	0.77	0	0.50
fbaA	0.77	0	0.49
tpiA	0.78	0	0.50

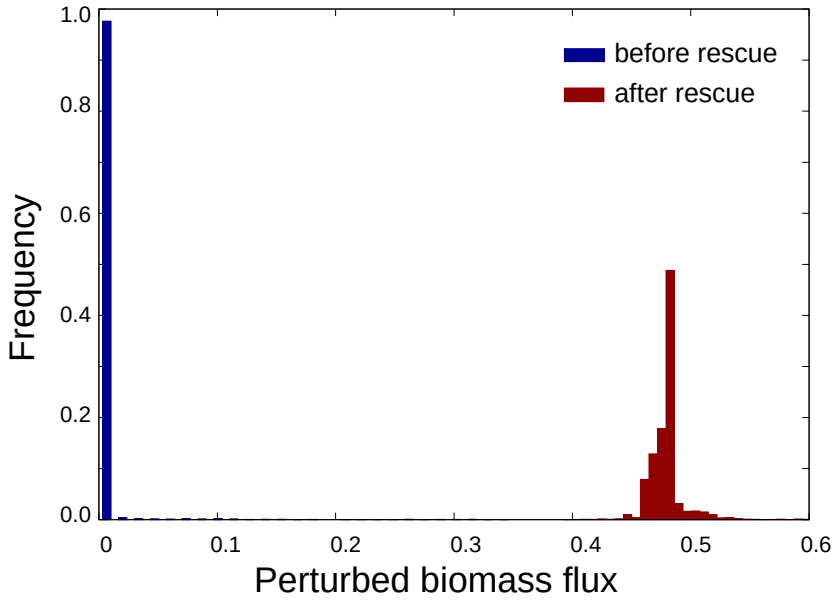

Figure S1: Histogram of the predicted biomass fluxes $\bar{G}^1_{MOMA}$ (blue) and $\bar{G}^2_{MOMA}$ (red) for *tpiA*-deficient *E. coli* mutants in glucose minimal medium when randomly chosen fluxes of the wild-type organism are perturbed by up to 20% and the wild-type growth rate changes up to 5%. The sample size is 4000.

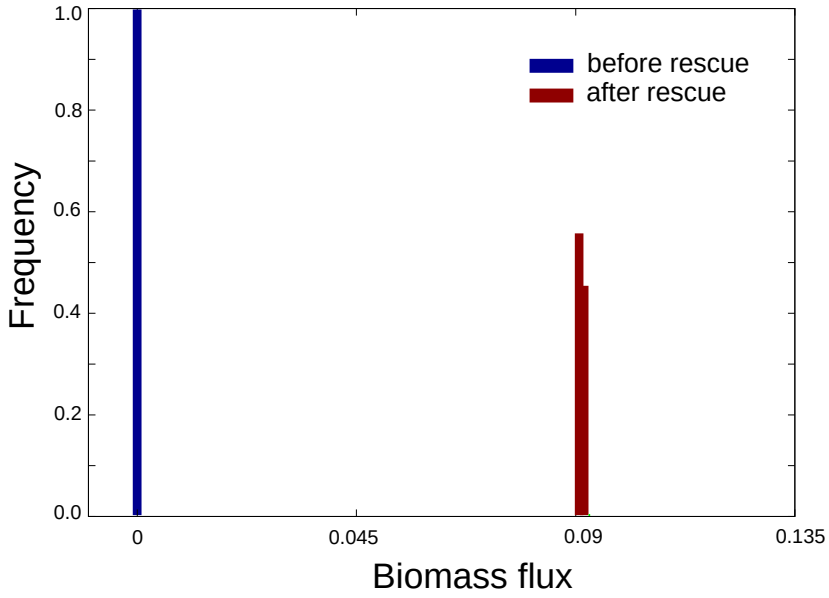

Figure S 2: Histogram of the biomass fluxes $\bar{G}^1_{MOMA}$ (blue) and $\bar{G}^2_{MOMA}$ (red) for *tpiA*-deficient *E. coli* mutants for which alternate optimal wild-type solutions are chosen as MOMA reference points. The mutant is rescued with the deletion of a 10-gene set. The carbon source is glucose and the number of sampled alternate solutions is 1000. Note that for all sampled alternate solutions, the rescued mutant growth is more than 9% of the wild-type strain growth.

