Genome-scale architecture of small molecule regulatory networks and the fundamental trade-off between regulation and enzymatic activity

Ed Reznik[1,*,†], Dimitris Christodoulou[2,3,*], Joshua E. Goldford[4], Emma Briars[4], Uwe Sauer[2], Daniel Segrè[4,5,6], Elad Noor[2,*,†,^]

[1] Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY

[2] Institute of Molecular Systems Biology, ETH Zurich, Zurich, Switzerland

[3] Systems Biology Graduate School, Zurich 8057, Switzerland

[4] Bioinformatics Programs, Boston University, Boston MA

[5] Department of Biology, Boston University, Boston MA

[6] Department of Biomedical Engineering, Boston University, Boston MA

^ lead contact

* equal contribution

† corresponding authors (reznike@mskcc.org, noor@imsb.biol.ethz.ch)

Abstract

Metabolic flux is in part regulated by endogenous small molecules that modulate the catalytic activity of an enzyme, *e.g.* allosteric inhibition. In contrast to transcriptional regulation of enzymes, technical difficulties have hindered the production of a genome-scale atlas of small molecule/enzyme regulatory interactions. Here, we develop a framework leveraging the vast, but fragmented, biochemical literature to reconstruct and analyze the small molecule regulatory network (SMRN) of the model organism *Escherichia coli*, including the primary metabolite regulators and enzyme targets. Using metabolic control analysis, we prove a fundamental trade-off between regulation and enzymatic activity, and combine it with metabolomic measurements and the SMRN to make inferences on the sensitivity of enzymes to their regulators. Generalizing the analysis to other organisms, we identify highly conserved regulatory interactions across evolutionarily divergent species, further emphasizing a critical role for small molecule interactions in the maintenance of metabolic homeostasis.

Introduction

Despite nearly a century of accumulated experimental data on the identity, biophysical nature, and structural basis of small molecule regulation of enzymes, there has been little progress in elucidating at genome-scale the regulation of enzymes by small molecules. Such interactions, *e.g.* allosteric regulation or competitive inhibition, play an essential role in homeostasis and in fast adaptations to abrupt environmental changes. For instance, feedback inhibition of amino-acid biosynthesis pathways by end-products preserves anabolic resources when sufficient levels of amino acids are available [1]. In contrast, the ultrasensitive feed-forward activation of PEP carboxylase by fructose bisphosphate in glycolysis enables *Escherichia coli* to rapidly import glucose following periods of carbon starvation [2].

Recently, it has been established that properly accounting for the activation/inhibition of enzymes by endogenous small molecules can lead to metabolic models that explain experimental data better [2]–[5], facilitate engineering of novel metabolic pathways [6], [7], and improve our understanding of metabolic phenomena in health and disease [8]. So far, high-throughput experimental assays for discovering small molecule regulatory interactions have been technically limited [9]–[14], while hybrid approaches that integrate experimental data with computational models are not scalable and typically focus on central carbon metabolism [3], [4], [15], [16]

An alternative strategy for studying small molecule regulation is to leverage the vast record of biochemical studies to informatically reconstruct a genome-scale small molecule regulatory network (SMRN). Such an approach would produce a network of interactions between enzymes and metabolites/small molecules (terms we will use interchangeably) that mirrors the native interactions of metabolites as substrates for enzymes, and could be naturally integrated with genome-scale metabolic models (GSMMs, *e.g.* BIGG models [17]) in wide use today. An informatic approach would likely cover a larger swath of metabolism, including peripheral and rarely-studied pathways, by aggregating experimental data from many reports. Furthermore, independent reports of small molecule/enzyme regulation from separate publications could provide additional, independent evidence for such an interaction. Finally, an informatically reconstructed SMRN would also offer a window to distilling how the critical regulatory components of the cell, i.e. the regulating metabolites and the regulated enzymes, fit into the broader hierarchy of processes controlling metabolic flux, from thermodynamics to transcriptional regulation.

