



Proteome Regulation Patterns Determine *Escherichia coli* Wild-Type and Mutant Phenotypes

 Tobias B. Alter,^{a*} Lars M. Blank,^a  Birgitta E. Ebert^{a,b}

^aInstitute of Applied Microbiology (IAMB), Aachen Biology and Biotechnology (ABBT), RWTH Aachen University, Aachen, Germany

^bAustralian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, Brisbane, Australia

ABSTRACT It is generally recognized that proteins constitute the key cellular component in shaping microbial phenotypes. Due to limited cellular resources and space, optimal allocation of proteins is crucial for microbes to facilitate maximum proliferation rates while allowing a flexible response to environmental changes. To account for the growth condition-dependent proteome in the constraint-based metabolic modeling of *Escherichia coli*, we consolidated a coarse-grained protein allocation approach with the explicit consideration of enzymatic constraints on reaction fluxes. Besides representing physiologically relevant wild-type phenotypes and flux distributions, the resulting protein allocation model (PAM) advances the predictability of the metabolic responses to genetic perturbations. A main driver of mutant phenotypes was ascribed to inherited regulation patterns in protein distribution among metabolic enzymes. Moreover, the PAM correctly reflected metabolic responses to an augmented protein burden imposed by the heterologous expression of green fluorescent protein. In summary, we were able to model the effects of important and frequently applied metabolic engineering approaches on microbial metabolism. Therefore, we want to promote the integration of protein allocation constraints into classical constraint-based models to foster their predictive capabilities and application for strain analysis and engineering purposes.

IMPORTANCE Predictive metabolic models are important, e.g., for generating biological knowledge and designing microbes with superior performance for target compound production. Yet today's whole-cell models either show insufficient predictive capabilities or are computationally too expensive to be applied to metabolic engineering purposes. By linking the inherent genotype-phenotype relationship to a complete representation of the proteome, the PAM advances the accuracy of simulated phenotypes and intracellular flux distributions of *E. coli*. Being equally computationally lightweight as classical stoichiometric models and allowing for the application of established *in silico* tools, the PAM and related simulation approaches will foster the use of a model-driven metabolic research. Applications range from the investigation of mechanisms of microbial evolution to the determination of optimal strain design strategies in metabolic engineering, thus supporting basic scientists and engineers alike.

KEYWORDS constraint-based modeling, enzyme kinetics, metabolic engineering, protein allocation, transcriptional control, *Escherichia coli*

For many decades, metabolic models have been developed to describe, unravel, and understand the drivers of microbial phenotypes. In their simplest forms, these models quantitatively connect observable phenomena such as carbon source consumption and biomass formation, leading to seminal empirical growth laws such as the Monod equation (1). In general, coarse-grained models aid in explaining the dependencies between intracellular processes and corresponding phenotypes (2–7). Constraint-based modeling techniques facilitate the prediction of growth rates and by-product secretion, as well as the investigation of metabolic flux distributions solely based on

Citation Alter TB, Blank LM, Ebert BE. 2021. Proteome regulation patterns determine *Escherichia coli* wild-type and mutant phenotypes. *mSystems* 6:e00625-20. <https://doi.org/10.1128/mSystems.00625-20>.

Editor Joshua E. Elias, Chan Zuckerberg Biohub
The review history of this article can be read [here](#).

Copyright © 2021 Alter et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Birgitta E. Ebert, birgitta.ebert@uq.edu.au.

* Present address: Tobias B. Alter, Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark.

Received 9 July 2020

Accepted 21 January 2021

Published 9 March 2021

the stoichiometry of biochemical reaction networks and an appropriate cellular objective function (8–12). The development of genome-scale constraint-based models (GEM) fostered the investigation of fundamental biological phenomena (9, 13, 14), the systematic analysis of complex omics data sets (15–18) and the suggestion of favorable genetic perturbations for the overproduction of desired chemicals (19–23).

The utilization of GEMs has made valuable contributions to systems biological analyses and metabolic engineering. However, the GEM's predictive capabilities of cellular phenotypes strongly rely on *ad hoc* capacity bounds on key reactions (15, 24), without which basic phenomena such as overflow metabolism are not observable *in silico*. The consideration of additional cellular processes and properties in metabolic reconstructions resolved these predictive insufficiencies of GEMs. In this manner, macromolecular expression (ME) models couple metabolism to gene expression by linking enzyme concentrations to metabolic reactions and accounting for the transcriptional and translational processes leading to enzyme expression (25). ME models simultaneously simulate maximum growth and substrate uptake rates and the underlying responses on the mRNA level, as well as the corresponding gene expression profiles at metabolic steady state (26). Thus, ME models facilitate holistic insights into intracellular processes and how they are affected by environmental, biochemical, or genetic perturbations (27–29), while reliably informing about corresponding flux distributions. As advantageous ME models are for correct predictions at the flux or phenotypic level, their detail and complexity can be cumbersome for future applications in strain design approaches. Many other approaches exist that add various resource or spatial constraints to classical stoichiometric models to investigate otherwise undisclosed growth laws (30–36). However, most of them are either similarly computationally expensive or demand a tremendous number of mostly unknown parameters.

Cellular protein allocation and its regulation have previously been suggested as the main drivers of metabolic phenomena and a key process behind bacterial growth laws (3, 5–7, 37–39). The incorporation of protein constraints in GEMs exploits the principles of protein allocation as a fundamental growth law. It simultaneously allows for the use of established, tractable, and intuitive constraint-based modeling methods. By dividing the limited proteome into three growth-variant sectors representing (i) ribosomal proteins, (ii) anabolic enzymes, and (iii) catabolic enzymes and one invariant housekeeping protein sector, the constrained allocation flux balance analysis (CAFBA) framework computes the optimal partitioning between these protein sectors and the correspondingly weighted flux rates to reach maximum growth (37, 39). Thus, CAFBA accounts for the trade-off between a limited protein availability for biosynthesis and growth and enables a quantitative prediction of pathway usages under various conditions, particularly suboptimal growth yields during overflow metabolism. The integration of enzyme kinetics into a GEM of *Saccharomyces cerevisiae* and *Escherichia coli* in the form of explicit enzymatic constraints on flux rates facilitated the utilization of proteomics data (40). The respective GECKO framework (GEM with enzymatic constraints using kinetic and omics data) gives detailed insights into metabolic realizations based on proteome measurements and predicts growth phenomena even without augmented protein concentration data.

Here, we introduce an approach that consolidates protein allocation and enzymatic constraints on metabolic fluxes of an *E. coli* GEM. The resulting protein allocation model (PAM) represents the major protein sectors and their various ratios with changing growth conditions (41). The PAM predicts experimentally observed phenotypes and intracellular flux distributions at maximum growth and on various growth media. Besides the wild-type behavior, the PAM's predictive ability is demonstrated for strains harboring gene deletions or expressing heterologous proteins. In line with previous studies, we emphasize the fundamental role of protein allocation in steering microbial metabolism. Therefore, the PAM framework is envisioned as a practical approach to improve the rational design of microbial production strains, as it enables rapid dry-lab screening of design and cultivation strategies with improved reliability.

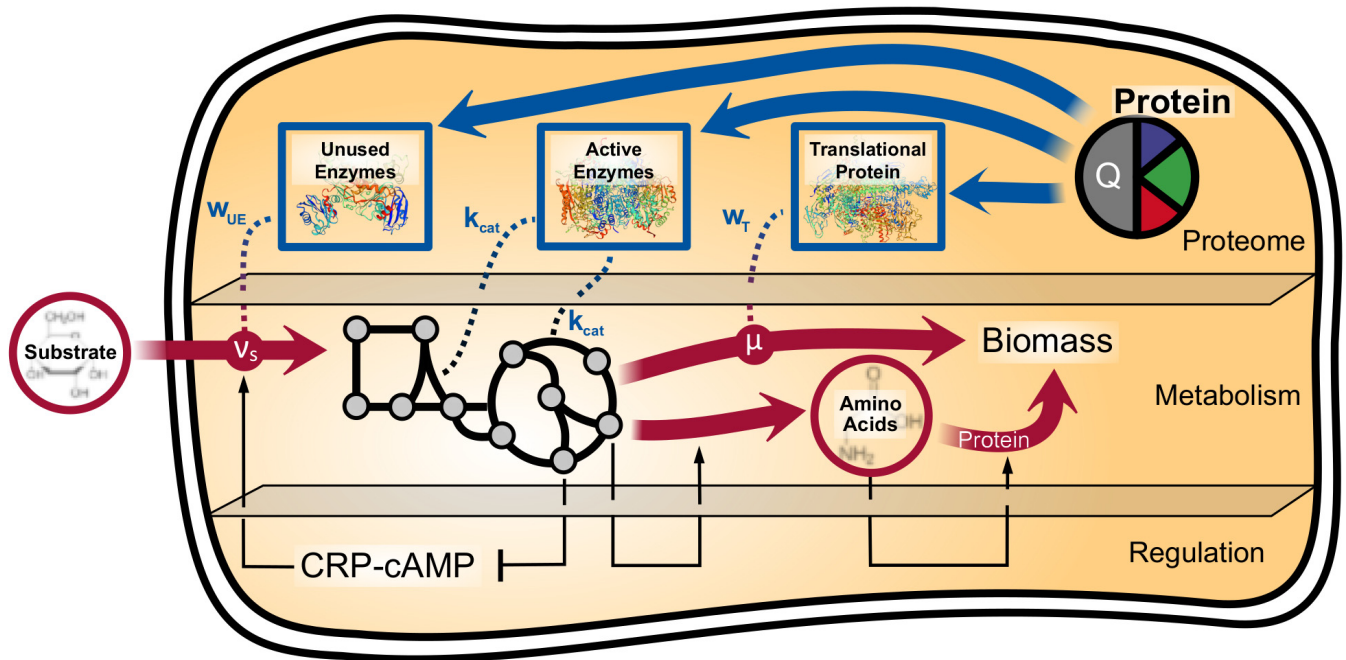


FIG 1 Scheme of the protein allocation model, including the classical metabolic as well as the added proteome level. w_T , w_{UE} , and k_{cat} represent the linear correlation factors between model variables and the translational, unused enzyme, and active enzyme sectors, respectively. Q denotes the housekeeping protein sector, whose concentration is constant. The dotted lines mark the model-inherent, linear relations between protein sectors and metabolic rates. A simplified mechanism regulating the microbial proteome, which is not part of the model, is also shown in accordance with the theory proposed in reference 4. It depicts the activation of catabolic enzyme expressions by the CRP-cAMP complex, whose synthesis is hampered by elevated amino acid precursor levels. Additionally, the activating effects of the concentration of amino acids and their precursors on protein and amino acid synthesis, respectively, are sketched.

RESULTS

Accounting for the total proteome in genome-scale metabolic models. A key challenge for microbes is distributing the limited proteins to the intracellular processes in a way that allows for maximum growth under a given environmental condition. While the concentration of ribosomes efficiently meets the demand of protein biosynthesis under mildly nutrient-limited to unlimited conditions (5, 42, 43), the protein household for energy and biomass precursor production generally contains unutilized and underutilized enzymes (44) to allow for a flexible response to environmental changes. Three major protein sectors cover this condition-dependent proteome and represent (i) translational protein, including ribosomes, (ii) metabolically active enzymes, and (iii) un- as well as underutilized enzymes (3). We will refer to the latter as the unused enzyme sector for simplicity. The fourth protein sector covers the housekeeping proteins, whose abundance is constant under any growth condition. Consequently, this sector cannot cause condition-dependent phenotypes. The biomass synthesis equation covers its constant demand, but the protein allocation toward the housekeeping sector does not need to be modeled separately.

To quantitatively account for the condition-dependent protein allocation, we modeled and added each relevant protein sector independently to the *E. coli* K-12 MG1655 GEM *iML1515* (45) (Fig. 1). We refer to the resulting model as the protein allocation model (PAM). The mass concentration, ϕ , of each sector is a linear function of one or more inherited variables or fluxes of *iML1515*. They have been partly fitted to experimental proteomic data as depicted in Table 1 and are described in detail in Materials and Methods.

The active enzyme sector. The active enzymes sector covers the protein demand of all enzymatic reactions of the *iML1515* model in a GECKO fashion (40). GECKO introduces

TABLE 1 Modeled mass concentrations ϕ of protein sectors and their linear dependencies on inherited variables of the *E. coli* K-12 MG1655 GEM iML1515^a

Protein mass concn	Symbol	Linear dependency
Active enzymes	ϕ_{AE}	Flux rates ν of metabolic reactions
Unused enzymes	ϕ_{UE}	Substrate uptake rate ν_s
Translational protein	ϕ_T	Growth rate μ
Total protein	$\phi_{P,C}$	$\phi_{AE}, \phi_{UE}, \phi_T$

^aParameters within these linear dependencies and their (data) sources are listed in Table 2.

enzyme mass balances to classical stoichiometric models and couples the manifestation of metabolic fluxes to enzyme abundances. That is, each modeled enzyme concentration within this sector is linearly dependent on the flux rate of the catalyzed metabolic reaction. The linear relations are based on a simplified rate law of reversible Michaelis-Menten reactions, thus on simple mass action kinetics, which neglects the nonlinear reversibility, enzyme saturation, and regulation factors (38). With this simplification, enzymes are assumed to operate at their maximal velocity, at which the turnover number k_{cat} describes the relation between enzyme concentration and flux rate. Hence, *in silico* metabolic fluxes are only limited by the enzyme's maximum capacity. While this formulation can lead to biologically more feasible solutions (40), enzymes do not generally operate at their maximum velocity (46, 47). Therefore, in the PAM context, the turnover number describes an enzyme's maximum capacity within a metabolic network operating under unlimited growth conditions, that is, under maximum growth and substrate uptake rate. The additional protein burden due to an incompletely used capacity of enzymes, e.g., under carbon-limited growth conditions, is accounted for by the unused enzyme sector introduced in the following section.