	aceE	atpI	crr	cyoA	eno	gdhA	glyA	gnd	lpdA	nuoA	pgi	ppc	sdh	tpiA	zwf
iJE660[1]															
iJE660[2]															
iJR904															

Figure S 3: Comparison of our predictions in glucose medium using *E. coli* model *i*JE660a (first row), model *i*JE660a with the Entner-Doudoroff pathway (second row), and model *i*JR904 (third row). The difference $\Delta\bar{G}$ between G^2_{MOMA} and G^1_{MOMA} normalized by G^{wt}_{FBA} is shown for the mutants that are eligible for improvement in all three models. Color coding indicates the degree of improvement obtained in biomass flux: $\Delta\bar{G} > 25\%$ (red), $25\% \geq \Delta\bar{G} > 15\%$ (orange), and $15\% \geq \Delta\bar{G} > 0\%$ (yellow). Note that the three models strongly agree in most cases.

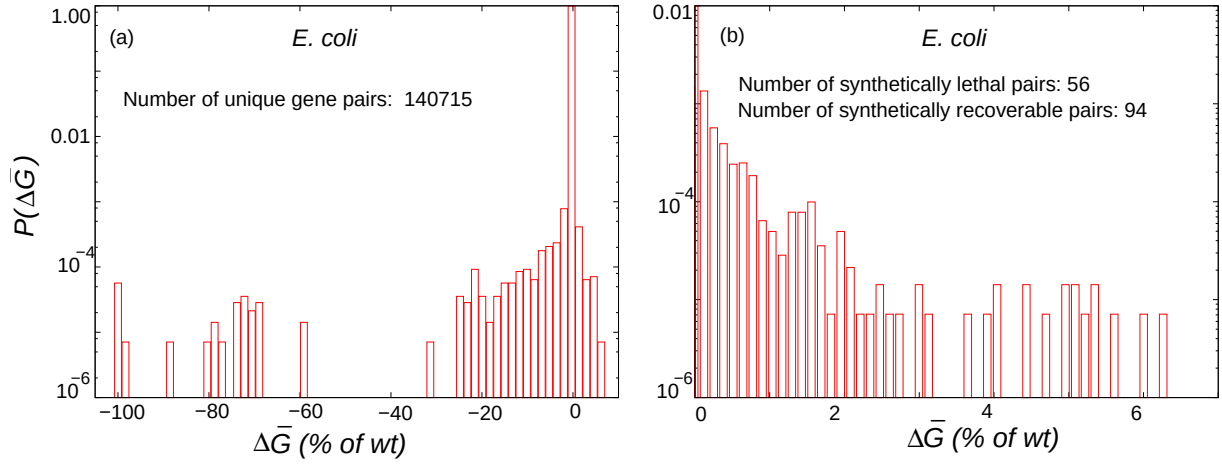

Figure S 4: Double gene deletion results for all gene pairs of *E. coli* in glucose minimal medium. The histogram of $\Delta\bar{G} = \bar{G}_{ij} - \min\{\bar{G}_i, \bar{G}_j\}$ is shown, where $\bar{G}_{ij}$ is the MOMA predicted growth rate after the concurrent deletion of genes i and j, and $\bar{G}_i$ and $\bar{G}_j$ are the MOMA predicted growth rates after the single deletion of gene i and gene j, respectively. A magnification of the histogram for all synthetically recoverable pairs is shown in (b). The growth rates are determined relative to the wild-type growth rate $G^{wt}_{FBA} = 0.908$ mmol/g DW-h.