Here, we report the development of a computational framework to investigating small molecule regulation across the complete metabolism of an organism. Using *E. coli* as a model, we assemble the first genome-scale SMRN by mining the BRENDA [18] and BioCyc [19] databases. The resulting atlas of small molecule regulation captures widespread inhibition and activation of metabolic enzymes by endogenous metabolites. Overlaying this network onto a genome-scale metabolic model of *E. coli* enables a direct comparison between the topology of metabolism and its regulatory scaffolding. Integrating the SMRN with experimentally determined metabolite concentrations and binding affinities exposes how cells balance between the dual roles of small molecules (i.e. as substrates or inhibitors) as well as their condition-dependent contribution to metabolic flux regulation. Finally, by a natural extension of our approach, we compare the incidence of small molecule regulatory interactions across phylogenetic taxa spanning all kingdoms of life, revealing a handful of canonical regulatory interactions that permeate the metabolism of widely divergent species.

Results

Assembling a Small Molecule Regulatory Network (SMRN)

In contrast to the proliferation [20], [21] and automated reconstruction [22] of genome-scale metabolic models, no analogous computational pipeline is available for constructing and analyzing the network of enzyme regulation by small molecules. Here, we report a framework that includes a pipeline for mining existing data on small molecule regulation from public repositories and computational tools for integrating it with a curated, genome-scale metabolic model (Figure 1).

Our general approach relies on the BRENDA and BioCyc databases for data concerning the inhibition or activation of enzymes by small molecules in a particular organism. To facilitate comparison across species, information in these databases is organized along Enzyme Commission (EC) numbers, which

functionally classifies enzymes according to the reaction they catalyze. Thus, for every EC number, we obtained a list of possible regulating small molecules, the type of interaction (activation vs. inhibition) and the interaction constant (K_I), if available. In addition to gathering data on the presence of a small molecule interaction, we also compiled the available information on the Michaelis-Menten constants for the substrates of each metabolic reaction. The computational pipeline is freely available for download at http://github.com/eladnoor/small molecule-regulation.

Because of its well-defined genome, highly curated metabolic network, and heavily studied metabolism, we decided to focus on the model bacterium *Escherichia. coli*. The assembled data described above was mapped onto a genome scale metabolic reconstruction of *E. coli* [20], producing a small molecule regulatory network (SMRN, Supplementary Material - table S1). Importantly, our computational framework can be extended with minimal effort to reconstruct and analyze the SMRN of other organisms besides *E. coli*, provided that a genome scale metabolic model is available and that there is sufficient data in the BRENDA and BioCyc databases.

Landscape of Interactions in the *E. coli* SMRN

The computationally reconstructed *E. coli* SMRN contains 1669 unique interactions between 321 unique endogenous metabolites and 364 unique enzymes (EC numbers) (SI Table 1). The vast majority (83%) of these interactions are inhibitory. Out of the ~700 unique EC numbers in the *E. coli* model, about half are regulated by at least one native metabolite in *E. coli*. Similarly, ~320 distinct native metabolites (out of the total ~1000) regulate at least one enzyme (Figure 2a,b). Figure 1 provides a tally of interactions and kinetic constants recovered from the BRENDA and BioCyc databases.

Certain metabolites and EC numbers participated in exceptionally high numbers of regulatory interactions (Figure 2c,d). In particular, the cofactors ATP, AMP, ADP, PI, PPI, NADPH, and GTP together with the metabolites cysteine, pyruvate, and phosphoenolpyruvate (PEP) were the most frequent metabolite regulators, participating in at least 15 interactions (both inhibitory and activating). Notably, ATP was found to regulate 57 different reactions, reflecting its important role as a global reporter of the energetic status of the cell [23]. Although not strictly molecules, metal ions were found to comprise a significant fraction of the group of small molecule regulators. Interestingly, potassium was the most frequently reported activator, likely reflecting the ability of monovalent cations to activate a broad group of enzymes [24]. Divalent cations, on the other hand, are among the most recurrent inhibitors, with zinc having more than 50 reported interactions and copper, calcium, manganese, mercury, iron, and magnesium each reported to inhibit more than 20 different reactions.

