Supplementary Material for "Characterizing post-translational modifications in prokaryotic metabolism using a multi-scale workflow"

Authors: Elizabeth Brunk, Roger L. Chang, Jing Xia, Hooman Hefzi, James T. Yurkovich, Donghyuk Kim, Evan Buckmiller, Harris Wang, Byung-Kwan Cho, Chen Yang, Bernhard O. Palsson, George M. Church, Nathan E. Lewis

Table of contents

Supplementary Methods

Regulated Metabolic Branch Analysis (RuMBA)

Metabolic regulation is a rapid means to redirect flux in a metabolic network, while transcriptional regulation and regulation of enzyme abundance are processes that act on a longer time scale. Therefore, it is expected that following a shift to a new growth condition, allosteric regulation and post-translational enzyme modification will redirect flux at important branch points. The rationale for this response is that, *in vivo*, there are regular fluctuations in the cellular microenvironment and frequent environmental changes (1–3). It would be advantageous for the cell to have a means to rapidly regulate metabolic pathway usage using reversible mechanisms while slower and more permanent regulatory mechanisms are being activated. The relative costs and timescale of a few types of regulation are given in SI Appendix, Fig. S1.

Two methods have been developed to predict which enzymes will require significant changes in activity level following a change in carbon substrate for shorter and longer timescales, called these RuMBA and FSS, respectively. Code, compatible with the COBRA Toolbox is provided at doi:10.7303/syn15667323.

FSS has been used previously (4–6). Another method similar to FSS has also been published, showing its conceptual accuracy (7). A brief discussion of FSS provides a conceptual basis to understand RuMBA. Constraint-based modeling, the framework upon which both RuMBA and FSS are based, uses the metabolic network topology to define a space of possible phenotypes by adding a series of known biologically-relevant governing constraints (e.g., uptake rates for media components, by-product secretion rates, growth rates, etc.) (8, 9). This space of possible phenotypes represents all possible combinations of metabolic steady-state pathway usage that a cell can use in the given growth conditions. Assuming the constraints are accurate, the actual steady state flux distribution (or pathway usage) should be within the *in silico* solution space (SI Appendix, Fig. S2.a). The range and distribution of flux through each reaction within these solution spaces are dependent on the constraints, such as reaction thermodynamics, metabolite uptake rates, etc. Therefore, the space is condition-specific, i.e., the various dimensions of the space might move when the model is simulated under two different growth conditions. For example, as shown in SI Appendix, Fig. S2.b-c, the flux may be significantly higher in the second growth condition (reaction 2), or show no significant change between the two growth conditions (reaction 1).

The predicted changes in pathway usage from FSS represent the changes that lead to the optimal pathway usage in different growth conditions. However, to achieve this optimality, the activity of numerous enzymes must be fine-tuned, and often, many proteins need to be up- or down-regulated to meet this requirement. These adjustments require significant changes in transcription and translation, which can take a generation or two for entire pathways. On a shorter time scale, when changes in enzyme level are either less efficient (e.g., protein degradation) and/or not feasible to obtain, a more reasonable adaptive response involves a temporary suppression of the activity of an enzyme to avoid sending metabolites down less efficient pathways, or to boost the activity of present enzymes that will be needed in the new growth conditions. Thus, regulation at metabolic branch-points becomes of great importance, so that metabolites can be shuttled down the most efficient pathways.

RuMBA leverages this idea to compute the shift of the solution space for short-time scale changes in metabolic pathway activity at metabolic branch points. To do this, Markov chain Monte Carlo sampling of the metabolic solution space is used to obtain a uniformly distributed assessment of feasible flux values each reaction can have at steady state. To assess each branch point metabolite in the network, all reactions that can produce or consume it are identified. For example, aconitase produces isocitrate, while isocitrate dehydrogenase and isocitrate lyase both consume it (SI Appendix, Fig. S3a). Flux through each branch point metabolite in the network with a connectivity less than 30 is assessed. For each sample point in the solution space (SI Appendix, Fig. S3b), all incoming fluxes are summed up, as are all outgoing fluxes. Then, for each ith reaction, the fraction of total flux through the metabolite, v_{met}, that is contributed by the reaction of interest, is computed as follows:

$$f_i = \frac{v_i}{v_{met}}$$

where v_i is the flux through reaction i and f_i is the fraction of all flux passing through the metabolite of interest, that is passing through reaction i. Since this is done for many random feasible sets of flux values through all of the reactions at the branch point, a distribution of f_i fractions is computed for each reaction for the two growth conditions of interest (SI Appendix, Fig. S3a). Therefore, a p-value can be computed that measures the overlap of the f_i values for that reaction under the given growth condition, thus quantifying how significantly the flux changes from one enzyme to another when environmental conditions change (SI Appendix, Fig. S3b-d). The function of a phosphorylation event can subsequently be predicted if the change in phosphorylation is also known (SI Appendix, Fig. S4a).

A small fraction of reactions can show miniscule, but significant changes due mostly to slight differences in predicted growth rates. Thus, the list of the regulated reactions and their associated enzymes is filtered to focus on the more significant results. Reactions that change their predicted magnitude of flux by less than 50% are filtered out from the list of reactions requiring regulation. This was done by simulating changes in reaction flux occurring in a shift between two conditions, as done previously (4, 5).

Once the flux values were normalized, the changes of fluxes between two conditions were determined as previously described (5). Briefly, calls on differential reaction activity were made when the distributions of feasible flux states (obtained from MCMC sampling) under two different conditions did not significantly overlap. For each metabolic reaction, a p-value was obtained by computing the probability of finding a flux value for a reaction in one condition that is equal to or more extreme than a given flux value in the second condition. Significance of p-values was adjusted for multiple hypotheses (FDR = 0.01). When the magnitude of flux changed less than 50% of the initial flux magnitude, these reactions were filtered out from the set of predicted sites of regulation and excluded from further analysis. However, results were robust for a wide range of filter levels.

To test if this method can predict the function of PTMs, three *E. coli* enzymes were identified from the literature, that undergo differential protein phosphorylation between growth on glucose and acetate. RuMBA was employed to predict the effect of phosphorylation on these three enzymes (SI Appendix, Fig. S4a). At late log phase, enolase has been shown to have seven times higher phosphorylation when *E. coli* was grown on glucose than when grown on acetate (10). *In silico*, RuMBA predicts that enolase will have a reduced flux level on acetate. Therefore, one may predict that the phosphorylation event would activate its forward flux. It was determined that when treated with acid phosphatase, enolase was inhibited (10). Similarly, RuMBA predicts that on acetate, the flux through isocitrate dehydrogenase (ICDHyr) decreases, while the flux through isocitrate lyase (ICL) should increase. Experimentally, the phosphorylation of ICDHyr increases and may increase for ICL (phosphorylation is high when grown on acetate, but has not been rigorously tested on glucose). Thus, it is predicted that phosphorylation of ICDHyr inhibits enzyme activity, while it activates ICL. Both of these predictions are consistent with published data (11, 12).

Markov chain Monte Carlo sampling
The distribution of feasible fluxes for each reaction in the models used here were determined using Markov chain Monte Carlo (MCMC) sampling (13), as previously described (5, 14), and was

implemented with the COBRA Toolbox v2.0 (15). Uptake rates were used to constrain the models as detailed above. To model more realistic growth conditions (16), suboptimal growth was modeled. Specifically, the biomass objective function (a proxy for growth rate) was provided a lower bound of 90% of the optimal growth rate as computed by flux balance analysis (17). Thus, the sampled flux distributions represented sub-optimal flux-distributions, while still modeling fluxes relevant to cell growth and maintenance.

MCMC sampling was used to simulate thousands of feasible flux distributions (referred to here as "points") using the artificially centered hit-and-run algorithm with slight modifications, as described previously (5, 14). Briefly, a set of non-uniform points was generated. Each point was subsequently moved in random directions, while remaining within the feasible flux space. To do this, a random direction is first chosen. Next, the limit for how far the point can travel in the randomly-chosen direction is calculated. Lastly, a new random point on this line is selected. This process is repeated until the set of points approaches a uniform sample of the solution space, as measured using the mixed fraction metric, which measures uniformity by measuring how many of the sample points pass through the middle line of the solution space (18). A mixed fraction of approximately 0.50 was obtained, suggesting that the space of all possible flux distributions is nearly uniformly sampled.

The distributions of sampled fluxes for each reaction were compared between two media conditions. First, flux magnitudes were normalized between each pair of media conditions (media A and B). To do this, a ratio of total flux through the metabolic network was computed and used to normalize each sample point. To compute this ratio, each sample point was taken and the magnitudes of all n non-loop-associated reaction fluxes were summed to acquire a value for the total network flux. For both media conditions, the median total network flux was taken and used to normalize each reaction flux for all sample points in medium B, as follows:

$$v^*_{i,j,B} = v_{i,j,B} \frac{median(\{\sum_{r=1}^{n}|v_{r,1,A}|,...,\sum_{r=1}^{n}|v_{r,j,A}|,...,\sum_{r=1}^{n}|v_{r,p,A}|\})}{median(\{\sum_{r=1}^{n}|v_{r,1,B}|,...,\sum_{r=1}^{n}|v_{r,j,B}|,...,\sum_{r=1}^{n}|v_{r,p,B}|\})},$$

where $v^*_{i,j,B}$, is the normalized flux through reaction i in sample point j under media condition B, obtained after multiplying the sampled flux $v_{i,j,B}$, by the ratio of the median total flux magnitude for the reaction for all p sample points under growth on medium A to the median total flux magnitude for the reaction for all p sample points under growth on medium B. Code for RuMBA is available in the COBRA Toolbox 3.0 (19) and at doi:10.7303/syn15667323.

Metabolic model parameterization
The genome-scale metabolic model of *E. coli* was used with published uptake and secretion rates (20). A few irreversible reactions were removed because they had reversible duplicates in the model. These include: GLCtexi, URIt2pp, URAt2pp, THMDt2pp, KAT1, INSt2pp, INDOLEt2pp, ICHORSi, CYTDt2pp, and ADNt2pp.

To identify all possible simulated media formulations in *E. coli* (SI Appendix, Dataset S4), glucose uptake was set to zero in the model, and flux balance analysis was used to find which of all other

carbon sources could support growth in M9 minimal media. For each of the 174 growth-supporting carbon sources, an uptake rate was set, which was consistent with uptake rate of glucose in the published model (i.e., 8 mmol gDW^{-1} hr^{-1}), normalized by the number of carbons in the metabolite. For example, since glucose has 6 carbons, the uptake rate of glycerol, with 3 carbons, was set as 16 mmol gDW^{-1} hr^{-1}, which is similar to the actual reported glycerol uptake rate in M9 minimal media (21). While this was used to standardize the media conditions, variations in carbon uptake rates did not significantly impact the results presented in this work (SI Appendix, Fig. S16).

Clustering of reaction changes

An m x n matrix with m gene-reactions pairs (predicted to be regulated in at least one media shift; m = 1814) and n total media shifts (n = 15,051) was made, detailing in which shifts each gene-reaction pair is predicted to require regulation (FDR < 0.01). All gene-reaction pairs with at least one significantly regulated enzyme were subjected to k-means clustering (k = 3). Clustering was repeated 100 times with different seed values to find consensus clusters.

Determination of expressed genes

For the analysis in Figure 2b, expression profiles were obtained from previous studies (22–25). The Affymetrix CEL files were normalized using GCRMA, implemented in R. Genes were considered not expressed if they did not have a mean expression level across biological replicates that were significantly higher than the five highest-expression non-*E. coli* negative control probe sets on the array (1-tail t-test; FDR = 0.05). The sets of expressed genes from each study were used to estimate the number of expressed proteins.

Gene expression profiling on glucose and acetate minimal media

Gene expression profiling data were generated from cultures of exponentially-growing E. coli under aerobic conditions on glucose, or acetate M9 minimal media. *E. coli* K12 MG1655 was grown and expression profiled on 0.2% (w/v) glucose M9 minimal media or 0.2% (w/v) acetate M9 minimal media at 37°C. Expression profiling was done using Affymetrix *E. coli* Genome 2 Arrays. Each experimental condition was tested in triplicate in the respective carbon sources using independent cultures and processed following the manufacturer-recommended protocols. Cultures were grown to mid-exponential growth phase aerobically (OD600 = 0.3) in minimal media, supplemented with the appropriate carbon source. Three ml of culture was added to 2 volumes of RNAprotect Bacteria Reagent (Qiagen) and total RNA was then isolated using RNeasy columns (Qiagen) with DNase I treatment. cDNA synthesis, fragmentation, end-terminus biotin labeling, and array hybridization were performed as recommended by the Affymetrix standard protocol.

The Affymetrix CEL files were normalized using GCRMA (version 2.20.0) implemented in R (version 2.11.1). Genes were considered not expressed if their median expression level across replicates was lower than the median value of intergenic (IG) probes. Genes were subsequently removed from further analysis if they were not expressed in any conditions. Differentially expressed genes were determined using the Serial Analysis of Microarrays (SAM) method followed by false discovery rate (FDR) P-value adjustment (FDR = 0.01). Data can be acquired from NCBI/GEO, Accession number GSE108871.

Residue conservation analysis

All protein sequences of 1057 prokaryotic species were acquired from the KEGG database (Release 58.0). Homologs to all *E. coli* proteins containing at least one known PTM were identified by using the Smith-Waterman algorithm. SSEARCH35 of the FASTA suite (26) was used to determine a PID conservation for each post-translationally modified iAF1260 gene in all other genomes. The flags used in SSEARCH35 were '–m9 –E 1 –q –H'. When more than two proteins in one species had the same percent identity, the protein with the lowest e-value was chosen. In the rare case in which multiple proteins from a species had identical % identity scores and e-values, all qualifying proteins were included.

Each metabolic *E. coli* protein with a PTM (n=109) was then grouped with its homologs (median number of homologs for a protein = 911, 25th percentile = 706, 75th percentile = 1000), and the pairwise Smith Waterman alignment between the individual *E. coli* protein and each of the homologs was used to quantify the conservation of post-translationally modified residues, as calculated (i.e., the percent of pairwise comparisons where the aligned residue was identical in the homolog). Conservation of non-modified residues for these amino acids was calculated in an identical fashion. Relative conservation of the PTM residues on each protein was calculated by comparing their conservation to the conservation of non-PTM residues on the same protein, and a statistically significant enrichment of higher conservation was seen for PTM sites on proteins that were predicted to be regulated by RuMBA. From this we found that PTM residues on RuMBA-regulated proteins are highly conserved across the 1057 prokaryotic genomes (compared to non-modified S/T/Y/K residues in the same proteins, $p = 0.0041$; rank-sum test), suggesting that many PTMs are biologically important and therefore conserved.

Salt bridge prediction and measurement of distance from PTMs to active site residues

Protein structures for modified enzymes were obtained from the Protein Data Bank. Potential salt bridges that could be disrupted by a PTM were determined by finding all residues within 4Å of a lysine or serine that could form a salt bridge. Potential new salt bridges were found by searching for basic residues within 8Å of a phosphorylated serine, threonine, or tyrosine. Distances between modified residues and all other amino acids were calculated between centroids of each amino acid. These were used to compare distance between random residues and modified residues with distances between modified residues and functional residues. Functional residues are defined as active sites on proteins, substrate binding sites, and residues which modulate enzyme activity if replaced, and were all acquired from Ecocyc, Uniprot, and the literature. An analysis of the PTMs on 62 proteins with available crystallographic structures (SI Appendix, Dataset S3) showed many PTMs were within 10Å of catalytic site residues (SI Appendix, Fig. S4d-e), and more than half of the 62 metabolic proteins have PTMs that potentially disrupt or create salt bridging interactions (SI Appendix, Dataset S3).