The unused enzyme sector. Particularly under nutrient-limited conditions, microbial cells put enzymes on hold to react quickly to changing environmental conditions (27, 44). This behavior contrasts with the general assumption that cells make efficient use of limited resources but reflects a necessary trade-off between maximal resource efficiency and quick adaptability. However, reversible enzyme kinetics prohibit the simultaneous manifestation of maximum velocity for all enzymes since the substrate of one enzyme is the product of another, inflicting opposing driving forces on both enzymes (38). As shown by metabolomics analyses, enzymes operate at substrate concentrations in their K_M range under most physiological situations, thus well below their maximum velocity, and are considered underutilized (46). Moreover, unutilized enzymes of inactive, mostly catabolic metabolic pathways may be expressed to allow for swift metabolic adjustments upon changes in the environmental conditions or substrate availability.

With a detailed analysis of experimental proteomic data of *E. coli* (41), O'Brien et al. (44) showed the existence of such under- and unutilized enzymes, which we summarized in the PAM as the unused enzyme sector. More specifically, O'Brien augmented an *E. coli* ME model with a comprehensive proteomic data set covering all enzymes of the iML1515 model (Fig. 2A) and assessed the environment-specific proteome utilization *in silico*. They showed that the mass concentration of the active enzyme sector decreases with increasing growth rates, despite the increased metabolic activity and rising protein requirements. Conversely, enzymes, which are not catalytically active, accumulate the stronger the carbon limitation and the lower the growth rate. This phenomenon is regulated by the cAMP signaling pathway, which senses the carbon influx (4). We linked the resulting negative feedback loop between protein expression and substrate uptake caused by the cAMP signaling pathway (Fig. 1) to the functional representation of the unused enzyme sector (see Materials and Methods for details). As a result, protein allocation toward the unused enzyme sector decreases with increasing substrate uptake rate and diminishes at the maximum substrate uptake rate. In this way, the PAM allows for the adaptation of substrate-specific allocation characteristics of the unused enzyme sector.

The translational protein sector. The quantitative description of the translational protein sector was empirically derived from global *E. coli* proteome measurements.

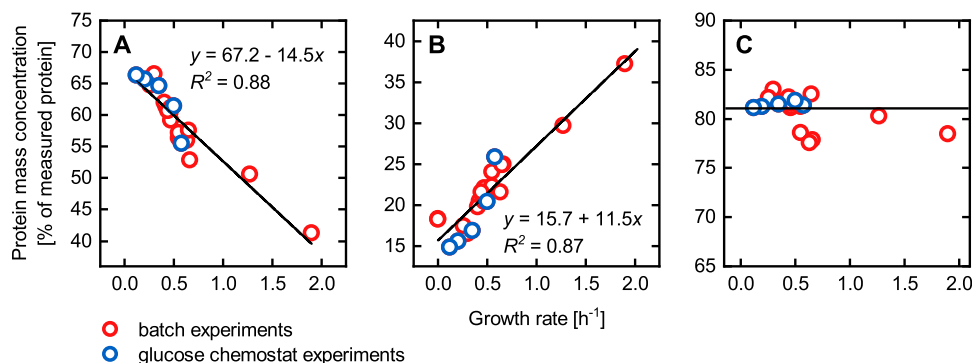


FIG 2 Protein mass data of distinct protein sectors under diverse growth conditions taken from Schmidt et al. (41). (A) All proteins represented by the *i*ML1515 GEM and found in the proteomic data set, comprising the active and unused enzymes sectors. (B) Translational sector covering proteins assigned to the COG (Clusters of Orthologous Groups) class “translation, ribosomal structure, and biogenesis,” which are not included in the *i*ML1515. The black lines in panels A and B are linear fits of the data points resulting in the shown equations and coefficients of determination R^2 . (C) Sum of the protein mass concentrations shown in panels A and B. The horizontal black line marks the 81% of measured protein mass used to constrain protein availability in the PAM. Glucose chemostat and batch experiments are highlighted in blue and red, respectively. Note that protein concentrations relative to the cell dry weight (cdw) were related to the total protein concentration of $0.32 \text{ g g}_{\text{cdw}}^{-1}$ measured by Schmidt et al. (41).

The experimental proteomic data set was taken from Schmidt et al. (41), who investigated the quantitative dependencies between protein allocation and growth for a wide range of conditions, substrates, and different *E. coli* strains. We processed the measured protein mass data (fg cell^{-1}) to arrive at model-relevant mass units relative to the cell dry weight (cdw) ($\text{g g}_{\text{cdw}}^{-1}$). We found that the translational protein mass concentration correlates linearly with the growth rate and modeled this protein sector accordingly (Fig. 2B). This linear growth dependency depicts a resource-efficient regulation of the translational apparatus to justly provide the proteins necessary for maintaining maximal division rates in a particular environmental condition, as reported previously (48, 49).

The total protein concentration constraint. Under most conditions, the total protein content of an *E. coli* cell is approximately $0.55 \text{ g g}_{\text{cdw}}^{-1}$ (50). A proteomic analysis of various *E. coli* strains grown under different conditions (41) covered a constant protein mass concentration of $0.32 \text{ g g}_{\text{cdw}}^{-1}$ (58% of the total protein content) across conditions. Therefore, a constant 81%, or $0.26 \text{ g g}_{\text{cdw}}^{-1}$ ($\phi_{p,c}$), is allocated to the three protein sectors, that is, the active enzymes, unused enzymes, and translational protein sectors (Fig. 2C). The unassigned proteins of the proteomic data set and those not covered by the experimental data were assigned to the housekeeping protein sector. The protein fraction $\phi_{p,c}$ of $0.26 \text{ g g}_{\text{cdw}}^{-1}$ considered by the PAM is in agreement with the fact that the growth-dependent part of the proteome constitutes roughly half of *E. coli*'s total protein content (3–5, 37, 39, 51). Note that the metabolic cost of the total protein biosynthesis in terms of energy and amino acid demand is fully covered by the biomass equation inherited from *i*ML1515.

The proteins within the data sets of Schmidt et al. (41) not covered by the PAM are mainly poorly characterized proteins (51% of all the protein fractions within the proteomic data set not covered by the PAM) or are assigned to transcription (8%), replication (6%), and posttranslational modification (6%). In turn, approximately one-third of the gene products considered by the PAM have not been quantified by Schmidt et al. (41). Those proteins are mainly inner and outer membrane proteins (58% of all gene products in the PAM not covered by the proteomic data set) or allocated to the glycerophospholipid (9%) and alternate carbon (8%) metabolism.

As experimental proteomic data suggest a consistent protein mass concentration across conditions, we modeled and added the invariable sum of active enzymes, unused enzymes, and translation protein to the PAM:

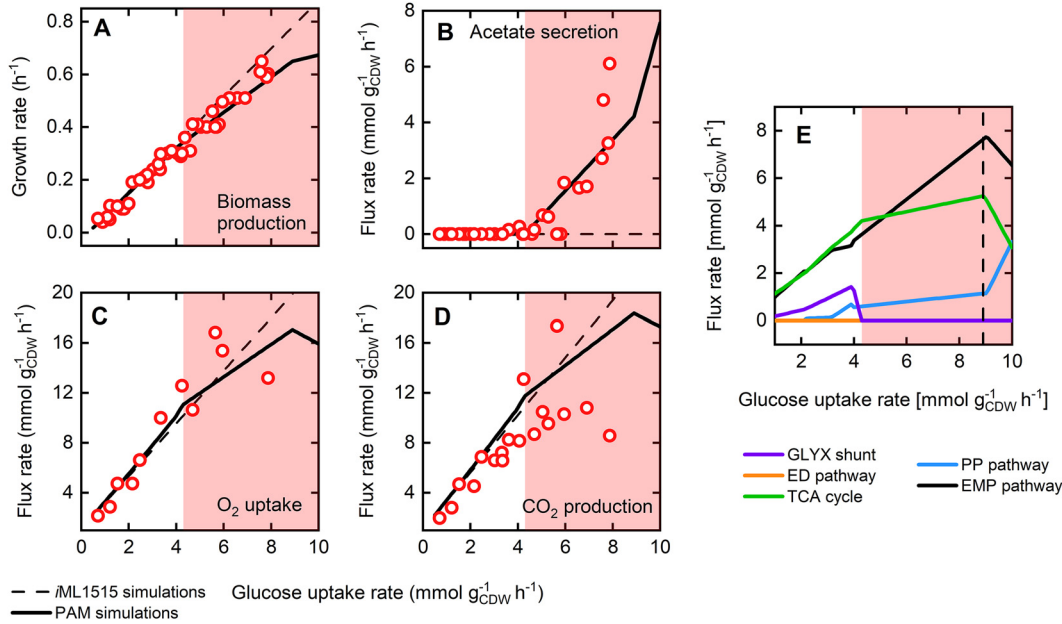


FIG 3 PAM predictions for *E. coli* phenotypes for a range of physiologically relevant glucose uptake rates in comparison to data from the literature. (A to D) Predictions (black lines) and experimental data (41, 52–56, 58, 59) (red dots) of (A) growth, (B) acetate secretion, (C) oxygen uptake, and (D) carbon dioxide production rates are compared. Simulation results using the original stoichiometric iML1515 model are additionally shown (dashed lines). (E) Simulated fluxes through central metabolic pathways. The dashed line marks the maximum glucose uptake rate of 8.9 mmol g_{CDW}⁻¹ h⁻¹, where the unused enzyme concentration ϕ_{UE} was defined to be zero. In all panels, the red shaded area marks the occurrence of fermentation resulting in the secretion of acetate.

$$\phi_{P,c} = \phi_T + \phi_{AE} + \phi_{UE} = 0.26 \text{ g}_{\text{CDW}}^{-1} \quad (1)$$

Here, ϕ_T , ϕ_{AE} , and ϕ_{UE} are the variable protein mass fractions of the translational protein, active enzymes, and unused enzyme sector, respectively. $\phi_{P,c}$ is the constant total protein mass concentration. Equation 1 introduces a concise representation of the capabilities and limits of *E. coli*'s metabolism and its condition-dependent protein allocation to stoichiometric modeling approaches. Due to the linear nature of the added protein allocation and enzyme activity constraints, classical flux balance analysis (FBA) and other established constraint-based modeling approaches can be applied to the PAM without a significant loss in computation speed compared to purely stoichiometric models.

The protein allocation model predicts wild-type phenotypes. To benchmark the PAM's predictive capabilities, we simulated the wild-type phenotypic behavior on glucose minimal medium and compared the results to extensive literature data (52–59). The maximum glucose uptake rate, a parameter for the unused enzyme sector, was set to 8.9 mmol g_{CDW}⁻¹ h⁻¹, which supported an observed maximal growth rate of 0.65 h⁻¹ (52). The simulated data derived with this constraint is in good agreement with experimentally observed phenotypes for a range of carbon-limited conditions and depicts significant improvements compared to the purely stoichiometric iML1515 model (Fig. 3). In particular, the acetate secretion trend correctly mirrors the metabolic overflow characteristics of *E. coli* above glucose uptake rates of 4.3 mmol g_{CDW}⁻¹ h⁻¹ (Fig. 3B). Despite reasonable projections of growth, acetate secretion, and oxygen uptake rates, the PAM as well as the iML1515 model overestimate carbon dioxide secretion rates, pointing to potential inconsistencies in the biomass synthesis equation used in both models.

The phenotypes simulated by the PAM are also mirrored in the fluxes through central metabolic pathways as shown in Fig. 3E. The distribution of the carbon flux among the metabolic pathways qualitatively follows previous findings (53). Under carbon-

limited and fully respiratory conditions, flux rates through the Embden-Meyerhof-Parnas (EMP) pathway, the tricarboxylic acid (TCA) cycle, the pentose phosphate (PP) pathway, and the glyoxylate shunt (GLYXS) scale linearly with the glucose uptake rate. The GLYXS activity reflects the need for anaplerosis to compensate for the drainage of TCA cycle intermediates under slow growth conditions, which has been experimentally demonstrated (60).