Small molecules can regulate enzymes for which they are not native substrates or products, potentially as a mechanism for long-distance signalling between metabolic subnetworks that have no direct connection via their reactants. Therefore, we investigated the distance covered by small molecule regulation (*i.e.* the shortest graph-distance between each regulatory metabolite and its targeted enzyme), using the genome-scale stoichiometric network of *E. coli* as a scaffold. As described in the Methods, this distance corresponds to the number of reactions a metabolite must traverse in order to reach the target enzyme of interest. We found that enzymes are typically regulated by metabolites that are in their close vicinity. Specifically, 17% of inhibitory interactions are also reactants of the corresponding enzyme, 25% are only one enzymatic step away, and 35% are two steps away. Activating interactions tend to have slightly longer range interactions with only 8% of activators regulating enzymes that utilize them as substrates (Supplementary Figure S1).

To obtain specific insight on how small molecule – enzyme interactions operate across different metabolic pathways, we used the genome-scale *E. coli* metabolic model [25] to classify each reaction to functional metabolic subsystems. We found that most interactions in the *E. coli* SMRN seem to target seven main subsystems/pathways: cofactor biosynthesis, alternate carbon metabolism, nucleotide salvage pathway, arginine/proline metabolism, nucleotide biosynthesis, cell envelope biosynthesis, and glycolysis/gluconeogenesis (Supplementary Figure S3). Interestingly, some canonically high-flux pathways (e.g. the TCA cycle and the pentose phosphate pathway) were regulated by comparatively few metabolites, and the quantity of total regulators in these pathways was substantially lower than in the seven highly-regulated pathways listed above.

Further mining of this rich dataset revealed pathway-specific preferences for the regulatory targets of certain small molecules. In particular, PEP, citrate, and AMP each activated 3 or more reactions in the glycolysis/gluconeogenesis pathway, suggesting that these three metabolites act as critical sensors controlling the overall rate and direction of glucose metabolism. Similarly, a group of nucleoside triphosphates, deoxynucleoside triphosphates, and adenosine appear to specifically inhibit enzymes in three pathways related to nucleotide metabolism (nucleotide salvage, purine/pyrimidine biosynthesis, and prosthetic group biosynthesis), perhaps reflecting a negative feedback loop for the maintenance of adequate levels of various nucleotides (Supplementary Figure S3).

Design Principles in the Regulation of Central Carbon Metabolism

Central carbon metabolism, encompassing glycolysis, the pentose phosphate pathway, and the TCA cycle, provides all the energetic and biosynthetic precursors for the cell, and is known to be highly transcriptionally, post-translationally, and allosterically regulated [26]. The central metabolism of *E. coli* is also one of the few parts of metabolism where *in vivo* evidence is available to support the functional role of small molecule regulation, *e.g.* in order to induce flux reversal [3]. *In silico* efforts to model the response of the central metabolism to nutrient perturbations have highlighted the fact that our understanding of the intricate regulation of central metabolism is incomplete [2]–[4], [27], [28].

The majority of enzymes in *E. coli*'s central carbon metabolism are regulated (Figure 3, Supplementary Figure S9), and interact with more small molecules than average in the SMRN (Supplementary Figure S2), reflecting the heavy research attention these pathways have historically attracted. Interestingly, some of the enzymes in central metabolism are very heavily regulated - specifically those in upper glycolysis (*e.g. fbp, pfk*, and *fba*), terminal glycolysis (*pck, ppc, pps,* and *pyk*), and the branching reactions of the TCA cycle (*mae*, *aceA,* and *icd*). Conversely, some metabolites seem to have a more central role in certain regions of central metabolism; PEP, for instance, regulates six reactions in glycolysis (*pfk, pgi, fbp, fba, pps, pyk*).