Mutant Growth assays

Wild type *E. coli* and several mutants missing kinases, phosphatases, or acetyltransferases (Δ*aceK*, Δ*cobB*, Δ*pphA*, Δ*yeaG*, Δ*yfiQ*, Δ*yiaC*, Δ*yihE*, and Δ*ynbD*) were obtained from the Keio collection (27). Gene deletion was verified by PCR of the scar region, and strains were subsequently grown

overnight M9 media, supplemented in 2g/L glucose, L-lactate, or inosine in a seeding culture. An aliquot of culture was returned to fresh media such that the OD600 was ~0.03. Cultures were subsequently grown at 37°C with constant stirring. Turbidity was periodically measured at OD600 as a proxy for cell count, and growth rates were computed from OD measurements at mid-exponential phase.

Strains and culture condition for MAGE

The EcNR2 strain (28) used here was a mutant of WT MG1655 *E. coli* in which the λ prophage with the bla gene was introduced via P1 transduction at the bioA/bioB gene locus and selected on ampicillin. In the strain, mutS was also replaced with a chloramphenicol resistance gene (cmR cassette). To enhance electroporation efficiency, EcNR2 was grown in LB-Lennox medium, a low salt LB-min medium with 10 g tryptone, 5 g yeast extract, 5 g NaCl, dissolved in 1 L ddH$_2$O, with 50 μg/ml carbenicillin. For growth screens following MAGE, M9 minimal media was used (Teknova, catalog #M8005), supplemented with 0.1 μM biotin and carbon sources of 1.77 g/L glucose, 4 g/L NaAc*3H$_2$O, or 1.58 g/L inosine. For growth selection, Azure media was also acquired from Teknova (catalog #3H5000) and supplemented with 1.77 g/L glucose. LB-Lennox was used for all LB experiments.

Oligonucleotide design for MAGE

A panel of phosphorylation and acetylation sites were identified from previous studies (29–31), and codons for the phosphorylation sites on serine and threonine or lysine acetylation were changed. Serine and threonine were changed to glutamate to mimic the phosphorylation and an asparagine to mimic the unphosphorylated residue. Lysine was converted to glutamine to mimic the acetylated state and arginine to inhibit acetylation. Codons were selected to require at least two point mutations to the gene sequence in order to ensure that subsequent sequencing of the wild-type and mutant forms would not be masked by sequencing errors. All 90-mer MAGE oligonucleotide sequences are provided for the subset of genes studied (SI Appendix, Datasets S5-S6). MAGE oligonucleotides were synthesized by Integrated DNA Technologies with standard purification. Oligos were designed to target the lagging strand and to minimize secondary structure. MAGE Oligonucleotides also contained four phosphorothioate bases at the 5' end to enhance efficiency as described previously (28). Additional primers were designed to validate a subset of the targets using MASC-PCR (SI Appendix, Dataset S9). Two sets of primers were designed to enable a two-step amplification and library preparation for amplicon sequencing and barcoding of libraries for each sample (SI Appendix, Dataset S10-11). Specifically, the first set of primers were designed to amplify 99 regions containing all mutation sites targeted in our screen. At the 5' end, each forward primer also contained the sequence 5'-CCTACACGACGCTCTTCCGATCTNNNN-3' and each reverse primer contained the sequence 5'-GAGTTCAGACGTGTGCTCTTCCGATCT-3'. The second set of primers were designed to add the remaining sequenced needed for barcoding and next-generation sequencing.

MAGE

MAGE was conducted as previously described (28). Specifically, cultures were initially inoculated with EcNR2 cells into 3 mL of LB-Lennox medium, and cells were grown in sterilized 10-ml polystyrene tubes at 30°C in a rotating incubator under gentle agitation until they reached an OD of 0.4 at 600nm. Cells were then heat shocked at 42°C in a shaking water bath (300 rpm) for 15

minutes. The cells were then chilled at 4°C to make them electrocompetent. One mL of cells was subsequently gently washed through several rounds of centrifugation, buffer exchanges with ice-cold ddH$_2$O, and resuspension. The washed cell suspension was then mixed with 50μL single-stranded MAGE oligos (total concentration of 10μM), which were then electroporated into cells in a 1 mm gap conductive cuvette with the following setup: 1.8 kV, 200 Ω, and 25 μF. The cells were then resuspended with LB-Lennox media in preparation for further rounds of MAGE. Four rounds of MAGE were conducted. Multiplex allele-specific colony PCR (MASC-PCR) was used as previously described (32) to verify mutations and to identify specific mutants for phenotyping.

Screen for PTM mutation fitness

We used pooled screens to assess any changes in cell fitness for each of the 268 genetic changes across multiple media conditions (e.g., LB, Azure defined rich + glucose, Glucose M9, Acetate M9, and Inosine M9) at 30°C as well as for two oscillating conditions (Azure and glucose M9, or glucose and acetate M9). The screens were sampled at 2-4 time points (SI Appendix, Dataset S12) and allele frequencies were quantified by amplifying the genes with PTM sites from the genomic DNA and sequencing the amplicons with next-generation sequencing (NGS). To obtain the final pool with all MAGE mutants, MAGE was conducted in 5 batches, each with approximately 46 different MAGE oligos. MAGE oligos were grouped to ensure that no two oligos targeted within 100 base pairs of each other, to avoid competition between oligos in any one pool. The batches of mutants were combined and subjected to phenotypic selections.

Measurements of the allele frequency were made at three hours after electroporation and pooling and overnight storage at 4°C. Cells pellets were subsequently washed with the medium used in the screen. Cells were maintained at 30°C at exponential growth by serial dilution at regular intervals (about every three doublings; see SI Appendix, Dataset S12 for values). Aliquots were saved at each dilution, and time points were selected for subsequent sequencing and allele frequency analysis at each PTM site.

In addition, oscillatory experiments were designed to test the fitness of the mutants when subjected to periodic changes in the nutritional environment. The oscillating conditions tested here were (i) glucose M9 and glucose-supplemented Azure chemically defined rich media and (ii) glucose M9 and acetate M9 minimal media. The experimental details are as follow. After the initial expansion of cells after the final electroporation and pooling of MAGE batches, the 24 hour time point cells were washed with the starting medium for the oscillation and allowed to grow to an OD of 0.3 at 600nm. At that point, the cell pellet was washed with the second medium and grown therein. Media were then periodically alternated after every 1-2 doublings (see SI Appendix, Dataset S12 for details). More data on doubling times and the results from MAGE screen are found in SI Appendix, Datasets S12-16.

Sequencing, alignment and quantification of variants

Cells were pelleted and DNA was isolated with the MasterPure DNA purification kit (Epicentre), and quantified using Qubit Fluorometric quantification. Sequencing libraries were prepared as follows. Genomic regions targeted by the MAGE oligos were amplified by PCR with KAPA HiFi HotStart DNA polymerase and primers in SI Appendix, Dataset S10. Amplicons were gel-quantified using ImageJ.

For each sample, amplicons were pooled and a second set of PCR primers added barcodes to each sample (SI Appendix, Dataset S11). Samples were gel purified, Qubit quantified, and paired-end sequenced on a HiSeq 2500.

We developed a custom DNA sequence aligner tailored to our MAGE sequencing data to map the reads to the genome and to quantify the MAGE mutants. This was done with our algorithm called K-meR-based Alignment for Multiple mismatchEs per Read (KRAMER; see code at doi:10.7303/syn15667323). This Python-based DNA sequence aligner allows the alignment of sequencing reads with high mismatch frequency to be aligned to a predetermined set of genomic loci. The aligner takes in these loci as input and aligns k-mers derived from each sequencing read. That is, each sequencing read is broken into a set of k-mers of length k (default = 8). Of these k-mers, m (default = 8) must map to a particular locus in order for that read to be mapped. The reads can be broken into overlapping k-mers by specifying o (default = no overlap). For the results shown here, values for k and m were varied to provide the best results; k=8 and m=8 were chosen after sensitivity analysis

After assigning a locus to each sequencing read, each read is compared to the wild type locus to determine if a particular target site in the locus perfectly matches a site in the read. Specifically, in this study, we used MAGE to change at least one codon in each gene. Thus, we searched for perfect matches surrounding the site of the modified codon, and then also looked to see if the site of the modification had the WT codon (Ser,Thr, or Lys), the codon for the PTM mimic (Glu or Gln), or the codon for the amino acid that cannot be post-translationally modified (Asp or Arg). The algorithm uses a parameter called targetsize t (default = 9), which is specified to be the length of the stretch of target DNA that will be matched; in this implementation, t is an odd number from 5 to 43 (e.g., t = 5 would have the target codon with one flanking nucleotide on both ends). To aid in quality control assessment, reads that map to the *E. coli* genome but that do not map to the targeted loci are saved to a separate file, thus allowing further analysis and identification of potential contaminants. Similarly, reads that do not map to the *E. coli* genome or MAGE target sites are written to a file for quality assessment. The implementation provided in this work allows for other optional arguments:

-h, --help   display arguments
-k K   Length of k-mer (default = 8)
-m NUMBERMATCHES   How many k-mers must align for read to map (default = 8)
-t TARGETSIZE   Size of mutation region to map (default = 9)
-o OVERLAP   Should k-mers overlap within a read (default = n)
-d DIRECTORY   Directory of fastqfiles (default = cwd)
-l LOGFILE   Creates a logfile with runtime information
-c NUMCORES   Number of cores (default = 4)

After quantifying the allele frequency for each sample, the allele frequencies were median-normalized, and fold change in frequency was determined by log transforming the allele frequencies and after subtracting the mean frequencies of the control samples (hour 3, pre and post incubation at 4°C).

Identification of covariates modulating the impact of loss of PTM switching

We first identified several biological features for each experiment, gene, and modification site for the MAGE screens. These included the following phenotypic features for experimental samples: 1) whether the experiment was a steady growth condition or oscillating, 2) if the media included glucose or an alternative poor carbon substrate, 3) if the media was M9 minimal media or a rich medium, and 4) the number of doublings seen by the sample after the start of the time course. In addition, we considered, for each PTM, if the modification was phosphorylation or acetylation, and if the modification was on a gene that is predicted to be essential for the given growth condition, based on flux balance analysis simulations (9).

GEE analysis of MAGE screen data

To identify features that best explained the variation in phenotypic impacts of the MAGE mutations, the generalized estimating equation was used with Markov correlation structure using the GEEQBOX package in MATLAB (33). This model identified features that best explained the variation in phenotypic impacts of the MAGE mutations. This model was used to control for the multiple measurements of each experiment while controlling for variation in number of doublings across the samples.

A univariate pre-screening was conducted to assess the contribution of each experimental and biological feature. Since each sample was measured at multiple time points, the generalized estimating equation was used with the Markov correlation structure (34) to account for correlation between time points. Biological features that were not significant in the univariate pre-screening were eliminated from further analysis. Significant variables were subsequently assessed for multicollinearity to eliminate redundant variables. Following the pre-screening, a few features were identified as providing a significant contribution to fitness of mutants in the screen. These included 1) whether the cells were grown in a single growth condition or oscillating media, 2) whether the media contained glucose or a poor carbon source, 3) whether the media was rich or minimal media, 4) if the PTMs were on essential genes for the given growth condition, 5) the proximity of the PTM to active site residues, and 6) whether the PTM is predicted to modulate salt bridges. The significant media conditions were multicollinear and two models were analyzed including only one of the two correlating features. In the final models, analyses comparing poor vs. rich carbon sources and minimal vs. complex media were correlated and therefore were analyzed in separate models.

Molecular Dynamics Simulations

Classical molecular dynamics simulations were performed starting from the crystal structure of all proteins. The individual mutations were manually changed according to the post-translational modification of interest. Parameters for the phosphorylated amino acids were based on the parametrization of Homeyer et al. (35). Using PROPKA (36–38) we estimated that all of the residues adopt the default protonation states. All other non-standard parameters (i.e. for substrates) were calculated per procedures used for the generation of the parm99 parameters and recommended in the AMBER manual. RESP (39–41) charges were generated by performing a three stage RESP fit on two HF/6-31G* optimized structures. Simulations were performed for both substrate-bound and substrate-free states. Each structure was solvated with TIP3P water and, depending on the total

charge of the system, either 14, 7, 20, Na$^+$ ions, achieved system neutrality, (for apo serine hydroxymethyltransferase (GlyA), transaldolase and enolase, respectively), in an orthothrombic periodic box (dimensions for GlyA: 94 × 95 × 114 Å; dimensions for transaldolase: 89 x 74 x 75 Å; dimensions for enolase: 110 x 101 x 89 Å). The particle mesh Ewald (PME) method (42, 43), with a nonbonded cutoff of 12 Å, was used with periodic boundary conditions and the Langevin piston Nosé–Hoover method (44–46) to ensure constant pressure and temperature conditions. For each system, GPU-enabled PMEMD molecular dynamics was performed (47), using the AMBER 99sb force field (48, 49) for 50-120 ns per protein state (i.e., substrate-bound versus substrate-free in wild-type or modified variant proteins).

Supplementary Analysis

Comparison of competitive and allosteric regulation

The RuMBA algorithm aims to identify the enzymes that must be suppressed or activated to divert flux from one pathway to another when the environment shifts from one primary nutrient to a second. Decades of detailed biochemical research has shown that this shift of flux can be achieved through the regulation of metabolic enzymes by small metabolites. Thus, RuMBA should be able to predict the enzymes that must be regulated to adapt quickly to common metabolic shifts. Consistent with this hypothesis, when RuMBA is applied to study the glucose-acetate diauxy, the enzymes that are predicted to be regulated for the shift are significantly enriched in known metabolite-mediated regulation. Indeed, RuMBA predicts twice as many known metabolite-mediated regulated enzymes than expected by chance ($p = 6.5 \times 10^{-10}$; SI Appendix, Fig. S3g). This provides an additional validation of the algorithm. Here we further test the ability for RuMBA to capture more nuanced differences between modes of metabolite-mediated regulation. First, we describe the differences between competitive and allosteric regulation.

There are two fundamental differences between competitive inhibition and allosteric regulation. First, as its name implies, competitive inhibition only inhibits, while allosteric regulation can activate or repress an enzymatic activity. Second in competitive inhibition, the V_{max} (i.e., the maximum reaction rate) is not affected. Rather the dissociation constant, K_d, increases (i.e., there is a decrease in the affinity of the primary substrate to the enzyme). Thus, in a network in which flux is going through the network, a competitive inhibitor might weakly divert flux, but the inhibition will likely increase substrate concentration which will partially or completely overcome the inhibition eventually. Therefore, competitive inhibition is more effective at stabilizing flux through an enzyme.

In allosteric regulation, on the other hand, the binding of an effector molecule occurs at a site that the substrate does not bind. This binding event distally modifies the conformation of the enzyme active site. Thus, the lack of competition from the normal substrate for the binding site of the allosteric makes it such that the V_{max} significantly changes (i.e., the maximum reaction rate either increases or decreases), and the Km can also change.

For these reasons allosteric regulation is important in feedback and feedforward regulation. Moreover, since a simple increase in the substrate concentration cannot overcome the regulation, it is a far better mechanism than competitive regulation in diverting flux to a new steady state flux distribution when the cell experiences a change in its nutritional environment. Therefore, competitive inhibition likely serves more often as a mechanism to stabilize a current steady state, while allosteric regulation may play more of an important role in diverting flux.

Given the nature of how allosteric regulation differs from competitive inhibition in that allosteric regulation is better suited for diverting flux, we anticipate that RuMBA will be better at predicting enzymes that undergo allosteric regulation, as opposed to competitive inhibition. Consistent with this, as described in greater detail in the subsequent section, when RuMBA is applied to study the glucose-acetate diauxic shift (SI Appendix, Fig. S3), its predicted regulated branch points includes

four times as many allosterically-regulated reactions than expected by chance ($p = 2 \times 10^{-8}$). The results are also enriched in reactions that are competitively regulated, to a lesser degree (twice as many as expected, $p = 3 \times 10^{-3}$). Thus, this provides further support that RuMBA is able to identify the enzymes that would need to be regulated to divert flux from one pathway to another following a sudden change in the nutritional environment.