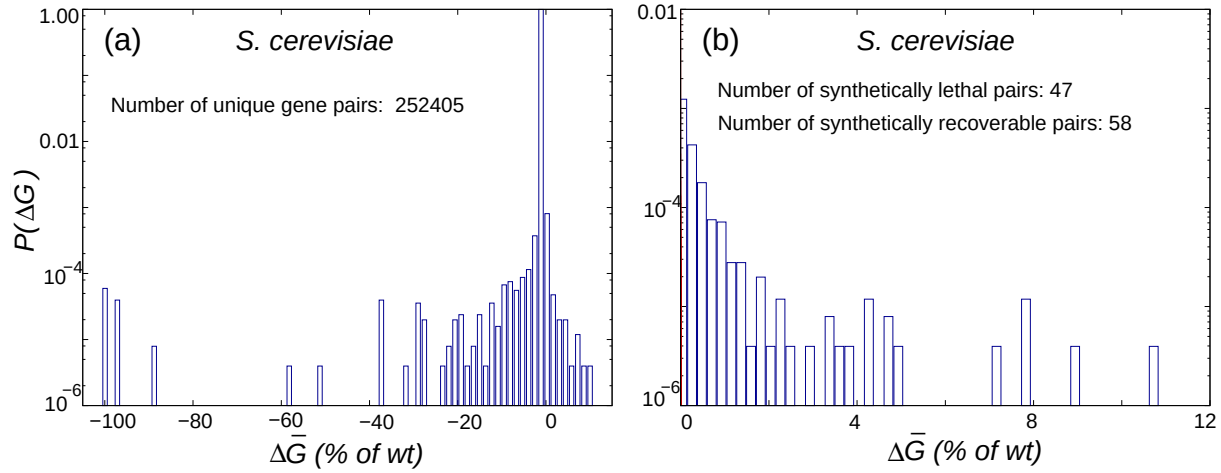

Figure S 5: Double gene deletion results for all gene pairs of *S. cerevisiae* in glucose minimal medium. The wild-type growth rate is $G^{wt}_{FBA} = 0.819$ mmol/g DW-h. The conventions are the same used in Figure S 4.

Table SII: Minimum set of rescue genes to recover *E. coli* mutants in glucose medium for a growth increase of at least 3% of the wild-type growth ($G^{wt}_{FBA} = 0.908$ mmol/g DW-h). Columns correspond to gene-deficient mutants, number of genes in the rescue set (n_c), observed growth increase ($\Delta\bar{G}$), and one choice (not necessarily unique) of genes in the rescue set, respectively.

mutant	n_c	$\Delta\bar{G}(\%wt)$	rescues
aceE	1	5	*panF*
atpA	15	6	*aceA, gadA, gadB, gltP, gltS, lysP, panF, pat, pheP, purU, putP, sucAB tnaB, tynA, tyrP*
crr	1	3	*gpt*
cycBC	19	5	*araFGH, aspA, cycA, gltJKL, gpt, gsk, hpt, kbl, lysP, panF, pdxK, ppsA pstABCS, putP, tdcC, tdh, trkAEHG, u27, u36*
cydAB	19	5	*araFGH, aspA, cycA, gltJKL, gpt, gsk, hpt, kbl, lysP, panF, pdxK, ppsA pstABCS, putP, tdcC, tdh, trkAEHG, u27, u36*
cyoA	1	5	*cycA*
eno	5	3	*maeA, pat, pckA, sucAB, tynA*
fbaA	10	5	*aceA, gadA, gadB, gltP, gltS, gpt, lpdA, pat, tynA, xapB*
gdhA	4	3	*glpABC, glpD, pntAB, putA*
glyA	4	3	*cycA, pheP, tnaB, tyrP*
gnd	3	3	*gpt, gsk, xapB*
lpdA	1	4	*cycA*
nuoA	1	6	*cycA*
pfkA	10	7	*aceA, gadA, gadB, gpt, gltP, gltS, lpdA, pat, tynA, xapB*
pgi	3	3	*pat, putP, tynA*
pgl	3	3	*gpt, gsk, xapB*
ppc	3	6	*pheP, tnaB, tyrP*
ptsG	1	3	*gpt*
rpe	2	3	*tnaB, xapB*
sdhABCD	7	3	*gadA, gadB, gltP, gltS, kgtP, putP, sucAB*
talB	1	5	*xapB*
tpiA	8	5	*aceA, gadA, gadB, lpdA, tynA, gpt, putA, pat*
zwf	3	3	*gpt, gsk, xapB*