A glance at the structure of small molecule regulation in *E. coli's* central metabolism strongly suggests that the distribution of regulatory interactions is non-random, and has likely been shaped by evolutionary selection. What are the selective pressures selecting for regulatory interactions in *E. coli*'s SMRN? In this regard, the theoretical and experimental literature has proposed a variety of thermodynamic and economic arguments to explain patterns of SMR interactions in central carbon metabolism. Below, we evaluate the consistency of each hypothesis with data from the *E. coli* SMRN.

One frequently cited hypothesis is that small molecule regulation is concentrated in those reactions exhibiting a large drop in free energy [1]. To evaluate this possibility, we acquired thermodynamic data for most metabolic reactions using the component contribution method [29]. Using reactions' ΔG° together

with reaction stoichiometry and standard physiological metabolite concentrations of substrates or products, we calculated a reversibility index (denoted Γ) quantifying the extent to which each reaction is thermodynamically reversible [30]. We found that distributions of Γ values for regulated and not regulated reactions in central metabolism were not statistically different (Figure 4b). The same result was observed when repeating the analysis on all available reactions in *E. coli* (Figure 4a). We found that some reactions, like 6-phosphogluconolactonase in the PP pathway (EC 3.1.1.31), exhibit a large drop in free energy (Γ=9.6), and do not have any reported small molecule interaction. On the other hand, many reversible reactions have several regulators, as in the case of succinyl-CoA synthetase (EC 6.2.1.5, Γ=0.22), which is inhibited by NADH and alpha-ketoglutarate. Thus, our data does not support the hypothesis that thermodynamic reversibility is a determinant for the likelihood that a reaction is regulated by a small molecule.

A second proposed principle underlying the incidence of SMR interactions is that cells use these regulatory interactions to conserve precious metabolic resources, *e.g.* by avoiding wasteful biosynthetic overproduction and preventing futile cycling. We observe several examples of interactions in the *E. coli* SMRN supporting these design principles. For example, *fbp* and *pfk* catalyze opposing reactions, and their simultaneous operation would lead to futile cycling [31]. Two metabolites (citrate and PEP) serve as activators for *fbp* and inhibitors for *pfk*, thus curbing this futile cycle. A similar regulatory architecture can be found in the regulation of the four reactions metabolizing PEP: acetyl-CoA is an activator of *pck* but an inhibitor of the reverse reaction catalyzed by *ppc.* Similarly, PEP itself activates pyruvate kinase but inhibits the reverse reaction PEP synthase. Moving beyond anecdotal observations, we used a genome scale metabolic model (iJO1366) to detect reactions that can lead to futile cycling. Combining this information with our small molecule regulatory network, we show that there is no statistically significant overrepresentation of small molecule regulation in reactions that take part in futile cycling (Supplementary Text S1).

Apart from prevention of futile cycling, SMR interactions can also prevent large fluctuations/instability in the concentrations of biosynthetic end products via feedback inhibition. In fact, the majority of precursors for synthesizing cellular biomass are provided by central metabolism, in a process that is essential for growth and at the same time requires a substantial investment of metabolic resources. It is, therefore, often hypothesized that allosteric regulation of the branching reactions from central metabolism leading to amino acid biosynthesis may prevent unnecessary waste. Interactions in the *E. coli* SMRN are consistent with this hypothesis: analysis of the SMRN reveals that 16 out of the 20 amino acids regulate their own biosynthesis using a negative feedback loop - i.e. by inhibiting the first enzyme of their biosynthetic pathway (Supplementary Figure S8). The remaining non-feedback-inhibitory amino acids (glycine, alanine, aspartate, and glutamate) are 4 of the 5 cheapest ones in terms of the energetic investment required to produce them [32], [33].