Analysis of glucose-acetate diauxie

In RuMBA, metabolic network flux is simulated using Markov Chain Monte Carlo (MCMC) sampling (13) before and after a change in media. Predicted flux is then analyzed to identify enzymes that, if regulated, could rapidly force flux from one pathway to another to achieve the steady-state flux for the new nutritional environment. This method predicts immediate regulation of metabolic enzymes that could rapidly reroutes flux between pathways, as opposed to slower modes of regulation that would aim to change the topology of entire pathways (e.g., transcriptional regulation and protein degradation).

Using RuMBA, we first studied *E. coli's* canonical diauxic shift from glucose to acetate metabolism. As *E. coli* grows on glucose, acetate and other fermentation products are secreted; as glucose is exhausted, the cells begin metabolizing acetate (50). Under this shift, branch-point enzymes are regulated to direct flux toward pathways that increase cell fitness. A key branch point is the split between the TCA cycle and the glyoxylate shunt (50, 51), where isocitrate is consumed by isocitrate dehydrogenase (ICDH) to synthesize alpha-ketoglutarate or isocitrate lyase (ICL) to synthesize glyoxylate (SI Appendix, Fig. S3a). Our model predicts a significant diversion of flux from ICDH to ICL during the shift from glucose to acetate metabolism, suggesting that this branch point is likely regulated ($p \ll 1 \times 10^{-5}$; SI Appendix, Fig. S3e). Consistent with RuMBA predictions, ICDH is used predominantly during glucose metabolism, while growth on acetate uses ICL to support anaplerosis. In addition to ICDH and ICL, RuMBA predicts that 131 additional proteins could be regulated to aid in the diauxic shift (SI Appendix, Dataset S1). We further compared RuMBA predictions to 1219 experimentally-validated cases in which metabolites regulate enzymes in *E. coli* (SI Appendix, Dataset S2). Most enzymes predicted to require regulation are regulated allosterically by metabolites (SI Appendix, Fig. S3g).

Characterization of growth conditions associated with PTM-based regulation

K-means clustering of *in silico* regulated enzymes

All reactions with predicted regulation that also had known PTMs were clustered (k = 4) to find groups of reactions with similar regulation. For each of these clusters, the 15,051 shifts were also clustered (k = 5) to identify groups of environmental shifts that were most similar in which reactions they regulate. All k-means clustering was replicated 100 times to identify consensus clusters (SI Appendix, Fig. S7).

Reaction cluster 1 accounts for regulation of glycolysis. The environmental shifts clustered into groups that differed in how glycolysis is used and therefore must be regulated. Reaction cluster 2 accounts for regulation necessary to balance flux between the glyoxylate shunt and the TCA cycle

(i.e., primarily between the metabolism of fermentation products and other carbon sources). The 3rd reaction cluster is more sparse and shows some more complex features. The primary feature is that two of the clusters show a clear difference in model-predicted growth rate, following a normalization of the total flux through the metabolic network. Growth conditions enriched in the 1st and 5th environmental clusters show considerably lower growth rates than on the condition they are shifted to (90% and 87% of the shifts, respectively). Several of the substrates in the conditions showing much lower growth are known to decrease the growth of *E. coli* K12 (52).

An assessment of aerobic vs. anaerobic shifts

PTMs are particularly enriched among most changes in media carbon source. Are aerobic/anaerobic shifts subject to the same level of post-translational regulation? To test this we simulated growth under 135 minimal media conditions that could sustain in silico growth under micro-aerobic conditions (O_2 uptake less than 0.001 mmol gDW^{-1} hr^{-1}). Shifts between these and their aerobic analogues were computed, and regulation was predicted as described above. Surprisingly, only 22% of the aerobic/anaerobic shifts were enriched in PTMs ($p < 0.01$). Significantly enriched aerobic/anaerobic shifts are dominant among minimal media with nucleotides or common glycolytic sugars, such as fructose and glucose.

An assessment of reactions that are never predicted to be regulated

According to RuMBA, 623 reactions are reported to never need regulation across all changes in nutritional environment. A comprehensive analysis was done to detail why these reactions are predicted to never require regulation. We find that 2% are reactions that are only connected to metabolites that are connected to more than 30 reactions and are therefore filtered from this work. Furthermore, 10% are reactions that participate in loops and 90% may never participate in flux splits (e.g., they mainly occur in linear pathways).

Correlation of metabolite structural similarity with PTM enrichment

There is a high level of enrichment of PTMs in most nutritional environment shifts modeled in this study. However, the question remains as if some of the shifts fail to show enrichment because the primary carbon substrates in the media are similar, thereby requiring flux diversion at few branch points. To test this, pairwise Tanimoto coefficients were computed for 154 of the 174 consumable carbon substrates in the *E. coli* iAF1260 model. The remaining 20 metabolites did not have appropriate identifiers to unambiguously identify a metabolite structure. Pairwise Tanimoto coefficients were calculated using the software tool Pipeline Pilot (Accelrys Software Inc.) with the FCFP_6 fingerprint descriptor. A coefficient of zero suggests that the two primary carbon substrates are completely different, while a coefficient of 1 suggests that they have the same structure.

There was a significant correlation between the Tanimoto coefficients of the primary carbon substrates in each shift and the PTM enrichment p-value ($R_s = 0.18$; $p = 1\times10^{-178}$). Subsequently, the correlation was computed for subsets of the data to look at local correlation between enrichment of PTMs and Tanimoto coefficients. A moving interval was taken, spanning one order of magnitude of

the PTM enrichment p-values and the Spearman rank correlation was computed and plotted against the geometric mean enrichment p-value (SI Appendix, Fig. S17.a). When the average enrichment p-value is greater than 0.1, a significant correlation between the enrichment p-value and the Tanimoto correlation emerges (SI Appendix, Fig. S17.b), and the Tanimoto coefficients dramatically increase (SI Appendix, Fig. S17.c). Thus, on average, when a shift exhibits a poor enrichment of PTMs among the model-predicted regulated enzymes, this results from high structural similarity between the primary carbon substrates. This high structural similarity likely leads to similarity in how the carbon substrates are metabolized. Therefore, few nodes will require significant diversion of flux.

Deleting genes encoding modifying proteins impacts *in vivo* cell fitness

Another way to demonstrate the important role that PTMs play in regulating bacterial metabolism is to assess the impact of removing the biological machinery used to modify enzymes. It is widely recognized that protein kinases, phosphatases and acetyltransferases are the primary drivers for regulation through protein modification (53). To investigate whether different types of nutrient environments require the activity of specific kinases, phosphatases, and acetyltransferases, we tested the impact of deleting specific modifying enzymes on growth rate in different nutrient conditions.

Mutant strains of *E. coli* lacking kinases, phosphatases, and acetyltransferases (Δ*aceK*, Δ*cobB*, Δ*pphA*, Δ*yeaG*, Δ*yfiQ*, Δ*yiaC*, Δ*yihE*, and Δ*ynbD)* were grown on M9 minimal media with different supplementation (e.g., glucose, L-lactate, or inosine). We find that the cellular fitness of these strains depends strongly on the nutrient environment, as certain mutants exhibited faster growth (relative to wild-type) in certain nutrient conditions but slower growth on others (SI Appendix, Fig. S18). Although deleting modifying proteins has an overall marginal effect of organism fitness (the magnitude of the difference in growth rate is around 7%-9% of the WT growth rate), we observe a unique relationship between nutrient condition and the modifying proteins themselves that regulate downstream post-translational regulation processes.

Protein modifications influence *in vivo* fitness in dynamic environments

What properties determine whether a PTM influences cellular fitness? To address this, we performed a global analysis of the MAGE screen data using a generalized estimating equation (GEE) to identify how modifications impact fitness in response to attributes of the modified proteins or the environmental conditions. The GEE is a semiparametric regression technique and controls for correlation across samples with multiple time points. First, the GEE showed that modifications impacting growth rate are often located on proteins that are essential for growth in each medium tested (SI Appendix, Fig. S9b). Indeed, the modification of PTM sites associated with essential genes impacted growth rate, on average, by 5%, demonstrating that many PTMs locate to proteins that directly impact growth (p = $9x10^{-7}$). Second, MAGE modifications impacting growth often occur at structurally-relevant positions in proteins (e.g., salt bridge residues or near active site residues, p $<< 1x10^{-10}$ and p=$9x10^{-7}$, respectively), which could impact growth by controlling enzyme activity. Modifications of PTM sites associated with salt bridges impacted the growth rate by 5%, and for

every 10Å closer a PTM site is to an active site residue, there is an additional 2% impact on growth rate. Thus, on average, the PTM sites are positioned on proteins to be able to regulate growth.

Modifications should be able to influence the cell's ability to adapt to environmental shifts. Specifically, the effects of modifications were tested within two types of nutrient environments, static (a single medium) and oscillating (media conditions were changed periodically, e.g. between glucose and acetate). The GEE demonstrated that when MAGE mutations force proteins to remain in a single PTM state (thereby preventing transient control at these sites), they more significantly impact the growth rate in oscillating environments compared to static environments (SI Appendix, Fig. S9b). Specifically, freezing the PTM state with a MAGE mutation decreased the growth rate on oscillating conditions by an average of 3% more than in static conditions ($p = 3\times10^{-3}$), with many individual sites showing a greater impact.

Altogether, the genetic screens show that PTMs are functionally relevant and that they regulate specific enzymes *in vivo*. Specifically, the GEE analysis of the MAGE data showed that PTMs are best positioned on enzymes in the metabolic network to control cell physiology under the growth conditions tested. Furthermore, the modifications impact *in vivo* cellular fitness most when the primary nutrients change, consistent with the computational predictions, wherein PTMs localize to model-predicted branch points to facilitate pathway switching when media changes. This reinforces the notion that PTMs may be used by the cell to quickly adapt to rapidly changing nutrient conditions.

Molecular Dynamics Simulations

We mainly considered: (i) the availability of PTM data (e.g. proteins with known PTMs); (ii) the conservation of the residues that are modified, when comparing to homologous proteins; (iii) availability of crystallographic structures; (iv) the number of total shifts where proteins required regulation, as predicted by RuMBA. In each case, we performed long-time MD simulations (on the order of 100-120 nanoseconds) for (i) the wild-type (WT) protein and (ii) each of the associated modified states by phosphorylation or acetylation. Further, we compared the protein conformational landscapes for WT and each of the modified proteins, both in the presence and absence of bound substrates and cofactors. Information for all three proteins that were simulated in this study is provided in SI Appendix, Dataset S7.

Serine hydroxymethyl transferase

Starting from the crystallographic structure of SHMT (PDB entry 1DFO), we performed explicit MD simulations of wild-type (WT) protein, in both monomeric and dimeric forms (SI Appendix, Fig. S11) as well as in substrate-bound and substrate-free complexes. The total number of protein atoms 12,241 solvated in 73,359 explicit solvent TIP3P water molecules. In addition, we modeled proteins with different post-translationally modified lysines (K54, K250, and K354) to understand the effect of these changes with respect to overall structure, ligand binding and dimer formation. To

probe local changes in protein structure, we computed intermolecular interaction distances between key catalytic residues throughout the combined MD trajectory time of 0.5 μs.

Large-scale changes in protein structure are seen in the N-terminal domain (residues 1–34), which mediates intersubunit contacts and folds into two α-helices and one β-strand (SI Appendix, Fig. S11). The second N-terminal domain or the "large" domain (residues 34 to 290), which binds PLP, has most of the active site residues and folds into an α–β–α sandwich containing nine α-helices wrapped around a seven-stranded mixed β-sheet. RMSD was calculated over 30-120 nanosecond throughout molecular dynamic trajectories in both enzyme-substrate complex (PLG and FFO) and substrate-free states (SI Appendix, Fig. S11a-b). In each case, the reference frame is taken to be the equilibrated protein.

A Z-score-based analysis was carried out to determine the most significantly shifting interaction distances between wild-type and modified proteins. The Z-score calculation was repeated 1,000 times, and the mean value is reported. Z-scores were computed to determine which shifts in intermolecular interaction distances (between wild-type SHMT and modified variants) were ranked significant over others (if z-scores are greater than 1.9). We find that acetylation of K54 (in substrate-free protein) shifts the interaction distances of H203-K229 (z=4.1) most significantly. Furthermore, in the enzyme-substrate complex, we find that acetylation of K308 affects the intermolecular distance between H203-T227 (z=3.1), which are key active site residues in the vicinity of PLG and FFO (SI Appendix, Fig. S11a). Interestingly, the modification site that most influences *in vivo* and *in vitro* enzyme activity (K54) shows the greatest shift (2-3 Å, on average) in intermolecular interaction distance for the intermolecular interaction between K229 and H203 (SI Appendix, Fig. S12a (right plot)). Modification of K54 also completely disrupts a persistent salt bridge interaction with E36 of the neighboring subunit. In the WT protein, the interaction distance between K54 and E36 is, on average, 3.8 Å, whereas, upon acetylation of K54, it shifts to 7.5Å (SI Appendix, Fig. S12b). This shift also influences the orientation of other nearby active site residues, H203, H228 and K229 and Y55, which is essential for correct positioning of the covalently-attached pyridoxal 5'-phosphate (PLP) cofactor (54). Finally, as SHMT functions as a homodimer, acetylation of K54 may also decrease the efficacy of dimerization, which would inhibit catalysis by decreasing the affinity of PLP for the enzyme (55).

Transaldolase

We examined the crystal structure of transaldolase from *Thermoplasma acidophilum* (PDB entry 3S1U) to find that many of the residues surrounding the binding channel were in structurally equivalent positions (SI Appendix, Fig. S13a). Most importantly, the outermost C-terminal helix (α11), which runs across the barrel of a neighboring subunit packs against helix α2 and partially covers the active site. In this organism, the α11 and the α2 form a part of the dimer interface, yet the interactions (and locations) of key residues in this interaction network (e.g., P36 and F302 as well as K308-D305-S37 in *E. coli*) are conserved. As shown in SI Appendix, Fig. S13a, the catalytic lysine (Lys86, located in strand β4) corresponds to K132 in *E. coli* and is solvent-accessible and points toward the interior of the barrel. As for the case in *E. coli*, the active site resembles previously characterized transaldolases. As discussed previously, the presumed catalytic water is bound within hydrogen-bonding distance to Nζ of K86. As illustrated here, the "closed" form is

characterized by a narrowing of the substrate channel by up to 4 Å compared to the open form. In this conformation, F208 (which belongs to domain-swapping helix α11 of the neighboring subunit and corresponds to F302 in *E. coli*), is found to hydrophobically interact with P29, hindering free exchange of the catalytic water with bulk solvent. We also note that the placement of F132 is conserved across these organisms (F178 in *E. coli*), which is found to consistently interact with F302 (F208 in *T. acidophilum*) in the open state. Interestingly, certain post-translational modifications appear to induce changes in the intermolecular interaction network such that the interactions between F178 and F302 occur more frequently, shifting the state of the enzyme to the "open" conformation. One such modification is the phosphorylation of S226 in *E. coli* (which corresponds to S167 in *T. acidophilum*).

To study the effect of several known PTMs on the structure and reactivity of transaldolase, we performed 10 independent atomistic simulations for five different post-translational modifications and sampled protein conformational space over 50-100 nanoseconds per protein (a combined MD simulation time of 1.05 μs for all WT and modified proteins). The total number of protein atoms is 5,874, which were solvated in 44,760 explicit solvent TIP3P water molecules. We find that the root-mean-squared deviation (RMSD) relative to the crystallographic (WT) protein varies among modified proteins, ranging from 1.7 to 2.6 Å. We observe most deviation resulting from complexation with its native substrate, D-sedoheptulose-7-phosphate (s7p). The modified proteins with the largest degree of deviation from the crystallographic structure include those with modifications on T33 and S37 (SI Appendix, Fig. S13b).

