Protein cost allocation explains metabolic strategies in *Escherichia coli*

Pranas Grigaitis[1,2], Brett G. Olivier[1,2], Tomas Fiedler[3], Bas Teusink[2], Ursula Kummer[1], and Nadine Veith[1,✉]

[1]Modeling of Biological Processes, BioQuant/Center for Organismal Studies Heidelberg, Heidelberg University, Im Neuenheimer Feld 267, D-69120 Heidelberg, Germany
[2]Systems Biology Lab, Amsterdam Institute of Molecular and Life Sciences, VU Amsterdam, De Boelelaan 1085, NL-1081HZ Amsterdam, the Netherlands
[3]Institute of Medical Microbiology, Virology, and Hygiene, Rostock University Medical Center, Schillingallee 70, D-18055 Rostock, Germany

In-depth understanding of microbial growth is crucial for the development of new advances in biotechnology and for combating microbial pathogens. Condition-specific proteome expression is central to microbial physiology and growth. A multitude of processes are dependent on the protein expression, thus, whole-cell analysis of microbial metabolism using genome-scale metabolic models is an attractive toolset to investigate the behaviour of microorganisms and their communities. However, genome-scale models that incorporate macromolecular expression are still inhibitory complex: the conceptual and computational complexity of these models severely limits their potential applications. In the need for alternatives, here we revisit some of the previous attempts to create genome-scale models of metabolism and macromolecular expression to develop a novel framework for integrating protein abundance and turnover costs to conventional genome-scale models. We show that such a model of *Escherichia coli* successfully reproduces experimentally determined adaptations of metabolism in a growth condition-dependent manner. Moreover, the model can be used as means of investigating underutilization of the protein machinery among different growth settings. Notably, we obtained strongly improved predictions of flux distributions, considering the costs of protein translation explicitly. This finding in turn suggests protein translation being the main regulation hub for cellular growth.

Genome-scale models | Resource allocation | Quantitative proteomics | Microbial metabolism
Correspondence: ***nadine.veith@bioquant.uni-heidelberg.de***

1. Introduction

Constraint-based modelling techniques are nowadays the most frequently used methods for studying metabolism at genome-scale. These stoichiometric models include information of all metabolic reactions for which enzymes are encoded in the genome of a particular organism. Such whole-cell models are used to compute and investigate intracellular flux distributions. Most commonly, this is done by using linear programming-based optimization methods like Flux Balance Analysis (FBA) and Flux Variability Analysis (FVA) (1). FBA allows calculating flux distributions within the stoichiometric model by optimizing (maximizing or minimizing) a pre-defined linear objective function. Common examples of these objective functions are maximization of the growth rate of an organism or a production rate of a certain metabolite of interest. In many cases, the flux distributions calculated by FBA are not unique and alternative solutions may exist. In these cases, FVA can be used to identify intervals of feasible flux values that each flux can obtain by still resulting in the same objective function value.

Both FBA and FVA rely on the definition of different sets of constraints that limit the resulting flux distributions to include only physiologically relevant ones. The definition of suitable constraints is however not trivial and the resulting flux distributions mostly also include physiologically less relevant solutions, and limits the predictability of such models (2).

Limited predictability mainly arises from three major aspects: First, in conventional genome-scale metabolic models, all metabolic reactions are active concurrently, which implies a simultaneous expression of the whole proteome. This is a rather unrealistic scenario as organisms are subject to different economic and physical constraints. These limitations confine the gene expression towards an environment- and time-dependent behaviour (3). Examples of limitations include limited availability of nutrients or energy as well as physical limits of cell size and volume. Second, the underlying universal constraints that govern microbial growth are still largely unknown, despite ample studies on microbial growth and various physiological parameters (4). Finally, kinetic effects in the metabolic network, e.g. inhibition of certain enzymes, cannot be reproduced due to the absence of kinetic information in genome-scale models.

The predictability of genome-scale models could be considerably improved by explicitly accounting for resource allocation strategies. Resource allocation describes the partitioning of limited resources between different cellular processes (5). Microbial growth, for instance, is often observed to be directly related to the amount of ribosomes (6). The production of ribosomes provides the machinery required for the metabolic reactions. However, at the same time, ribosome biosynthesis is a major competitor for resources with other cellular processes (7). The metabolic network, in turn, is responsible for energy production and adaptation to frequently changing environmental conditions. Thus, the cellular growth and adaptation to changing environments is highly impacted by optimal resource allocation (8).

In recent years, proteome-related information was employed in genome-scale models to further narrow physiologically-relevant solution space. For instance, integration of -omics data could be used to compute flux bounds for metabolic reaction (9). Constraints, representing physical limitations of the cell, such as maximal cell volume (FBA with Molecular Crowding, FBAwMC,

(10)) or plasma membrane surface area (11) available to the proteins, were shown to improve flux distributions. Following these, several approaches rely on extensive descriptions of gene expression and protein biosynthesis (Genome-scale Models of Metabolism and Macromolecular Expression, ME-models, (12)), resource allocation (Resource Balance Analysis, RBA, (13)) and demand of macromolecular machinery to sustain metabolic fluxes via integrating kinetic data (GEM with Enzymatic Constraints using Kinetic and Omics data, GECKO, (14)). Following any of these approaches, the capacity of the metabolic network becomes defined by the abundance of the macromolecular machinery that is required to operate it.

However, the applicability of these modelling approaches (ME-models, RBA and GECKO) is still rather cumbersome. First, the fine-grained description of every step in the gene expression processes (e.g. transcription of mRNA in ME-models) brings little additional information (especially under balanced growth, (15)) at a cost of considerable increase in model complexity. Further, these approaches introduce growth rate-dependent variables, meaning that the optimization problem has to be reformulated depending on the growth rate μ for every iteration of the solving process. To obtain the solutions in a sane amount of wall time, usage of high-performance computing resources becomes inevitable. Finally, contrary to ME-models and RBA, the GECKO approach provides a framework for a straightforward integration of quantitative proteomics data, yet lacks description of protein turnover costs.

The overall complexity of the models, high computational demands and complicated interpretation of the results calls for a new framework to overcome some of the limitations of the conventional genome-scale models without an enormous increase in model complexity. Integrating the reasoning of several previously introduced methods, in this study, we present an improved strategy for augmenting genome-scale models to include proteome-related constraints. Moreover, we explicitly account for resource allocation mechanisms by introducing the descriptions of protein biosynthesis and turnover costs.

Here, we analyse scenarios in which our model outperforms conventional genome-scale metabolic models (M-models). The model is also used to identify and analyse underutilized circuits of the metabolic network by integrating quantitative proteomics measurements. We show that the model successfully predicts condition-dependent flux distributions, and also captures experimentally described metabolic traits. Thus, we expect the model to become a handy tool for predicting context-dependent behaviour of microorganisms in a computationally inexpensive and easily interpretable manner.

2. Materials and Methods

2.1. Model generation. The here presented extended genome-scale metabolic model is based on the previously published genome-scale model of *E. coli* iML1515 (16) that was used as template. The descriptions of exchange reactions were altered in order to be able to account for different nutrient availability in different growth media. Previously irreversible exchange reactions were assumed to be reversible, although preserving the originally defined values of the flux bounds.

Metabolic reactions were assigned to metabolic clusters ("subsystems") as described below according to the annotations in the original model. These annotations were further used in the analyses of either individual fluxes through the metabolic reactions, or for comparison of sums of fluxes through clusters in a condition-dependent manner.

In order to describe the protein turnover in the model, reactions for protein biosynthesis, folding and degradation were introduced into the model. Information on the full reference proteome sequences and protein properties (e.g. molecular weight, Gene Ontology terms etc.) were acquired from UniProt (17). Information on the classification (EC numbers) of enzymes in the reference proteome was collected from ENZYME (18).

2.2. Experimental datasets. Different datasets were used to constrain the extended genome-scale model. These datasets included metabolite composition of the growth media as well as condition-dependent absolute protein abundances in *E. coli* (5, 19) and were used for the incorporation of quantitative proteomic data. In order to mimic the growth conditions of the different media that have been used in the experimental studies, environmental constraints are introduced into the model.

The information about the metabolite composition of the chemically defined media were used to calculate maximum uptake and production fluxes in the unit of $mmol(L)^{-1}(h)^{-1}$ (corresponding to units of fluxes through exchange reactions in the model). These fluxes were used to constrain the respective exchange reactions by setting lower flux bounds (maximum uptake flux to the model) to the values calculated, and the upper flux bound (maximum production flux) was set to 1000 $mmol(L)^{-1}(h)^{-1}$.

The protein abundances were used for the incorporation of quantitative proteomics data. Therefore, absolute protein abundances in the unit $mmol(gDW)^{-1}$ of cells were recomputed from protein molecule counts ($(cell)^{-1}$) (see "Proteomics data used for constraining the models" in Supplementary Note 1 for further information).

2.3. Kinetic data. The incorporation of protein biosynthesis and turnover processes into the model allowed the introduction of additional constraints to define flux capacities for these processes. Information about flux capacities were derived from kinetic data, since the maximum achievable flux through one reaction corresponds to the $V_{max,\,i}$ of the respective enzyme E_i, defined as $k_{cat,\,i} \times [E_i]$. Therefore, maximum k_{cat} values (corresponding to wild-type enzymes) were acquired from SABIO-RK (20) and BRENDA (21).

For some enzymes, no experimentally determined k_{cat} value could be retrieved. In these cases, data from very homologous proteins were used. The homology search was performed using Protein BLAST+ (22) against the SwissProt protein database.

We considered proteins, irregardless of their taxon, to be highly homologous given $\geq$70% amino acid sequence identity and $\geq$70% of sequence coverage with the query (23). In some cases, still no homologous proteins with a measured k_{cat} value were found. Then, we collected the median k_{cat} values of such enzymes from other species.
In very few cases, however, no k_{cat} values could be retrieved. Here, we assumed a default value of 65.9 s^{-1}. The value of 65.9 s^{-1} corresponds to the median k_{cat} computed for yeast enzymes with known turnover values (originally published in (14)). We made use of this value due to a high similarity in distribution of k_{cat} for both prokaryotic and eukaryotic domains of life, as suggested in (24). In some cases, the same enzyme catalyses different metabolic reactions. Then, the highest k_{cat} value was uniformly applied for all reactions.

2.4. Software. Genome-scale models were generated and simulated using PySCeS (25) CBMPy 0.7.21, running under the Python environment. Linear programming was performed using the IBM ILOG CPLEX Optimizer 12.8.0, using its Python API, which is callable by CBMPy.

3. Calculation

In this study, the previously published genome-scale model (GEM) of *E. coli* iML1515 (16) was used as a template. We extended the model by a detailed description of protein biosynthesis, folding and turnover. A detailed description of newly introduced reactions and constraints is provided in the Supplementary Note 1. The scripts to extend conventional GEMs, as well as supplementary materials for analysis of the model results are available on GitHub at pranasag/extendedEcoliGEM.
Building on the GECKO framework (14), we introduced protein species E_i ($i = 1, 2, 3, ..., N$ protein species in proteome) as model metabolite species. Enzyme species enter metabolic reactions as substrates, and a scaling is introduced for defining the amount of flux in which a unit enzyme can participate in. For any enzyme-catalysed reaction, a stoichiometric coefficient equal to $(k_{cat,\,i})^{-1}$ is assigned to each enzyme. The maximal flux capacity of an enzyme E_i thus equals the product of its abundance and its catalytic constant $[E_i] \times k_{cat,\,i}$. In the GECKO framework, enzymes are being "consumed" (i.e. not retrieved) in reactions (Eq. 1). Furthermore, the supply of each protein to the model is provided by the respective source reactions (Eq. 2):

$$A + B + (k_{cat,\,i})^{-1} \times E_i \rightarrow C + D \quad \textbf{(1)}$$

$$\emptyset \rightarrow E_i \quad \textbf{(2)}$$

The (Eq. 1) in GECKO formalism implies that the protein is "consumed" by the metabolic reaction, together with (Eq. 2) suggesting cost-less protein production and turnover in this formalism. Contrary to GECKO notation of metabolic reactions (Eq. 1), we considered the enzyme species E_i (with explicitly defined turnover costs) to enter reactions as substrates and to be retrieved as "used" enzyme species E_i^* (see Supplementary Note 1 for detailed explanation):

$$A + B + (k_{cat,\,i})^{-1} \times E_i \rightarrow C + D + (k_{cat,\,i})^{-1} \times E_i^* \quad \textbf{(3)}$$

Retrieving protein species E_i^* after metabolic reactions permits to describe their degradation process, which is an integral part of the whole protein turnover. Moreover, instead of providing proteins from a generic source as in (Eq. 2), we provided a detailed description of protein biosynthesis, chaperone-mediated folding for each E_i and protease-mediated degradation for each E_i^*. A similar scheme as for (Eq. 3) applies to the processes, catalysed by macromolecular complexes, e.g. protein synthesis by ribosome:

$$[substrates] + n\ Ribo \rightarrow [products] + n\ Ribo^* \quad \textbf{(4)}$$

All the reactions, required to maintain the protein machinery, are catalyzed by macromolecular complexes, thus establishing a self-replicator-like model (7). For the biosynthesis, we considered costs both in terms of carbon and nitrogen (amino acids), as well as energy equivalents (ATP). The degradation process retrieves free amino acids from used protein species, which can be redirected either to synthesis of other protein species or growth (biomass) of the organism.
Contrary to the ME-models and RBA, these processes are growth rate-independent. Here, we assume a growth rate-independent protein elongation rate of a ribosome, as well as average requirements for chaperone machinery and proteases. This way, we preserve the linearity of the optimization problem.
Thereby, with these two major considerations, we satisfy the requirement to explicitly account for cellular protein abundance and turnover costs. Thus, the model structure now allows to study resource allocation between metabolism and protein biosynthesis.
Finally, an additional constraint was defined that accounts for the total protein abundance that a cell can accommodate. This new constraint is defined by the sum of fluxes through the protein synthesis reactions, adopted from (26):

$$b \times \sum_i M_{w,\,i} \times v_{synthesis,\,E_i} \leq \frac{f_P}{C} \quad \textbf{(5)}$$

Thus, this constraint establishes a relationship between the flux through the protein synthesis reaction of a protein E_i with its molar volume, considering $b = 0.73\ mL(g)^{-1}$ as the empirical relation between the molar volume of a protein and its molecular mass $M_{w,\,i}$ (in $g/mmol$) (26). On the right-hand side, the maximal protein volume in one gram of cell dry weight is computed from the fraction of protein per gDW $f_P = 0.62\ g(gDW)^{-1}$ (27) and the estimated total protein density in one gram wet weight of *E. coli* $C = 0.34\ g(mL)^{-1}$ (26).