In summary, we find that one of the common (thermodynamic) hypotheses regarding the incidence of small molecule regulation is not supported by data from the *E. coli* SMRN. We do, however, find compelling support for an economic role of SMR interactions via feedback inhibition of biosynthesis pathways, at least for 16 out of the 20 amino acids. The four exceptions to this rule, however, might prove to be interesting cases where having a regulatory interaction is more costly than the benefit it provides to the cell. In the following sections, we elaborate on this cost, and quantify it using metabolic control analysis.

How is a Metabolite's Role as Regulator and Substrate Balanced?

Small molecule metabolites serve two fundamentally distinct roles in the cell: one, as substrates for metabolic reactions, and another, as regulatory molecules affecting the activity of enzymes and transcription factors [34]. How does the cell balance between these two responsibilities, especially in bacteria that have no intracellular compartments that could offer spatial separation? More specifically, are the cellular concentrations of these metabolites and the affinities of their interactions with the different enzymes in *E. coli* tuned such that they can inhibit some reactions while efficiently serving as substrates for others? To address this question, we gathered all reported Michaelis-Menten constants (K_M) and inhibitory half-saturation constants (K_I) (activation constants were not available in BRENDA). We decided to use these K_M and K_I values as approximate indicators of a small molecule's "metabolic operating point" and "regulatory operating point", respectively.

We first evaluated whether K_M and K_I values were quantitatively similar to each other (K_M values, or binding constants associated with activating interactions, were not available in BRENDA and were thus excluded). As reported by others [35], in general, K_I values tend to be higher than K_M values (Figure 4d, Mann-Whitney U test p-value < 0.005). A more informative approach to evaluating differences in K_M and K_I values is by a direct comparison with physiological metabolite concentrations. For example, if a metabolite's concentration is much higher than all of its associated K_M and K_I values, then all interactions related to this metabolite are approximately fully saturated, and any differences between K_M and K_I are not physiologically meaningful. Here, we quantify the level of saturation using the formula $\frac{s}{s + K_S}$, where K_S is a binding constant (either K_M or K_I) and S is the concentration of the substrate or inhibitor. For a substrate, the saturation represents the relative activity of an enzyme compared to a case where S is very high and the enzyme is fully activated (assuming all other parameters are kept constant). We use this definition to identify physiologically relevant differences between substrate and inhibitor affinities.

To apply the notion of saturation to our data, we obtained previously published metabolite concentrations in exponentially growing *E. coli* cultures on 13 different carbon sources [36]. For each unique binding constant/metabolite concentration pair, we calculated a saturation (1604 enzyme-inhibitor-condition triplets and 402 enzyme-substrate-condition triplets) (Figure 4e). A comparison between substrate and inhibitor saturation levels yielded a significant difference in the saturation of inhibitor and substrate binding sites (Mann-Whitney U test p-value < 10^{-72}). This suggests that at physiologically relevant concentrations of metabolites, the majority of substrate binding sites are at or near saturation, while inhibitor sites are occupied but largely far from being saturated, as was reported in the past [35], [37].

Quantifying the metabolic response to small molecules across different conditions and the trade off between regulation and enzymatic activity

The results presented in Figure 4e indicate that approximately one third of the inhibitory interactions involved metabolites whose concentrations were higher than the associated K_I (saturation level at least 0.5). This suggests that many enzymes are operating below their maximal catalytic potential. Why would metabolic enzymes be poised at such a point, well below maximal activity? As we show below and in the Supplementary Text, one possible explanation is a fundamental tradeoff enzymes face: in order to be responsive to the abundance of metabolite regulators, enzymes must sacrifice some of their catalytic activity. Put another way, we find that there is an inherent cost associated with small molecule regulation: to effectively regulate the activity of an enzyme, an inhibitor must be at sufficiently high concentration.

Using theoretical arguments from metabolic control analysis (MCA), we prove that this trade-off between activity and regulation is valid for a general class of kinetic rate laws (*e.g.* competitive, noncompetitive, and uncompetitive inhibition), applying both to inhibitors and activators (see Supplementary Text S1).