Using a statistical analysis, we find that the interaction distances between F178-D305 and D305-K132 are anti-correlated in wild-type transaldolase (Pearson coefficient of -0.74), whereas when transaldolase is phosphorylated at S226 or T33 or acetylated at K187 or K308, these interaction distances are no longer correlated. Furthermore, we find that interaction distances between D305-K308 and F302-K308 are strongly correlated in wild-type (0.8) and acetylation of K08 strongly disrupts the dynamic of these pairs of residues. SI Appendix, Dataset S8 provides a list of all per-residue interaction distances that are no longer correlated as a result of residue modification. A Z-score-based analysis was carried out to determine the most significantly shifting interaction distances between wild-type and modified proteins. The Z-score calculation was repeated 1,000 times, and the mean value is reported. Z-scores were computed to determine which shifts in intermolecular interaction distances (between wild-type transaldolase and modified variants) were ranked significant over others (if z-scores are greater than 2.0). We find that phosphorylation of S226 (in substrate-free protein) shifts the interaction distances of P36-F302 (z=3.9), P36-F178 (z=4.5) and S37-K132 (z=2) most significantly. Furthermore, in the enzyme-substrate complex, we find that acetylation of K308 affects the intermolecular distance between F178-D305 (z=2.2), which spans the opening of the active site. These interaction distances correspond to characteristic intermolecular interactions important in the "open" and "closed" states of the enzyme. The changes in intermolecular bonding distances for key residues lining the substrate binding channel are plotted in SI Appendix, Fig. S13c (e.g., F178-D305 (left), P36-F178 (center), and P36-F302 (right)). Interestingly, the modification site that most influences *in vivo* and *in vitro* enzyme activity (S226) shows the greatest shift (4-5 Å, on average) in intermolecular interaction distance for two out of three of the intermolecular interactions (P36-F178 and P36-F302).

Enolase

We performed six independent atomistic simulations for two different post-translational modifications (phosphorylation of T375 and T379), considering substrate-free (apo) and enzyme-substrate (bound to D-Glycerate 2-phosphate (2pg)) complexes. For all simulations, we simulated enolase in its homodimeric state, starting from the crystallographic structure (PDB 3h8a). We sampled protein conformational space over 50-100 nanoseconds per protein (a combined MD simulation time of .45 μs for all WT and modified enolase proteins). The total number of protein atoms is 15,108, which were solvated in 77,244 explicit solvent TIP3P water molecules. To simulate the activated complex, we used restraints to maintain the interatomic distances of residues in the active site with two magnesium ions. In the substrate-free complex, we used restraints to maintain the interatomic distances of residues in the active site with one magnesium ion. We find that the root-mean-squared deviation (RMSD) relative to the crystallographic (WT) protein ranges from 0.8 to 1.4 Å, for wild-type and modified enolase variants. We observe minimal deviation resulting from phosphorylation of T375 and/or T379 (SI Appendix, Fig. S13) in both apo and complexed states.

We find that certain PTMs in enolase are likely to modulate activities beyond a protein's metabolic role. An illustrative example is the phosphorylation of two residues, T375 and T379, in enolase. Classical molecular dynamics simulations of enolase modified at each of these sites do not show significant changes in protein structure or molecular interaction networks with respect to wild-type protein (SI Appendix, Fig. S14). However, the phosphorylation of these residues likely affect the complexation of enolase with RNAse E, which binds in a conserved cleft within the interface homodimer (PDB entry 2fym). As shown in SI Appendix, Fig. S10 (bottom panel), T375 is located within 5Å of a negatively charged aspartate residue in the conserved RNAse E binding domain. Upon phosphorylation, it is likely that the repulsion of these two negatively charged moieties impact the binding stability of the RNAse E mediated assembly and, thus, the recruitment of enolase via the RNA degradosome (56, 57). Consistent with this hypothesis are the effects of modifying T375 *in vivo,* which appear to positively impact the overall organism's fitness in certain substrate shifts.

Robustness of metabolic modeling analyses with the addition of novel PTMs

Shortly after the completion of the MAGE-Seq experiments, a few new phosphoproteomic data sets were published that captured serine, threonine, and tyrosine phosphorylation sites (58–60), increasing the number of published phosphorylated metabolic enzymes in *E. coli* by 98. To evaluate the robustness of our results, we tested if the addition of these 98 newly identified phosphoproteins impacted the conclusions from the RuMBA analysis. Thus, we evaluated the enrichment of post-translationally modified proteins within the RuMBA predictions across all 15051 nutrient shifts. Prior to their addition, we found that 91.8% of the shifts were significantly enriched in enzymes were a PTM was detected. When accounting for the 98 additional phosphoproteins, this increased slightly to 92.0%. Thus, our results in this study are robust to the addition of additional post-translationally modified enzymes.

Supplementary Figures

Legend of Supplementary Figures

Fig. S1: Conceptual categorization of different types of enzyme regulation

Fig. S2: Conceptual illustration of condition-specific shifts in the flux solution space

Fig. S3: RuMBA predicts regulatory nodes in metabolism, and predictions are regulated using small molecule-based regulation.

Fig. S4: PTMs enriched on enzymes that are predicted to require regulation

Fig. S5: Enzymes that are predicted to be regulated are associated with known regulation by metabolite-mediated allosteric regulation and measured PTMs

Fig. S6: Enzymes that are predicted to be regulated are differentially expressed between glucose and acetate minimal media

Fig. S7: Clusters of PTM-associated RuMBA predictions provide insight into environmental shifts that are associated with the regulation of different pathways

Fig. S8: GlyA K54R increased growth rate following the diauxic shift after glucose was exhausted

Fig. S9: MAGE analysis across conditions

Fig. S10: Atomistic details for GlyA, transaldolase and enolase

Fig. S11: Characteristic structural changes in serine hydroxymethyl transferase and modified variants (PTMs)

Fig. S12: Changing intermolecular interaction distances in modified SHMT proteins.

Fig. S13: Characteristic structural changes in transaldolase and modified variants (PTMs).

Fig. S14: Characteristic structural changes in enolase and modified variants (PTMs).

Fig. S15: Enzyme *in vitro* assays for activity measurements of proteins and modified variants.

Fig S16: Moderate variations in metabolite uptake rate in the metabolic model have little impact on RuMBA predictions

Fig. S17. Conditions with high metabolite structural similarity are depleted of PTMs.

Fig. S18. Deletion of kinases, phosphatases, and acetyltransferases can have positive or negative impacts on growth, depending on the media.

Fig. S19. Inclusion of new phosphoproteins shows that PTMs continue to be enriched in RuMBA results.
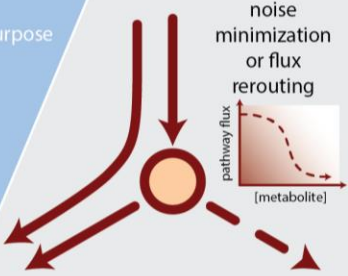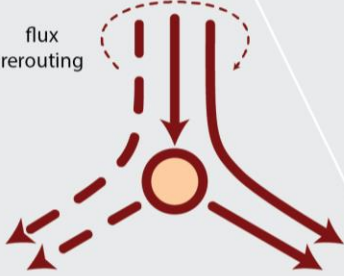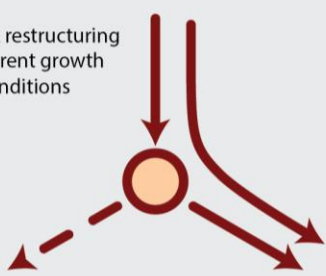
	Small molecule binding (allosteric/competitive)	Post-translational modification of enzymes	Regulation through protein degradation	Transcriptional regulation
Time scale	Immediate	Fast	Slow	Slow (minutes)
Reversibility	Fully reversible	Often reversible	Irreversible	Irreversible
Cost	~ none	1. Maintenance of modifying enzyme 2. Modification	1. Modification to signal for degradation 2. Degradation machinery maintenance 3. Synthesis of mRNA/protein when microenvironment changes again	1. Activation of signaling cascade 2. Usage of transcription / translation machinery 3. Synthesis of mRNA/protein when microenvironment changes again 4. Synthesis of other components in regulon
purpose	noise minimization or flux rerouting	flux rerouting	network restructuring for current growth conditions	