In the vicinity of the turning point from a fully respiratory to partially fermentative metabolism, the activity of the GLYXS diminishes completely. Beyond glucose uptake rates of $4.3 \text{ mmol g}_{\text{cdw}}^{-1} \text{ h}^{-1}$, limitations in the global protein household impede exclusive ATP production via respiration. As a consequence, the carbon flux is partly diverted from the NADH-yielding, and thus respiration-fueling, TCA cycle toward acetate. Simultaneously, the split between the EMP and PP pathways starts to increase with elevated glucose uptake rates in favor of glycolysis. While the NADPH supply through the PP pathway follows the growth-rate-dependent demand for biomass synthesis, the glycolytic flux (EMP) is accelerated to compensate for the carbon loss toward acetate.

The PAM's stoichiometric and protein allocation constraints support feasible growth states beyond the assumed maximum glucose uptake rate (Fig. 3). However, the increase in growth rate is at the expense of biomass yield, which becomes evident by a drastic decrease of the flux through the EMP pathway and the TCA cycle and a progressively increasing acetate secretion rate. These simulation results do not necessarily indicate an underestimated maximum glucose uptake rate since PAM parameterization stays feasible even for maximum glucose uptake rates far beyond physiologically relevant values (data not shown). In fact, this suggests that the main metabolically limiting factor for *E. coli* is neither stoichiometry nor protein allocation. As we will show in the next section, maximal growth rates may be attributed to a limited, maximum molar protein synthesis rate, thus indicating transcriptional restrictions as the determining growth-limiting factor.

Computing maximum substrate uptake rates. Since glucose is the preferred carbon source of *E. coli*, its metabolism and regulation are adapted for an effective glucose utilization, which led us to assume that there are no unused enzymes under glucose-excess conditions. However, it is reasonable to assume that such a state of adaption does not hold for alternative carbon sources. Thus, experimentally observed substrate uptake rates may not reflect growth states with (near-)zero unused enzymes. To enable the PAM to simulate growth on alternative carbon sources without the need for cultivation data, we approximated maximum substrate uptake rates assuming a maximum total protein synthesis rate $N_{\text{p,max}}$. $N_{\text{p,max}}$ represents a global cellular constraint for protein biosynthesis, applicable for any substrate. It manifests at the maximum substrate uptake rate and therefore at maximum growth at which the unused enzyme sector approaches zero. In the PAM, $N_{\text{p,max}}$ represents the sum of molar synthesis rates of proteins from the active enzymes and the translational sectors.

Based on the PAM and phenotypic data of *E. coli* grown on several substrates (61), we found that an $N_{\text{p,max}}$ of $2.04 \text{ } \mu\text{mol g}_{\text{cdw}}^{-1} \text{ h}^{-1}$ defines maximum substrate uptake rates (Table S1). For glucose as the sole carbon source, an uptake rate of $9.82 \text{ mmol g}_{\text{cdw}}^{-1} \text{ h}^{-1}$ is predicted accordingly. Interestingly, experimental values suggest that the maximum growth rate on acetate is approached for an uptake rate of $19.6 \text{ mmol g}_{\text{cdw}}^{-1} \text{ h}^{-1}$ and a protein synthesis rate below $N_{\text{p,max}}$. This phenomenon indicates that protein allocation limits the maximum growth on acetate. Acetate is metabolized through the TCA cycle and the GLYXS (61). Energy-yielding routes other than respiration are biochemically impossible since substrate phosphorylation via the protein-efficient fermentation pathway results in acetate formation and, hence, a metabolic cycle with zero net carbon uptake. As a consequence, ATP generation solely relies on protein-costly respiration. We assume that the allocation of protein to the respiratory pathway is ultimately limited by the total protein concentration $\phi_{\text{p,c}}$. Thus, ATP supply and cofactor regeneration become the limiting factors before the maximum molar protein synthesis capability is reached.

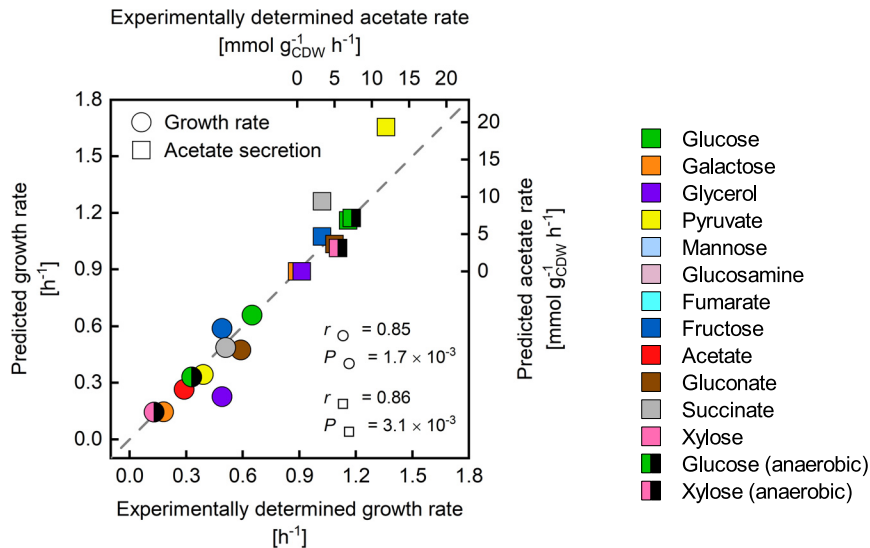


FIG 4 Comparison of experimentally determined (61) and predicted growth and acetate secretion rates on different, single carbon sources. Substrate uptake rates were constrained according to the reported values. Maximum substrate uptake rates were approximated according to a maximum total protein synthesis rate $N_{P,max}$ (Table S1). The goodness of the correlations between simulations and experiments for growth and acetate secretion rates were determined using the Pearson correlation coefficient r and the corresponding P value.

By parameterizing the unused enzyme sector with the substrate-specific maximum substrate uptake rates, PAM simulations of phenotypes of *E. coli* grown on alternative carbon sources showed a good correlation with experimentally observed data (Fig. 4). The overestimation of growth and acetate secretion rates at the determined maximum substrate uptake rates (Fig. S1) indicates an unused potential to adapt *E. coli* to these alternative carbon sources. This assumption is, for instance, supported by a study by Fong et al. (62), in which the glycerol uptake rate was evolved to around $15 \text{ mmol g}_{cdw}^{-1} \text{ h}^{-1}$, which is close to the PAM's prediction. Interestingly, the shift from glucose to glycerol at the beginning of the evolutionary adaption of *E. coli* caused immediate, significant transcriptional responses. Regulatory mechanisms appeared to significantly affect the expression of genes associated with catabolism, such as *cra* and *crp*, presumably to nonspecifically scavenge and prepare for alternate substrates. During the adaptive evolution process, metabolism was focused on the available substrate, glycerol, and the basal transcriptional state was approached again. Thus, optimal glycerol uptake and utilization were mediated by changes in transcriptional regulation resulting in the adaption of catabolism to this sole carbon source. Moreover, flagellar and motility gene expression decreased during the adaptation, probably increasing the protein availability for growth-related processes.

Prediction of *E. coli* flux distributions. By exploiting the optimality principles of microbial growth, GEMs give quantitative insights into the intracellular flux distribution and pathway usage, based purely on stoichiometric constraints. For a particular environmental condition of interest, the prediction accuracy of the stoichiometric GEM generally scales with the amount and accuracy of experimental data introduced to the model in the form of flux constraints. With a minimum need for such data, i.e., the substrate uptake rate, the PAM allows for an accurate blueprint of the intracellular metabolic processes. A comparison with fluxomics data from multiple studies shows that flux distributions of the central carbon metabolism of *E. coli* grown on a minimal glucose medium are well predicted by the PAM, which is indicated by Pearson correlation coefficients up to 0.97 (Fig. 5). Here, the PAM was constrained with the measured glucose uptake rates, and the unused enzyme sector was parameterized according to the

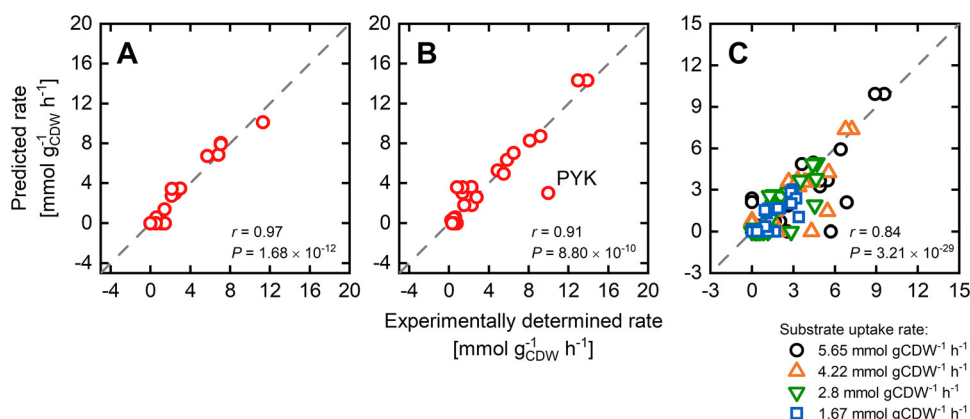


FIG 5 PAM predictions of intracellular fluxes of the central carbon metabolism of *E. coli* grown on a glucose minimal medium. The glucose uptake rate was constrained with experimentally determined values. The unused enzymes sector was parameterized according to the computationally determined maximum value of $9.82 \text{ mmol g}_{\text{cdw}}^{-1} \text{ h}^{-1}$ based on a maximum total protein synthesis rate N_p of $2.04 \mu\text{mol g}_{\text{cdw}}^{-1} \text{ h}^{-1}$. (A to C) The predictions are compared with experimental flux data from panels A (61), B (63), and C (53). The goodness of the correlations was computed based on the Pearson correlation coefficient r and the corresponding P value.

methodically determined maximum glucose uptake rate, which, in turn, was a direct model output (cf. the previous section). Thus, protein allocation and enzymatic constraints alone enhance constraint-based modeling as a metabolic prediction tool, making this modeling approach particularly useful under data scarcity. The relatively large discrepancies between simulated and experimental flux data for the pyruvate kinase (53, 63) (Fig. 5B and C) result from the neglect of the phosphotransferase system (PTS) for glucose uptake in the ^{13}C metabolic flux analysis models. The PTS system transfers the phosphate group from phosphoenolpyruvate (PEP) to glucose during its import, thereby producing glucose-6-phosphate and pyruvate. The omission of this alternative uptake system necessitates an elevated pyruvate kinase flux to balance PEP. In traditional ^{13}C -based metabolic flux analysis, glucose uptake via the PTS and the ABC transporter cannot be dissected, as they do not affect the ^{13}C isotope patterns of metabolites. It is therefore common to incorporate only one pathway in these models.

We further used the methodically determined maximum substrate uptake rates and constrained the PAM with measured uptake rates to predict flux distributions for non-glucose carbon sources (Fig. S2). Here, the prediction capabilities were more diverse, resulting in high correlations for acetate, galactose, and succinate ($r > 0.92$) but intermediate to weak predictions, e.g., for fructose or gluconate ($r > 0.65$). In the case of gluconate consumption, the Entner-Doudoroff (ED) pathway was experimentally observed to be the main catabolic route, whereas the simulated carbon flux was exclusively channeled toward pyruvate through the pentose phosphate (PP) pathway. The flux split between the PP and ED pathways is highly sensitive to the ratio in the protein demands of both pathways (data not shown). Consequently, inconsistencies in the applied k_{cat} values, particularly of backward reactions, but also the substrate-dependent differences in the biomass compositions (13) may cause these observed discrepancies. Further in-depth investigations are needed to validate one or the other assumption.

PAM explains the growth defect upon heterologous protein expression. Enzyme overproduction or the expression of heterologous genes is a common strategy in many biotechnological disciplines. General purposes are introducing novel cellular functionalities, flux enforcement through a specific pathway, or investigation on cellular processes via reporter proteins. In any case, the introduced pull of proteins from the limited native protein household and the resulting metabolic burden inevitably cause a growth defect.