The relevant quantity in MCA which leads to the finding above is the scaled elasticity $\varepsilon_s{}^v$, which quantifies how fluctuations in the concentration of a metabolite *S* affect the flux through reaction *v*. For substrates of reactions described by irreversible Michaelis-Menten kinetics, the scaled elasticity is given by $\varepsilon_s{}^v = \frac{\partial v}{\partial s} \cdot \frac{s}{v} = 1 - \frac{s}{s + K_M}$, i.e. it is maximized (equal to 1) when the substrate binding site is unsaturated ($S \ll K_M$); in this regime, a fluctuation in the concentration of substrate leads to a linearly proportional change in the flux of the enzyme (for an isolated enzyme, like in an *in vitro* assay). On the other hand, the elasticity of non-cooperative, non-competitive inhibitors *I* is described by $\varepsilon_I{}^v = \frac{\partial v}{\partial I} \cdot \frac{I}{v} = -\frac{I}{I + K_I}$ (see Supplementary Figure S4). Counter-intuitively, this means that a reaction is most sensitive to an inhibitor *I* when *I* is at high concentration and the enzyme is strongly inhibited ($I \gg K_I$). Conversely, when the inhibitor concentration is low, its elasticity approaches zero and the flux cannot respond to changes in *I*. Therefore, substrates and inhibitors are subject to opposite quantitative relationships describing their potential to regulate reactions (Figure 5A). In summary, the saturation level has two consequences: first, it affects the elasticity (how sensitive the reaction rate would be to changes of the regulator level), and second, it determines the necessary amount of enzyme (because enzyme that is not used at its full capacity needs to be compensated by higher enzyme levels).

Thus, an alternative way of examining the difference between K_M and K_I data is by translating estimates of saturation into elasticities (which we treat as proxies for the metabolic response coefficient, see Discussion and SI Text). A detailed examination of elasticities across the 13 different growth conditions, further reveals the regulatory contribution of a set of central metabolites as substrates and inhibitors (Figure 5, Supplementary Figure S5). Interestingly, metabolites like IMP, ATP and ADP have very low elasticities as substrates (since they are typically at saturating levels), whereas others (GMP, AKG) have high substrate elasticities spanning from 0.7 to 0.9. As inhibitors, many metabolites are poised between elasticities of 0.2-0.5, whereas some metabolites (e.g. GDP, ADP) have high elasticities across all conditions. A higher elasticity could increase the flux response to changes in these metabolites, which is supported by prior work *e.g.* it has been reported that ADP strongly regulates PRPP synthase (*prpps*, EC 2.7.6.1) as a form of biosynthetic feedback inhibition [38], as well as fructose-1-phosphate kinase (*fruk*, EC 2.7.1.56) as a mediator of end-product inhibition [39]

Notably, the elasticity of several highly connected metabolites in the SMRN (*i.e.* PEP and FDP) changes substantially from one environmental condition to the next. FDP has high elasticities when cells grow in a glycolytic mode (e.g. growth on glucose, fructose, mannitol), suggesting that it has an important role in regulating the corresponding reaction fluxes only in glycolytic conditions. This is consistent with the role of FDP as a flux sensor in glycolytic conditions [28]. Interestingly, PEP as an inhibitor has high (absolute) elasticities on gluconeogenic carbon sources (e.g. pyruvate, acetate, succinate), operating anti-symmetrically to FDP. This regulatory design is critical for adaptation to environmental conditions, for example for the control of the flux through phosphofructokinase (*pfk*), which PEP is known to inhibit [40]. Under gluconeogenic conditions, it is important that the irreversible glycolytic enzyme *pfk* is turned off and tightly inhibited in order to avoid futile cycling. This inhibition is not needed though when cells grow on glycolytic carbon sources like glucose. Interestingly, we find that on glycolytic carbons that can only support a slow growth rate, PEP has still high elasticities (e.g. -0.38 and -0.36 for *pfk* when cells grow on galactose or mannose respectively).