Fig. S1. Conceptual categorization of different types of enzyme regulation. The different timescales, and the reversibility of different types of regulation conceptually lend themselves to different types of responses. Small molecule-mediated competitive inhibition is particularly well-suited to stabilize flux for a given metabolic state or temporarily reroute flux. PTMs, however, can rapidly shift flux from one pathway to the next by activating or inhibiting enzymes at branch points. This rerouting can be maintained until upstream signals cause the removal of the PTM or the enzyme is degraded. The speed of PTMs allow for immediate response to add or remove a PTM. Protein degradation and transcriptional regulation take longer to implement, are costlier in the sense of bioenergetics and required building blocks, and are irreversible. Thus, they are ideal for long and sustained changes in the nutritional environment. It is anticipated that due to the cost of maintaining unnecessary proteins, it is likely that PTMs will more often suppress the activity of highly used proteins to divert flux to pathways that show lower levels of expression.
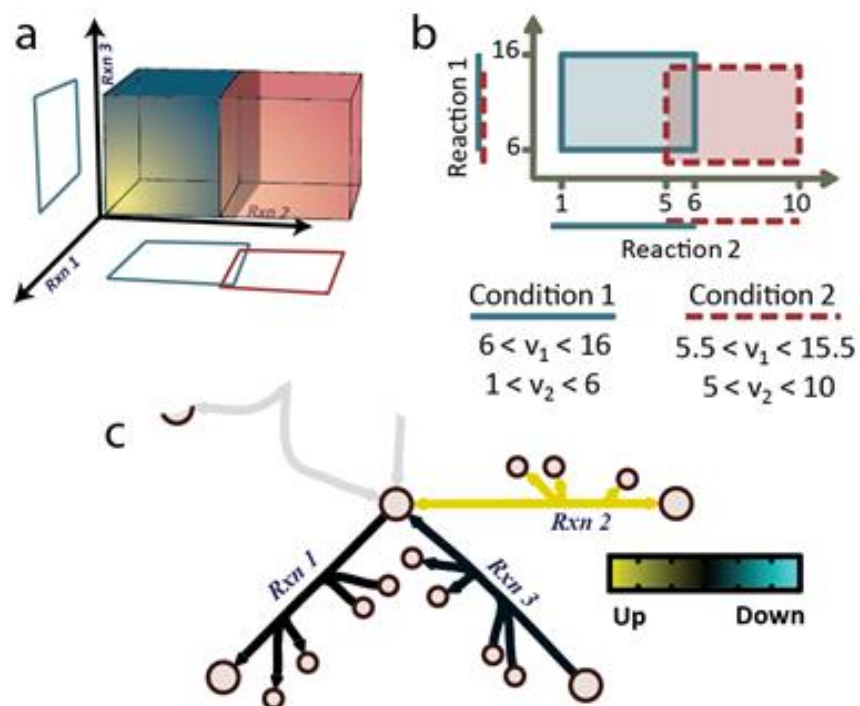
Fig. S2. Conceptual illustration of condition-specific shifts in the flux solution space. (a) Constraint-based modeling employs governing constraints to define a space of feasible phenotypes, which are represented by allowable steady-state fluxes for each reaction. When growth conditions change (e.g., a change in carbon source, or aerobicity), the space of feasible fluxes can change. **(b)** For example, reaction 2 shows a change in the range of feasible flux levels under the new growth condition, which can be shown in the metabolic map **(c)**. These changes can be mapped back to the genes and proteins associated with each reaction. This approach is further described elsewhere (6).
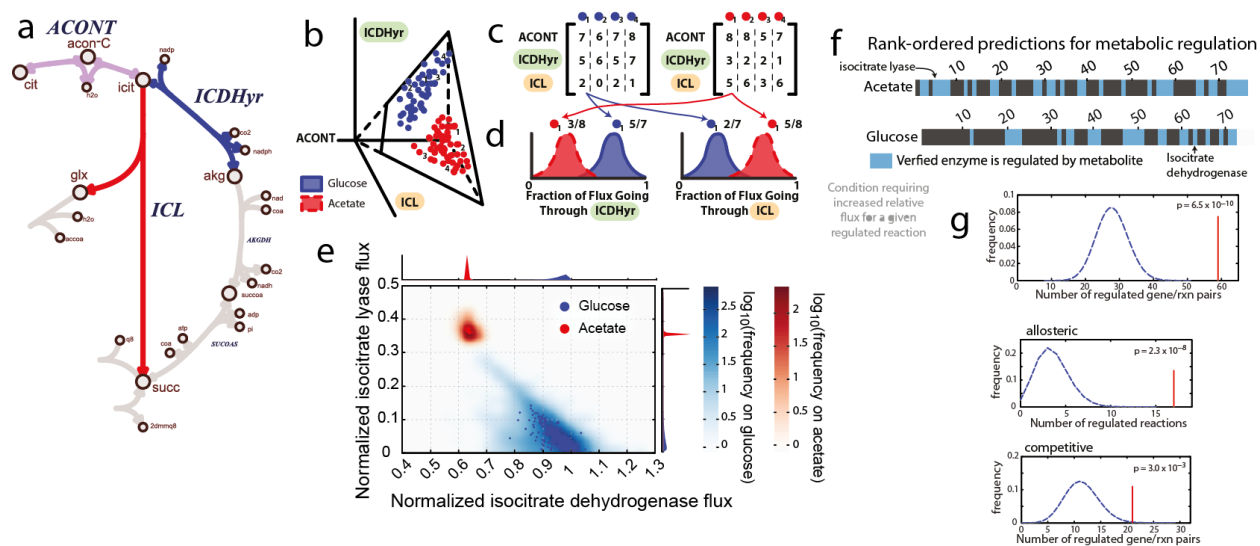
Fig. S3. RuMBA predicts regulatory nodes in metabolism, and predictions are regulated using small molecule-based regulation. (a) For the split between isocitrate dehydrogenase (ICDHyr) and isocitrate lyase (ICL), RuMBA predicts that different branches will be used under different growth conditions. **(b)** When *E. coli* is grown on glucose, RuMBA predicts that most of the flux through aconitase (ACONT) continues in the TCA cycle through (ICDHyr). However, when switched to acetate minimal media, a significant amount of flux is siphoned off into the glyoxylate shunt through ICL. **(c)** RuMBA predicts this shift by using MCMC sampling to compute a uniform sample of feasible steady-state flux values (points) for all reactions that produce or consume a metabolite of interest, such as isocitrate. **(d)** The fraction of flux that goes through each branch is computed for each point, yielding a distribution of fractional split values for both conditions. RuMBA then identifies all reactions where the fractional split values significantly change, suggesting that there is a high likelihood that regulation of the enzyme would enable rapid and efficient shifting of the flux to a new pathway. **(e)** Randomly-sampled *in silico* flux distributions for growth on acetate and glucose show that isocitrate dehydrogenase is used for growth on glucose minimal media, while flux is diverted to isocitrate lyase when the cell is metabolizing acetate. **(f)** The RuMBA algorithm identifies a rank-ordered list of reactions and their associated enzymes that require significant regulation to redirect flux when rapidly shifting from one carbon substrate to another. Many of these predictions are enzymes that are known to undergo metabolic regulation (blue; see SI Appendix, Dataset S1 for identities of enzymes). **(g)** The RuMBA predictions are significantly enriched in known metabolically-regulated enzymes, particularly for allosteric regulation (hypergeometric test).
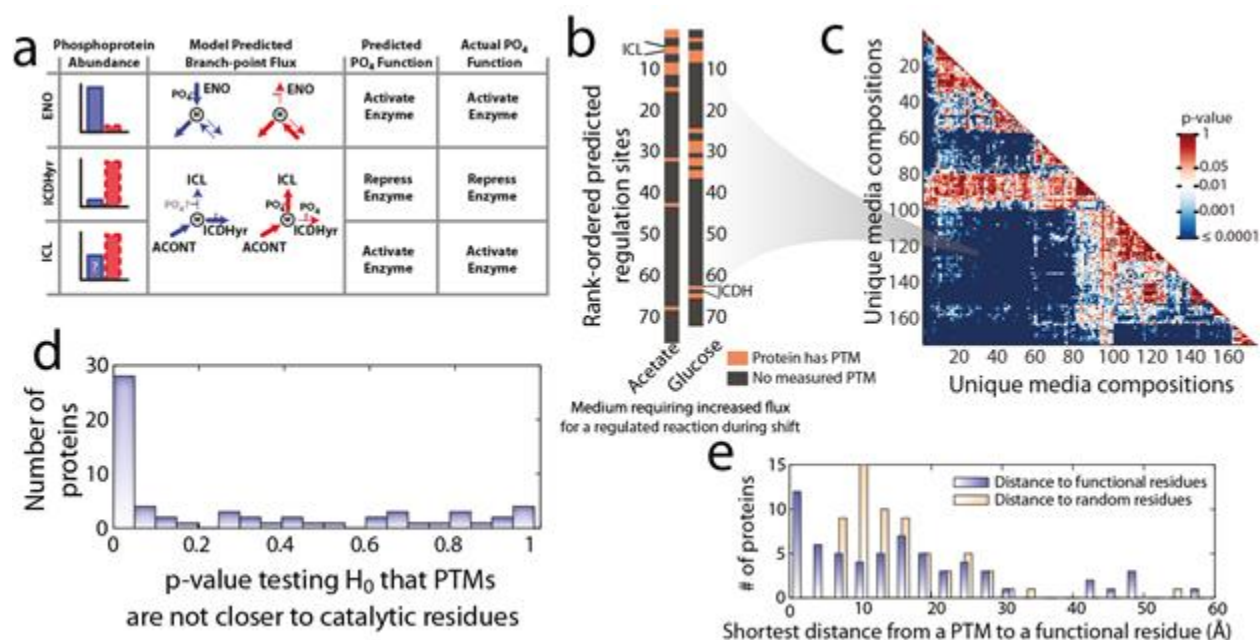
Fig. S4. PTMs enriched on enzymes that are predicted to require regulation. **(a)** RuMBA results can be compared to experimentally measured changes in enzyme activity for the given growth conditions, yielding predictions for the metabolic regulatory function of phosphorylation events. When compared to known functional phosphorylation events, RuMBA accurately predicts experimentally measured effects. **(b)** Beyond the three examples with known regulation, the RuMBA algorithm provides a rank-ordered list of additional enzymes that should be regulated to redirect flux when rapidly shifting between nutrients. PTMs (salmon) are enriched in RuMBA predictions for the glucose-acetate diauxie, and for **(c)** 92% of the pairwise shifts between 174 media conditions (color indicates the significance of enrichment of PTMs on proteins in the RuMBA predicted list of regulated enzymes). **(d)** Across 62 metabolic proteins with known structures, PTMs were significantly closer to functional residues (i.e., catalytic or substrate binding residues) than expected >48% of the proteins, and **(e)** 37% of the proteins had PTMs closer than 10Å to functional residues.
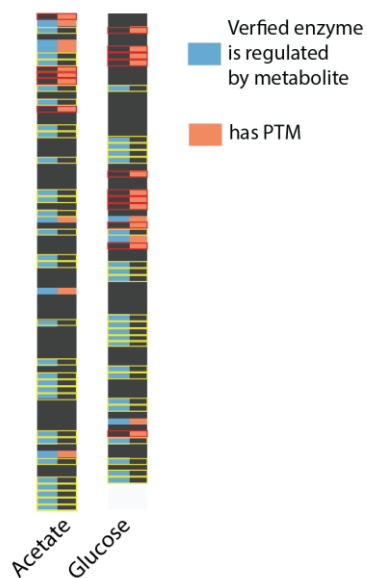
Fig. S5. Enzymes that are predicted to be regulated are associated with known regulation by metabolite-mediated allosteric regulation and measured PTMs. RuMBA-predicted changes for the glucose-acetate diauxie are enriched for known instances of metabolite-mediated allosteric regulation. PTMs are also enriched among the most highly regulated enzymes. While there is some overlap of enzymes that are regulated by small molecules and PTMs, the two are more complementary, with either metabolites regulating the enzyme (yellow outline) or PTMs (red outline).
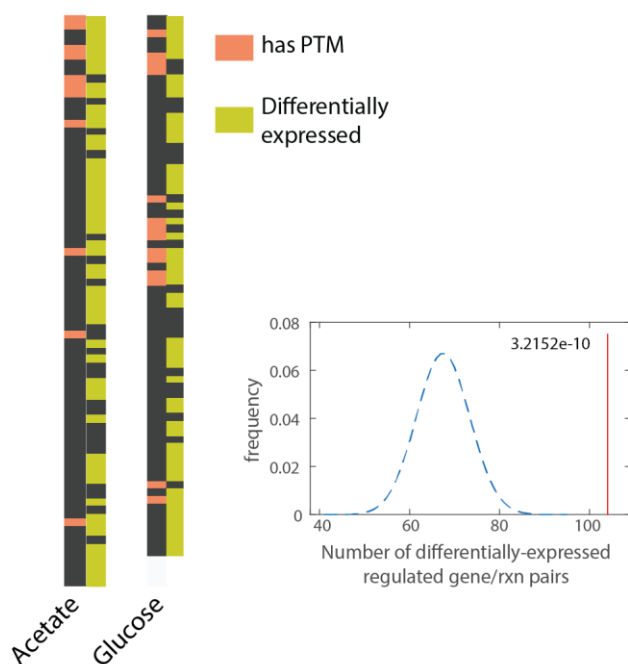
Fig. S6. Enzymes that are predicted to be regulated are differentially expressed between glucose and acetate minimal media. Proteins with PTMs were enriched among enzymes that were predicted to require regulation in the glucose-acetate shift, based on RuMBA. Furthermore, the PTMs were found more often on the enzymes that showed the strongest need for regulation (near the top of the list, in which each row of the column is a different enzyme-reaction pair). See Supplementary SI Appendix, Dataset S1 for list of enzymes and their relative need for regulation). Similarly, when comparing gene expression levels of E. coli grown on glucose or acetate minimal media, the RuMBA-regulated enzymes were significantly enriched in differentially expressed genes (p = 3 x 10^{-10}). However, the differentially expressed genes were not significantly associated with the enzymes that had the highest RuMBA scores (p > 0.15). Thus, the PTMs are positioned more towards the most critical enzymes, but the cells eventually differentially express most of the RuMBA predicted regulated enzymes when shifted from glucose to acetate minimal media.
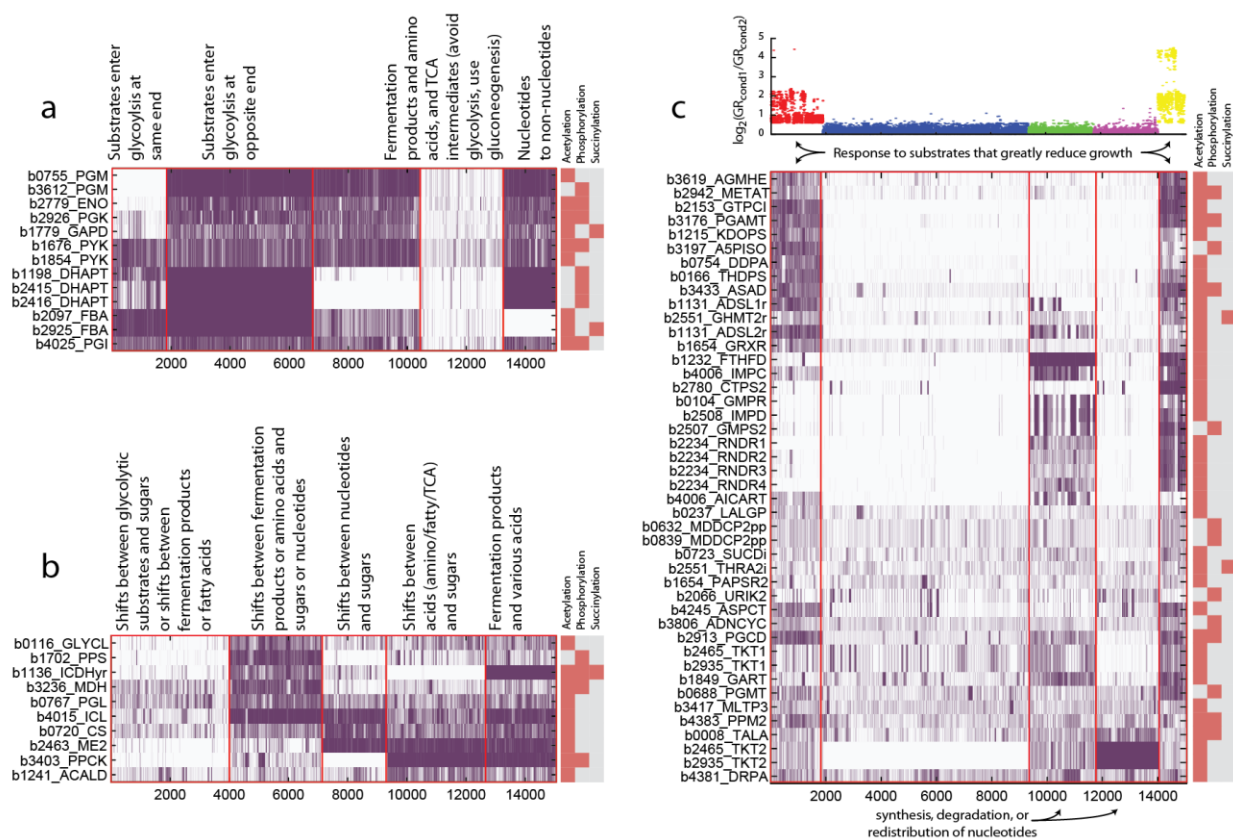

Fig. S7. Clusters of PTM-associated RuMBA predictions provide insight into environmental shifts that are associated with the regulation of different pathways. Clustering of RuMBA predictions identifies modules with similar regulation patterns, including glycolysis, the glyoxylate shunt, and nucleotide metabolism and the pentose-phosphate pathway. Regulation of nucleotide metabolism is particularly high when model-predicted growth rates significantly change between two media conditions. Within these clusters, **(a)** shifts that significantly changed glycolysis included shifts between sugars and organic and amino acids. **(b)** The glyoxylate shunt was usually regulated in shifts between fermentation products and amino acids, sugars, or nucleotides. **(c)** Nucleotide

metabolism and the pentose phosphate pathway were frequently required regulation when shifts involved nucleotides and various acids, and often involved significant changes in growth rates between the substrates. Purple elements in each heatmap signify that the enzyme (for the associated reaction) is predicted by RuMBA to be regulated. Red boxes to the left signify if the enzyme has a known PTM site.
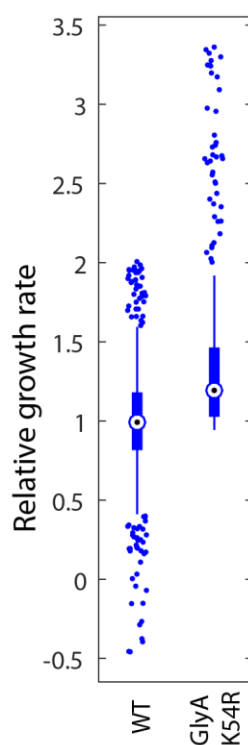