To investigate and evaluate the response of the PAM to such an induced protein

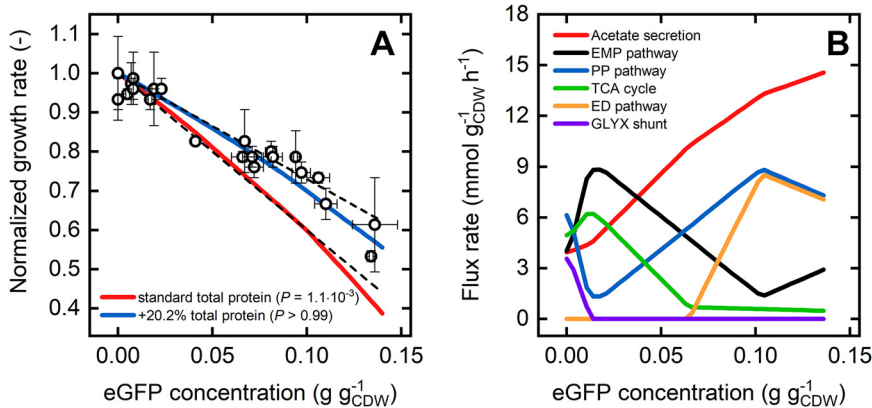


FIG 6 (A) Simulated growth rates relative to the maximum growth rate are shown for a range of intracellular eGFP concentrations using the PAM with a standard (red line) and increased (blue line) total condition-dependent protein concentration $\phi_{P,C}$ ($0.26 \text{ g g}_{CDW}^{-1}$ and $0.31 \text{ g g}_{CDW}^{-1}$, respectively). Experimental data are taken from reference 64. P values are derived from Student's t -test and indicate how well model predictions explain the experimental data. The dashed lines illustrate the predicted range of values based on theoretical considerations by Bienick et al. (64). (B) Simulated fluxes through central metabolic pathways are shown for a range of intracellular eGFP concentrations and a $\phi_{P,C}$ of $0.31 \text{ g g}_{CDW}^{-1}$.

demand, we simulated growth for a range of expression levels of an enhanced green fluorescent protein (eGFP) and compared relative growth rates with experimental data from Bienick et al. (64). For the standard concentration of total condition-dependent protein $\phi_{P,C}$ of $0.26 \text{ g g}_{CDW}^{-1}$, the simulated relative growth defect is much more pronounced than to experimentally observed growth for intracellular eGFP concentrations beyond $0.03 \text{ g g}_{CDW}^{-1}$ (Fig. 6A). The use of an *E. coli* TUNER strain, a derivative of the genome-reduced BL21 strain optimized for protein expression, in this study, explains this discrepancy. In this strain, the protein requirements for the growth rate-independent housekeeping sector are reduced, and more resources are available for the metabolic, condition-dependent protein sectors, which can be reflected in the PAM by increasing the total protein concentration $\phi_{P,C}$. Accordingly, increasing the protein availability by expanding $\phi_{P,C}$ about 20% attenuates the detrimental effects of eGFP expression on growth and results in an excellent reproduction of experimentally observed phenotypes (Fig. 6A).

Interestingly, the relation between eGFP expression strength and growth predicted by the PAM is nonlinear, which contrasts with a previous theoretical postulation (64). The nonlinearity arises from a combined effect of a protein drain from the translational and the active metabolic enzyme sectors. The enforced protein drain toward eGFP causes a decline in the ribosome concentration, resulting in a reduced translation rate. Under the assumption of a constant total protein concentration, the protein allocation to the metabolic sectors producing biomass precursors, amino acids, and energy decreases, which, in a vicious cycle, further decelerates the protein production rate and growth. This intracellular tug-of-war for proteins is intrinsically manifested in the PAM, leading to outperformance in predicting protein overexpression phenotypes over too simplified coarse-grained models.

The PAM also discloses relevant effects of the protein drain on the pathway flux level. For an increasing eGFP expression strength and the accompanied protein deficiency, central carbon fluxes are progressively diverted to fermentation pathways (acetate secretion) and eventually to the ED pathway. Both routes are more protein efficient but yield fewer energy equivalents per substrate molecule than respiration or the EMP pathway (39, 65, 66). The PAM confirms this diversion of ATP generation from respiratory to fermentation pathways. The proportion of ATP generated by respiration, more precisely by the ATP synthase, in the total cellular ATP generation rate

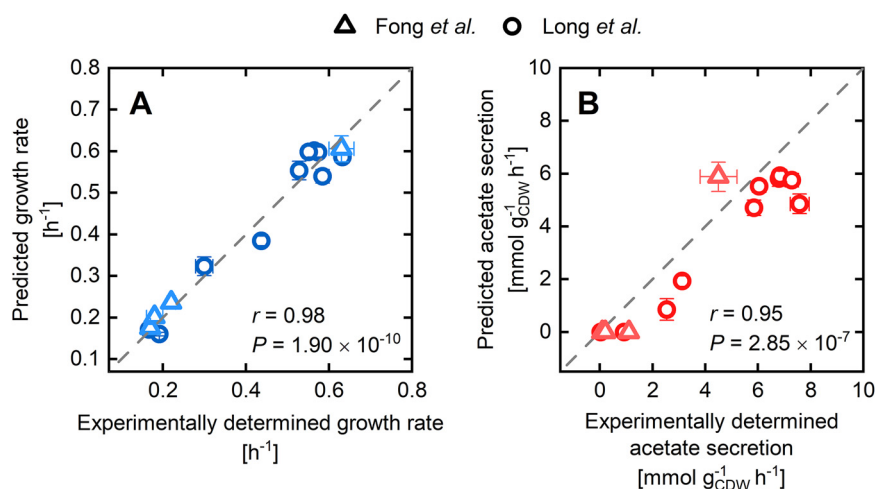


FIG 7 (A and B) Comparison of predicted growth (A) and acetate secretion rates (B) of gene-deletion mutant strains with experimentally determined values taken from references 70 (circles) and 72 (triangles). Predictions were made applying the PAM and constraining the glucose uptake rates to observed values.

significantly decreases with an increasing eGFP expression (Fig. S3). Thus, the excelled protein burden enforces a shift toward protein-efficient, substrate-level phosphorylation in *E. coli*.

Limitations in the protein allocation of single enzymes lead to gene deletion mutant phenotypes. Alongside the over- and heterologous expression of genes, rearrangement of metabolic networks and flux distributions by gene deletions is a core instrument in metabolic engineering. In recent years, many computational strain design methods have been developed to accelerate and rationalize the engineering of microbial cell factories. However, in contrast to the vast number of model-driven strain design and optimization methods (67), constraint-based methods have often proven unreliable in predicting phenotypes of gene deletion mutant strains (GMSs). While GMSs have been shown to evolve toward FBA-predicted phenotypes (68), observed growth defects and intracellular fluxes of nonevolved GMSs cannot be explained by stoichiometry and a cellular growth objective alone (69–71).

First, we tested the impact of 10 single-gene deletions on the PAM's FBA results by parameterizing the unused enzyme sector with the methodically determined maximum glucose uptake rate of $9.82 \text{ mmol g}_{\text{cdw}}^{-1} \text{ h}^{-1}$. The calculated phenotypes did not significantly differ from the unperturbed wild-type solutions and, hence, did not compare to experimental data (70, 72) (Fig. S4). Only after constraining the glucose uptake rate to the measured values did the simulated growth and acetate secretion rates correlate well with the experiments (Fig. 7). Furthermore, a blind test in the form of FBA simulations augmented with *in vivo* glucose uptake rates of the GMS, but with an *intact* target gene, yielded the same phenotypes (Fig. S5). These results led us to conclude that the main driver for the observed growth defects of GMSs is a naturally orchestrated catabolite uptake repression induced by the respective network perturbations. A gene disruption resulting in a growth phenotype implicates a decrease in the levels of one or more amino acid precursors since tight proteome coordination (4) hampers enzyme allocation toward alternative precursor synthesis routes. According to the coarse-grained model of You et al. (4), which considers the general mechanism of rRNA transcription (73), low amino acid levels stall ribosomal protein synthesis via ppGpp. Moreover, low precursor levels, such as oxaloacetate or α -ketoglutarate, lead to an increased cAMP synthesis and elevated CRP-cAMP levels, which foster the expression of catabolic enzymes for alternative substrates adding to growth reduction. Elevated cAMP concentrations, which were experimentally observed in multiple GMSs (74), indicate increased CRP-cAMP activities and support this view on gene deletion triggering

the integral feedback of metabolic control. Therefore, we argue that the degree of metabolic rewiring in GMSs necessary to overcome growth defects is often low. Indeed, restoration of growth can be achieved by evolutionary adaptation processes that alter these regulatory patterns (74).

The assumption that a regulatory substrate uptake inhibition shapes GMS' phenotypes raised the question if the PAM can quantitatively predict substrate uptake modes. To tackle this question, we recalled computational frameworks such as minimization of metabolic adjustment (MOMA) (75), regulatory on/off minimization (ROOM) (76), and relative change (RELATCH) (69), which determine the metabolic impact of gene knockout strains. The common principle behind all three methods is the minimization of the metabolic response to genetic perturbations due to an unchanged regulatory system that forces the GMS' flux distribution toward the original steady state. In the context of protein allocation, the minimal response principle can be translated as follows: upon a network perturbation, a GMS establishes a substrate uptake rate so that increases in protein allocation toward single enzymatic reactions are minimal compared to genetically unperturbed strains. The cellular objective is to allow for maximum metabolic activity in the face of knockout-induced flux rerouting and a hampered reallocation of protein due to a strict (wild-type) regulatory regime.

We applied the minimal response principle to the PAM for mutant phenotype predictions, and we simulated growth optimal flux distributions for a range of substrate uptake rates (cf. Materials and Methods for a detailed description). For each flux distribution, the difference in the synthesis rate ΔN_e between a reference, wild-type state at maximum growth and a mutant state is calculated for each enzyme within the PAM. The GMS's maximum substrate uptake rate is determined from the flux distribution for which ΔN_e meets a defined upper bound ΔN_e^{crit} for at least one enzyme within that flux distribution. In doing so, we assign a certain flexibility to the overexpression capacity of single enzymes, which may be attributed to the activation of underutilized enzyme capacities. Moreover, if ΔN_e^{crit} is met for one enzyme, flux rerouting to circumvent the saturated enzyme reaction is impossible. Refusing substantial metabolic rewiring conforms to experimental observations that unevolved GMSs do generally not show a large amount of metabolic adjustment to the metabolic network perturbations (68, 75). The strict upper bound ΔN_e^{crit} reflects this metabolic rigidity on changes of enzyme synthesis rates in GMSs compared to the parental wild-type strain.

Figure 8 shows the prediction results in comparison to experimentally determined phenotypic data (70–72) for an ΔN_e^{crit} of $16 \text{ nmol g}_{\text{cdw}}^{-1} \text{ h}^{-1}$. ΔN_e^{crit} was derived from minimizing the discrepancy between simulated and experimentally observed phenotypes for various single-gene deletion mutants (cf. Materials and Methods for a detailed description). With one exception, there is a general agreement between predicted and experimentally determined phenotypes (Pearson correlation: $r = 0.81$, $P = 7.76 \times 10^{-11}$). By disregarding the Δrpe deletion mutant, the Pearson correlation improves to $r = 0.97$ with a P value of 1.5×10^{-23} . For the omitted outlier, simulations show a near-wild-type behavior, whereas growth is significantly reduced in the experiment. The *in vivo* observed growth defect of the Δrpe GMS is peculiar since knockouts of enzymes adjacent to ribulose-phosphate 3-epimerase, such as *gnd*, do not show a similar phenotype. Even though the gene products of *gnd* and *rpe* catalyze consecutive reactions transforming 6-phospho-D-gluconate to D-xylulose 5-phosphate, an important intermediate in the PP pathway, deletion of *gnd* has a significantly smaller effect on growth, as does the deletion of *rpe*. Moreover, the experimentally determined phenotype of a Δrpe GMS from reference 77 was similar to the wild-type *E. coli* strain and thus obscures a plausible judgment of the simulated phenotype.

The considerable agreement between experimental and simulated phenotypic data also applies to intracellular flux data. We computed flux distributions for single-gene deletions that result in an experimentally proven inactivation of the corresponding reaction (i.e., we excluded single-gene deletions of isozymes) (71). The predicted flux responses correlated well with experimentally determined flux data (Fig. 9), noticeable

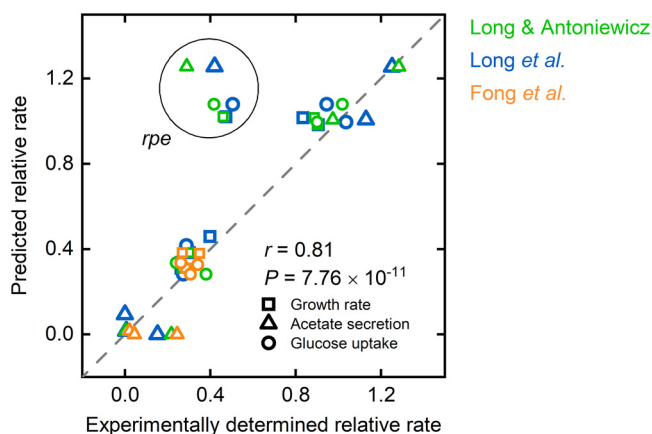


FIG 8 Comparison between simulated and experimentally determined phenotypes of GMSs without considering observed glucose uptake rates as model constraints. Predicted growth, glucose uptake, and acetate secretion rates are plotted against experimentally determined values taken from references 70 (blue), 71 (green), and 72 (orange). All values were normalized with corresponding wild-type data. Predictions were made applying the PAM and constraining the unused enzyme sector according to a maximum glucose uptake rate of $9.82 \text{ mmol g}_{\text{cdw}}^{-1} \text{ h}^{-1}$. The maximum overexpression capacity of single enzymes $\Delta N_{\text{e}}^{\text{crit}}$ was set to $16 \text{ nmol g}_{\text{cdw}}^{-1} \text{ h}^{-1}$.

on Pearson's correlation coefficients of $r > 0.93$, for all tested single deletions, except for the $\Delta tpiA$ mutant ($r = 0.67$). For the $\Delta tpiA$ GMS, the experimentally observed rerouting of the glycolytic flux through the methylglyoxal pathway toward pyruvate (71, 78) contrasts with the PAM results, which suggests the activation of the ED pathway surpasses the blocked glycolysis. The simulated behavior can be traced back to an underestimated protein demand of the ED pathway, as predictions were improved after decreasing the k_{cat} values of the two central ED pathway reaction steps phosphoglucuronate dehydratase and 2-dehydro-3-deoxyphosphoglucuronate aldolase to 2.5% of the original values. After reevaluating the maximum glucose uptake rate, the simulations showed a diversion of the glycolytic flux into the methylglyoxal pathway and acetate secretion (Fig. S6). This selective adaption of turnover numbers, and thus of protein costs for the corresponding reactions, resulted in a significant convergence of simulated flux data toward measured values for the $\Delta tpiA$ and also the Δrpe GMS (r of 0.90 and 0.99, respectively) (Fig. S7). However, at the same time, flux predictions for the Δgnd mutant were deteriorated ($r = 0.95$). Here, flux was shifted from the ED pathway toward glycolysis, which contradicts experimental data. Nevertheless, the k_{cat} value adaption only slightly lessened the PAM's overall predictive capabilities (Fig. S7) and thus generally points to the need for fine-tuning kinetic parameters to improve the flux split between pathways across different conditions.