4. Results

4.1. Description of macromolecular machinery improves flux distributions. First, we studied how the new model structure affects the flux distributions and compared the results to the conventional GEM of *E. coli* iML1515. Here, no additional constraints were applied and only effects that arise from the new model structure (i.e. description of protein expression and metabolic coupling) were analysed.

We computed flux distributions for both models for three different growth media, a minimal MOPS growth medium without amino acid supplement (MOPS -AA), medium supplemented with amino acids except methionine (MOPS -Met) and one with the full amino acid supplement (MOPS +AA). All media contain glucose as the main carbon source as reported in the study of Li *et al.* (5).

We used the descriptions of these chemically-defined media to define constraints that are imposed by different media and adjusted the flux bounds of the metabolite exchange reactions (as described in Materials and Methods). Then, we used FBA and FVA to compute flux distributions through both models.

The original GEM iML1515 includes annotations of metabolic reactions, assigning them to different metabolic clusters ("subsystems") of *E. coli* (see Materials and Methods), which we further used. To compare growth on media with different amino acid supplementation, we selected the following clusters of the metabolic network: transport through the inner membrane and folate metabolism, Thr and Lys metabolism, as well as Val, Leu and Ile (VLI) metabolism (Fig. 1).

In the original iML1515 model, all three medium variants resulted in large flux intervals (spanning up to some 4000 $mmol(gDW)^{-1}h^{-1}$) for the selected metabolic clusters (Fig. 1A). The extended model (Fig. 1B), in contrast, resulted in considerably decreased sums of fluxes through the selected clusters, for example, with respect to amino acid metabolism.

In the extended model, the computed fluxes for MOPS -Met and MOPS +AA for VLI and Thr/Lys metabolism were considerably lower than for MOPS -AA. Also, these amino acids were predicted to be taken up from the growth medium. The decrease in intracellular metabolism of these amino acids was also demonstrated experimentally. It was shown that the specific activity of some enzymes of VLI or Thr/Lys metabolism was decreased upon growth of *E. coli* in a medium containing Val (28), as well as Thr or Lys (29).

Moreover, contrary to iML1515, our model suggests little flux variability in the folate pathway. In our model, the sums of fluxes through this cluster for MOPS -Met and MOPS +AA were lower than for MOPS -AA. Experimentally, ^{13}C-labelled glycine was shown to be taken up from growth medium and directed to the glycine cleavage system (30) and increase the activity of enzymes in the folate pathway (31). Yet, we observed a higher sum of fluxes for MOPS -Met than MOPS +AA through the folate metabolism. This is consistent with experimental data, suggesting that the activity of single-carbon metabolism enzymes is diminished when grown on methionine-supplemented media (32).

In the case of iML1515, there were no differences in sums of fluxes through the transport of metabolites through the inner membrane. Contrary, for our model, the lowest sum of fluxes for this cluster was computed for MOPS -AA, while MOPS -Met and MOPS +AA had increasingly higher sums of fluxes. It was shown that in amino acid-rich medium these nutrients are taken up by *E. coli* (33, 34). However, experiments of growth medium supplementation with single amino acid, have reported repression of both the transport (35) and the growth of *E. coli* (36). Yet, the latter effect was not captured by neither iML1515 nor our model since such effects are usually attributed to regulatory pathways which are not present in stoichiometric models.

Overall, our model shows a more precise agreement with existing experimental knowledge, compared to iML1515. Thus, we continued to explore the predictions of our model further. The sums of fluxes through central metabolic pathways, such as pentose phosphate pathway (PPP) or pyruvate metabolism varied between MOPS -AA $vs.$ MOPS -Met or MOPS +AA. In the MOPS +AA medium, the sum of fluxes through PPP was lower than in MOPS -AA. This is an indication of cells entering respiro-fermentative energy harvesting in MOPS +AA, rather than a more respirative-like regime upon MOPS -AA (discussed in more detail in Section 4.2).

The sum of fluxes through pyruvate metabolism was the highest for MOPS -AA medium. This might be an indication of less flux through reactions withdrawing carbon from the lower glycolytic pathway. One of these pathways is serine biosynthesis from 3-phosphoglycerate. The enzyme phosphoserine dehydrogenase was shown to be downregulated in growth on medium supplemented with amino acids (Thr, Met, Leu, Ile) (37). A higher maximal sum of fluxes through pyruvate metabolism for MOPS +AA medium $vs.$ MOPS -Met agrees well with experimentally shown inhibition of the aforementioned pathway by methionine (38).

4.2. Proteome capacity constraint determines the acetate switch. A recurrent issue of GEMs is prediction of more complex metabolic phenomena, like overflow metabolism (2). Overflow metabolism describes a phenomenon, when metabolic

fluxes are re-routed towards low-yield metabolic pathways. In *E. coli*, this phenomenon is commonly observed in excess of glucose and high growth rates, when the glucose consumption and energy generation are shifted from respiration towards acetate formation (39).

Conventional GEMs often represent conditions of balanced microbial growth. Experimentally, this is usually ensured by limiting the abundance of a particular carbon source, amino acid or energy. At nutrient-rich conditions, however, a set of additional constraints is required in many cases to correctly predict growth rates and to compute physiologically relevant flux distributions.

For our model, one of the additional constraints is the global protein abundance constraint (see (Eq. 5) and Supplementary Note 1 for a detailed description). Under nutrient-rich conditions, protein synthesis rates are high and this constraint is expected to be hit. Alongside with the nutrient abundance limitation, these two constraints now allow the successful prediction of the acetate switch in *E. coli* (40).

For this purpose, we simulated the acetate switch for glucose-fuelled MOPS -AA, MOPS -Met and MOPS +AA media. For the simulation, we varied the available external concentration of glucose and amino acids, which leads to increased uptake fluxes of glucose and amino acids (and the growth rate μ as well). As shown in Fig. 2A, the switch from respiration to acetate fermentation was predicted for all three media variants at $\mu \geq 0.4h^{-1}$. This result is in close agreement with experimental measurements of glucose-grown *E. coli* (41). However, the amount of acetate flux is overpredicted, compared to reports of (42–44).

For all the media (Fig. 2A), mixed respiro-fermentative energy harvesting was observed due to limitation by the proteome capacity constraint. To investigate the re-routing of carbon from respiration towards acetate fermentation, we performed FVA at five different glucose uptake rates, corresponding to $0.10 \leq \mu \leq 0.42h^{-1}$. All three media variants in fully-respiratory states (Fig. 2B, first three panels), were found to have comparable sums of fluxes through the TCA cycle reactions and pyruvate metabolism. Further increase in growth rate ($\mu \geq 0.40\ h^{-1}$) and entering respiro-fermentative energy harvesting resulted in a decrease in sum of fluxes through TCA cycle reactions, which was the most remarkable for MOPS -AA. At $\mu = 0.42h^{-1}$ on MOPS -Met and MOPS +AA, we observed a higher sum of fluxes through TCA cycle, compared to MOPS -AA, with a decrease in the fluxes through pyruvate metabolism. This suggests that while glucose is redirected towards acetate-producing fermentation, the amino acids enter the TCA cycle. Reports from ^{13}C flux tracking experiments suggested that carbon from ^{13}C-labelled alanine was mostly accumulated in TCA intermediates (45).

4.3. Identification of core proteome for glucose-growing *E. coli*.

In general, microbes invest a high amount of cellular energy in protein biosynthesis (46). As energetic costs are important determinants for fitness, protein expression is expected to be highly optimized towards the synthesis of proteins that are crucially needed.

Since the metabolic fluxes in our model demand expression of respective proteins, we can now compute flux distributions and identify the essential protein species in condition-dependent manner. In our model, an essential protein could be determined as the one, having non-zero flux through its synthesis reaction. Using this criterion, we probed the flux distributions for MOPS -AA, MOPS -Met and MOPS +AA media.

We identified non-zero flux values for 391, 378 and 358 protein synthesis reactions for MOPS -AA, MOPS -Met and MOPS +AA media, respectively (Table S1A). Consistently with the observation that protein synthesis demands high amounts of energy and resources, fluxes through protein synthesis reactions were determined to have none or negligibly small variability (Fig. S1). Comparing the pools of the essential proteins in all three media conditions, we observed 307 of these protein synthesis reactions, which were carrying fluxes in all three medium conditions (Table 1).

Many of the non-zero fluxes through protein synthesis reactions were obtained for proteins of the protein turnover machinery. The expression of these proteins is required for any experimental condition. A number of metabolic (e.g. carbohydrate or amino acid metabolism) enzymes were also identified as essential (Table 1). These proteins represent a kind of core proteome which comprises all proteins that are essential to obtain the optimal growth under different conditions. We also compared this core proteome with the predictions of gene essentiality in the conventional GEM without protein cost allocation iML1515 (Table

Table 1. Numbers of protein entities, present in the core proteome of all MOPS media (n = 307).

Macromolecular machinery	No. of reactions
Chaperone subunits	4
Protease subunits	4
Ribosomal subunits	56
ATP synthase subunits	8
Amino acid biosynthesis	16
Cofactor biosynthesis	29
Carbohydrate metabolism	18
Lipid metabolism	15
Total metabolic enzymes:	**243**

S1B).

In the iML1515, deletion of 196, 152 and 150 genes were identified as lethal for MOPS -AA, MOPS -Met and MOPS +AA media, respectively. 149 of these genes were essential to all three growth conditions. 118 of these 149 genes were present in the core proteome we obtained. The majority of the 31 genes that were essential for iML1515 and not for our model are involved in biosynthesis of various cofactors and transport of inorganic ions. We suggest that the essentiality of these genes for the iML1515 model is due to the optimal stoichiometry of the reactions they catalyse. In our model, however, a combination of optimality of stoichiometry (i.e. highest yield) and protein turnover costs are assessed to identify the non-essential pathways, thus we argue that our model predicts employment of different pathways compared to the ones predicted by iML1515.

Albeit protein expression consumes high amounts of the available cellular energy, the amount of proteins expressed in living cells are far higher than the core proteome. The dataset of Li *et al.* (5) reported expression levels of approximately 3000 proteins for MOPS -AA, MOPS -Met and MOPS +AA media. Both the number of species and their abundance were higher, compared to the here predicted minimal core proteome with only about 400 proteins (experimentally determined $vs.$ predicted abundance correlated with Pearson's $r \approx 0.1$, Fig. S2). This finding was expected, as we compute the *minimal* requirement of the proteins to sustain growth, albeit it is well known that most of the *E. coli* enzymes work in undersaturated regimen (47). Given little agreement between minimal protein requirement and the experimentally observed protein expression, we further addressed the impact of the integration of proteomic data to the model predictions.

4.4. Input of quantitative proteomic data affects flux variability in the metabolic network. Multiple experimental studies have reported expression of a high number of protein species (5, 19, 48). Protein synthesis in microbes is indeed so high that it is a major consumer of energy produced in the cell (46) and can eventually be in the limit of molecular crowding (49). As protein synthesis cannot be predicted on the basis of the direct necessity in neither conventional or extended GEMs, integration of quantitative proteomic data should provide a more realistic view on the proteome expression.

Li and colleagues (5) quantified the entire proteome of *E. coli* in copy numbers per cell for MOPS -AA, MOPS -Met and MOPS +AA media. Using this information, we formulated a minimal protein abundance requirement for 1 gram of cell dry weight ($mmol(gDW)^{-1}$) with an assumed growth rate $\mu = 1\ h^{-1}$. For each protein species quantified we set the lower flux bound for the respective protein synthesis reaction to 90% of the experimentally determined abundance. With applying the 90% correction, we aimed to adjust the limited proteome space for the proteins which might have been underrepresented in the experimental measurements. Using the minimal proteome abundance requirement as an additional set of constraints, we analysed the resulting flux distributions.

We performed FVA to compute the flux intervals through individual reactions of the metabolic network for MOPS -AA, MOPS -Met and MOPS +AA media. Then, for each scenario, we ranked the flux intervals by size. The here calculated FVA intervals differed considerably from the intervals computed without integrating the quantitative proteomic data, where only a low number of reactions had non-zero flux variability. To compare the flux variability with and without proteomic data input, we acquired cumulative flux distributions by summing up the flux intervals over all the reactions of the metabolic network (Fig. 3A). The largest sum of flux intervals was assigned the cumulative value of 1.0, and the others were scaled accordingly.