Fig. S8. GlyA K54R increased growth rate following the diauxic shift after glucose was exhausted. After ~15 hours, all glucose was exhausted for wildtype and GlyA K54R mutants, and so cells shifted to consume the small amount of remaining fermentation products (Figure 3b). At each time point following the shift, a growth rate was computed, based on change in OD. For each time point growth rates were normalized based on the average growth rate for the wildtype cells at that time point (relative growth rate of 1 = average WT growth rate for a given time point). At almost all time points, the GlyA K54R showed a higher growth rate than the average WT growth rate after glucose was depleted.
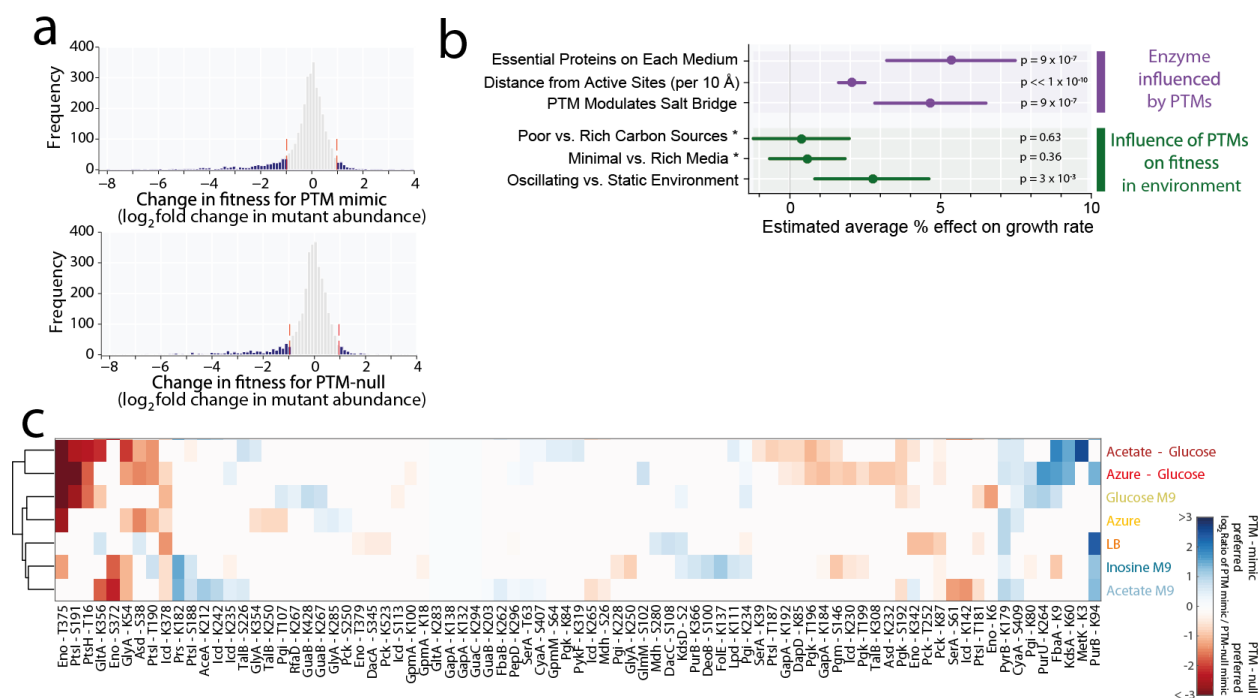
Fig. S9 MAGE analysis across conditions. (a) Many combinations of PTM mutations and media conditions showed >2 fold changes in fitness, and 88% of PTM sites showed a phenotype (>2 standard deviations) in at least one condition. **(b)** A GEE analysis quantified factors influencing cell fitness after MAGE inhibited the change of PTM state. MAGE mutations did not significantly impact growth differently on rich or minimal media. In comparison, on average, oscillating media led to a 5% greater decrease in growth rate (compared to static growth conditions) when the PTM state was frozen in the PTM-mimic or PTM-null state. The localization of PTMs near active site residues on essential proteins significantly influenced cell growth, especially when predicted to modulate salt bridges. Thus, the PTMs are poised to regulate the key proteins to immediately adapt to changes in growth condition. **(c)** Full heatmap showing that most mimic/null preferences were condition specific, with oscillating media showing more extensive preferences (only preferences with p <0.1 shown).
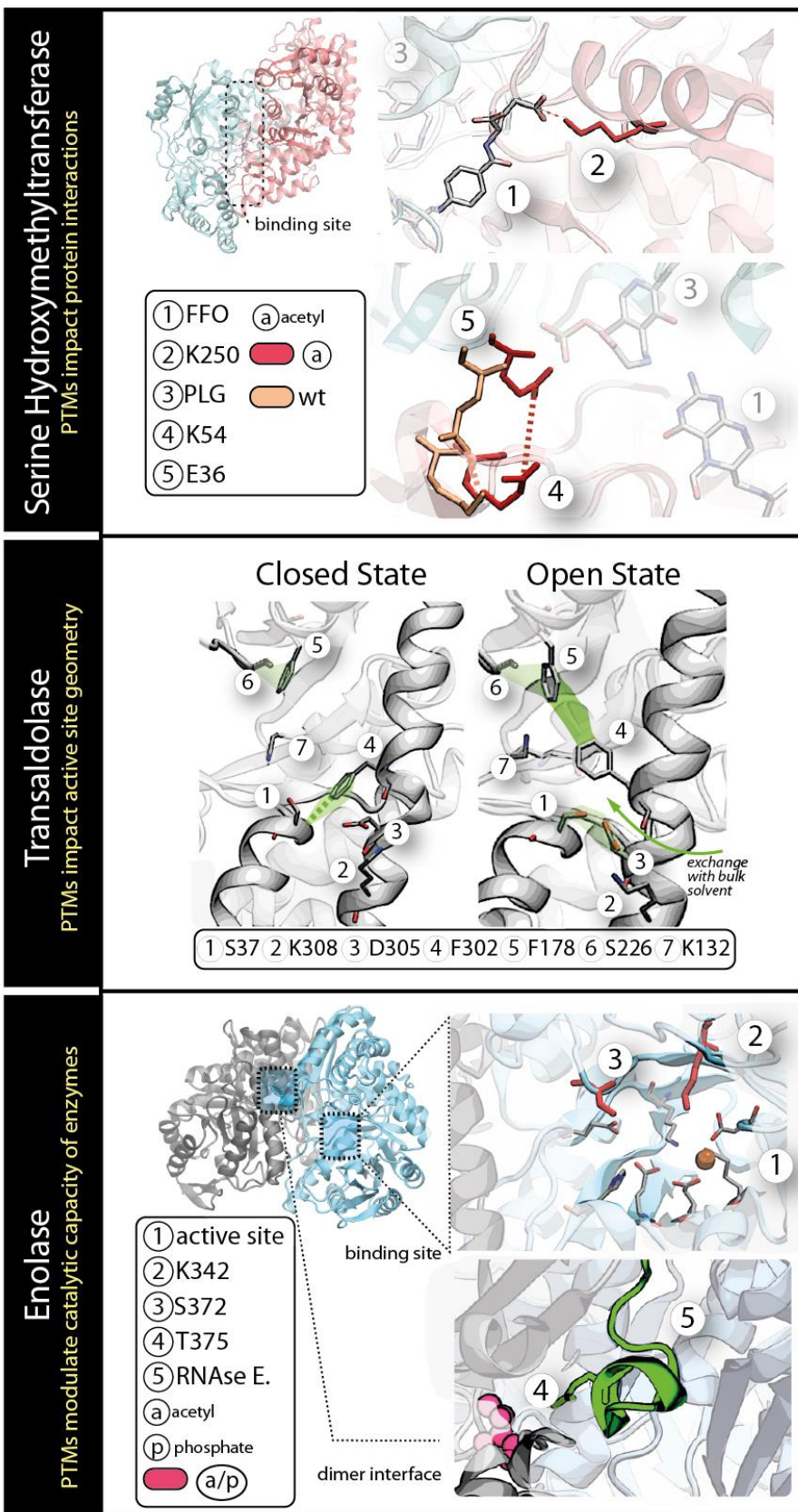
Serine Hydroxymethyltransferase
PTMs impact protein interactions

1. FFO — (a) acetyl
2. K250 — (red) (a)
3. PLG — (orange) wt
4. K54
5. E36

Transaldolase
PTMs impact active site geometry

Closed State — Open State

exchange with bulk solvent

1 S37 2 K308 3 D305 4 F302 5 F178 6 S226 7 K132

Enolase
PTMs modulate catalytic capacity of enzymes

1. active site
2. K342
3. S372
4. T375
5. RNAse E.
(a) acetyl
(p) phosphate
(pink) (a/p)