DISCUSSION

Proteins are the major molecular class in cells, and because they are the catalyst for global cellular functionalities, the importance of the mutual connection between microbial metabolic behaviors and protein allocation is broadly accepted (3, 5–7, 37–39). Based on existing techniques for implementing protein allocation and enzymatic constraints in constraint-based modeling (37, 40), we implemented a methodology to account for the total condition-dependent proteome in an *E. coli* GEM. Besides the integration of basic enzyme kinetics in the form of turnover numbers to model the concentrations of active enzymes, the resulting PAM considers simple, linear relations between microbial growth and the translational protein as well as the unused enzyme sector to describe 81% of the total protein mass concentration.

The PAM's accurate predictions confirmed the prominent role of protein allocation in shaping microbial metabolism. Nevertheless, protein allocation itself appears to be

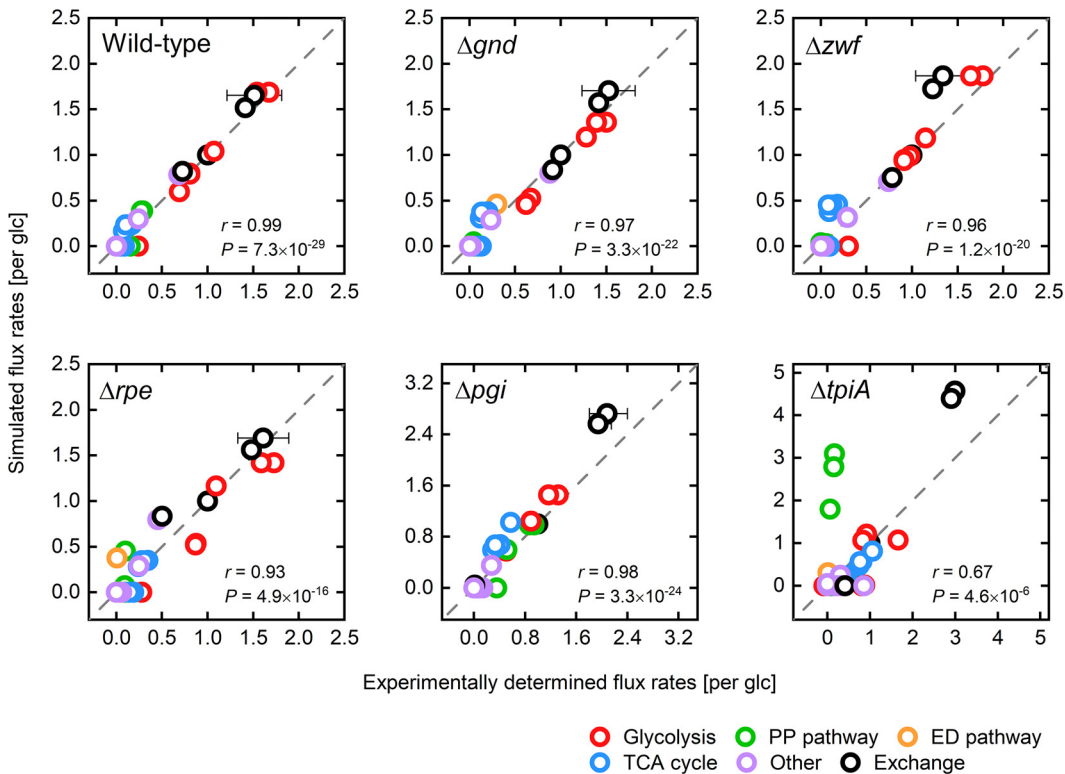


FIG 9 Comparison between experimentally determined and simulated flux distributions of GMSs without considering observed glucose uptake rates as model constraints. Predicted intracellular flux rates are plotted against experimentally determined values taken from Long and Antoniewicz (71). All data were normalized with corresponding glucose uptake rates. Predictions were made applying the PAM and constraining the unused enzyme sector according to a maximum glucose uptake rate of $9.82 \text{ mmol g}_{\text{cdw}}^{-1} \text{ h}^{-1}$. The maximum overexpression capacity of single enzymes $\Delta N_{\text{e}}^{\text{crit}}$ was set to $16 \text{ nmol g}_{\text{cdw}}^{-1} \text{ h}^{-1}$.

regulated by biochemical limits. Genetic sequencing results of *E. coli* strains that have undergone extensive adaptive laboratory evolution (ALE) (79) suggest a causal link between transcription limitations and maximum cell proliferation rates. The adapted strains, all exhibiting a fitness increase of up to 1.6-fold, showed mutations in the *rpoB* or *rpoC* gene leading to single amino acid substitutions in the β/β' subunit of the RNA polymerase. This subunit is part of the enzyme's active center. Thus, the observed mutations globally affect transcription (79). We introduced the total molar synthesis rate of proteins as a proxy for the transcription capability of *E. coli* since the PAM does not represent any transcriptional processes. We found that maximum substrate uptake rates, and consequently maximum growth rates, are dictated by an ultimately limited total molar synthesis rate of condition-dependent proteins. Based on our study, we presume that a reduction of transcriptional limitations in evolved *E. coli* strains allows for an increased protein allocation toward the metabolically active enzymes and translational protein sector. A detailed proteomics study is necessary to uncover changing protein allocation principles in growth-optimized microbial strains and to test this hypothesis with the PAM.

By recognizing transcriptional limitations as hard constraints for the microbial metabolism and recalling that the presented PAM predictions represent growth optimal flux states, we deduce a cellular principle, which is supported by previous findings (39); particularly under substrate-limited conditions, *E. coli* regulates central cellular processes to maintain a Pareto-optimum between growth rate and the ability to adapt its metabolism flexibly to changing environmental conditions. The degree of flexibility is directly linked to the amount of allocated unused enzymes, which is hard-coded in the

PAM. Interestingly, the cell retains a strictly substrate uptake-orientated allocation of protein to the unused enzyme sector even when protein becomes a metabolically limiting factor, apparent from the onset of overflow metabolism. This preservation of flexibility could allow for an evolutionary advantage in the original environmental niche. However, it may be a promising target for engineering *E. coli* or any microbe toward a high-performance cell factory.

Recently, systems metabolic engineering was emphasized as an integral part of the development and optimization of microbial cell factories (67, 80–83). Thus, we want to put the PAM forth to highlight the advantages of considering total condition-dependent proteome allocation for constraint-based modeling techniques that favor more accurate mutant strain predictions. We showed sound predictions of growth upon a range of overexpression levels of a nonenzymatic protein. The predictive capability for the overexpression of enzymes participating in metabolic reactions, e.g., in heterologously expressed pathways, still needs to be verified. However, the strict regulation of protein allocation, represented by the functional description of translational protein and unused enzyme sectors in the PAM, appeared to shape metabolic responses and is insensitive to genetic interferences. We also confirmed a rather inflexible protein allocation behavior for gene deletion mutant strains. Coherent prediction results were obtained by allowing only small divergences from a wild-type expression rate of single enzymes. Hence, in addition to the restrictions mediated by CRP-cAMP, we propose a link to transcriptional limitations, similar to our observation of maximum growth of wild-type strains mentioned above. This hypothesis is supported by ALE experiments in which adaption of single-gene deletion mutants frequently generated mutations affecting the regulation of global and pathway-specific transcription (74). These mutations possibly eliminate transcriptional hurdles and (partly) restore growth rates.

In summary, we want to stress the importance of protein allocation constraints in GEMs for the systematic constraint-based reconstruction and analysis (COBRA) and the design of microbial metabolism without having to sacrifice computational speed or applicability of established COBRA methods (84). However, the limited availability and credibility of enzymatic kinetic data, particularly of turnover numbers, still poses a significant obstacle in providing PAMs for any microorganism. Therefore, we join the call to establish a thorough kcatome as part of an accessible, genome-wide kinetome (85).

MATERIALS AND METHODS

Formulating the protein allocation model. The protein allocation model (PAM) is based on the *E. coli* K-12 MG1655 GEM iML1515 reconstruction (45). It includes linear representations of the protein mass concentrations of the active enzymes, the unused enzymes, and the translational protein sector as additional constraints. The sum of the protein mass concentrations of all sectors is kept constant according to Equation 1. The modeled protein sectors are described in detail in the following.

The translational protein sector. In agreement with previous studies, a linear correlation between the translational protein sector ϕ_T and the growth rate μ (5, 41) was implemented. Therefore, the inverse of the maximum ribosomal elongation rate w_T and a measure $\phi_{T,0}$ for an increasing overcapacity of ribosomes with a decreasing growth rate (43, 86) were used as the slope and intercept, respectively:

$$\phi_T = \phi_{T,0} + w_T \mu \quad (2)$$

Both parameters w_T and $\phi_{T,0}$ were determined for *E. coli* by fitting Equation 2 to measured, cross-conditional concentration data of the translational protein sector (41) (Fig. 2B), resulting in values of 50.0 mg g_{cdw}^{-1} (19% of the total protein mass concentration $\phi_{P,c}$ of 0.26 g g_{cdw}^{-1}) and 36.8 mg h g_{cdw}^{-1} for $\phi_{T,0}$ and w_T , respectively. The relative value for $\phi_{T,0}$ differs from the visual intercept in Fig. 2B, since protein mass concentrations in Fig. 2 were related to the total protein concentration of 0.32 g g_{cdw}^{-1} measured by Schmidt et al. (41).

The unused enzyme sector. An evolutionary characteristic of microbes facing unforeseeable changes in environmental conditions is the synthesis of unused proteins. This protein hedging empowers the cell to quickly increase central carbon metabolism fluxes upon a sudden increase in substrate availability or to immediately catabolize a new carbon source (44). Enzyme overabundance also facilitates metabolic robustness against genetic perturbations (87). However, these protein reserves significantly reduce the growth rate, discussed in depth by O'Brien et al. (44) and Scott et al. (3). Protein expression and (over)allocation is generally coordinated by the cyclic AMP (cAMP)-dependent signaling pathway via the cAMP-activated global transcriptional receptor protein (CRP) (4, 88) (Fig. 1). The CRP-cAMP complex enhances the transcription of over 100 genes by attaching near or at their promoter

regions, thereby mediating the binding of RNA polymerase for transcription initiation (89). The target genes are mainly associated with catabolism, and thus, CRP-cAMP aids in stimulating the carbon influx leading to the accumulation of precursors for amino acids and stimulation of protein synthesis. Amino acid precursors, such as oxaloacetate or α -ketoacids, inhibit the cAMP synthesis, thus lowering the CRP-cAMP level and closing the negative feedback loop between protein precursors and substrate uptake. We refer to You et al. (4) for respective insights into the cAMP signaling pathway. In this way, CRP-cAMP indirectly coordinates the global protein allocation, including the allocation of unused enzymes, even though the transcription of only a small fraction of genes is directly affected. Considering the cAMP-controlled signaling pathway as a blueprint for protein synthesis regulation, we modeled the unused enzyme sector as a negative linear function of the substrate uptake flux, which we mathematically expressed as

$$\phi_{UE} = \phi_{UE,0} - w_{UE} \nu_s \quad (3)$$

where $\phi_{UE,0}$ is the unused enzyme concentration at zero substrate uptake, and ν_s is the substrate uptake rate. w_{UE} relates the decrease of the unused enzymes' concentration to the increase in ν_s . $\phi_{UE,0}$ was determined from ME model simulations of un- and underutilized enzymes (44), resulting in a value of $0.17 \text{ g g}_{\text{cdw}}^{-1}$ equivalent to 65% of the total protein mass concentration $\phi_{p,c}$. At zero substrate uptake rate ($\nu_s = 0$), and therefore at zero growth, the sum of unused enzymes and translational protein $\phi_{UE,0} + \phi_{T,0}$ is $0.22 \text{ g g}_{\text{cdw}}^{-1}$. Thus, near $\nu_s = 0$ the PAM explains only a fraction (84%) of the total protein mass concentration $\phi_{p,c}$ which is, however, assumed to be constant under any given condition. We expect this discrepancy to be caused by deviations from an otherwise constant total protein household in severely starving cells or specific protein requirements during non-growth conditions. The fraction not explained by the PAM is compensated for by nonphysiological, random activation of metabolic fluxes when calculating growth-optimal solutions near substrate uptake rates of zero. Thus, the PAM may produce feasible solutions at low metabolic activity but hardly represent physiologically relevant flux states.