Small Molecule Regulation Across Kingdoms of Life

While some central pathways of metabolism are nearly ubiquitous, whole-cell metabolism varies substantially between different organisms. At the coarsest resolution, some phylogenetic taxa (e.g. bacteria and plants) can fix inorganic carbon, while others (e.g. animals) cannot. On the other hand, the architecture of central carbon metabolism is broadly conserved across all kingdoms of life [41], [42]. However, there is little understanding of the extent to which small molecule regulatory interactions are conserved across evolutionarily distant taxa. Therefore, we proceeded to investigate whether patterns of small molecule regulation observed in *E. coli* were evident in other, evolutionarily distant, species.

To study small molecule regulation across kingdoms of life, we analyzed all available data on small molecule activators and inhibitors available in the BRENDA database, stratifying by the species in which the interaction was reported. We mined this data for recurrence of a regulatory interaction between a small molecule and EC number across different species. We then focused on analyzing interactions which satisfied two criteria: (1) they were evident in at least 10 different species, and (2) they were supported by at least 10 different reports in the literature. A full list describing the 253 such interactions is available in SI Table 2.

Because of the high interest in modeling and understanding of flux through central carbon metabolism, we focused our efforts on understanding recurrent regulatory interactions in this pathway (Figure 6). After excluding small ions and non-endogenous metabolites, we identified 18 small molecule regulatory interactions evident broadly across several phylogenetic taxa, which converged on the regulation of a small number of enzymes: four nodes in glycolysis, phosphofructokinase (6 recurrent regulators), fructose bisphosphatase (2 recurrent regulators), PEP carboxylase (3 recurrent regulators) and pyruvate kinase (3 recurrent regulators), were subject to particularly extensive regulation. Surprisingly, we found few conserved interactions in the TCA cycle, glyoxylate shunt and the non oxidative branch of the PP pathway. In line with its role as a committing step in glycolysis, phosphofructokinase was subject to negative feedback control by two metabolites far downstream in glucose catabolism, citrate and phosphoenolpyruvate, as well as regulation by energy-related cofactors ATP, ADP, and AMP. In contrast, pyruvate kinase was negatively regulated by ATP, but also subject to feedforward activation by FDP (fructose-1,6 diphosphate). In *E. coli*, the counter-intuitive activation of pyruvate kinase by FDP has been shown [43] to enable the cell to rapidly respond to changes in environmental glucose levels, by driving accumulation of PEP for future phosphorylation of glucose in glucose-depleted conditions, while the inhibition of phosphofructokinase by PEP has also been shown to be of critical importance in dynamic perturbations [3]

Interestingly, several of the recurrent regulatory interactions we identified were evident in only a subset of phylogenetic taxa. In some cases, this effect was due to the absence of the regulated enzyme in a taxon (*e.g.* PEP carboxylase is only present in archaea, bacteria, and plants, and the pentose phosphate pathway is not present in archaea). In other cases (where the enzyme was present in all taxa), small molecule regulation was simply different across taxa, with potentially interesting implications. For example, pyruvate kinase was inhibited by L-alanine in fungi and animals, but not in other species. In humans, this differential regulation plays a role in disease: the inhibition of one splice isoform of pyruvate kinase (PKM2) by L-alanine (but not the other, PKM1) contributes to the cancer-associated shift to aerobic glycolysis by promoting the shunting of glucose-derived carbon into biosynthetic pathways [44].