binding site

dimer interface

Fig. S10. Atomistic details for GlyA, transaldolase and enolase. (top) Acetylation of K250 of serine hydroxymethyl transferase (GlyA) disrupts a salt-bridge, leading to decreased THF binding **(middle)** Interactions between six active site residues determine the accessibility of the active site in transaldolase. A closed state involves D305-K308 and P36-F302 interactions. In an open state, S37 interacts with D305 and K308, while F302 interacts with F178, to allow the exchange of catalytic water. **(bottom)** The catalytic site of enolase (PDB 3h8a) and (bottom) enolase in complex with a minimal binding segment of RNase E (PDB 2yfm).
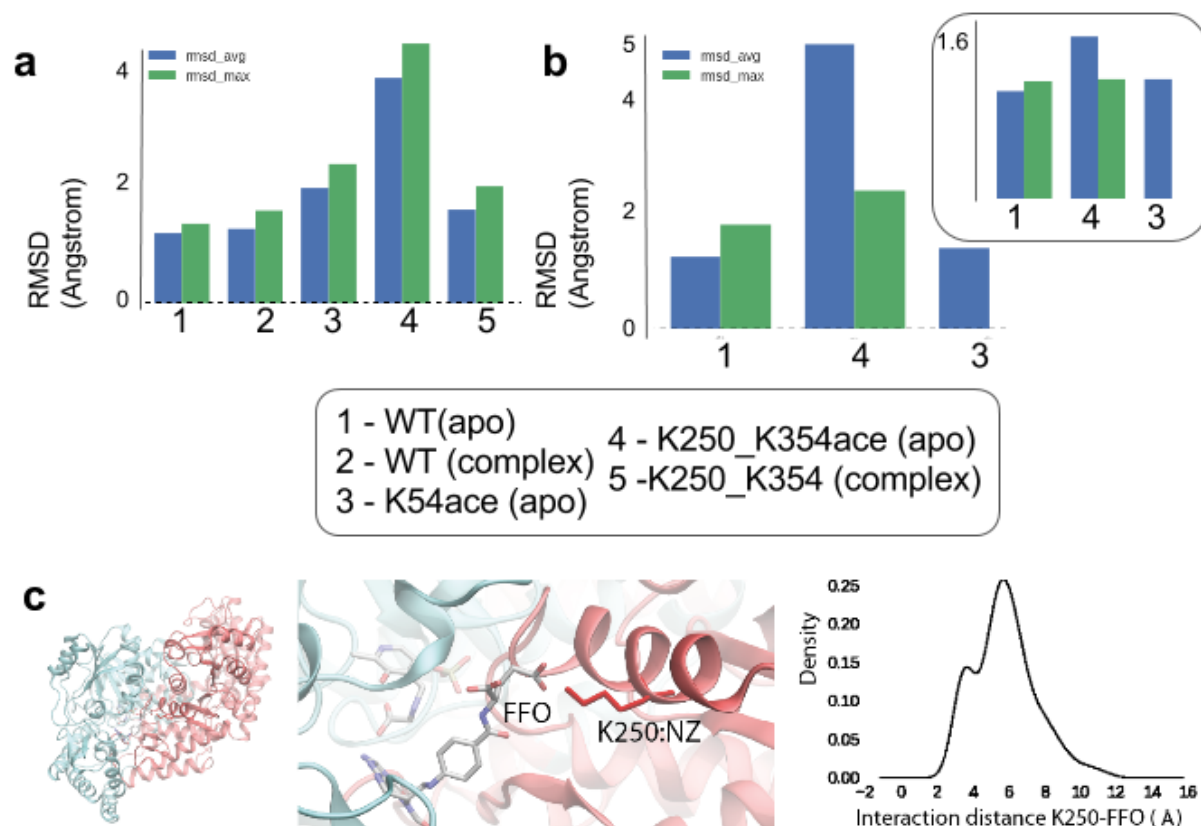
Fig. S11. Characteristic structural changes in serine hydroxymethyl transferase and modified variants (PTMs). (a) Root-mean squared deviation (RMSD) of the entire protein backbone SHMT and its associated post-translationally modified variants, with and without bound substrate. RMSD was calculated from 37-120 nanosecond molecular dynamic simulations in both enzyme-substrate complex, bound to analogs of pyridoxal-5-phosphate and tetrahydrate folate (PLG and FFO, respectively) and substrate-free states. In each case, the reference frame is taken to be the equilibrated protein. Acetylation of K250 and K354 undergoes the most structural change in free enzyme. **(b)** RMSD of the N-terminal domain of SHMT and their associated post-translationally modified variants (K250/K354 versus K54), with and without bound substrate. In the inset is the RMSD plotted for the second N-terminal domain (residues 34 to 290), which bind PLP. **(c)** Interatomic distances between K250 and substrate/cofactor tetrahydrofolate (crystallized as FFO). **(a)** the homodimer protein in complex with PLG and FFO substrates. **(b)** A zoomed-in view of residue K250 interacting with the acetic acid group on FFO. **(c)** Distances are plotted between atoms NZ on K250 and the terminal carbon of the acetic acid moiety of FFO throughout the MD trajectory. The average interaction distance is 5.8 Angstroms, indicating that for part of the trajectory, K250 and FFO participate in a salt-bridge interaction (20% of the time). Such a salt-bridge is disrupted when K250 is acetylated.
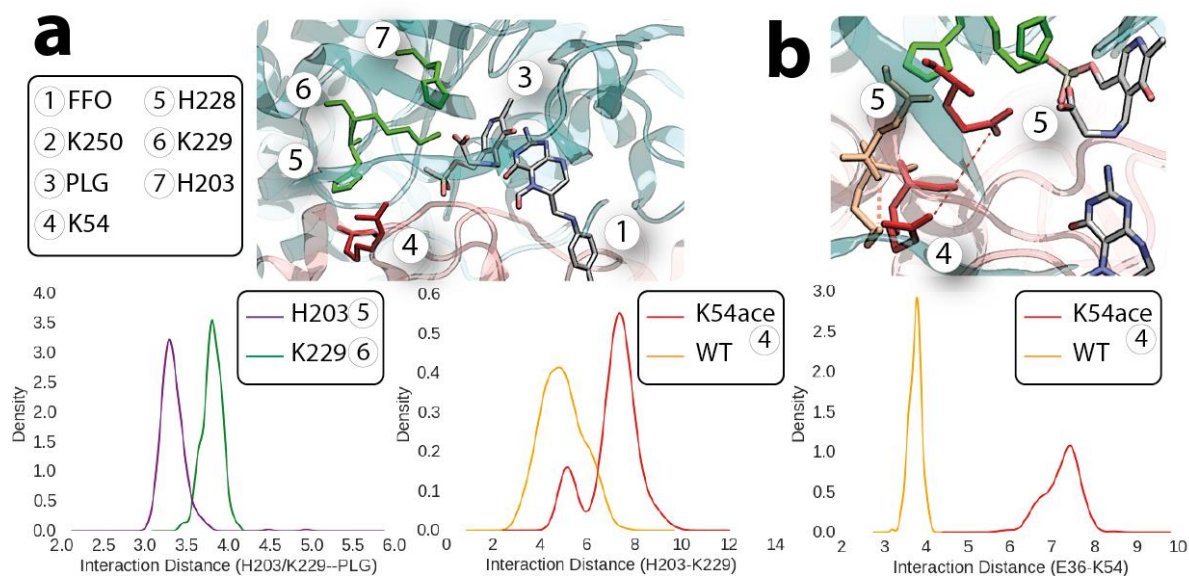
Fig. S12. Changing intermolecular interaction distances in modified SHMT proteins. (a) Intermolecular distances are plotted between H203 and K229 and substrate (PLG) in the active site pocket of wild-type SHMT (bottom left). Distances are given between atoms, NZ on K229 (green) vs NE2 on H203 (purple) and the carbon of the terminal acetic acid moiety of PLG throughout the MD trajectory for wild-type SHMT. (bottom right) When K54 is acetylated, the intermolecular distance between K229 (NZ) and H203 (NE2) shifts from 4.9Å, in wild-type, to 7.1Å, in post-translationally modified SHMT, suggesting that acetylation of K54 significantly shifts (p < 0.01 using a two-tailed test) the positions of key, conserved, catalytic residues. (b) Other changes, upon acetylation of K54, occurring near the active site (which is also the homodimer interface) are shifts in the intermolecular distance of K54 and E36, which participate in a strong salt bridge throughout the MD trajectory in wild-type SHMT. Distances shift, on average, more than 4 Å in modified protein.
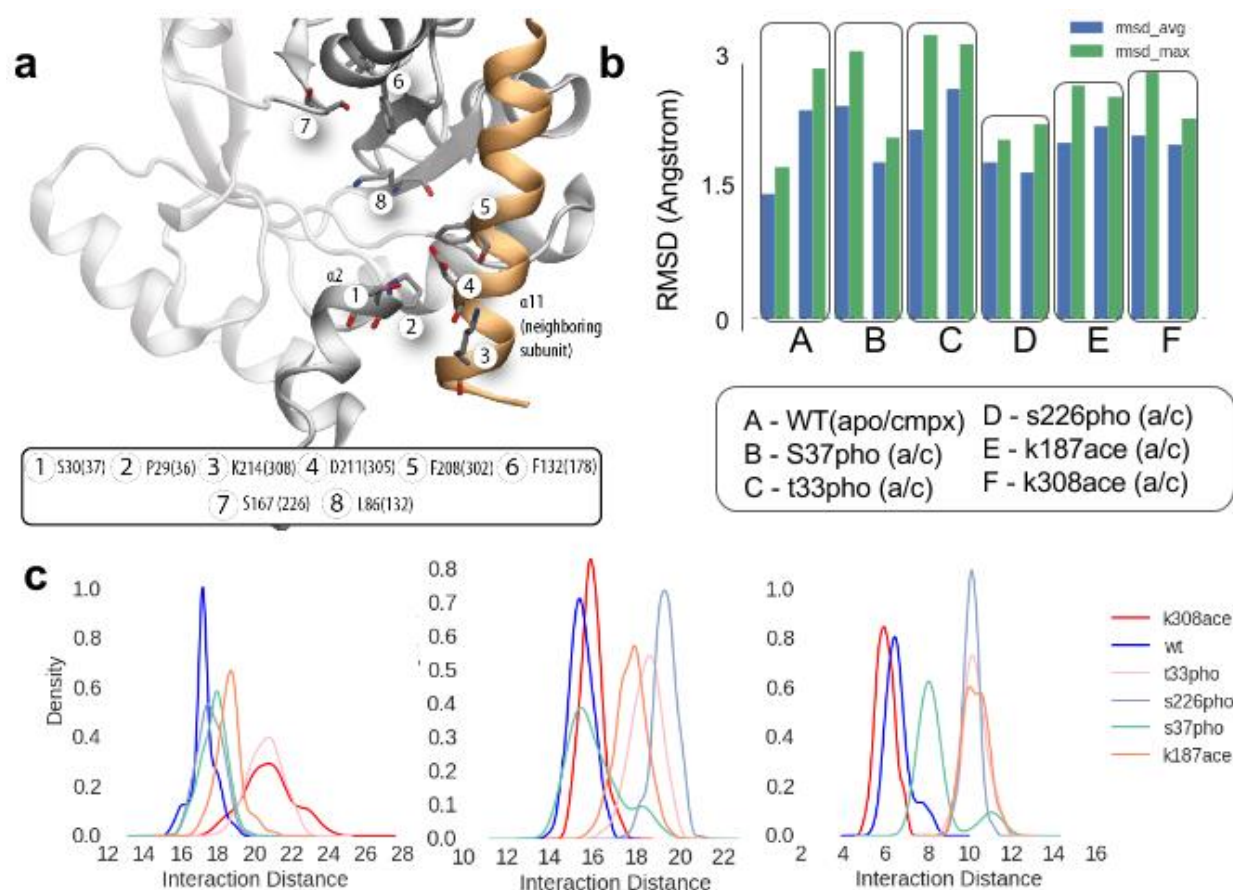
Fig. S13. Characteristic structural changes in transaldolase and modified variants (PTMs). (a) The crystal structure from *Thermoplasma acidophilum* (PDB entry 3s1u) shows a conserved set of residues in the substrate binding channel and their associated *E. coli* counterparts (in parentheses). Stacking interaction is shown between Pro29 at α2 and Phe208 at the C-terminal helix α11 of the neighboring subunit in the "closed conformation". The essential Lys86 (Lys132 in *E. coli*), in the active site is also shown. The close interaction between residues #5 and #2 has been characterized in the closed conformation in this enzyme. **(b)** RMSD was calculated from 50-100 nanosecond molecular dynamic simulations in both apo (substrate-free) and complex (enzyme-substrate complex, where the substrate is taken to be sedoheptulose 7-phosphate (s7p)) states. In each case, the reference frame is taken to be the equilibrated protein. Both maximum and average RMSD is shown and the apo protein are to the left of the enzyme-substrate complex for all pairs in the plot. **(c)** Significantly shifting molecular interaction distances resulting from post-translational modification. **(left)** Per-residue interaction differences, relative to wild-type, for Phe178--Asp305 across variant transaldolases co-complexed with s7p. **(center)** Interaction differences for Pro36--Phe178, relative to wild-type in substrate-free variant transaldolases. **(right)** Per-residue interaction differences relative to wild-type for Pro36--Phe302. In each of the three interaction distance distributions, we find distinct "states" that cluster variants based on how different they are from wild-type behavior. Significantly shifted states ($p > 0.05$ in a two-tailed test) are found for specific variants, namely k308ace (178_305), s226pho (36_178, 36_302, 36_132), t33pho (37_178, 37_302) and k187ace (37_178, 37_302). All distances are given in Angstrom.
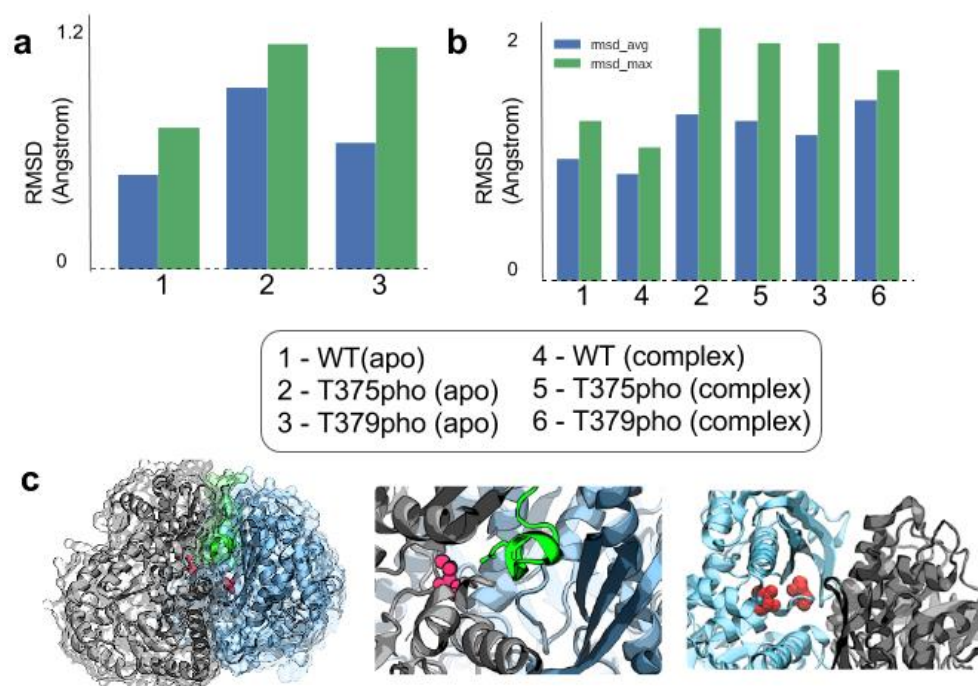
Fig. S14. Characteristic structural changes in enolase and modified variants (PTMs). (a) RMSD was computed (i.e., for the protein backbone atoms) from 50-100 nanosecond molecular dynamic trajectories in both enzyme-substrate complex and substrate-free states. In each case, the reference frame is taken to be the equilibrated protein. Relatively minimal changes in protein structure occur in the modified enolase proteins, compared to wild-type protein. Both the average and maximum RMSD is shown. **(b)** RMSD computed for all atoms in the 5Å vicinity of T375 and T379 residues during 50 nanoseconds, where the reference frame is taken to be the equilibrated protein. **(c) (left)** The biological assembly of enolase, as a homodimer and the location of the modified residues, T375 and T379, at the interface of this complex. (center) Enolase in complex with a minimal binding segment of RNase E (PDB entry 2yfm). The residues in a natively unstructured segment of RNase E. (shown here in green) bind to a cleft in the interface of the enolase dimer, less than 6Å from the modification site. (right) A close-up view of the enolase dimer interface and both modified residues, T375 and T379, shown in van der Waals representation (red).
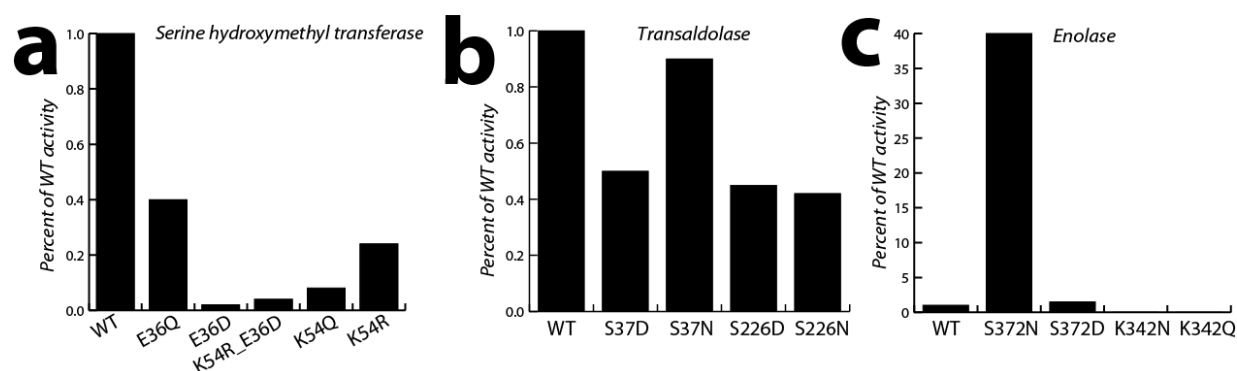
Fig. S15. Enzyme *in vitro* assays for activity measurements of proteins and modified variants. **(a)** Comparison of enzyme activity in WT protein with modified serine hydroxymethyl transferase proteins (mutated at position 36 and 54 in the protein). The enzyme activity was based on the following reaction: L-serine + a tetrahydrofolate ↔ glycine + a 5,10-methylene-tetrahydrofolate + H_2O. **(b)** Comparison of enzyme activity in WT transaldolase protein and modified variants (at positions 37 and 226). The enzyme activity was based on the following reaction: D-glyceraldehyde 3-phosphate + D-sedoheptulose 7-phosphate ↔ β-D-fructofuranose 6-phosphate + D-erythrose 4-phosphate. **(c)** Comparison of enzyme activity in WT enolase protein and modified variants (at positions 372 and 342). The enzyme activity was based on the following reaction: 2-phospho-D-glycerate ↔ phospho*enol*pyruvate + H_2O. All activities are reported as percentage of wild-type protein activity.
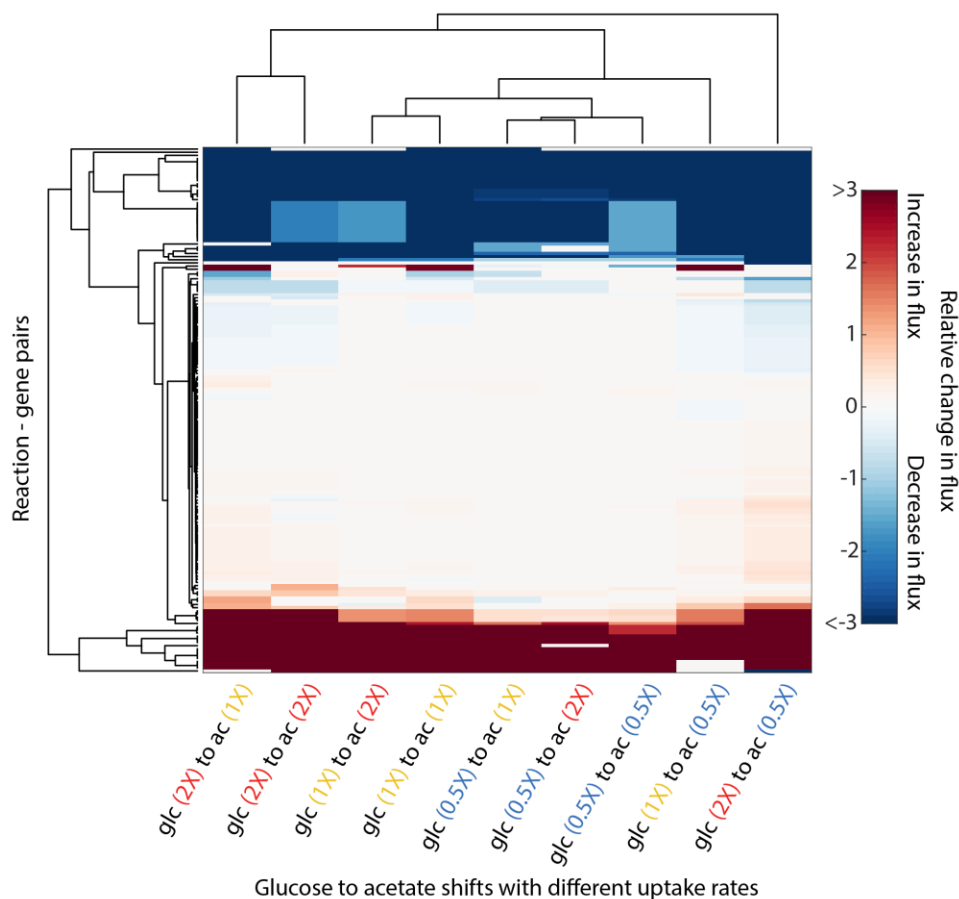
Fig. S16. Moderate variations in metabolite uptake rate in the metabolic model have little impact on RuMBA predictions. In this study, we simulated growth on 174 different minimal media conditions. To parameterize the simulations, we set up each media condition to take up 48 mmol carbon equivalents per gram DW per hour, which is consistent with published experimental measurements on glucose, lactate, and glycerol (21, 24, 61). Given that these were estimates, it was important to test if RuMBA results were sensitive to sizable variations in substrate uptake rates. Therefore, RuMBA was run on the glucose-acetate diauxy at different glucose and acetate uptake rates, ranging from half of the standard uptake rate to double the rate (4, 8, and 16 mmol glucose per gDW per hour and 12, 24, and 48 mmol acetate per gDW per hour). RuMBA was computed for each pairwise shift (e.g., double glucose vs. half acetate). This heatmap shows the fold change relative flux for each reaction and their associated proteins (rows) and each comparison (columns). We have plotted all reactions with at least a 10% change in relative flux. As demonstrated here, RuMBA results are quite robust against variations in the uptake rates. This is because RuMBA depends less on absolute flux levels, and instead the relative flux levels between reactions in a flux split. Thus, even though estimated substrate uptake rates were used here, the results in this study should not be strongly affected by the assumption of the uptake rates.
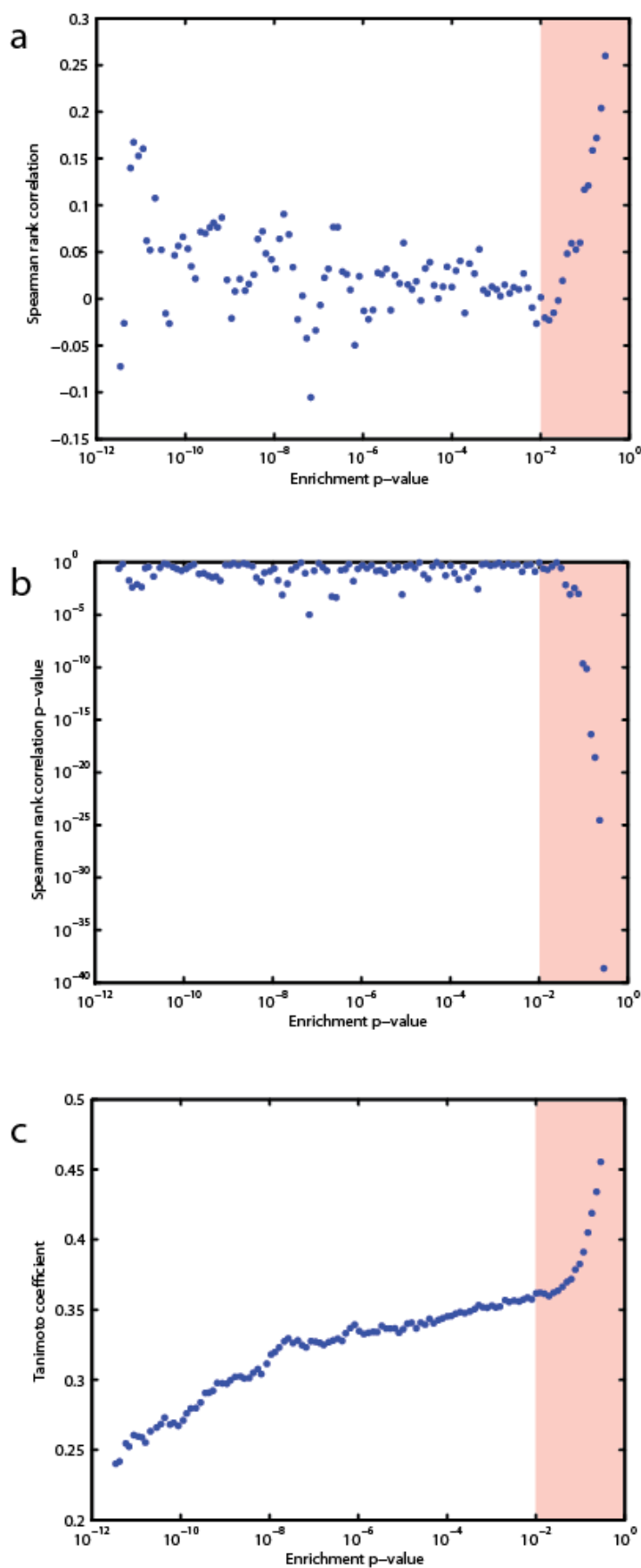
Fig. S17. Conditions with high metabolite structural similarity are depleted of PTMs. A subset of substrate shifts was predicted to have few enzymes that are regulated, and were not enriched in PTMs. **(a)** The Spearman rank correlation was computed between the enrichment p-values and Tanimoto coefficients for the subsets of the shifts having a p-value of enrichment of PTMs within half of an order of magnitude of each point shown. Once the median p-value within each subset passed p = 0.1, there was a clear correlation between the enrichment p-value and the Tanimoto coefficient, which measures the similarity between the primary carbon substrates in the two media conditions. **(b)** The correlation of the flux vectors is increasingly significant with increasing enrichment p-value. **(c)** This increasing correlation is accompanied by a sharp increase in Tanimoto coefficient, demonstrating that shifts that fail to be enriched in PTMs tend to show a high structural similarity between the main carbon substrates in the media formulations participating in the shift. P-values > 0.01 are highlighted in pink.
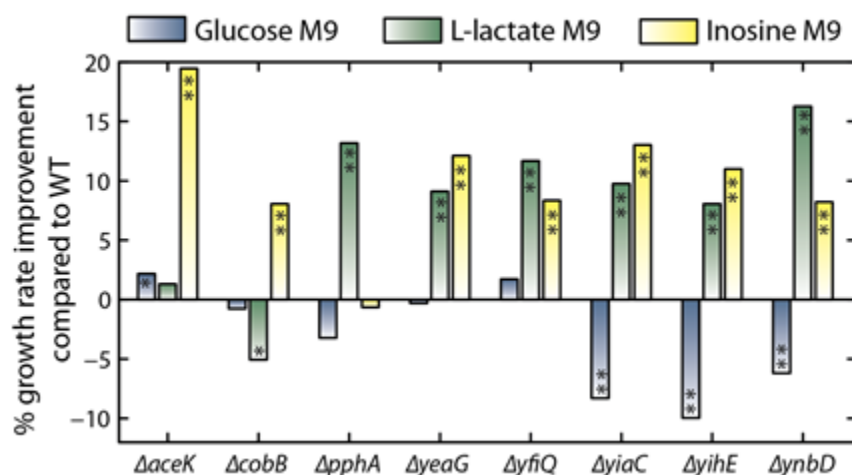
Fig. S18. Deletion of kinases, phosphatases, and acetyltransferases can have positive or negative impacts on growth, depending on the media. Mutant strains of *E. coli* lacking a kinase, phosphatase, or acetyltransferase (Δ*aceK*, Δ*cobB*, Δ*pphA*, Δ*yeaG*, Δ*yfiQ*, Δ*yiaC*, Δ*yihE*, and Δ*ynbD)* were grown on M9 minimal media with glucose, L-lactate, or inosine. Cellular fitness of these strains depends strongly on the nutrient environment, as certain mutants exhibited faster growth (relative to wild-type) in certain nutrient conditions but slower growth on others.