The slope w_{UE} is a measure of the increase in enzyme usage efficiency with increasing substrate uptake rate. It is individually assigned for each substrate under the assumption that the unused enzyme concentration is zero at the maximum substrate uptake rate. Thus, w_{UE} is calculated from

$$w_{UE} = \frac{\phi_{UE,0}}{\nu_{s,\text{max}}} \quad (4)$$

The maximum substrate uptake rate $\nu_{s,\text{max}}$ needs to be provided as an observable but can be directly inferred from the PAM by assuming an upper limit in the transcriptional capacity, as shown in "Determination of Maximum Substrate Uptake Rates."

The active enzyme sector. To account for the enzyme demand of metabolic fluxes, we integrated enzyme mass balances for all relevant metabolic reactions in the stoichiometric matrix of *iML1515* according to the GECKO framework (40), as explained in the following. The employed enzyme mass balance formulation (Equation 5) relates the flux ν_e to the minimally required concentration ρ_e of the enzyme catalyzing reaction e by the enzyme's turnover number $k_{\text{cat},e}$. Since the PAM accounts for the total proteome by mass, we transformed molar to mass concentrations using the molar mass M_e of each enzyme in the GEM.

$$\rho_e = \nu_e \frac{M_e}{k_{\text{cat},e}} \quad (5)$$

Eventually, the sum of the concentration of all E enzymes constitutes the metabolically active enzyme sector ϕ_{AE} and is expressed as

$$\phi_{AE} = \sum_e^E \rho_e \quad (6)$$

Parameterization of the active enzyme sector is important to facilitate a meaningful relation between fluxes and enzyme concentrations. For the PAM, we determined k_{cat} values for 2,843 enzymes of the *iML1515* model from queries of the BRENDA (90) database following the protocol of Sánchez et al. (40).

Since simulated maximum growth rates were unreasonably low when applying the initial k_{cat} set, the k_{cat} data set was manually curated to allow for the computation of reasonable phenotypes (cf. Data Set S1, sheet 1 for the final data set). Therefore, enzymes were ranked according to the sensitivity of FBA-derived, maximal growth rates toward their k_{cat} values. Approximately 150 k_{cat} parameters that were found to be most influential on computed growth rates were reevaluated. For most of the underlying enzymes, no specific k_{cat} entry could be found in BRENDA, and k_{cat} values had been taken from close enzyme classes during the automated database queries. This k_{cat} approximation was corrected by manually querying SABIO-RK (91), UniProt (92), and primary literature for more sound values regarding growth optimality. If no specific data could be found for an enzyme, k_{cat} values from adjacent enzyme classes were manually selected and assigned to the PAM.

The manual curation effort transformed the primary *in vitro* k_{cat} estimates to effective (or apparent) *in vivo* turnover numbers k_{app} (48, 93), relating enzyme concentrations to metabolic fluxes under unlimited growth conditions. Recently, measurements of $k_{\text{app,max}}$, the maximum k_{app} values across conditions, were extrapolated to genome scale using machine learning models informed with biochemical and

enzymatic data (94). Interestingly, although the distribution of $k_{app,max}$ values is significantly different from our curated set (Fig. S8), simulations using the PAM parameterized with either or both $k_{app,max}$ sets yielded comparable phenotypes (data not shown). Thus, it appears that the ratio of k_{cat} values between different reactions or pathways is an essential factor for the predictive capability of PAMs in general.

Processing of experimental proteomic data from literature. The experimental proteomic data set generated by Schmidt et al. (41) was used in this study to parameterize the translational and the unused enzymes sector of the PAM. To fit the units of the data set to the model variables, the protein mass per cell measured by Schmidt was converted to dry cell weight-based concentration values:

$$\phi_p = \frac{M_p}{V_\mu c_{cdw}} \quad (7)$$

Here, ϕ_p is the intracellular concentration (g g_{cdw}^{-1}), and M_p is the experimentally determined mass per cell of a protein or a protein sector P . V_μ is the growth rate-dependent cellular volume, which was thoroughly determined by Volkmer and Heinemann (95). The cell dry weight (cdw) concentration per cell volume c_{cdw} was assumed to be $268.36 \text{ g}_{cdw} \text{ liter}^{-1}$ (50). Refer to Data Set S1, sheets 2 and 3 for the complete data set.

UniProt identifiers (92) and the COG classification (Clusters of Orthologous Groups) were used to match the proteins and protein sectors of the PAM to the proteomic data.

Solving the protein allocation model problem. All flux solutions and corresponding phenotypes in this work represent growth-optimal solutions of the following, classical FBA-problem amended with additional protein allocation and enzymatic constraints:

$$\begin{aligned} & \max_{\substack{\vec{v} \in \mathbb{R}^N, \vec{\rho} \in \mathbb{R}^E}} \mu \\ & \text{s.t. Classical FBA constraints} \\ & \overline{S} \vec{v} = \vec{0} \left. \begin{array}{l} v_i^{lb} \leq v_i \leq v_i^{ub} \\ v_i \geq 0 \end{array} \right\} \forall i \in N \\ & \text{Protein allocation constraints} \\ & \rho_e = v_e \frac{M_e}{k_{cat,e}} \left. \right\} \forall e \in E \\ & \rho_e \geq 0 \\ & \phi_T = \phi_{T,0} + w_T \mu \\ & \phi_{UE} = \phi_{UE,0} - w_{UE} v_s \\ & \phi_{P,c} = \phi_T + \phi_{UE} + \sum_e \rho_e \end{aligned} \quad (8)$$

Here, S is the stoichiometric matrix of the original GEM, and v_i is the flux variable of reaction i from the metabolic reaction pool N . Note that each reversible reaction in the original GEM is split into irreversible forward and backward reactions to only allow for positive flux values; hence, the lower and upper flux bounds v_i^{lb} and v_i^{ub} are equal to or greater than zero. The additional protein allocation and enzymatic constraints comprise the mass concentrations ρ_e for each considered enzyme or enzyme complex, as well as the mass concentrations of the ribosomal and unused enzyme sector ϕ_T and ϕ_{UE} , respectively. All protein allocation constraints in Equation 8 were added to the stoichiometric matrix S of the GEM iML1515 representing the *E. coli* K-12 MG1655 strain (45).

Each reaction e from pool E , comprising all reactions linked to one or more genes via a gene-protein-reaction (GPR) relation, is assigned to precisely one protein with a unique turnover number $k_{cat,e}$ and molar mass M_e . In case a reaction e is catalyzed by an enzyme complex (multiple genes are connected via logical AND operators in the GPR relation), the molar mass M_e is the sum of molar masses of the participating gene products. If two or more enzymes can catalyze the same reaction independently from each other (multiple genes are connected via logical OR operators in the GPR relation), the isozymes are merged into one hypothetical protein with a molar mass equal to the mean of the molar masses of the merged isozymes. Molar masses of enzymes were calculated as the sum of molar masses of the amino acids constituting the respective primary sequences reduced by the mass of one water molecule per peptide bond. The amino acid sequences were retrieved from the KEGG database (96) by queries with the GEM-inherent, KEGG-specific gene identifiers. The k_{cat} values for all enzymatic reactions in the model were derived as described in the previous section. The final set of curated k_{cat} values and molar masses of enzymes used throughout this study can be found in Data Set S1, sheet 1.

The translational and unused enzyme sectors ϕ_T and ϕ_{UE} are linearly related to the growth rate μ and the substrate uptake rate v_s , respectively. The linear equations of both sectors (Equation 8) are parameterized with condition-dependent proteomic data (41, 44) as discussed in the respective Materials and Methods sections and, except for the maximum substrate uptake rate v_s , parameters are maintained for any simulation in this work. Finally, the total mass concentration of condition-dependent protein $\phi_{P,c}$ is kept constant according to Equation 1 and comprises the sum of the translational sector ϕ_T , unused enzyme sector ϕ_{UE} , and active enzyme sector ρ_e . An overview of the applied parameters is given in Table 2.

Determination of maximum substrate uptake rates. By applying the PAM with parameterized protein sectors and minimal medium constraints, maximum uptake rates for single substrates were

TABLE 2 Source and values of universal PAM parameters applied for all simulations in this study

Parameter	Symbol	Value	Source
Total protein concentration	$\phi_{p,c}$	258.0 mg g _{cdw} ⁻¹	Derived from the sum of experimentally determined protein masses under different environmental conditions (41)
Intercept translational protein sector	$\phi_{t,0}$	49.9 mg g _{cdw} ⁻¹	Derived from fitting and extrapolating proteomic data of translational protein (41)
Slope translational protein sector		36.8 mg h g _{cdw} ⁻¹	Derived from fitting proteomic data of translational protein (41)
Intercept unused enzymes sector	$\phi_{ue,0}$	171.1 mg g _{cdw} ⁻¹	Derived from fitting and extrapolating proteomic data of all proteins included in the iML1515 metabolic model (41, 44)

determined assuming a maximally allowable total protein synthesis rate $N_{p,max}$. N_p represents the sum of molar synthesis rates of proteins from the active enzymes, unused enzymes, and the translational sector. At the maximum substrate uptake rate, and therefore at maximum growth, the unused enzyme sector is zero. Thus, the maximum total protein synthesis rate $N_{p,max}$ is

$$N_{p,max} = \mu \left(\frac{\phi_T}{M_R} + \sum_e^E \nu_e / k_{cat,e} \right) \quad (9)$$

Here, M_R is the sum of the molar masses of all 21 ribosomal subunits of 3.5×10^5 g mol⁻¹, and ν_e is the fluxes of each enzymatically catalyzed reaction e .

We hypothesized that there is a unique maximum total protein synthesis rate $N_{p,max}$ which determines maximum uptake rates for any given substrate. Knowledge of $N_{p,max}$ would allow the determination of maximum substrate uptake rates and the parameterization of the unused enzyme sector (cf. Equations 3 and 4) solely based on model simulations. A correspondingly parameterized PAM should therefore be able to correctly predict phenotypes for substrate-limited conditions.

To test this hypothesis, we simulated $N_{p,max}$ values for substrates considered by Gerosa et al. (61) using PAMs parameterized with a wide range of maximum substrate uptake rates. More specifically, the unused enzyme sector was parameterized at each of the tested maximum substrate uptake rates according to Equation 4. Growth rates were simulated by optimizing the parameterized PAMs for growth and constraining the substrate uptake rate to the experimentally determined value. These simulated growth rates were then compared with the observed values by computing the absolute difference. At an $N_{p,max}$ of $2.04 \mu\text{mol g}_{cdw}^{-1} \text{h}^{-1}$, the sum of absolute differences between simulated and measured growth rates was minimal. For each of the considered substrates, maximum uptake rates (parameter of the PAM) leading to $N_{p,max}$ by applying the experimentally observed substrate uptake rate (constraint) were extracted and are shown in Table S1.

eGFP overexpression. Expression of eGFP was simulated by introducing an additional column to the stoichiometric matrix of the PAM representing a protein with a mass of 2.8×10^4 g mol⁻¹. Thus, the expression strength of eGFP (g h⁻¹ g_{cdw}⁻¹) is controlled by the respective model variable describing the protein's intracellular concentration and the growth rate. To identify the total protein concentration that represents the elevated protein availability in the *E. coli* TUNER strain used by Bienick et al. (64), the correlation between eGFP concentrations and growth rates relative to the wild-type was computed with the PAM for a range of total protein concentration values $\phi_{p,c}$. For each $\phi_{p,c}$, the summed difference between simulated and measured relative growth rates was calculated as a measure for the agreement between simulation and experiment. Student's t-test was applied to determine the $\phi_{p,c}$ yielding the best possible correlation between measurements and simulations.

Phenotype determination of gene deletion mutants. The deletion of a gene was simulated by identifying all reactions connected to this gene via the model-inherent GPR rules and setting the respective upper bounds to zero. To predict maximum substrate uptake rates of GMSs, enzyme synthesis rates were calculated from growth-optimal flux distributions for a wide range of substrate uptake rates. The synthesis rate N_e of an enzyme e was computed from flux distributions by

$$N_e = \rho_e \cdot \mu \quad (10)$$

where μ and ρ_e are the growth rate and the molar concentration of enzyme e , both being optimization variables to the PAM.