We compared the cumulative distributions of flux variability for each of the growth media between the flux distributions with and without the quantitative proteomic data. To quantify the distance between paired (i.e. at the same position of respective cumulative distribution) elements in the distributions of two samples, the Kolmogorov-Smirnov statistic D can be computed. Subsequently, the Kolmogorov-Smirnov test is used to determine whether the difference between samples is significant. Using the two-sample Kolmogorov-Smirnov test, we acquired $D = 1.00$, 0.8867 and 1.00 for MOPS -AA, MOPS -Met and MOPS +AA, respectively (null hypothesis rejected with $p < 0.001$ for all comparisons). The results of the test thus show that the input of quantitative proteomic data to the extended model considerably changes the flux variability in the metabolic network and therefore, the behaviour of the model.

4.5. Proteomics input-dependent flux variability guides identification of unused metabolic pathways. The emergence of non-zero flux variability in the metabolic network likely represents some reserve capacity (i.e. abundance of more enzymatic machinery than is consumed by corresponding metabolic fluxes) of the enzymatic machinery. Notably, this indicated that the expression of the corresponding enzyme machinery is higher than the levels in the essentially needed proteome, suggesting that these pathways might be unused or underutilized. As an example of such phenotypes, we focused on the amino acid metabolism (Fig. 3B).

The different MOPS media in the study of Li *et al.* (5) differed in terms of the availability of amino acids in the medium, including methionine (Met). The energy costs of Met biosynthesis were previously determined to be the highest per unit of amino acid produced (50). Thus, *E. coli* could benefit considerably from transporting Met from its growth environment instead of utilizing the pathways of Met biosynthesis. Based on this, we would expect that in presence of enough Met, no Met biosynthesis and no expression of the respective enzymes is taking place.

However, experimental studies indicated the expression of Met biosynthetic enzymes and active biosynthesis even in presence of sufficient Met in the growth medium. Given high investments in this pathway, this seems to be counter-intuitive. To gain more insights, we explored the effects of integrating proteomic data to the variability to pathways amino acid metabolism.

Initially, we observed the computed minimal demand for these processes, as described in Section 4.3. The sums of flux variability intervals through reactions in the cluster of methionine metabolism were 0.00018, 0.1843 and 0.000 $mmol(gDW)^{-1}h^{-1}$ for MOPS -AA, MOPS -Met and MOPS +AA, respectively. Then, we performed simulations with integration of proteomic data (see Section 4.4). Notably, an increase in sums of flux variability intervals were observed with the introduction of proteomics data: 0.3403, 0.64833 and 0.7882 $mmol(gDW)^{-1}h^{-1}$ for MOPS -AA, MOPS -Met and MOPS +AA, respectively (Fig. 3B). Thus, an increase of flux variability in Met metabolism suggests a underutilization of the enzyme machinery expressed, even in the MOPS +AA medium. The experimental measurements of *E. coli*, where a comparable expression was determined for methionine metabolism-related proteins grown in both MOPS -Met and Met-rich MOPS +AA media (5).

Other amino acid metabolic pathways followed the same pattern (Fig. 3B, Fig. S3). Integration of quantitative proteomic data resulted in high flux variability intervals in clusters of Thr/Lys metabolism, as well as VLI metabolism. A high reserve capacity of the metabolic pathways for these amino acids was observed in any variant of the medium tested.Thus, the underutilization of the protein machinery is likely to be a universal trait of the metabolic network.

4.6. Prediction of condition-dependent traits of metabolic network. *E. coli* can sustain its growth on a wide spectrum of different carbon sources. This often requires to relocate its resources in a context-dependent manner to sustain optimal growth. In the following, we investigated whether the model can adapt the flux distributions through the metabolic network for making use of different carbon sources. For this, we used the previously published dataset from (19). The entire proteome of *E. coli* was quantified in copy numbers per cell in 16 different growth media with distinct points-of-entry of the carbon sources into the metabolic network. To mimic the different growth conditions, we applied both nutrient availability- and proteomic data-related constraints to the model, as described in Sections 4.1 and 4.4.

We analysed flux distributions for twelve different growth conditions, using FBA and FVA. For the analysis, we selected several clusters of the metabolic network, which show some characteristic traits of condition-dependent metabolism (Fig. 4 and Fig. S4).

Fundamental metabolic pathways, such as glycolysis, TCA cycle or the pentose phosphate pathway (PPP) exhibited rather robust flux distributions (Fig. 4). For all three clusters, only the minimal sums of fluxes were condition-dependent, while the maximal sums of fluxes through these clusters were comparable for all conditions.

The minimal sums were considerably lower for substrates, entering the metabolic network through the TCA cycle, e.g. acetate or pyruvate. Experimentally, a decrease of the specific activity of phosphofructokinase was observed for *E. coli* grown on the TCA cycle intermediates (51). In line with consistent expression of the glycolytic enzymes, an overexpression of glycolysis/gluconeogenesis enzymes was shown to improve *E. coli* growth on succinate (52).

The TCA intermediates are primed for gluconeogenesis by entering the glyoxylate shunt. The sums of fluxes through this pathway were non-zero in the case of four TCA intermediates (acetate, fumarate, pyruvate and succinate). The transfer of glucose-pregrown *E. coli* to acetate- (53) or fatty acid-fuelled (54) growth medium was shown to cause a considerable increase in the activity of the glyoxylate shunt enzymes isocitrate lyase and malate synthase.

The minimal sum of fluxes through the cluster of TCA cycle was 25 times larger for growth on acetate than for growth on mannose. Thus, the point-of-entry to the metabolic network is of key importance here. Notably, larger variability intervals for fluxes through this cluster were observed for the upper glycolysis substrates: growth on hexose sugars and the PPP intermediate xylose.

A high minimal sum of fluxes through PPP was a notable trait for xylose-fuelled media. Xylose-grown *E. coli* was experimentally observed to downregulate the expression of pyruvate metabolism-related genes (55). There is experimental evidence that *E. coli* growing on glucose possesses excess PPP activity for adapting to changing redox environment (56). Thus, our result validates that PPP has reserve capacity under different growth conditions.

Growth on glycerol (Glyc) led to a higher minimal sum of fluxes through TCA cycle reactions, compared to e.g. glucose-fuelled media. This in line with an experimental observation that *E. coli* grown on glycerol accumulates succinate, acetate and formate (57). Yet, only for Glyc + AA (glycerol with amino acid supplement) growth medium there was a non-zero minimal sum of fluxes predicted through the methylglyoxal metabolic pathway.

Methylglyoxal is toxic to microorganisms, such as *E. coli* (58) and the pathway of its production begins with dihydroxyacetone phosphate (DHAP). In the limiting conditions, Glyc enters glycolysis via phosphorylation to Glyc-3-phosphate and oxidation to DHAP. Growth of *E. coli* in excess Glyc (59) or impaired feedback inhibition in the Glyc-3-phosphate oxidation (60) resulted in production of methylglyoxal. Consistently with these results, our model was able to capture the overflow of excess Glyc towards the methylglyoxal pathway.

The costs of amino acid transporters $vs.$ precursor-based biosynthesis were significantly reflected in growth on the Glyc + AA medium. For instance, very little flux was observed for the reactions in the cluster of VLI metabolism, as well as clusters of other amino acid metabolism (Fig. S4). Consistently with the predictions, experimental supplementation of growth media with these branched-chain amino acids was shown to repress the synthesis of the respective biosynthetic proteins (61) and acyl-tRNA synthethases (62). It is, however, important to stress that the proteins of the VLI metabolism are still expressed, as discussed in Section 4.5.

5. Discussion

In this study, we developed a new type of genome-scale model that explicitly accounts for protein expression and extends the scope of conventional genome-scale metabolic models by enabling to study resource allocation. The new model includes detailed descriptions of protein turnover processes. The model structure allows for the integration of quantitative high-throughput proteomic data at a single protein level. Yet, the model retains its linear properties and can thus be easily investigated at low computational cost by common methods like FBA and FVA.

Current extensions of conventional GEMs provide different levels of complexity in describing protein turnover processes. ME-models (12) include a fine-grained description of gene expression processes, including transcription of mRNA and replication of genomic DNA. In conventional GEMs, the expenditures on genomic DNA and mRNA production are formulated as a demand of nucleotides in biomass equation. In contrast, we did not include description of transcriptional processes in our approach. Notably, we did not alter the nucleotide demands in biomass equation either. Since protein expression and turnover consumes considerable amounts of cellular energy, we assumed protein expression to be the limiting factor for balanced growth. In favour of our argument, we were able to successfully identify a number of condition-dependent traits in the metabolic network of *E. coli* (Fig. 4). Furthermore, in balanced growth, gene and protein expression are in a tight relationship (i.e. fixed stoichiometry) (15).

Thus, considering only the protein (and not nucleic acid) turnover costs results in lesser inflation of the model structure, compared to extending the conventional GEMs to ME-models. With a fewer number of reactions and constraints, the computation of flux distributions for these models becomes less computationally expensive and can be done at sufficient accuracy using conventional LP solvers. Another modelling strategy, RBA, has successfully been applied to study resource allocation in *Bacillus subtilis* (13). In contrast to our model framework, RBA requires parameter estimation and introduction of growth rate-dependent parameters. Growth rate-dependent parameters are introduced in ME-models and GECKO models as well. Thus, for these approaches, usage of standardized methods, such as FBA and FVA, becomes complicated and computationally expensive. Meanwhile, another kind of genome-scale model of *E. coli* is proposed by Alter and colleagues and is available as a preprint (63). Yet, the protein allocation model (PAM) they proposed considers only optimal partitioning of a limited proteome to different protein entities and not their turnover costs. Using PAM, Alter and colleagues were able to reproduce scenarios like the acetate switch or compare model predictions from ^{13}C-flux tracking data. However, unlike our approach, setting up the PAM to predict distinct scenarios is subject to a lot of calibration, using experimental measurements. While this is not an issue for a well-studied organism, such as *E. coli*, the requirement of large amounts of data might be critical for transferring the PAM implementation to the GEMs of other species.

Finally, our strategy permits straightforward integration of high-throughput quantitative proteomic data. This benefit arises from a clear and direct description of protein turnover processes. For instance, the intrinsic flux coupling is introduced by the presence of protein species next to metabolites in the respective metabolic reactions. It resembles the GECKO strategy (as described in the Supplementary Note 1) which has been successfully applied to *Saccharomyces cerevisiae* (14) and *B. subtilis* (64). In the model of *B. subtilis*, input of quantitative proteomic data for 17 proteins led to reduction of flux variability for reactions of the central carbon metabolism, reinforcing their essentiality to the metabolic network.

We have thus shown that, in contrast to conventional GEMs, the flux distributions of the extended model are more confined to physiologically relevant states, as discussed in Section 4.1 (Fig. 1). Predictions by conventional GEMs are usually limited to the prediction of efficient metabolic states (2, 65). In this model structure, we, however, used a combination of two global constraints (66) to predict metabolically less efficient states without the presence of hard constraints at the respective, less efficient metabolic pathway. The extended model thus allows the prediction of overflow metabolism like the acetate switch in glucose-grown *E. coli* (Fig. 2) or the overflow of excess glycerol towards the methylglyoxalate pathway (Fig. 4). The latter was experimentally observed decades ago (62) but never mechanistically studied in detail.

We also computed the core proteome for glucose-growing *E. coli* (Table 1). A core proteome is of particular importance for synthetic biology as it comprises the minimal set of enzymes and proteins required for a functional, living cell. Using our model, we identified the core proteome of about 300 proteins. The essentiality of proteins for *E. coli* was also previously assessed using ME-models by Yang *et al.* (67). They studied the core proteome at the cluster level at different environmental conditions (19). Building on this, here we included subsequent analysis at a single protein level.

The model can also capture various bacterial metabolic strategies, e.g. reserve capacity for methionine metabolism (Section 4.5). As protein expression consumes considerable amounts of the available cellular energy (46, 68), microbes are usually forced into an economic balance between the "useful" proteome and investment in proteins that are not required at certain environmental conditions. Consequently, many biosynthetic enzymes are not expressed when the respective metabolites are present in the environment. For the amino acids, however, integration of quantitative proteomic data revealed a graspable reserve flux capacity through these biosynthetic pathways (Fig. 3B).

We observed the reserve flux capacity to be especially prominent in the case of L-methionine metabolism due to its high costs of biosynthesis (50). The expression and turnover of enzymes that are not essentially required at the current environmental condition might bring evolutionary benefit for microbes (69). In such way, *E. coli* gains advantage over competing species by successfully adapting to changing environments. Microbes are constantly subjected to quickly changing environmental

conditions, therefore, the expression of non-essential proteins and enzymes are commonly observed in other microbial species as well (70).

Importantly, only given non-changing conditions, the excess protein is subject to proteome reshuffling. The protein quantities are then expected to arrive to a more adapted state, which is represented in our model by the minimal protein requirement. Such behaviour was also observed experimentally: when cultivated in a steady environment for a prolonged time, *E. coli* does adapt to accommodate more useful proteins. Notably, this was shown both on transcriptome (71) and proteome (72) levels for *E. coli*, kept under glucose-limited conditions for over 200 generations.

6. Conclusions

To summarize, we have shown that our extended model confines the flux distributions towards more physiologically-relevant solutions, which are in good agreement with experimentally determined adaptations of *E. coli*. Thus, we expect this tool to find its applications in investigations of microbial growth.

ACKNOWLEDGEMENTS

This work was performed on the computational resource bwUniCluster funded by the Ministry of Science, Research and the Arts Baden-Württemberg and the Universities of the State of Baden-Württemberg, Germany, within the framework program bwHPC. PG was supported by MSCA ITN "SynCrop" (grant agreement no. 764591), NV and TF was supported by DFG (grant numbers VE1075/2-1 and FI 1588/2-1). We would like to thank Sven Sahle, Bernd Kreikemeyer and Eunice van Pelt-KleinJan for fruitful discussions.