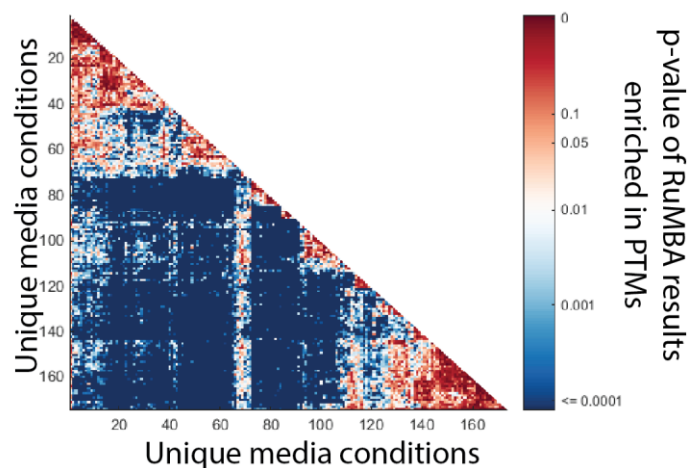**Fig. S19. Inclusion of new phosphoproteins shows that PTMs continue to be enriched in RuMBA results.** Three large phosphoproteomic studies were published after the MAGE-Seq screen was conducted. Thus, the screen of 15051 simulated nutrient shifts was redone to include the novel phosphoproteins identified. For most nutrient shifts (92%), the RuMBA results were enriched in proteins with measured PTMs.

Supplementary Datasets (Tables) & Captions

[Datasets (tables) S1-S16 are found in Supplementary Datasets S1-S17.xlsx]

Dataset S1: Genes and reactions predicted by RuMBA to be regulated in the glucose - acetate diauxie. Columns are defined as follows: (broken up into 2x3 columns for enzymes that show increased activity in glucose or acetate M9 minimal media conditions): the gene and catalyzing reaction name, metabolites involved at the branch point reaction, the activity increase (fold increase in flux).

Dataset S2: Experimentally validated metabolite-mediated metabolic regulation events in *E. coli* obtained from Ecocyc. Columns are defined as follows: the source of the information, the model reaction names, uniprot accession number (where available), all genes involved in the reaction, given by Blattner ID (enzyme stoichiometry in parentheses), mechanism of regulation, mode of action (+/- for activation or repression, respectively), associated metabolite names, associated KEGG identifiers, Pubchem identifiers and SMILES format of the metabolites.

Dataset S3: Structural properties of 62 proteins and their respective PTMs. Columns are defined as follows: the gene id (given by Blattner code), uniprot accession number, PDB structure and chain associated with the uniprot ID, length of the chain, all chains in the full protein, type of known modification, where the modification occurs in the protein, the corresponding location of the residue in the PDB file, residues where potential salt-bridging interaction may have been disrupted by the PTM, residues where potential salt-bridging interaction may have been newly formed by the PTM, any cases that were not computed in distance analysis due to missing annotation information.

Dataset S4: List of different media compositions tested with RuMBA. All media are M9 media, as described previously (20), except that glucose is replaced by each of these carbon sources. Metabolite identifiers are from the BiGG Models database (62).

Dataset S5: List of all PTM sites and their corresponding MAGE oligonucleotide sequences. Columns are defined as follows: gene identifiers, associated sequence, type of modification (PTM), type of genetic modification made *in vivo.*

Dataset S6: Genes (Blattner IDs) targeted by the MAGE oligos.

Dataset S7. A description of the proteins and their associated post-translationally modified residues (in blue are cases selected for analysis with MD simulations). A large percent conservation of the specific residue indicates that its role may be pertinent in catalytic function, given a high degree of homology compared to up to 1057 species. A z-score analysis was performed to identify whether the percent conservation was statistically significant ($z > 1.9$, given 10-100 degrees of freedom). The number of shifts indicates the number of cases that the activity of the protein was significantly altered across changing nutrient environments. We probed the three-dimensional (3D) crystallographic structure of the protein, given in the column labeled "PDB" to understand whether the residue is likely to participate in salt-bridging interactions with nearby negatively charged residues.

Dataset S8. Highly correlated interaction distances between two pairs of residues were probed in transaldolase and modified variants. Cases highlighted in the table are variant proteins whose per-residue interacting distances are no longer correlated as a result of post-translational modification (using Pearson correlation, r>0.5). The residue numbers are given for interaction amino acids, in which distances are measured between the alpha carbon atoms in Angstroms. The Pearson correlation for the corresponding interaction distance in WT is provided.

Dataset S9: List of primers were designed to validate a subset of the targets using MASC-PCR. Columns are defined as follows: gene name, Blattner code, type of PTM, location (amino acid), modification id, direction, sequence

Dataset S10: List of primers designed to enable a two-step amplification and library preparation for amplicon sequencing. Columns are defined as follows: forward primer sequence and reverse primer sequence

Dataset S11: List of PCR primers used to add barcodes to each sample. Columns are defined as follows: sample ID, sequence of forward primer with barcode, barcode sequence

Dataset S12: Number of cell doublings in each sample

Dataset S13: Fold change of MAGE mutant (PTM-mimic) across all conditions

Dataset S14: Fold change of MAGE mutant (PTM-null) across all conditions

Dataset S15: P-value of fold change of MAGE mutant (PTM-mimic) across all conditions

Dataset S16: P-value of fold change of MAGE mutant (PTM-null) across all conditions

Dataset S17: PTMs showing differential abundance between glucose and acetate, and comparison to RuMBA results. Peptides and fold change taken from (59)

Supplementary References

1. Mitchell A, et al. (2009) Adaptive prediction of environmental changes by microorganisms. *Nature* 460(7252):220–224.

2. Savageau MA (1983) Escherichia coli Habitats, Cell Types, and Molecular Mechanisms of Gene Control. *Am Nat* 122(6):732–744.

3. Savageau MA (1998) Demand theory of gene regulation. II. Quantitative application to the lactose and maltose operons of Escherichia coli. *Genetics* 149(4):1677–1691.

4. Bar-Even A, Noor E, Lewis NE, Milo R (2010) Design and analysis of synthetic carbon fixation pathways. *Proc Natl Acad Sci U S A* 107(19):8889–94.

5. Bordbar A, Lewis NE, Schellenberger J, Palsson BO, Jamshidi N (2010) Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. *Mol Syst Biol* 6:422.

6. Nam H, et al. (2012) Network context and selection in the evolution to enzyme specificity. *Science (80-)* 337(6098):1101–4.

7. Bordel S, Agren R, Nielsen J (2010) Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Comput Biol* 6(7):e1000859.

8. Bordbar A, Monk JM, King ZA, Palsson BO (2014) Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet* 15(2):107–20.

9. Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol* 10(4):291–305.

10. Dannelly HK, Duclos B, Cozzone AJ, Reeves HC (1989) Phosphorylation of Escherichia coli enolase. *Biochimie* 71(9–10):1095–1100.

11. Dean AM, Lee MH, Koshland Jr DE (1989) Phosphorylation inactivates Escherichia coli isocitrate dehydrogenase by preventing isocitrate binding. *J Biol Chem* 264(34):20482–20486.

12. Hoyt JC, Reeves HC (1988) In vivo phosphorylation of isocitrate lyase from Escherichia coli D5H3G7. *Biochem Biophys Res Commun* 153(2):875–880.

13. Schellenberger J, Palsson BO (2009) Use of randomized sampling for analysis of metabolic networks. *J Biol Chem* 284(9):5457–5461.

14. Lewis NE, et al. (2010) Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nat Biotechnol* 28(12):1279–85.

15. Schellenberger J, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* 6(9):1290–307.

16. Schuster S, Pfeiffer T, Fell DA (2008) Is maximization of molar yield in metabolic networks favoured by evolution? *J Theor Biol* 252(3):497–504.

17. Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? *Nat Biotechnol* 28(3):245–8.

18. Schellenberger J, Lewis NE, Palsson BO (2011) Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys J* 100(3):544–53.

19. Heirendt L, et al. (2018) Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. *Nat Protoc.*

20. Feist AM, et al. (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 3:121.

21. Hua Q, Joyce AR, Fong SS, Palsson BO (2006) Metabolic analysis of adaptive evolution for in silico-designed lactate-producing strains. *Biotechnol Bioeng* 95(5):992–1002.

22. Cho BK, et al. (2009) The transcription unit architecture of the Escherichia coli genome. *Nat Biotechnol* 27(11):1043–1049.

23. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 429(6987):92–96.

24. Fong SS, Joyce AR, Palsson BO (2005) Parallel adaptive evolution cultures of Escherichia coli lead to convergent growth phenotypes with different gene expression states. *Genome Res* 15(10):1365–1372.

25.   Lewis NE, Cho B-KK, Knight EM, Palsson BO (2009) Gene expression profiling and the use of genome-scale in silico models of Escherichia coli for analysis: providing context for content. *J Bacteriol* 191(11):3437–44.

26.   Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* 85(8):2444–8.

27.   Baba T, et al. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2:2006.0008.

28.   Wang HH, et al. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460(7257):894–898.

29.   Macek B, et al. (2008) Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Mol Cell Proteomics* 7(2):299–307.

30.   Yu BJ, Kim JA, Moon JH, Ryu SE, Pan JG (2008) The diversity of lysine-acetylated proteins in Escherichia coli. *J Microbiol Biotechnol* 18(9):1529–1536.

31.   Zhang J, et al. (2009) Lysine acetylation is a highly abundant and evolutionarily conserved modification in Escherichia coli. *Mol Cell Proteomics* 8(2):215–225.

32.   Wang HH, et al. (2012) Genome-scale promoter engineering by coselection MAGE. *Nat Methods* 9(6):591–3.

33.   Ratcliffe SJ, Shults J (2008) GEEQBOX : A MATLAB Toolbox for Generalized Estimating Equations and Quasi-Least Squares. *J Stat Softw* 25(14):1–14.

34.   Hanley JA, Negassa A, Edwardes MD deB, Forrester JE (2003) Statistical analysis of correlated data using generalized estimating equations: an orientation. *Am J Epidemiol* 157(4):364–75.

35.   Homeyer N, Horn AHC, Lanig H, Sticht H (2006) AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *J Mol Model* 12(3):281–9.

36.   Bas DC, Rogers DM, Jensen JH (2008) Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins* 73(3):765–83.

37.   Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pKa values. *Proteins Struct Funct Bioinforma* 61(4):704–721.

38.   Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput* 7(2):525–37.

39.   Bayly CI, Cieplak P, Cornell W, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem* 97(40):10269–10280.

40.   Cieplak P, Cornell WD, Bayly C, Kollman PA (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J Comput Chem* 16(11):1357–1377.

41. Cornell WD, Cieplak P, Bayly CI, Kollmann PA (1993) Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J Am Chem Soc* 115(21):9620–9631.

42. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys* 98(12):10089.

43. Essmann U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103(19):8577.

44. Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys* 103(11):4613.

45. Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. *J Chem Phys* 101(5):4177.

46. Nosé S (1984) A unified formulation of the constant temperature molecular dynamics methods. *J Chem Phys* 81(1):511.

47. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC (2013) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* 9(9):3878–88.

48. Hornak V, et al. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65(3):712–25.

49. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* 21(12):1049–1074.

50. Holms WH, Bennett PM (1971) Regulation of isocitrate dehydrogenase activity in Escherichia coli on adaptation to acetate. *J Gen Microbiol* 65(1):57–68.

51. Shinar G, Rabinowitz JD, Alon U (2009) Robustness in glyoxylate bypass regulation. *PLoS Comput Biol* 5(3):e1000297.

52. Fay JP, Farias RN (1975) The inhibitory action of fatty acids on the growth of Escherichia coli. *J Gen Microbiol* 91(2):233–240.

53. Johnson LN, Barford D (1993) The Effects of Phosphorylation on the Structure and Function of Proteins. *Annu Rev Biophys Biomol Struct* 22(1):199–232.

54. Vivoli M, et al. (2009) Role of a conserved active site cation-pi interaction in Escherichia coli serine hydroxymethyltransferase. *Biochemistry* 48(50):12034–12046.

55. Cai K, Schirch D, Schirch V (1995) The affinity of pyridoxal 5'-phosphate for folding intermediates of Escherichia coli serine hydroxymethyltransferase. *J Biol Chem* 270(33):19294–19299.

56. Chandran V, Luisi BF (2006) Recognition of Enolase in the Escherichia coli RNA Degradosome. *J Mol Biol* 358(1):8–15.

57. Kühnel K, Luisi BF (2001) Crystal structure of the Escherichia coli RNA degradosome component enolase. *J Mol Biol* 313(3):583–92.

58.   Soares NC, Spät P, Krug K, Macek B (2013) Global dynamics of the Escherichia coli proteome and phosphoproteome during growth in minimal medium. *J Proteome Res* 12(6):2611–21.

59.   Lim S, Marcellin E, Jacob S, Nielsen LK (2015) Global dynamics of Escherichia coli phosphoproteome in central carbon metabolism under changing culture conditions. *J Proteomics* 126:24–33.

60.   Hansen A-M, et al. (2013) The Escherichia coli phosphotyrosine proteome relates to core pathways and virulence. *PLoS Pathog* 9(6):e1003403.

61.   Lewis NE, et al. (2010) Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* 6:390.

62.   King ZA, et al. (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res* 44(D1):D515–D522.