For each tested substrate uptake rate, the maximum difference in the computed enzyme synthesis rate ΔN_e^{max} between the GMS and a reference state representing a wild-type strain grown under substrate-unlimited conditions was identified among all modeled enzymes. By scanning ΔN_e^{max} values from high to low substrate uptake rates, the maximum substrate uptake rate for a GMS is found as soon as ΔN_e^{max} meets a critical value ΔN_e^{crit} (also termed maximum overexpression capacity). Thus, the corresponding metabolic mode supports a maximally achievable growth rate under a limited flexibility to change or reallocate protein among metabolic pathways and their single reactions. We assumed the level of restriction in flexibility ΔN_e^{crit} to be the same in any GMS according to phenotypic data of GMS from Long et al. (70), Long and Antoniewicz (71), and Fong et al. (72). We found that a ΔN_e^{crit} of $16.0 \text{ nmol g}_{cdw}^{-1} \text{h}^{-1}$ leads to a minimal sum of errors between predicted and observed growth, substrate

uptake, and acetate secretion rates. The corresponding comparisons of experimentally determined and predicted phenotypes as well as flux distributions are shown in Fig. 8 and 9, respectively.

Implementation. All conducted simulations, model reconstructions, and data analyses were performed in MATLAB 2018a on a Windows 7 machine with 16 GB of RAM and an AMD FX-8350 eight-core (at 4.00 GHz) processor. COBRA toolbox functions (84) and the Gurobi Optimizer (8.0.0; Gurobi Optimization, Inc.) were utilized to process and solve the metabolic models. All MATLAB functions necessary to handle and build a protein allocation model (PAM) from a COBRA format-based, stoichiometric reconstruction are provided on GitHub (https://github.com/Spherotob/PAM_public).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, XLSX file, 0.3 MB.

FIG S1, EPS file, 0.3 MB.

FIG S2, EPS file, 0.3 MB.

FIG S3, EPS file, 0.1 MB.

FIG S4, EPS file, 0.2 MB.

FIG S5, EPS file, 0.2 MB.

FIG S6, EPS file, 0.1 MB.

FIG S7, EPS file, 0.4 MB.

FIG S8, EPS file, 0.5 MB.

TABLE S1, DOCX file, 0.02 MB.

ACKNOWLEDGMENTS

T.B.A. was partially funded by the Excellence Initiative of the German federal and state governments (grant ID PFLS015). B.E.E. acknowledges funding through the CSIRO-UQ Synthetic Biology Alliance. The laboratory of L.M.B. received funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy within the Clusters of Excellence "236 TMFB" and "2186 The Fuel Science Center." The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

REFERENCES

- Monod J. 1949. The growth of bacterial cultures. *Annu Rev Microbiol* 3:371–394. <https://doi.org/10.1146/annurev.mi.03.100149.002103>.
- Klump S, Zhang Z, Hwa T. 2009. Growth rate-dependent global effects on gene expression in bacteria. *Cell* 139:1366–1375. <https://doi.org/10.1016/j.cell.2009.12.001>.
- Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. 2010. Interdependence of cell growth and gene expression: origins and consequences. *Science* 330:1099–1102. <https://doi.org/10.1126/science.1192588>.
- You C, Okano H, Hui S, Zhang Z, Kim M, Gunderson CW, Wang YP, Lenz P, Yan D, Hwa T. 2013. Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* 500:301–306. <https://doi.org/10.1038/nature12446>.
- Scott M, Klump S, Mateescu EM, Hwa T. 2014. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol Syst Biol* 10:747. <https://doi.org/10.15252/msb.20145379>.
- Weiß AY, Oyarzún DA, Danos V, Swain PS. 2015. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc Natl Acad Sci U S A* 112:E1038–E1047. <https://doi.org/10.1073/pnas.1416533112>.
- Erickson DW, Schink SJ, Patsalo V, Williamson JR, Gerland U, Hwa T. 2017. A global resource allocation strategy governs growth transition kinetics of *Escherichia coli*. *Nature* 551:119–123. <https://doi.org/10.1038/nature24299>.
- Savinell JM, Palsson BO. 1992. Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *J Theor Biol* 154:421–454. [https://doi.org/10.1016/s0022-5193\(05\)80161-4](https://doi.org/10.1016/s0022-5193(05)80161-4).
- Varma A, Boesch BW, Palsson BO. 1993. Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl Environ Microbiol* 59:2465–2473. <https://doi.org/10.1128/AEM.59.8.2465-2473.1993>.
- Varma A, Palsson BO. 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol* 60:3724–3731. <https://doi.org/10.1128/AEM.60.10.3724-3731.1994>.
- Edwards JS, Palsson BO. 2000. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol Prog* 16:927–939. <https://doi.org/10.1021/bp0000712>.
- Schuetz R, Kuepfer L, Sauer U. 2007. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Syst Biol* 3:119. <https://doi.org/10.1038/msb4100162>.
- Pramanik J, Keasling JD. 1998. Effect of *Escherichia coli* biomass composition on central metabolic fluxes predicted by a stoichiometric model. *Biotechnol Bioeng* 60:230–238. [https://doi.org/10.1002/\(SICI\)1097-0290\(19981020\)60:2<230::AID-BIT10>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1097-0290(19981020)60:2<230::AID-BIT10>3.0.CO;2-Q).
- Mahadevan R, Edwards JS, Doyle FJ, III. 2002. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J* 83:1331–1340. [https://doi.org/10.1016/S0006-3495\(02\)73903-9](https://doi.org/10.1016/S0006-3495(02)73903-9).
- Palsson B. 2002. *In silico* biology through omics. *Nat Biotechnol* 20:649–650. <https://doi.org/10.1038/nbt0702-649>.
- Blank LM, Kuepfer L, Sauer U. 2005. Large-scale 13 C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol* 6:R49. <https://doi.org/10.1186/gb-2005-6-6-r49>.
- Becker SA, Palsson BO. 2008. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol* 4:e1000082. <https://doi.org/10.1371/journal.pcbi.1000082>.
- Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D, Weitz KK, Eils R, König R, Smith RD, Palsson BØ. 2010. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* 6:390. <https://doi.org/10.1038/msb.2010.47>.
- Burgard AP, Maranas CD. 2003. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol Bioeng* 82:670–677. <https://doi.org/10.1002/bit.10617>.
- Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP, Nielsen J, Patil KR, Ferreira EC, Rocha M. 2010. OptFlux: an open-source software

- platform for *in silico* metabolic engineering. *BMC Syst Biol* 4:45. <https://doi.org/10.1186/1752-0509-4-45>.
21. Cardoso JGR, Jensen K, Lieven C, Lærke Hansen AS, Galkina S, Beber M, Özdemir E, Herrgård MJ, Redestig H, Sonnenschein N. 2018. Cameo: a Python library for computer aided metabolic engineering and optimization of cell factories. *ACS Synth Biol* 7:1163–1166. <https://doi.org/10.1021/acssynbio.7b00423>.
 22. Alter TB, Blank LM, Ebert BE. 2018. Genetic optimization algorithm for metabolic engineering revisited. *Metabolites* 8:33. <https://doi.org/10.3390/metabo8020033>.
 23. Alter TB, Ebert BE. 2019. Determination of growth-coupling strategies and their underlying principles. *BMC Bioinformatics* 20:447. <https://doi.org/10.1186/s12859-019-2946-7>.
 24. Covert MW, Famili I, Palsson BO. 2003. Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology? *Biotechnol Bioeng* 84:763–772. <https://doi.org/10.1002/bit.10849>.
 25. Lerman JA, Hyduke DR, Latif H, Portnoy VA, Lewis NE, Orth JD, Schrimper-Rutledge AC, Smith RD, Adkins JN, Zengler K, Palsson BØ. 2012. *In silico* method for modelling metabolism and gene product expression at genome scale. *Nat Commun* 3:929. <https://doi.org/10.1038/ncomms1928>.
 26. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BØ. 2013. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol* 9:693. <https://doi.org/10.1038/msb.2013.52>.
 27. Yang L, Yurkovich JT, Lloyd CJ, Ebrahim A, Saunders MA, Palsson BØ. 2016. Principles of proteome allocation are revealed using proteomic data and genome-scale models. *Sci Rep* 6:36734. <https://doi.org/10.1038/srep36734>.
 28. Yang L, Mih N, Anand A, Park JH, Tan J, Yurkovich JT, Monk JM, Lloyd CJ, Sandberg TE, Seo SW, Kim D, Sastry AV, Phaneuf P, Gao Y, Broddrick JT, Chen K, Heckmann D, Szubin R, Hefner Y, Feist AM, Palsson BO. 2019. Cellular responses to reactive oxygen species are predicted from molecular mechanisms. *Proc Natl Acad Sci U S A* 116:14368–14373. <https://doi.org/10.1073/pnas.1905039116>.
 29. Chen Y, Nielsen J. 2019. Energy metabolism controls phenotypes by protein efficiency and allocation. *Proc Natl Acad Sci U S A* 116:17592–17597. <https://doi.org/10.1073/pnas.1906569116>.
 30. Niebel B, Leupold S, Heinemann M. 2019. An upper limit on Gibbs energy dissipation governs cellular metabolism. *Nat Metab* 1:125–132. <https://doi.org/10.1038/s42255-018-0006-7>.
 31. Beg QK, Vazquez A, Ernst J, de Menezes MA, Bar-Joseph Z, Barabasi AL, Oltvai ZN. 2007. Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proc Natl Acad Sci U S A* 104:12663–12668. <https://doi.org/10.1073/pnas.0609845104>.
 32. Vazquez A, Beg QK, Demenezes MA, Ernst J, Bar-Joseph Z, Barabási AL, Boros LG, Oltvai ZN. 2008. Impact of the solvent capacity constraint on *E. coli* metabolism. *BMC Syst Biol* 2:7. <https://doi.org/10.1186/1752-0509-2-7>.
 33. Szenk M, Dill KA, de Graff AMR. 2017. Why do fast-growing bacteria enter overflow metabolism? Testing the membrane real estate hypothesis. *Cell Syst* 5:95–104. <https://doi.org/10.1016/j.cels.2017.06.005>.
 34. Goelzer A, Fromion V. 2011. Bacterial growth rate reflects a bottleneck in resource allocation. *Biochim Biophys Acta* 1810:978–988. <https://doi.org/10.1016/j.bbagen.2011.05.014>.
 35. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B, Assad-Garcia N, Glass JI, Covert MW. 2012. A whole-cell computational model predicts phenotype from genotype. *Cell* 150:389–401. <https://doi.org/10.1016/j.cell.2012.05.044>.
 36. Molenaar D, van Berlo R, de Ridder D, Teusink B. 2009. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol* 5:323. <https://doi.org/10.1038/msb.2009.82>.
 37. Mori M, Hwa T, Martin OC, De Martino A, Marinari E. 2016. Constrained allocation flux balance analysis. *PLoS Comput Biol* 12:e1004913. <https://doi.org/10.1371/journal.pcbi.1004913>.
 38. Noor E, Flamholz A, Bar-Even A, Davidi D, Milo R, Liebermeister W. 2016. The protein cost of metabolic fluxes: prediction from enzymatic rate laws and cost minimization. *PLoS Comput Biol* 12:1–29. <https://doi.org/10.1371/journal.pcbi.1005167>.
 39. Mori M, Marinari E, De Martino A. 2019. A yield-cost tradeoff governs *Escherichia coli*'s decision between fermentation and respiration in carbon-limited growth. *NPJ Syst Biol Appl* 5:16. <https://doi.org/10.1038/s41540-019-0093-4>.
 40. Sánchez BJ, Zhang C, Nilsson A, Lahtvee P, Kerkhoven EJ, Nielsen J. 2017. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol Syst Biol* 13:935. <https://doi.org/10.15252/msb.20167411>.
 41. Schmidt A, Kochanowski K, Vedelaar S, Ahrné E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R, Heinemann M. 2016. The quantitative and condition-dependent *Escherichia coli* proteome. *Nat Biotechnol* 34:104–110. <https://doi.org/10.1038/nbt.3418>.
 42. Bosdriesz E, Molenaar D, Teusink B, Bruggeman FJ. 2015. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *FEBS J* 282:2029–2044. <https://doi.org/10.1111/febs.13258>.
 43. Mori M, Schink S, Erickson DW, Gerland U, Hwa T. 2017. Quantifying the benefit of a proteome reserve in fluctuating environments. *Nat Commun* 8:1–8. <https://doi.org/10.1038/s41467-017-01242-8>.
 44. O'Brien EJ, Utrilla J, Palsson BO. 2016. Quantification and classification of *E. coli* proteome utilization and unused protein costs across environments. *PLoS Comput Biol* 12:1–22. <https://doi.org/10.1371/journal.pcbi.1004998>.
 45. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, Takeuchi R, Nomura W, Zhang Z, Mori H, Feist AM, Palsson BO. 2017. iML1515, a knowledge-base that computes *Escherichia coli* traits. *Nat Biotechnol* 35:904–908. <https://doi.org/10.1038/nbt.3956>.
 46. Fendt S, Buescher JM, Rudroff F, Picotti P, Zamboni N, Sauer U. 2010. Tradeoff between enzyme and metabolite efficiency maintains metabolic homeostasis upon perturbations in enzyme capacity. *Mol Syst Biol* 6:356. <https://doi.org/10.1038/msb.2010.11>.
 47. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R. 2011. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50:4402–4410. <https://doi.org/10.1021/bi2002289>.
 48. Valgepea K, Adamberg K, Seiman A, Vilu R. 2013. *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. *Mol Biosyst* 9:2344–2358. <https://doi.org/10.1039/c3mb70119k>.
 49. Li SHJ, Li Z, Park JO, King CG, Rabinowitz JD, Wingreen NS, Gitai Z. 2018. *Escherichia coli* translation strategies differ across carbon, nitrogen and phosphorus limitation conditions. *Nat Microbiol* 3:939–947. <https://doi.org/10.1038/s41564-018-0199-2>.
 50. Milo R. 2013. What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays* 35:1050–1055. <https://doi.org/10.1002/bies.201300066>.
 51. Scott M, Hwa T. 2011. Bacterial growth laws and their applications. *Curr Opin Biotechnol* 22:559–565. <https://doi.org/10.1016/j.copbio.2011.04.014>.
 52. Perrenoud A, Sauer U. 2005. Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in *Escherichia coli*. *J Bacteriol* 187:3171–3179. <https://doi.org/10.1128/JB.187.9.3171-3179.2005>.
 53. Nanchen A, Schicker A, Sauer U. 2006. Nonlinear dependency of intracellular fluxes on growth rate in miniaturized continuous cultures of *Escherichia coli*. *Appl Environ Microbiol* 72:1164–1172. <https://doi.org/10.1128/AEM.72.2.1164-1172.2006>.
 54. Vemuri GN, Altman E, Sangurdekar DP, Khodursky AB, Eiteman MA. 2006. Overflow metabolism in *Escherichia coli* during steady-state growth: transcriptional regulation and effect of the redox ratio. *Appl Environ Microbiol* 72:3653–3661. <https://doi.org/10.1128/AEM.72.5.3653-3661.2006>.
 55. Valgepea K, Adamberg K, Nahku R, Lahtvee PJ, Arike L, Vilu R. 2010. Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase. *BMC Syst Biol* 4:166. <https://doi.org/10.1186/1752-0509-4-166>.
 56. Folsom JP, Parker AE, Carlson RP. 2014. Physiological and proteomic analysis of *Escherichia coli* iron-limited chemostat growth. *J Bacteriol* 196:2748–2761. <https://doi.org/10.1128/JB.01606-14>.
 57. McCloskey D, Gangoiti JA, King ZA, Naviaux RK, Barshop BA, Palsson BO, Feist AM. 2014. A model-driven quantitative metabolomics analysis of aerobic and anaerobic metabolism in *E. coli* K-12 MG1655 that is biochemically and thermodynamically consistent. *Biotechnol Bioeng* 111:803–815. <https://doi.org/10.1002/bit.25133>.
 58. Peebo K, Valgepea K, Maser A, Nahku R, Adamberg K, Vilu R. 2015. Proteome reallocation in *Escherichia coli* with increasing specific growth rate. *Mol Biosyst* 11:1184–1193. <https://doi.org/10.1039/c4mb00721b>.
 59. Folsom JP, Carlson RP. 2015. Physiological, biomass elemental composition and proteomic analyses of *Escherichia coli* ammonium-limited chemostat growth, and comparison with iron- and glucose-limited chemostat growth.