COMPETING FINANCIAL INTERESTS

Authors declare no conflicting interests present.

Bibliography

1. Jeffrey D. Orth, Ines Thiele, and Bernhard O. Palsson. What is flux balance analysis?, 2010. ISSN 10870156.
2. Bas Teusink, Anne Wiersma, Douwe Molenaar, Christof Francke, Willem M. De Vos, Roland J. Siezen, and Eddy J. Smid. Analysis of growth of Lactobacillus plantarum WCFS1 on a complex medium using a genome-scale metabolic model. *Journal of Biological Chemistry*, 2006. ISSN 00219258. doi: 10.1074/jbc.M606263200.
3. Xin Wang, Kang Xia, Xiaojing Yang, and Chao Tang. Growth strategy of microbes on mixed carbon sources. *Nature Communications*, 2019. ISSN 20411723. doi: 10.1038/s41467-019-09261-3.
4. Matthew Scott, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Science*, 2010. ISSN 00368075. doi: 10.1126/science.1192588.
5. Gene Wei Li, David Burkhardt, Carol Gross, and Jonathan S. Weissman. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3): 624–635, 2014. ISSN 10974172. doi: 10.1016/j.cell.2014.02.033.
6. A. N. Milne, W. Wai Nam Mak, and J. Tze Fei Wong. Variation of ribosomal proteins with bacterial growth rate. *Journal of Bacteriology*, 1975. ISSN 00219193.
7. Douwe Molenaar, Rogier Van Berlo, Dick De Ridder, and Bas Teusink. Shifts in growth strategies reflect tradeoffs in cellular economics, 2009. ISSN 17444292.
8. Karl Peebo, Kaspar Valgepea, Andres Maser, Ranno Nahku, Kaarel Adamberg, and Raivo Vilu. Proteome reallocation in Escherichia coli with increasing specific growth rate. *Molecular BioSystems*, 2015. ISSN 17422051. doi: 10.1039/c4mb00721b.
9. Mingyuan Tian and Jennifer L. Reed. Integrating proteomic or transcriptomic data into metabolic models using linear bound flux balance analysis. *Bioinformatics*, 2018. ISSN 14602059. doi: 10.1093/bioinformatics/bty445.
10. Q. K. Beg, A. Vazquez, J. Ernst, M. A. De Menezes, Z. Bar-Joseph, A. L. Barabási, and Z. N. Oltvai. Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. *Proceedings of the National Academy of Sciences of the United States of America*, 2007. ISSN 00278424. doi: 10.1073/pnas.0609845104.
11. Mariola Szenk, Ken A. Dill, and Adam M.R. de Graff. Why Do Fast-Growing Bacteria Enter Overflow Metabolism? Testing the Membrane Real Estate Hypothesis. *Cell Systems*, 2017. ISSN 24054720. doi: 10.1016/j.cels.2017.06.005.
12. E. J. O{middle dot}Brien, Joshua A Lerman, Roger L Chang, Daniel R Hyduke, and Bernhard O. Palsson. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, 9(1):693–693, apr 2014. ISSN 1744-4292. doi: 10.1038/msb.2013.52.
13. Anne Goelzer, Vincent Fromion, and Gérard Scorletti. Cell design in bacteria as a convex optimization problem. *Automatica*, 2011. ISSN 00051098. doi: 10.1016/j.automatica.2011.02.038.
14. Benjamín J Sánchez, Cheng Zhang, Avlant Nilsson, Petri-jaan Lahtvee, Eduard J Kerkhoven, and Jens Nielsen. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology*, 13(8):935, aug 2017. ISSN 1744-4292. doi: 10.15252/msb.20167411.
15. S. Proshkin, A. R. Rahmouni, A. Mironov, and E. Nudler. Cooperation Between Translating Ribosomes and RNA Polymerase in Transcription Elongation. *Science*, 328(5977):504–508, apr 2010. ISSN 0036-8075. doi: 10.1126/science.1184939.
16. Jonathan M Monk, Colton J Lloyd, Elizabeth Brunk, Nathan Mih, Anand Sastry, Zachary King, Rikiya Takeuchi, Wataru Nomura, Zhen Zhang, Hirotada Mori, Adam M Feist, and Bernhard O Palsson. iML1515, a knowledgebase that computes Escherichia coli traits. *Nature Biotechnology*, 35(10):904–908, 2017. ISSN 15461696. doi: 10.1038/nbt.3956.
17. Alex Bateman, Maria Jesus Martin, Claire O'Donovan, Michele Magrane, Emanuele Alpi, Ricardo Antunes, Benoit Bely, Mark Bingley, Carlos Bonilla, Ramona Britto, Borisas Bursteinas, Hema Bye-AJee, Andrew Cowley, Alan Da Silva, Maurizio De Giorgi, Tunca Dogan, Francesco Fazzini, Leyla Garcia Castro, Luis Figueira, Penelope Garmiri, George Georghiou, Daniel Gonzalez, Emma Hatton-Ellis, Weizhong Li, Wudong Liu, Rodrigo Lopez, Jie Luo, Yvonne Lussi, Alistair MacDougall, Andrew Nightingale, Barbara Palka, Klemens Pichler, Diego Poggioli, Sangya Pundir, Luis Pureza, Guoying Qi, Steven Rosanoff, Rabie Saidi, Tony Sawford, Aleksandra Shypitsyna, Elena Speretta, Edward Turner, Nidhi Tyagi, Vladimir Volynkin, Tony Wardell, Kate Warner, Xavier Watkins, Rossana Zaru, Hermann Zellner, Ioannis Xenarios, Lydie Bougueleret, Alan Bridge, Sylvain Poux, Nicole Redaschi, Lucila Aimo, Ghislaine ArgoudPuy, Andrea Auchincloss, Kristian Axelsen, Parit Bansal, Delphine Baratin, Marie Claude Blatter, Brigitte Boeckmann, Jerven Bolleman, Emmanuel Boutet, Lionel Breuza, Cristina Casal-Casas, Edouard De Castro, Elisabeth Coudert, Beatrice Cuche, Mikael Doche, Dolnide Dornevil, Severine Duvaud, Anne Estreicher, Livia Famiglietti, Marc Feuermann, Elisabeth Gasteiger, Sebastien Gehant, Vivienne Gerritsen, Arnaud Gos, Nadine Gruaz-Gumowski, Ursula Hinz, Chantal Hulo, Florence Jungo, Guillaume Keller, Vicente Lara, Philippe Lemercier, Damien Lieberherr, Thierry Lombardot, Xavier Martin, Patrick Masson, Anne Morgat, Teresa Neto, Nevila Nouspikel, Salvo Paesano, Ivo Pedruzzi, Sandrine Pilbout, Monica Pozzato, Manuela Pruess, Catherine Rivoire, Bernd Roechert, Michel Schneider, Christian Sigrist, Karin Sonesson, Sylvie Staehli, Andre Stutz, Shyamala Sundaram, Michael Tognolli, Laure Verbregue, Anne Lise Veuthey, Cathy H. Wu, Cecilia N. Arighi, Leslie Arminski, Chuming Chen, Yongxing Chen, John S. Garavelli, Hongzhan Huang, Kati Laiho, Peter McGarvey, Darren A. Natale, Karen Ross, C. R. Vinayaka, Qinghua Wang, Yuqi Wang, Lai Su Yeh, and Jian Zhang. UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, 45(D1):D158–D169, 2017. ISSN 13624962. doi: 10.1093/nar/gkw1099.
18. A. Bairoch. The ENZYME database in 2000. *Nucleic Acids Research*, 2000. ISSN 0305-1048. doi: 10.1093/nar/28.1.304.
19. Alexander Schmidt, Karl Kochanowski, Silke Vedelaar, Erik Ahrné, Benjamin Volkmer, Luciano Callipo, Kèvin Knoops, Manuel Bauer, Ruedi Aebersold, and Matthias Heinemann. The quantitative and condition-dependent Escherichia coli proteome. *Nature Biotechnology*, 34(1):104–110, jan 2016. ISSN 1087-0156. doi: 10.1038/nbt.3418.
20. Ulrike Wittig, Maja Rey, Andreas Weidemann, Renate Kania, and Wolfgang Müller. SABIO-RK: an updated resource for manually curated biochemical reaction kinetics. *Nucleic Acids Research*, 46 (D1):D656–D660, jan 2018. ISSN 0305-1048. doi: 10.1093/nar/gkx1065.
21. Lisa Jeske, Sandra Placzek, Ida Schomburg, Antje Chang, and Dietmar Schomburg. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Research*, 47(D1):D542–D549, jan 2019. ISSN 0305-1048. doi: 10.1093/nar/gky1048.
22. Christiam Camacho, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. BLAST+: Architecture and applications. *BMC Bioinformatics*, 2009. ISSN 14712105. doi: 10.1186/1471-2105-10-421.
23. Patrick Aloy, Hugo Ceulemans, Alexander Stark, and Robert B. Russell. The relationship between sequence and interaction divergence in proteins. *Journal of Molecular Biology*, 2003. ISSN 00222836. doi: 10.1016/j.jmb.2003.07.006.
24. Arren Bar-Even, Elad Noor, Yonatan Savir, Wolfram Liebermeister, Dan Davidi, Dan S. Tawfik, and Ron Milo. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry*, 50(21):4402–4410, may 2011. ISSN 0006-2960. doi: 10.1021/bi2002289.
25. Brett G. Olivier, Johann M. Rohwer, and Jan Hendrik S. Hofmeyr. Modelling cellular systems with PySCeS. *Bioinformatics*, 2005. ISSN 13674803. doi: 10.1093/bioinformatics/bti046.