Table SIII: Minimum set of rescue genes to recover *S. cerevisiae* mutants in glucose medium for a growth increase of at least 3% of the wild-type growth ($G^{wt}_{FBA} = 0.819$ mmol/g DW-h). Columns correspond to gene-deficient mutants, number of genes in the rescue set (n_c), observed growth increase ($\Delta\bar{G}$), and one choice (not necessarily unique) of genes in the rescue set, respectively.

mutant	n_c	$\Delta\bar{G}(\%wt)$	rescues
cor1	46	3	*aac1, aac3, amd1, amd3, amd5, asp1, asp31e, asp32e*
			asp33e, asp34e, bgl2, cdd1, dut1, fbp1, fbp26, fcy1, gph1
			hnt2, hor2, lcb4, lcb5, met3, mht1, nth1, nth2, pde1, pde2
			pet9m, pnp1, rhr2, sam4, shm2, u40, u50, u52, u57, u58, u62
			u84, u89, u93, u109, u133, u159, u161, zwf1
dic1	6	6	*agp1, bap2, bap3, gap1, tat1, tat2*
fba1	24	3	*agp1, amd1, amd3, amd5, bgl2, car1, dut1, gap1, hip1, lcb4, lcb5*
			mht1, nth1, nth2, pfk, pnp1, sam4, tat1, u50, u52, 109, u142, u159, urk1
gcv1	4	3	*lcb4, lcb5, mae1, u62*
gcv2	4	3	*lcb4, lcb5, mae1, u62*
gcv3	4	3	*lcb4, lcb5, mae1, u62*
lpd1	2	3	*mae1, car1*
mdh2	3	3	*u149, u160, u175*
mir1	3	3	*fdh1, fdh2, fdh3*
pfk	23	3	*agp1, amd1, amd3, amd5, bgl2, car1, dut1, gap1, hip1, lcb4, lcb5, mht1*
			nth1, nth2, pnp1, sam4, tat1, u50, u52, u109, u142, u159, urk1
pgi1	2	3	*nth1, nth2*
ser1	3	4	*lcb4, lcb5, fbp1*
ser2	3	4	*lcb4, lcb5, fbp1*
tpi1	1	3	*pck1*

Table SIV: Reaction activity for *E. coli* experimental data corresponding to Figure 5(a). Columns indicate respectively gene i, wild-type flux of reaction i (in mmol/g DW-h), wild-type flux of reaction i under optimal growth (in mmol/g DW-h), and predicted and experimental change in growth caused by the deletion of gene i.

Gene	Flux	Opt Flux	Prediction	Experiment
Akg				
ackr	7.45	0.03	increase	increase
frd	10.85	0.11	increase	increase
pck	9.70	2.70	increase	increase
ppc	6.95	0.01	increase	increase
Glc				
ackr	1.83	1.05	increase	increase
frd	3.88	0.25	increase	increase
pck	1.37	0.01	increase	decrease
ppc	2.34	1.52	increase	increase
tpi	4.67	5.20	decrease	decrease
zwf	5.17	3.84	increase	increase
Gly				
frd	15.77	0.15	increase	increase
zwf	8.39	0.00	increase	increase
Lac				
ackr	16.87	12.07	increase	increase
frd	23.94	22.09	increase	increase
tpi	6.12	4.77	increase	decrease
Mal				
ackr	5.01	0.01	increase	increase
pck	9.81	5.13	increase	increase
ppc	4.63	0.01	increase	increase
zwf	6.30	0.01	increase	increase
Rib				
pck	12.82	0.00	increase	increase
ppc	13.25	1.05	increase	increase
zwf	10.21	2.34	increase	increase