- Microbiology (Reading) 161:1659–1670. <https://doi.org/10.1099/mic.0.000118>.
60. Fischer E, Sauer U. 2003. Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. *Eur J Biochem* 270:880–891. <https://doi.org/10.1046/j.1432-1033.2003.03448.x>.
 61. Gerosa L, Haverkorn van Rijsewijk BRB, Christodoulou D, Kochanowski K, Schmidt TSB, Noor E, Sauer U. 2015. Pseudo-transition analysis identifies the key regulators of dynamic metabolic adaptations from steady-state data. *Cell Syst* 1:270–282. <https://doi.org/10.1016/j.cels.2015.09.008>.
 62. Fong SS, Joyce AR, Palsson BØ. 2005. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res* 15:1365–1372. <https://doi.org/10.1101/gr.3832305>.
 63. Haverkorn van Rijsewijk BRB, Nanchen A, Nallet S, Kleijn RJ, Sauer U. 2011. Large-scale ¹³C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*. *Mol Syst Biol* 7:477. <https://doi.org/10.1038/msb.2011.9>.
 64. Bienick MS, Young KW, Klesmith JR, Detwiler EE, Tomek KJ, Whitehead TA. 2014. The interrelationship between promoter strength, gene expression, and growth rate. *PLoS One* 9:e109105. <https://doi.org/10.1371/journal.pone.0109105>.
 65. Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R. 2013. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc Natl Acad Sci U S A* 110:10039–10044. <https://doi.org/10.1073/pnas.1215283110>.
 66. Ng CY, Wang L, Chowdhury A, Maranas CD. 2019. Pareto optimality explanation of the glycolytic alternatives in nature. *Sci Rep* 9:2633. <https://doi.org/10.1038/s41598-019-38836-9>.
 67. Maia P, Rocha M, Rocha I. 2016. *In silico* constraint-based strain optimization methods: the quest for optimal cell factories. *Microbiol Mol Biol Rev* 80:45–67. <https://doi.org/10.1128/MMBR.00014-15>.
 68. Fong SS, Palsson BØ. 2004. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat Genet* 36:1056–1058. <https://doi.org/10.1038/ng1432>.
 69. Kim J, Reed JL. 2012. RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. *Genome Biol* 13:R78. <https://doi.org/10.1186/gb-2012-13-9-r78>.
 70. Long CP, Gonzalez JE, Sandoval NR, Antoniewicz MR. 2016. Characterization of physiological responses to 22 gene knockouts in *Escherichia coli* central carbon metabolism. *Metab Eng* 37:102–113. <https://doi.org/10.1016/j.mben.2016.05.006>.
 71. Long CP, Antoniewicz MR. 2019. Metabolic flux responses to deletion of 20 core enzymes reveal flexibility and limits of *E. coli* metabolism. *Metab Eng* 55:249–257. <https://doi.org/10.1016/j.mben.2019.08.003>.
 72. Fong SS, Nanchen A, Palsson BO, Sauer U. 2006. Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. *J Biol Chem* 281:8024–8033. <https://doi.org/10.1074/jbc.M510016200>.
 73. Paul BJ, Ross W, Gaal T, Gourse RL. 2004. rRNA transcription in *Escherichia coli*. *Annu Rev Genet* 38:749–770. <https://doi.org/10.1146/annurev.genet.38.072902.091347>.
 74. McCloskey D, Xu S, Sandberg TE, Brunk E, Hefner Y, Szubin R, Feist AM, Palsson BO. 2018. Evolution of gene knockout strains of *E. coli* reveal regulatory architectures governed by metabolism. *Nat Commun* 9:3796. <https://doi.org/10.1038/s41467-018-06219-9>.
 75. Segrè D, Vitkup D, Church GM. 2002. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A* 99:15112–15117. <https://doi.org/10.1073/pnas.232349399>.
 76. Shlomi T, Berkman O, Ruppin E. 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci U S A* 102:7695–7700. <https://doi.org/10.1073/pnas.0406346102>.
 77. Nakahigashi K, Toya Y, Ishii N, Soga T, Hasegawa M, Watanabe H, Takai Y, Honma M, Mori H, Tomita M. 2009. Systematic phenome analysis of *Escherichia coli* multiple-knockout mutants reveals hidden reactions in central carbon metabolism. *Mol Syst Biol* 5:306–314. <https://doi.org/10.1038/msb.2009.65>.
 78. McCloskey D, Xu S, Sandberg TE, Brunk E, Hefner Y, Szubin R, Feist AM, Palsson BO. 2018. Adaptation to the coupling of glycolysis to toxic methylglyoxal production in *tpiA* deletion strains of *Escherichia coli* requires synchronized and counterintuitive genetic changes. *Metab Eng* 48:82–93. <https://doi.org/10.1016/j.mben.2018.05.012>.
 79. LaCroix RA, Sandberg TE, O'Brien EJ, Utrilla J, Ebrahim A, Guzman GI, Szubin R, Palsson BO, Feist AM. 2015. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of *Escherichia coli* K-12 MG1655 on glucose minimal medium. *Appl Environ Microbiol* 81:17–30. <https://doi.org/10.1128/AEM.02246-14>.
 80. Wittmann C, Lee SY. 2012. Systems metabolic engineering. Springer, Dordrecht The Netherlands. <https://doi.org/10.1007/978-94-007-4534-6>.
 81. Saa PA, Cortés MP, López J, Bustos D, Maass A, Agosin E. 2019. Expanding metabolic capabilities using novel pathway designs: computational tools and case studies. *Biotechnol J* 14:1800734. <https://doi.org/10.1002/biot.201800734>.
 82. Choi KR, Jang WD, Yang D, Cho JS, Park D, Lee SY. 2019. Systems metabolic engineering strategies: integrating systems and synthetic biology with metabolic engineering. *Trends Biotechnol* 37:817–837. <https://doi.org/10.1016/j.tibtech.2019.01.003>.
 83. Sarkar D, Maranas CD. 2019. Engineering microbial chemical factories using metabolic models. *BMC Chem Eng* 1:22. <https://doi.org/10.1186/s42480-019-0021-9>.
 84. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdóttir HS, Wachowiak J, Keating SM, Vlasov V, Magnúsdóttir S, Ng CY, Preciat G, Žagare A, Chan SHJ, Aurich MK, Clancy CM, Modamio J, Sauls JT, Noronha A, Bordbar A, Cousins B, El Assal DC, Valcarcel LV, Apaolaza I, Ghaderi S, Ahookhosh M, Ben Guebila M, Kostromins A, Sompairac N, Le HM, Ma D, Sun Y, Wang L, Yurkovich JT, Oliveira MAP, Vuong PT, El Assal LP, Kuperstein I, Zinoviyev A, Hinton HS, Bryant WA, Aragón Artacho FJ, Planes FJ, Stalidzins E, Maass A, Vempala S, Hucka M, Saunders MA, Maranas CD, et al. 2019. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc* 14:639–702. <https://doi.org/10.1038/s41596-018-0098-2>.
 85. Nilsson A, Nielsen J, Palsson BO. 2017. Metabolic models of protein allocation call for the kinetome. *Cell Syst* 5:538–541. <https://doi.org/10.1016/j.cels.2017.11.013>.
 86. Metz-Raz E, Kafri M, Yaakov G, Soifer I, Gurvich Y, Barkai N. 2017. Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *Elife* 6:1–21. <https://doi.org/10.7554/eLife.28034>.
 87. Sander T, Farke N, Diehl C, Kuntz M, Glatter T, Link H. 2019. Allosteric feedback inhibition enables robust amino acid biosynthesis in *E. coli* by enforcing enzyme overabundance. *Cell Syst* 8:66–75.e8. <https://doi.org/10.1016/j.cels.2018.12.005>.
 88. Kolb A, Busby S, Buc H, Garges S, Adhya S. 1993. Transcriptional regulation by cAMP and its receptor protein. *Annu Rev Biochem* 62:749–795. <https://doi.org/10.1146/annurev.bi.62.070193.003533>.
 89. Busby S, Ebricht RH. 1999. Transcription activation by catabolite activator protein (CAP). *J Mol Biol* 293:199–213. <https://doi.org/10.1006/jmbi.1999.3161>.
 90. Jeske L, Placzek S, Schomburg I, Chang A, Schomburg D. 2019. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res* 47:D542–D549. <https://doi.org/10.1093/nar/gky1048>.
 91. Wittig U, Kania R, Golebiewski M, Rey M, Shi L, Jong L, Alga E, Weidemann A, Sauer-Danzwith H, Mir S, Krebs O, Bittkowski M, Wetsch E, Rojas I, Müller W. 2012. SABIO-RK: database for biochemical reaction kinetics. *Nucleic Acids Res* 40:D790–D796. <https://doi.org/10.1093/nar/gkr1046>.
 92. Bateman A. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47:D506–D515. <https://doi.org/10.1093/nar/gky1049>.
 93. Davidi D, Noor E, Liebermeister W, Bar-Even A, Flamholz A, Tumbler K, Barenholz U, Goldenfeld M, Shlomi T, Milo R. 2016. Global characterization of *in vivo* enzyme catalytic rates and their correspondence to *in vitro* k_{cat} measurements. *Proc Natl Acad Sci U S A* 113:3401–3406. <https://doi.org/10.1073/pnas.1514240113>.
 94. Heckmann D, Lloyd CJ, Mih N, Ha Y, Zielinski DC, Haiman ZB, Desouki AA, Lercher MJ, Palsson BO. 2018. Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nat Commun* 9:5252. <https://doi.org/10.1038/s41467-018-07652-6>.
 95. Volkmer B, Heinemann M. 2011. Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. *PLoS One* 6:e23126-6. <https://doi.org/10.1371/journal.pone.0023126>.
 96. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 45:D353–D361. <https://doi.org/10.1093/nar/gkw1092>.