26. Alexei Vazquez, Qasim K Beg, Marcio A DeMenezes, Jason Ernst, Ziv Bar-Joseph, Albert-lászló Barabási, László G Boros, and Zoltán N Oltvai. Impact of the solvent capacity constraint on E. coli metabolism. *BMC Systems Biology*, 2(1):7, 2008. ISSN 1752-0509. doi: 10.1186/1752-0509-2-7.
27. Steven B Zimmerman and S O Trach. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *Journal of molecular biology*, 222(3):599–620, dec 1991. ISSN 0022-2836.
28. S. B. Dwyer and H. E. Umbarger. Isoleucine and valine metabolism of Escherichia coli. XVI. Pattern of multivalent repression in strain K-12. *Journal of Bacteriology*, 1968. ISSN 00219193.
29. E. Boy and J. C. Patte. Multivalent repression of aspartic semialdehyde dehydrogenase in Escherichia coli K-12. *Journal of Bacteriology*, 1972. ISSN 00219193.
30. L. Han, M. Doverskog, S. O. Enfors, and L. Häggström. Effect of glycine on the cell yield and growth rate of Escherichia coli: Evidence for cell-density-dependent glycine degradation as determined by 13C NMR spectroscopy. *Journal of Biotechnology*, 2002. ISSN 01681656. doi: 10.1016/S0168-1656(01)00373-X.
31. T. H. Meedel and L. I. Pizer. Regulation of one carbon biosynthesis and utilization in Escherichia coli. *Journal of Bacteriology*, 1974. ISSN 00219193.
32. H M KATZEN and J M BUCHANAN. ENZYMATIC SYNTHESIS OF THE METHYL GROUP OF METHIONINE. 8. REPRESSION-DEREPRESSION, PURIFICATION, AND PROPERTIES OF 5,10-METHYLENETETRAHYDROFOLATE REDUCTASE FROM ESCHERICHIA COLI. *The Journal of biological chemistry*, 1965. ISSN 0021-9258.
33. Rich Leavitt and H E Umbarger. ISOLEUCINE AND VALINE METABOLISM IN ESCHERICHIA COLI. *Journal of bacteriology*, 1961.
34. J. D. Friesen, N. P. Fiil, and K. Von Meyenburg. Synthesis and turnover of basal level guanosine tetraphosphate in Escherichia coli. *Journal of Biological Chemistry*, 1975. ISSN 00219258.
35. Y. INUI and H. AKEDO. AMINO ACID UPTAKE BY ESCHERICHIA COLI GROWN IN PRESENCE OF AMINO ACIDS. EVIDENCE FOR REPRESSIBILITY OF AMINO ACID UPTAKE. *Biochim. Biophys. Acta*, 94:143–152, Jan 1965.
36. Allen C. Rogerson and Martin Freundlich. Control of isoleucine, valine and leucine biosynthesis VIII. Mechanism of growth inhibition by leucine in relaxed and stringent strains of Escherichia coli K-12. *BBA - General Subjects*, 1970. ISSN 03044165. doi: 10.1016/0304-4165(70)90051-6.
37. J. C. McKitrick and L. I. Pizer. Regulation of phosphoglycerate dehydrogenase levels and effect on serine synthesis in Escherichia coli K-12. *Journal of Bacteriology*, 1980. ISSN 00219193.
38. R. C. Greene and C. Radovich. Role of methionine in the regulation of serine hydroxymethyltransferase in Escherichia coli. *Journal of Bacteriology*, 1975. ISSN 00219193.
39. Markus Basan, Sheng Hui, Hiroyuki Okano, Zhongge Zhang, Yang Shen, James R. Williamson, and Terence Hwa. Overflow metabolism in Escherichia coli results from efficient proteome allocation. *Nature*, 2015. ISSN 14764687. doi: 10.1038/nature15765.
40. A. J. Wolfe. The Acetate Switch. *Microbiology and Molecular Biology Reviews*, 2005. ISSN 1092-2172. doi: 10.1128/mmbr.69.1.12-50.2005.
41. Anke Kayser, Jan Weber, Volker Hecht, and Ursula Rinas. Metabolic flux analysis of Escherichia coli in glucose-limited continuous culture. I. Growth-rate-dependent metabolic efficiency at steady state. *Microbiology*, 2005. ISSN 13500872. doi: 10.1099/mic.0.27481-0.
42. Hans Peter Meyer, Christian Leist, and Armin Fiechter. Acetate formation in continuous culture of Escherichia coli K12 D1 on defined and complex media. *Journal of Biotechnology*, 1984. ISSN 01681656. doi: 10.1016/0168-1656(84)90027-0.
43. G. N. Vemuri, E. Altman, D. P. Sangurdekar, A. B. Khodursky, and M. A. Eiteman. Overflow metabolism in Escherichia coli during steady-state growth: Transcriptional regulation and effect of the redox ratio. *Applied and Environmental Microbiology*, 2006. ISSN 00992240. doi: 10.1128/AEM.72.5.3653-3661.2006.
44. A. Varma and B. O. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Applied and Environmental Microbiology*, 1994. ISSN 00992240.
45. Ohji IFUKU, Hiroaki MIYAOKA, Nobuyoshi KOGA, Jiro KISHIMOTO, Shinichiroui HAZE, Youji WACHI, and Masahiro KAJIWARA. Origin of the carbon atoms of biotin: 13C-NMR studies on biotin biosynthesis in Escherichia coli. *European Journal of Biochemistry*, 1994. ISSN 14321033. doi: 10.1111/j.1432-1033.1994.tb18659.x.
46. P.-J. Lahtvee, Andrus Seiman, Liisa Arike, Kaarel Adamberg, and Raivo Vilu. Protein turnover forms one of the highest maintenance costs in Lactococcus lactis. *Microbiology*, 160(Pt_7):1501–1512, jul 2014. ISSN 1350-0872. doi: 10.1099/mic.0.078089-0.
47. Kaspar Valgepea, Kaarel Adamberg, Andrus Seiman, and Raivo Vilu. Escherichia coli achieves faster growth by increasing catalytic and translation rates of proteins. *Mol. BioSyst.*, 9(9):2344–2358, 2013. doi: 10.1039/C3MB70119K.
48. Eduardo P.C. Rocha. The replication-related organization of bacterial genomes, 2004. ISSN 13500872.
49. S. Klumpp, M. Scott, S. Pedersen, and T. Hwa. Molecular crowding limits translation and cell growth. *Proceedings of the National Academy of Sciences*, 110(42):16754–16759, 2013. ISSN 0027-8424. doi: 10.1073/pnas.1310377110.
50. Christoph Kaleta, Sascha Schäuble, Ursula Rinas, and Stefan Schuster. Metabolic costs of amino acid and protein production in Escherichia coli. *Biotechnology Journal*, 2013. ISSN 18606768. doi: 10.1002/biot.201200267.
51. O. H. Lowry, J. Carter, J. B. Ward, and L. Glaser. The effect of carbon and nitrogen sources on the level of metabolic intermediates in Escherichia coli. *Journal of Biological Chemistry*, 1971. ISSN 00219258.
52. Y. P. Chao, R. Patnaik, W. D. Roof, R. F. Young, and J. C. Liao. Control of gluconeogenic growth by pps and pck in Escherichia coli. *Journal of Bacteriology*, 1993. ISSN 00219193. doi: 10.1128/jb.175.21.6939-6944.1993.
53. H. L. Kornberg. The role and control of the glyoxylate cycle in Escherichia coli., 1966. ISSN 02646021.
54. S. R. Maloy, M. Bohlander, and W. D. Nunn. Elevated levels of glyoxylate shunt enzymes in Escherichia coli strains constitutive for fatty acid degradation. *Journal of Bacteriology*, 1980. ISSN 00219193.
55. Ramon Gonzalez, Han Tao, K. T. Shanmugam, S. W. York, and L. O. Ingram. Global gene Expression differences associated with changes in glycolytic flux and growth rate in Escherichia coli during the fermentation of glucose and xylose. *Biotechnology Progress*, 2002. ISSN 87567938. doi: 10.1021/bp010121i.
56. Dimitris Christodoulou, Hannes Link, Tobias Fuhrer, Karl Kochanowski, Luca Gerosa, and Uwe Sauer. Reserve Flux Capacity in the Pentose Phosphate Pathway Enables Escherichia coli's Rapid Response to Oxidative Stress. *Cell Systems*, 2018. ISSN 24054720. doi: 10.1016/j.cels.2018.04.009.
57. Abhishek Murarka, Yandi Dharmadi, Syed Shams Yazdani, and Ramon Gonzalez. Fermentative utilization of glycerol by Escherichia coli and its implications for the production of fuels and chemicals. *Applied and Environmental Microbiology*, 2008. ISSN 00992240. doi: 10.1128/AEM.02192-07.
58. H. N.A. Fraval and D. C.H. Mc Brien. The effect of methyl glyoxal on cell division and the synthesis of protein and DNA in synchronous and asynchronous cultures of Escherichia coli B/r. *Journal of General Microbiology*, 1980. ISSN 00221287. doi: 10.1099/00221287-117-1-127.
59. S. Tötemeyer, N. A. Booth, W. W. Nichols, B. Dunbar, and I. R. Booth. From famine to feast: The role of methylglyoxal production in Escherichia coli. *Molecular Microbiology*, 1998. ISSN 0950382X. doi: 10.1046/j.1365-2958.1998.00700.x.
60. W. B. Freedberg, W. S. Kistler, and E. C. Lin. Lethal synthesis of methylglyoxal by Escherichia coli during unregulated glycerol metabolism. *Journal of Bacteriology*, 1971. ISSN 00219193.
61. J. J. Wasmuth and H. E. Umbarger. Effect of isoleucine, valine, or leucine starvation on the potential for formation of the branched chain amino acid biosynthetic enzymes. *Journal of Bacteriology*, 1973. ISSN 00219193.
62. E. McGinnis and L. S. Williams. Regulation of synthesis of the aminoacyl-transfer ribonucleic acid synthetases for the branched-chain amino acids of Escherichia coli. *Journal of Bacteriology*, 1971. ISSN 00219193.
63. Tobias B. Alter, Lars M. Blank, and Birgitta E. Ebert. Protein allocation and enzymatic constraints explain escherichia coli wildtype and mutant phenotypes. *bioRxiv*, 2020. doi: 10.1101/2020.02.10.941294.
64. Ilaria Massaiu, Lorenzo Pasotti, Nikolaus Sonnenschein, Erlinda Rama, Matteo Cavaletti, Paolo Magni, Cinzia Calvio, and Markus J. Herrgård. Integration of enzymatic data in Bacillus subtilis genome-scale metabolic model improves phenotype predictions and enables in silico design of poly-γ-glutamic acid production strains. *Microbial Cell Factories*, 2019. ISSN 14752859. doi: 10.1186/s12934-018-1052-2.
65. Margreet I. Pastink, Bas Teusink, Pascal Hols, Sanne Visser, Willem M. De Vos, and Jeroen Hugenholtz. Genome-scale model of Streptococcus thermophilus LMG18311 for metabolic comparison of lactic acid bacteria. *Applied and Environmental Microbiology*, 2009. ISSN 00992240. doi: 10.1128/AEM.00138-09.
66. Daan H. de Groot, Julia Lischke, Riccardo Muolo, Robert Planqué, Frank J. Bruggeman, and Bas Teusink. The common message of constraint-based optimization approaches: overflow metabolism is caused by two growth-limiting constraints. *Cellular and Molecular Life Sciences*, Nov 2019. ISSN 1420-9071. doi: 10.1007/s00018-019-03380-2.
67. Laurence Yang, James T. Yurkovich, Colton J. Lloyd, Ali Ebrahim, Michael A. Saunders, and Bernhard O. Palsson. Principles of proteome allocation are revealed using proteomic data and genome-scale models. *Scientific Reports*, 2016. ISSN 20452322. doi: 10.1038/srep36734.
68. Avi Flamholz, Elad Noor, Arren Bar-Even, Wolfram Liebermeister, and Ron Milo. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences of the United States of America*, 2013. ISSN 00278424. doi: 10.1073/pnas.1215283110.
69. Matteo Mori, Severin Schink, David W. Erickson, Ulrich Gerland, and Terence Hwa. Quantifying the benefit of a proteome reserve in fluctuating environments. *Nature Communications*, 2017. ISSN 20411723. doi: 10.1038/s41467-017-01242-8.
70. Jue Wang, Esha Atolia, Bo Hua, Yonatan Savir, Renan Escalante-Chong, and Michael Springer. Natural Variation in Preparation for Nutrient Depletion Reveals a Cost–Benefit Tradeoff. *PLoS Biology*, 2015. ISSN 15457885. doi: 10.1371/journal.pbio.1002041.
71. Alessandro G. Franchini and Thomas Egli. Global gene expression in Escherichia coli K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions. *Microbiology*, 2006. ISSN 13500872. doi: 10.1099/mic.0.28939-0.
72. Lukas M. Wick, Manfredo Quadroni, and Thomas Egli. Short- and long-term changes in proteome composition and kinetic properties in a culture of Escherichia coli during transition from glucose-excess to glucose-limited growth conditions in continuous culture and vice versa. *Environmental Microbiology*, 2001. ISSN 14622912. doi: 10.1046/j.1462-2920.2001.00231.x.

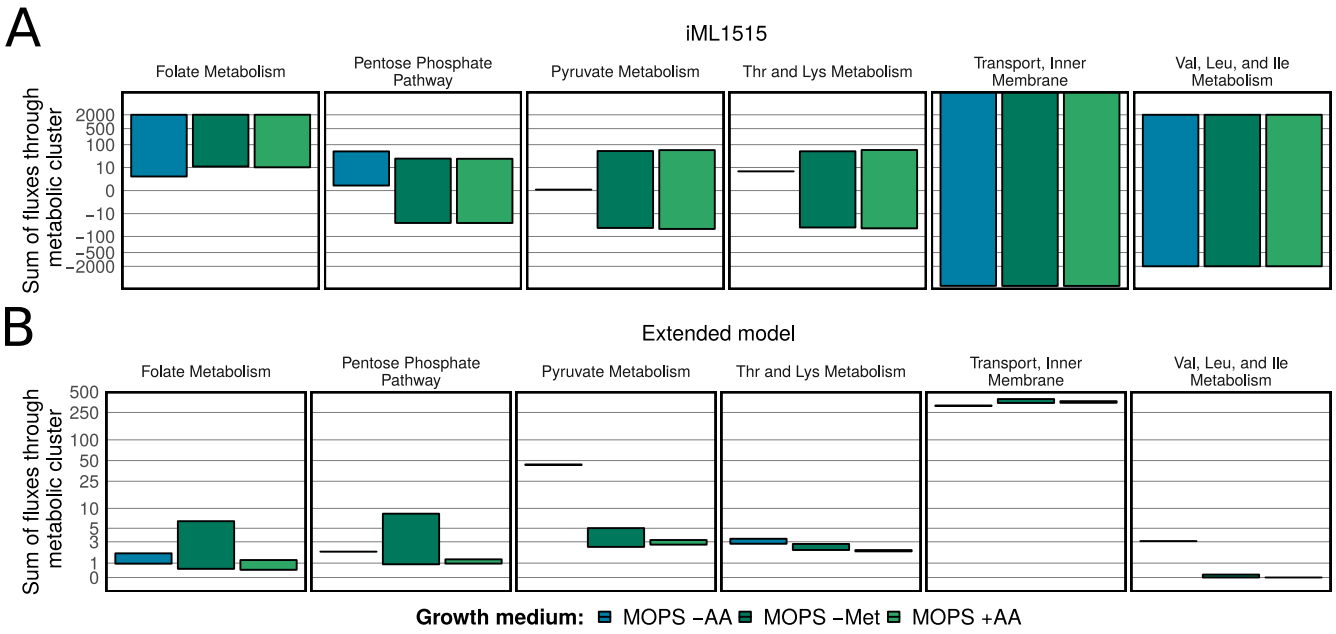

Fig. 1. Comparison of the sums of fluxes through representative clusters of the metabolic network for the genome-scale metabolic model iML1515 (A) and our extended genome-scale model (B).
For (A) and (B), Flux Variability Analysis (FVA) was used to compute flux intervals for all reactions in the model. Then the minimal and maximal values of the flux intervals (in $mmol(gDW)^{-1}h^{-1}$) were summed up to yield the lower and upper sum of fluxes, respectively.

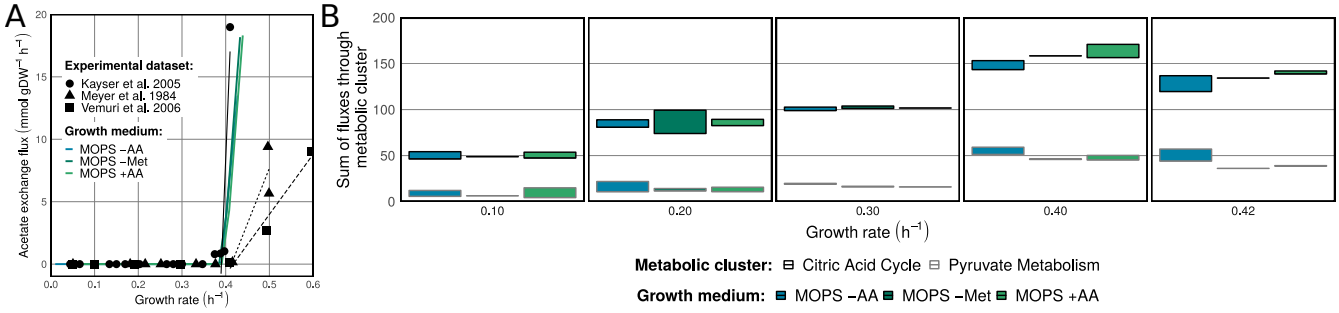

Fig. 2. Dependence of acetate production on the growth rate at different abundance of nutrients in MOPS media (A) and sums of fluxes through the TCA and pyruvate metabolism (B) reactions at given growth rates.
The initial media are MOPS -AA, MOPS -Met or MOPS +AA as indicated in (5). Then, the availability of carbon and nitrogen sources was step-wise changed by the same degree $0.0 < \gamma \leq 1.0$. Flux Variability Analysis (FVA) was used to compute flux intervals at every step for acetate exchange (A) or all reactions in the model (B). For (A), experimental data was obtained from (41–43). Black lines represent the linear fits of reported acetate excretion for each experimental dataset, starting at $\mu \geq 0.375\ h^{-1}$.
For (B), minimal and maximal values of the flux intervals (in $mmol(gDW)^{-1}h^{-1}$) were summed up to yield the lower and upper sum of fluxes, respectively.

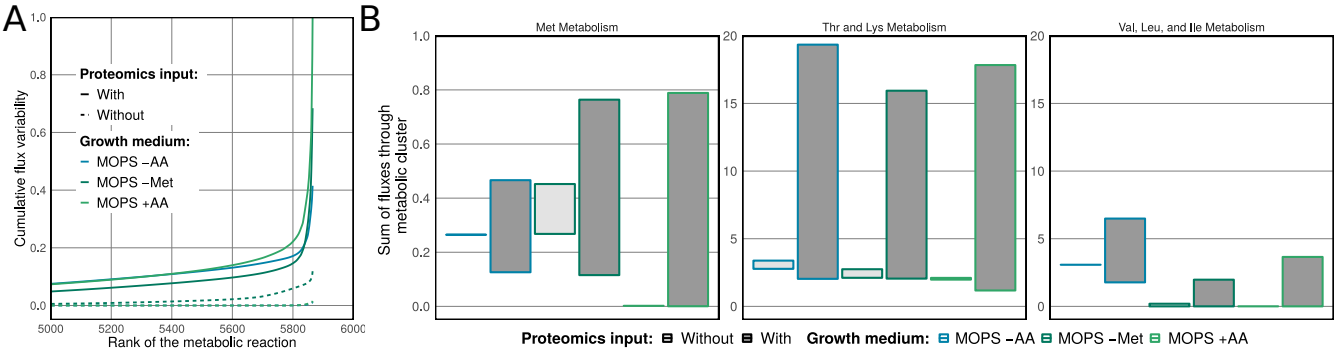

Fig. 3. Cumulative distributions of variability intervals of metabolic fluxes for different growth conditions (A) and flux distributions through clusters of selected amino acid metabolism (B).
Flux Variability Analysis (FVA) was used to compute flux intervals at different growth settings for all reactions in the model. For (A), the flux intervals acquired were ranked by increasing size. First 5000 out of 5865 values were omitted, and the rest of flux variability intervals were plotted as a cumulative function with respect to the condition with the highest cumulative flux variability (MOPS +AA, with proteomics input). For (B), the minimal and maximal values of the flux intervals acquired ($mmol(gDW)^{-1}h^{-1}$) were summed to yield the lower and upper sum of fluxes, respectively.

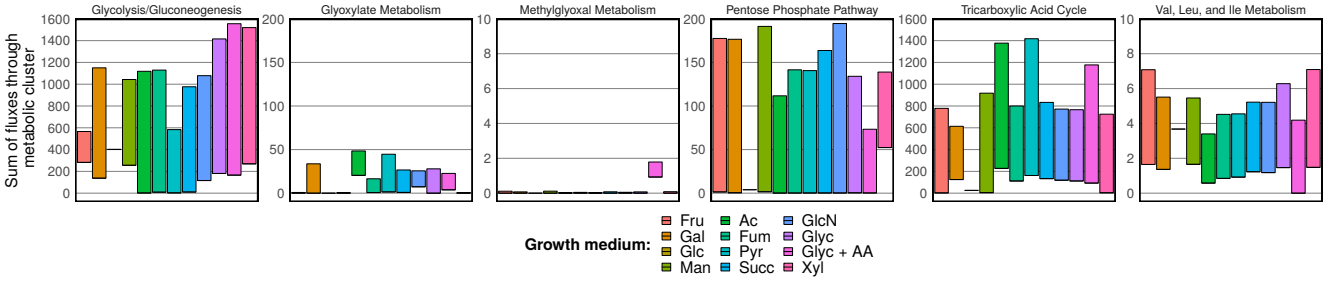

Fig. 4. Condition-dependent metabolic traits in representative clusters of the metabolic network.
Flux Variability Analysis (FVA) was used to compute flux intervals for all reactions in the model. The minimal and maximal values of the flux intervals acquired ($mmol(gDW)^{-1}h^{-1}$) were summed to yield the lower and upper sum of fluxes, respectively.

Figure Legends

Figure 1: Comparison of the sums of fluxes through representative clusters of the metabolic network for the genome-scale metabolic model iML1515 (A) and our extended genome-scale model (B).
For (A) and (B), Flux Variability Analysis (FVA) was used to compute flux intervals for all reactions in the model. Then the minimal and maximal values of the flux intervals (in $mmol(gDW)^{-1}h^{-1}$) were summed up to yield the lower and upper sum of fluxes, respectively.
Figure 2: Dependence of acetate production on the growth rate at different abundance of nutrients in MOPS media (A) and sums of fluxes through the TCA and pyruvate metabolism (B) reactions at given growth rates.
The initial media are MOPS -AA, MOPS -Met or MOPS +AA as indicated in (5). Then, the availability of carbon and nitrogen sources was step-wise changed by the same degree $0.0 < \gamma \leq 1.0$. Flux Variability Analysis (FVA) was used to compute flux intervals at every step for acetate exchange (A) or all reactions in the model (B). For (A), experimental data was obtained from (41–43). Black lines represent the linear fits of reported acetate excretion for each experimental dataset, starting at $\mu \geq 0.375\ h^{-1}$. For (B), minimal and maximal values of the flux intervals (in $mmol(gDW)^{-1}h^{-1}$) were summed up to yield the lower and upper sum of fluxes, respectively.
Figure 3: Cumulative distributions of variability intervals of metabolic fluxes for different growth conditions (A) and flux distributions through clusters of selected amino acid metabolism (B).
Flux Variability Analysis (FVA) was used to compute flux intervals at different growth settings for all reactions in the model. For (A), the flux intervals acquired were ranked by increasing size. First 5000 out of 5865 values were omitted, and the rest of flux variability intervals were plotted as a cumulative function with respect to the condition with the highest cumulative flux variability (MOPS +AA, with proteomics input). For (B), the minimal and maximal values of the flux intervals acquired ($mmol(gDW)^{-1}h^{-1}$) were summed to yield the lower and upper sum of fluxes, respectively.
Figure 4: Condition-dependent metabolic traits in representative clusters of the metabolic network.
Flux Variability Analysis (FVA) was used to compute flux intervals for all reactions in the model. The minimal and maximal values of the flux intervals acquired ($mmol(gDW)^{-1}h^{-1}$) were summed to yield the lower and upper sum of fluxes, respectively.

Supplementary Information

Supplementary Note 1. Description of the new model framework

A constraint-based model is a metabolic network, represented as stoichiometric matrix S with dimensionality of $m \times n$ for m molecular species and n reactions. Such a representation describes reaction fluxes through the metabolic network at a steady state.

The previously proposed approaches include strong influence of organism- (1) and growth condition-dependent coupling constraints (2) or do not account for protein turnover costs (3). Here, we propose a new approach which (i) preserves the linearity of the resulting problem and (ii) is detailed enough to describe protein costs and/or their abundance explicitly. This framework is expected to be highly flexible, so that distinct simulation scenarios (e.g. forced protein overexpression or metabolic transitions) do not require any adjustments in the model structure.

The model explicitly includes protein turnover costs via describing protein biosynthesis, folding and degradation in a detailed manner. Building on the GECKO approach (3), here we present one of the first attempts to extend GEMs to incorporate protein machinery costs without any condition-specific adjustments to the model structure. The structure of the previously proposed model frameworks were subject to variations depending on the experimental data, namely, the experimentally determined growth rate. Here, the experimental data is applied only for constraining the fluxes and does not present any changes to the structural properties of the model.

Starting metabolic network. Assuming a steady-state condition of the metabolic network:

$$S \times \mathbf{v} = 0 \tag{S1}$$

where S and $\mathbf{v}$ represent the stoichiometric matrix and metabolic flux vector, respectively. A linear optimization problem is acquired which subsequently is optimized with respect to a certain objective function. According to the nature of reactions (e.g. reversibility, kinetic data, experimental measurements etc.) the lower and upper flux values (bounds) are established for the reaction i to narrow down the available solution space:

$$LB \leq v_i \leq UB \tag{S2}$$

In this study, in order to further constrain the flux vector, we described the turnover of a proteome, consisting of N protein species. For this, as proposed in GECKO framework (3, 4), each enzyme E_i ($i = 1, 2, 3, ..., N$) is included in metabolic reactions with a stoichiometric coefficient n as any other metabolite species. In GECKO formalism, enzyme species are consumed in the reactions as substrates and are not retrieved as reaction products. As such a description violates the mass conservation of the enzyme species, we introduced another, in reality *non-existing*, type of metabolic species for proteins.

We called these species “used enzymes” (denoted by E_i^*), which are identical to the “normal” enzymes E_i in their amino acid composition and other parameters. In reality, functional enzymes are retrieved after reaction cycle, however, this is not possible to be described in stoichiometric models (see two contradicting examples in Eqs. S8-S10). Thus we introduced different metabolic species to couple metabolic fluxes to retrievable enzyme allocations (consumption and retrieval in reactions). Therefore, in the metabolic reactions, catalyzed by E_i, a "used" form of the enzyme E_i^* is obtained as a product:

$$A + B + n\ E_i \rightarrow C + D + n\ E_i^* \tag{S3}$$

In every row of the stoichiometric matrix the enzymatic species are represented with a single stoichiometric coefficient. Thus, all reversible reactions possessing gene-protein-reaction annotations (GPRs):

$$R_{reversible}:\ A + B \leftrightarrow C + D \tag{S4}$$

were split into two uni-directional reactions:

$$R_{fwd}:\ A + B + n\ E_i \rightarrow C + D + n\ E_i^*;\ 0 \leq v_{fwd} \leq UB \tag{S5}$$

$$R_{rev}:\ C + D + n\ E_i \rightarrow A + B + n\ E_i^*;\ 0 \leq v_{rev} \leq UB \tag{S6}$$

Also, the unused enzyme could be turned into its used form through a respective "futile usage" reaction instead:

$$E_i \rightarrow E_i^* \tag{S7}$$

The presence of "futile usage" reactions ensures that the steady state condition (Eq. S1) is not violated in the case when E_i does not catalyse any reactions in the metabolic network or when E_i is not used at its entirety.

Here, it should be noted that despite the enzymes, as any other catalysts, are not consumed during reactions, distinguishing between pre- and post-catalysis enzyme species is crucial (see below). **Note:** Some readers might want to skip the detailed explanation, which ends at the paragraph above (Eq. S11).
To show the contradiction between formalism and biochemical interpretation of enzyme-mediated catalysis, we consider a biochemical reaction which would retrieve a functional enzyme E_i as a product:

$$A + B + n\ E_i \rightarrow C + D + n\ E_i \tag{S8}$$

and noting down the stoichiometric coefficients $[-1; -1; +1; +1; [-n; +n]]$ in the stoichiometric matrix S for $[A,\ B,\ C,\ D,\ E_i]$ an ambiguity arises for the stoichiometric coefficient of E_i. However, only a single coefficient value is permitted for any column in a single row of the stoichiometric matrix. Thus, to overcome this limitation, splitting such a reaction into several steps could be considered. For instance, a reaction system could be formulated as follows: the substrates and the enzyme are combined into an intermediate substrate-enzyme complex Int, which releases the products and the functional enzyme in the second step:

$$A + B + n\ E_i \rightarrow Int \tag{S9}$$

$$Int \rightarrow C + D + n\ E_i \tag{S10}$$

At a first glance, this way a coupling between the metabolic network and protein machinery might be achieved. Under a steady state assumption, the fluxes through the reactions, depicted in (Eq. S9) and (Eq. S10) must be the same. Then the non-consumed metabolite E_i is essentially capable of participating in an indefinite quantity of fluxes given no limitation on other substrates. Thus, adopting such a construct yields a self-sustainable loop of the enzyme E_i flux. In other words, the metabolic flux becomes solely dependent on the abundance of substrates and not on the enzyme availability. As alternative formulations of catalytic cycles are not available, in the end, discrimination between a functional enzyme entity E_i and its "used" form E_i^* must be preserved. As argued above, this sort of notation ensures that the fluxes in the metabolic network become dependent on the protein machinery.
As the metabolic fluxes are now linked to the protein machinery, the requirement of E_i in metabolic reactions now can be controlled via altering the stoichiometric coefficient n in the reaction (Eq. S3). As suggested and derived in further detail in the GECKO framework (3), the coefficient n for the enzyme consumption in all the reactions catalysed by E_i is further assumed to be $(k_{cat,\ i})^{-1}$ for the enzyme E_i, which eventually brings the sum of all the fluxes through reactions which use enzyme E_i (denoted by R_i) to be:

$$\sum_{j \in R_i} v_j \leq k_{cat,\ i} \times [E_i] \tag{S11}$$

where $[E_i]$ is the concentration of enzyme in $mmol(gDW)^{-1}$. Since having enzyme molecules in the system is now a strict requirement, this simple principle offers a straightforward way to introduce the protein abundance as an additional set of constraints. Further, a description for the maintenance of protein homeostasis in the system is needed. Reactions, which govern the protein abundance will be constrained as follows: the lower flux bounds (the minimal requirement for protein abundance) for these reactions are established for each protein entity individually, according to experimental measurements of absolute protein abundance. The upper flux bound represents the limitation on total protein abundance and will be discussed below. In contrast to the GECKO framework, rather than computing the minimal proteome requirement to operate the metabolic network, here we explicitly take protein costs into account.

Explicit proteome constraint. The protein life cycle here is described in a detailed manner, including tRNA loading with amino acids, protein biosynthesis, folding and protease-catalyzed degradation of proteins. With the use of ATP, the latter of them yields free amino acids again. Amino acids a_j, $a \equiv$ [proteogenic amino acids], if not defined in the original model, are charged onto respective free tRNAs in the following reaction:

$$R_{a_j,\ loading} : a_j + tRNA_{a_j}^{un} + n\ E_{TRS,\ a_j} + ATP \rightarrow n\ E_{TRS,\ a_j}^* + tRNA_{a_j}^{ch} + AMP + PP_{in} \tag{S12}$$

where PP_{in} represents inorganic pyrophosphate. Then, the protein synthesis, folding and degradation $\forall E_i$, composed of m amino acids, is described as follows:

$$R_{synthesis,\ E_i} : \sum_{j \in a} n_j\ tRNA_{a_j}^{ch} + n\ Ribo + 4m\ ATP + 4m\ H_2O \rightarrow E_i^{uf} + \sum_{j \in a} n_j\ tRNA_{a_j}^{un} + n\ Ribo^* + 4m\ ADP + 4m\ P_{in} \tag{S13}$$

$$R_{folding,\ E_i} : E_i^{uf} + n\ Chap + 14\ ATP + 14\ H_2O \rightarrow E_i + n\ Chap^* + 14\ ADP + 14\ P_{in} \tag{S14}$$

$$R_{degradation,\ E_i^*} : E_i^* + n\ Prot + 6\ ATP + 6\ H_2O \rightarrow \sum_{j \in a} n_j a_j + n\ Prot^* + 6\ ADP + 6\ P_{in} \tag{S15}$$

where E_{TRS,a_j}, $tRNA_{a_j}^{un}$, $tRNA_{a_j}^{ch}$ and P_{in} denote the amino acid-specific acyl-tRNA synthases, uncharged and charged tRNAs for the amino acid a_j (with the number n_j of amino acid a_j in the peptide chain), and the inorganic phosphate, respectively; $Ribo$, $Chap$ and $Prot$ represent the ribosome, a generic chaperone and a generic protease, respectively; E_i^{uf} is the unfolded (also catalytically inactive) form of the protein E_i and m is the total number of amino acids in peptide chain.
The macromolecular complexes ($Ribo$, $Chap$ and $Prot$) are entering the reactions as distinctive model species and are retrieved as the "used" form of the complex. Thus for the macromolecular complexes we also have to define the reactions of complex (dis-)assembly, with stoichiometric coefficient n_j for every subunit of a protein species j in the macromolecular complex:

$$R_{assembly,\ Ribo}: \sum_{j \in Ribo} n_j E_j \rightarrow Ribo \tag{S16}$$

$$R_{disassembly,\ Ribo}: Ribo^* \rightarrow \sum_{j \in Ribo} n_j E_j^* \tag{S17}$$

The same reactions are defined for $Chap$ and $Prot$ (with $j \in Chap$ and $j \in Prot$, respectively).
Finally, the proteomic measurements are incorporated into the protein synthesis reaction to yield a lower constraint of:

$$v_{synthesis,\ E_i} \geq 0.9 \times [E_i] \tag{S18}$$

The life cycle of the macromolecular complexes ($Ribo$, $Chap$ and $Prot$), is defined as a combination of the protein subunits with the respective stoichiometric coefficients into a "bulk" complex molecule, which releases stand-alone and degradation-susceptible protein subunits after the reaction cycle.
In *E. coli* there are three main chaperone systems: (i) the trigger factor, (ii) $Hsp70$ and (iii) $GroE$. The trigger factor is a ribosome-bound particle which facilitates the formation of the tertiary structure of the nascent protein chain. Thus, instead of an extensive description of the chaperone machinery (5), here we implicitly accounted for the trigger factor-mediated folding. The requirement for the abundance of the trigger factor was introduced via inclusion of the trigger factor with $n = 1$ into the reactions of $Hsp70$ and $GroE$ complex formation.
We assumed a requirement of 7 folding cycles of $Hsp70$ or 1 folding cycle of $GroE$ per unfolded protein, resulting in the folding-associated energetic cost of 14 ATP per protein molecule. The $GroE$ complex of *E. coli* is assembled from single heptameric $groS$ ring (no ATPase activity, UniProtKB entry $P0A6F9$) and two heptameric $groL$ rings (possessing ATPase activity, UniProtKB entry $P0A6F5$). As each of the ATPase subunits hydrolyse 1 ATP per cycle, the demand of single protein folding cycle comes to 14 ATP. Meanwhile, the protein folding by the Hsp70 system, consisting of $dnaK - dnaJ - grpE$ (UniProtKB $P0A6Y8$, $P08622$ and $P09372$, respectively) requires 2 ATP molecules to be hydrolysed by $dnaK$ per folding cycle. As multiple folding cycles of $Hsp70$ are required, we assumed 7 folding cycles per protein to arrive at the same energetic cost (as for $GroE$) of 14 ATP per protein folded.
Moreover, the $GroE$ complex is thought to be able to accommodate only molecules with molecular weight $(M_w) \leq 70\ kDa$, due to its barrel-like tertiary structure (6). In the model, the bounds of $GroE$-mediated folding reactions for these protein entities were set to $[0;0]$ in order to block the $GroE$ from folding proteins with higher molecular weight.
Finally, the k_{cat} values for the protein synthesis, folding and degradation processes were assumed. The k_{cat} for folding processes was set to an arbitrary value of $k_{cat} = 1\ s^{-1}$. Meanwhile, for Lon- and Clp protease-mediated protein degradation, the turnover values k_{cat} were set to $k_{cat,\ Lon} = 0.25\ s^{-1}$ and $k_{cat,\ Clp} = 2\ s^{-1}$, respectively (according to the data of (14)). Meanwhile, the requirement of ribosomes for the production of E_i was calculated for every protein according to its amino acid chain length m, assuming a maximal rate of peptide chain elongation of $k_{cat} = \frac{20}{m}\ s^{-1}$ for *E. coli* (data from (7) and (15)).

Constraining the total protein abundance. Cells possess a finite volume. Therefore, when accounting for biological macromolecules in genome-scale models, it is important to consider a limitation of the physical cell volume which can be occupied by proteins. Based on the FBAwMC framework (8, 9), we formulated a constraint of the total cell volume, available for proteins in the genome-scale models. Similar to FBAwMC, the proteome volume is defined:

$$\sum_i V_i \times [E_i] \leq \frac{1}{C} \tag{S19}$$

where V_i denotes the molar volume of enzyme E_i and $C = 0.34\ g(mL)^{-1}$ is the cytosolic protein density of the *E. coli* (10). Thus, the right hand side of the inequality provides the upper value for the available proteome volume. Further, by assuming an empirical relation $b = 0.73\ mL(g)^{-1}$ between the molar weight and volume of the globular protein molecules, V_i is set to be:

$$V_i = b \times M_{w,\ i} \tag{S20}$$

where $M_{w,\ i}$ is the molar weight of a protein E_i (in $g/mmol$). Substituting (Eq. S20) into (Eq. S19) yields the following:

$$b \times \sum_i \times M_{w,\ i} \times [E_i] \leq \frac{1}{C} \tag{S21}$$

However, the right-hand side of the inequality still accounts for the wet weight (g) of cells. As C is the protein density in the cell wet weight, the protein amount must be scaled to the gDW in order to comply with the rest of the units in the model. For this, an appropriate relation of protein amount in dry weight of cells could be employed to convert the units. As reported in (11), an estimate of $f_P = 0.62\ g(gDW)^{-1}$ of proteins per one gDW of *E. coli* was used. Thus, the final notation of the total protein constraint is:

$$b \times \sum_i M_{w,\,i} \times [E_i] \ \leq \frac{f_P}{C} \tag{S22}$$

To use the constraint in the model, first, we considered the balance of the respective protein maintenance fluxes $\forall E_i$ under a steady state condition:

$$v_{synthesis,\,E_i} = v_{folding,\,E_i} = v_{futile\ usage,\,E_i} + \sum_{j \in R_i} v_j = v_{degradation,\,E_i^*} \tag{S23}$$

where j is a set of metabolic reactions, consuming E_i. Thus, under a steady state assumption, the constraint could be applied to any of the expressions in the equalities. However, for the sake of clarity, we decided to constrain the fluxes of protein synthesis reactions, bringing the expression of (Eq. S23) to:

$$b \times \sum_i M_{w,\,i} \times v_{synthesis,\,E_i} \leq \frac{f_P}{C} \tag{S24}$$

Proteomics data used for constraining the models. To constrain the models via introducing the minimal protein requirement in (Eq. S18), measurements of absolute protein abundances are needed. In this study, two datasets of condition-dependent quantification of *E. coli* proteome were considered (12, 13). In the dataset reported by Li *et al.* (5), ribosome profiling was employed to analyse proteome distribution in three variants of glucose-rich MOPS medium: the minimal, with a supplement of 19 proteinogenic amino acids (w/o methionine), and with a full supplement. In the case of Schmidt *et al.* (13), the bacteria were grown under 19 different growth conditions, namely, 4 chemostat cultures with medium dilution rates μ spanning $0.12 - 0.5\ h^{-1}$ and 15 batch culture conditions. These 15 cultures included 12 different carbon sources, and three additional cultures with environmental stresses, applied to *E. coli* grown in a glucose-containing medium (heat shock, osmotic shock and acid stress [$pH = 6$]).
The composition of all the media was reported in the studies and were used to constraint the nutrient availability. The concentrations in $mmol(L)^{-1}$ of the components in a chemically defined media were used to constrain the exchange reactions in the model. The data from these experimental studies were used to establish the minimal requirement of protein abundance and constraint the nutrient availability in the media. A model set-up was considered "condition-specific" iff both of these sets of constraints are applied simultaneously.

Supplementary References

1. J. A. Lerman, D. R. Hyduke, H. Latif, V. A. Portnoy, N. E. Lewis, J. D. Orth, A. C. Schrimpe-Rutledge, R. D. Smith, J. N. Adkins, K. Zengler, and B. O. Palsson, "In silico method for modelling metabolism and gene product expression at genome scale," *Nature Communications*, vol. 3, no. May, pp. 910–929, 2012.
2. E. J. O{middle dot}Brien, J. A. Lerman, R. L. Chang, D. R. Hyduke, and B. O. Palsson, "Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction," *Molecular Systems Biology*, vol. 9, pp. 693–693, apr 2014.
3. B. J. Sanchez, C. Zhang, A. Nilsson, P. Lahtvee, E. J. Kerkhoven, and J. Nielsen, "Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints," *Molecular Systems Biology*, 2017.
4. D. Machado, M. J. Herrgård, and I. Rocha, "Stoichiometric Representation of Gene–Protein–Reaction Associations Leverages Constraint-Based Analysis from Reaction to Gene-Level Phenotype Prediction," *PLOS Computational Biology*, vol. 12, p. e1005140, oct 2016.
5. K. Chen, Y. Gao, N. Mih, E. J. O'Brien, L. Yang, and B. O. Palsson, "Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation," *Proceedings of the National Academy of Sciences*, vol. 114, pp. 11548–11553, oct 2017.
6. M. Hayer-Hartl, A. Bracher, and F. U. Hartl, "The GroEL-GroES Chaperonin Machine: A Nano-Cage for Protein Folding," 2016.
7. T. V. Karpinets, D. J. Greenwood, C. E. Sams, and J. T. Ammons, "RNA: Protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis," *BMC Biology*, 2006.
8. Q. K. Beg, A. Vazquez, J. Ernst, M. A. de Menezes, Z. Bar-Joseph, A.-L. Barabasi, and Z. N. Oltvai, "Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity," *Proceedings of the National Academy of Sciences*, vol. 104, pp. 12663–12668, jul 2007.
9. A. Vazquez, Q. K. Beg, M. A. DeMenezes, J. Ernst, Z. Bar-Joseph, A. L. Barabasi, L. G. Boros, and Z. N. Oltvai, "Impact of the solvent capacity constraint on E. coli metabolism," *BMC Systems Biology*, 2008.
10. D. Ridgway, G. Broderick, A. Lopez-Campistrous, M. Ru'Aini, P. Winter, M. Hamilton, P. Boulanger, A. Kovalenko, and M. J. Ellison, "Coarse-grained molecular simulation of diffusion and reaction kinetics in a crowded virtual cytoplasm," *Biophysical Journal*, 2008.
11. S. B. Zimmerman and S. O. Trach, "Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli," *Journal of Molecular Biology*, 1991.
12. G. W. Li, D. Burkhardt, C. Gross, and J. S. Weissman, "Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources," *Cell*, vol. 157, no. 3, pp. 624–635, 2014.
13. A. Schmidt, K. Kochanowski, S. Vedelaar, E. Ahrné, B. Volkmer, L. Callipo, K. Knoops, M. Bauer, R. Aebersold, and M. Heinemann, "The quantitative and condition-dependent Escherichia coli proteome," *Nature Biotechnology*, vol. 34, pp. 104–110, jan 2016.
14. M. R. Maurizi. Proteases and protein degradation in Escherichia coli. *Experientia*, ISSN 00144754, doi:10.1007/BF01923511, 1992.
15. M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Science*, ISSN 00368075, doi:10.1126/science.1192588, 2010.

Supplementary Tables and Figures

Table S1 (available as a separate .xls document): Composition of predicted core proteomes of *E. coli* in extended model and iML1515 for different growth conditions (MOPS -AA, MOPS -Met and MOPS +AA).

Figure S1: Computed flux variability through protein synthesis reactions in extended model. Flux Variability Analysis (FVA) was used to compute flux intervals for all reactions in the model, and the variability intervals were defined as the difference between the maximal and the minimal flux values.

Figure S2: Comparison between the experimentally determined abundance of proteins and the model predictions.

Figure S3: Comparison of proteomics data integration-dependent flux variability in 10 metabolic clusters of amino acid metabolism. Flux Variability Analysis (FVA) was used to compute flux intervals for all reactions in the model, and the minimal and maximal values of the flux intervals acquired ($mmol(gDW)^{-1}h^{-1}$) were summed to yield the lower and upper sum of fluxes, respectively.

Figure S4: Growth condition-dependent variability in 10 metabolic clusters of amino acid metabolism. Flux Variability Analysis (FVA) was used to compute flux intervals for all reactions in the model, and the minimal and maximal values of the flux intervals acquired ($mmol(gDW)^{-1}h^{-1}$) were summed to yield the lower and upper sum of fluxes, respectively.

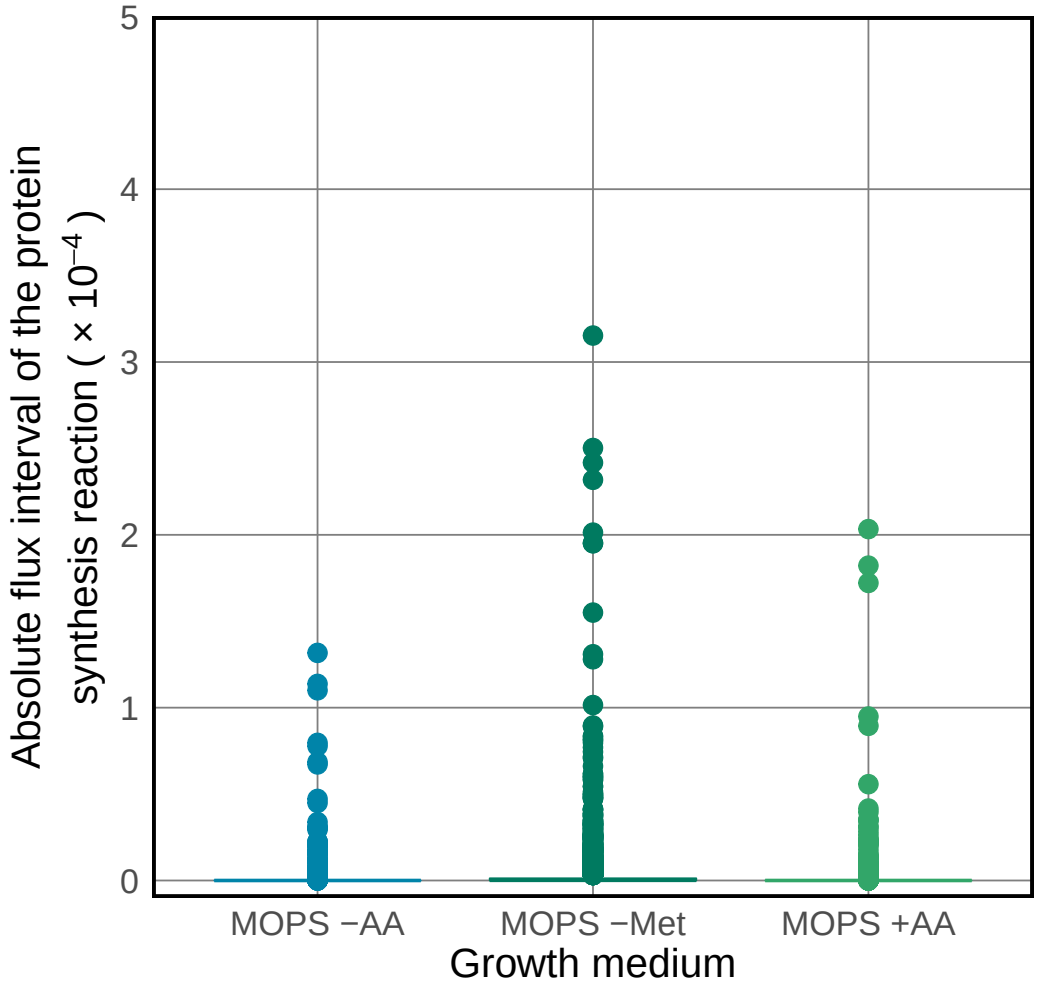
Absolute flux interval of the protein synthesis reaction (× 10−4)
5
4
3
2
1
0
MOPS −AA
MOPS −Met
MOPS +AA
Growth medium

Fig. S1

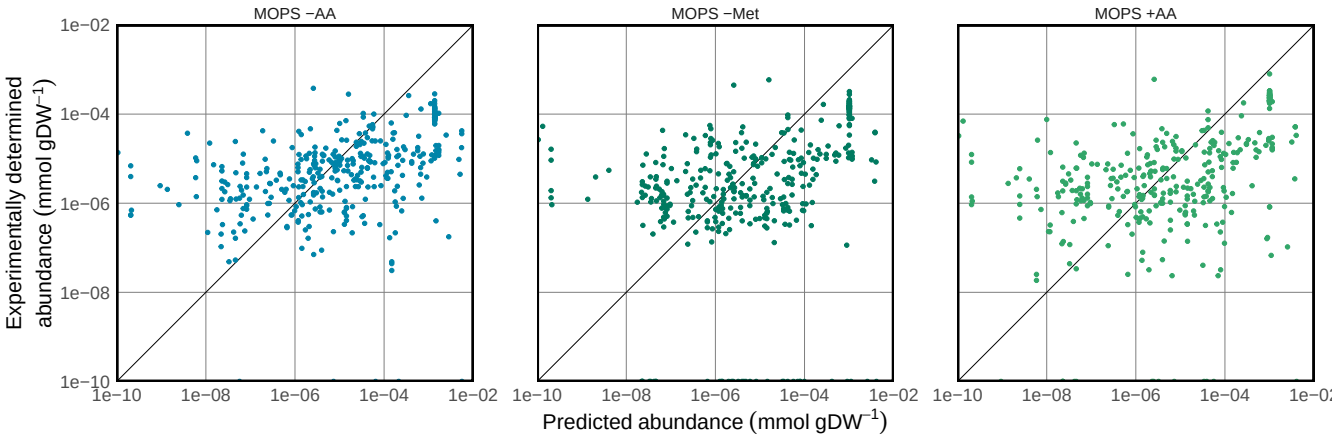
MOPS −AA
MOPS −Met
MOPS +AA
Experimentally determined abundance (mmol gDW^{-1})
Predicted abundance (mmol gDW^{-1})
1e−02
1e−04
1e−06
1e−08
1e−10

Fig. S2

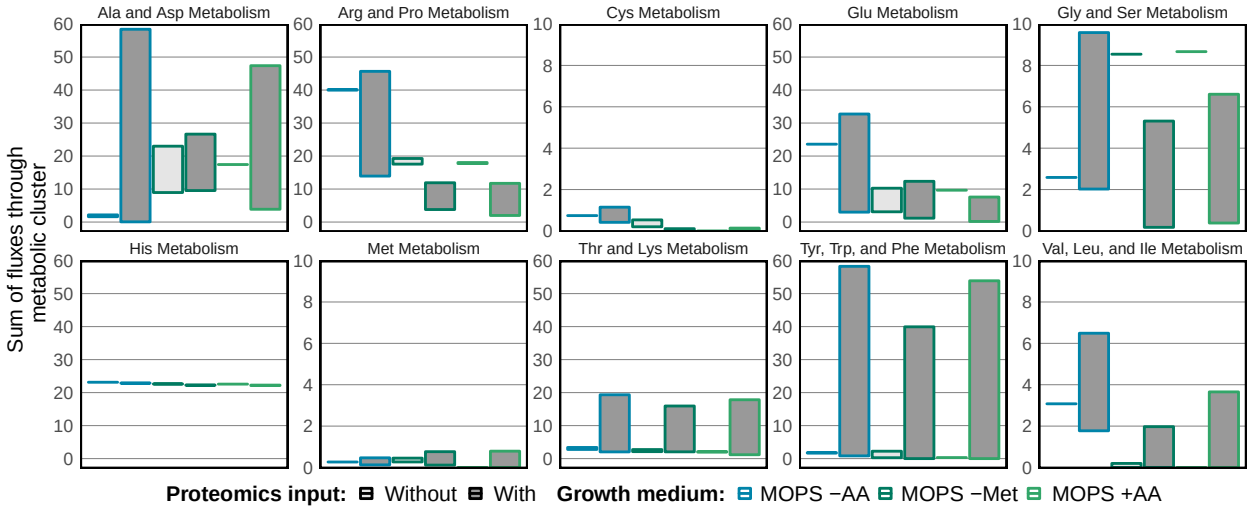
Ala and Asp Metabolism
Arg and Pro Metabolism
Cys Metabolism
Glu Metabolism
Gly and Ser Metabolism
His Metabolism
Met Metabolism
Thr and Lys Metabolism
Tyr, Trp, and Phe Metabolism
Val, Leu, and Ile Metabolism
Sum of fluxes through metabolic cluster
Proteomics input: Without With
Growth medium: MOPS −AA MOPS −Met MOPS +AA

Fig. S3

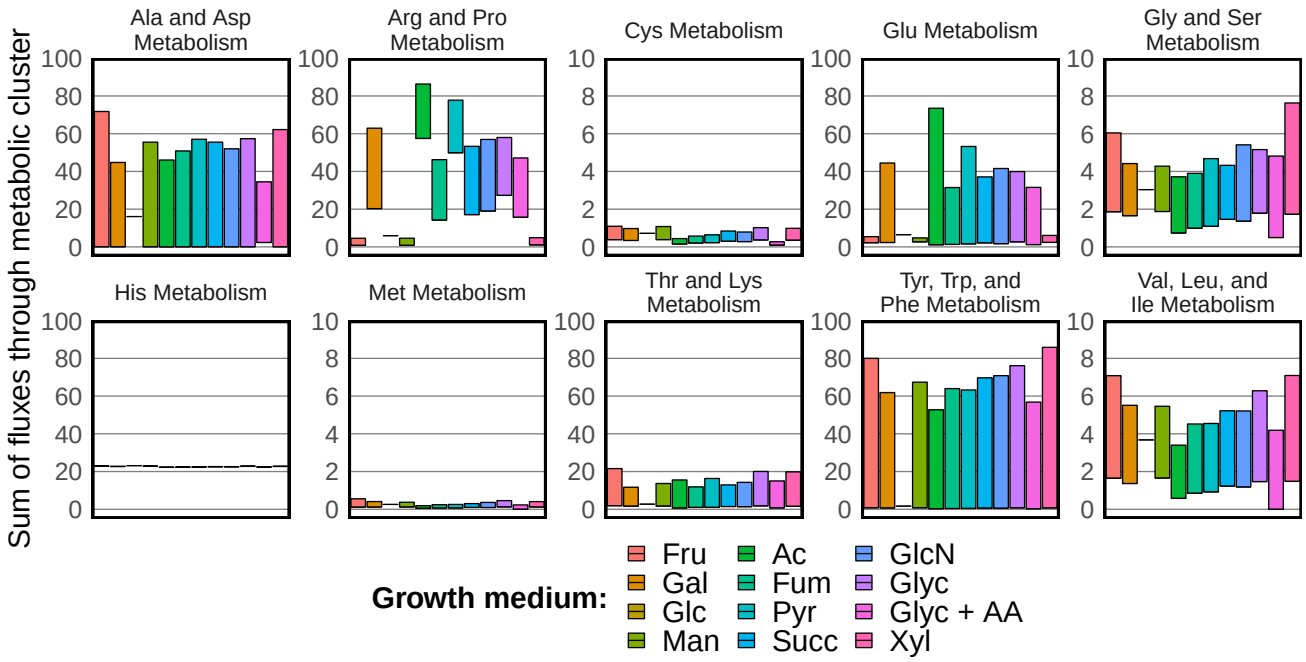
Ala and Asp Metabolism
Arg and Pro Metabolism
Cys Metabolism
Glu Metabolism
Gly and Ser Metabolism
His Metabolism
Met Metabolism
Thr and Lys Metabolism
Tyr, Trp, and Phe Metabolism
Val, Leu, and Ile Metabolism
Sum of fluxes through metabolic cluster
Growth medium:
Fru
Gal
Glc
Man
Ac
Fum
Pyr
Succ
GlcN
Glyc
Glyc + AA
Xyl

Fig. S4