Discussion

The regulatory action of small molecules on enzymes and other proteins ensures robust operation of metabolism upon dynamic changes. Indeed, at least for central metabolism, metabolites can directly (acting as effectors) or indirectly (acting as signals to transcription factors) regulate the flux of almost all enzymatic reactions (Figure 7)). While our understanding of transcriptional and - to a lesser extent - post-translational regulation of metabolism has benefited from technological advances in high-throughput sequencing and mass spectrometric technologies, experimental challenges have hindered similar breakthroughs in our understanding of regulation of enzyme activity by small molecule metabolites [48]. Our approach here has been to leverage the fragmented wealth of published biochemical data to generate an atlas of small molecule regulation informatically, *i.e.* without performing additional laboratory experiments. Our findings illustrate the dual architecture of small molecule regulation and the underlying metabolic network, allow us to test a common hypothesis about the connection between regulation and thermodynamics, and to compare between metabolite concentrations and their respective binding affinities to target enzymes.

Here, we report a resource of experimentally evaluated interactions between endogenous metabolites and enzymes. Importantly, the computational framework we developed is freely available (http://github.com/eladnoor/small-molecule-regulation) and can be readily applied to reconstruct the SMRN of an arbitrary organism of choice, given a suitable genome-scale metabolic model and adequate data in BRENDA/BIoCyc. Such a resource may guide future experiments designed to identify metabolite-protein interactions. For example, an SMRN may prioritize metabolite-enzyme pairs based on the number of independent literature reports supporting their existence, or based on the elasticity of the interactions in different conditions. In addition, the SMRN and its associated inhibitor constants can be used as prior information to parameterize dynamic models of metabolism, an approach which has found success in *E. coli* and yeast [3], [4].

A critical shortcoming of our approach is the inherent biases of the BRENDA and Ecocyc data, *i.e.* the tendency for well-studied enzymes, pathways, and organisms to be overrepresented in these databases. Indeed, one factor in selecting *E. coli* as the organism of choice for reconstruction was the breadth of studies conducted on its metabolism. Most of our results, however, are robust to this type of biases, since we focus mostly on quantitative differences among reported interactions. Where possible, *e.g.* for the thermodynamic analysis of regulated versus not-regulated reactions, we verify our results specifically for the central metabolic network, where all reactions have been studied extensively.

Our theoretical analysis of a wide class of inhibiting and activating small molecules shows that, in general, there is a direct trade-off between the elasticity of an effector, and the reduction in activity which is caused by its interaction with the enzyme. Mathematically, if θ is the relative activity of the enzyme (e.g. $\theta = 1 - \frac{x}{x + K_I}$ in the case of non-competitive inhibition), then the absolute scaled elasticity will be equal to $|\varepsilon_x^v| = 1 - \theta$ (see Supplementary Text S1). We show that this relationship between activity and elasticity holds for activators and inhibitors alike. Therefore, using a small molecule effector to regulate a flux always comes with the cost of effectively lowering the activity of the enzyme. This cost of regulation might be the reason why *E. coli* does not have end-product inhibition for 4 of its amino-acid biosynthesis pathways. Perhaps, the cost of regulation for these "metabolically cheap" amino-acids is larger than the energetic cost of overproduction.

In order to estimate the level of control imposed by a small molecule, we had to make several simplifications. First, we assumed that all substrates and inhibitors bind non-cooperatively and that inhibition is not competitive. Furthermore, according to metabolic control theory, a high elasticity is a necessary but not sufficient condition for high control of a flux by a metabolite. We treat interactions with high elasticity as cases where high regulatory capacity is possible; in contrast, interactions with low elasticity are likely to have little regulatory capacity. A tradeoff between enzyme activity and elasticity will therefore translate to a tradeoff between activity and regulation. In order to be fully consistent with MCA, one would have to calculate all control coefficients for the inhibited reactions and apply MCA to calculate response coefficients, which quantify the level of flux control of a metabolite, something that is not in the scope of our work. Nevertheless, future work may use such a large-scale approach to estimate the degree to which each regulatory interaction in the SMRN affects metabolic fluxes in a given experimental condition, and how multiple regulatory signals are "integrated" to control catalytic activity in the entire system. Beyond applications in MCA, the SMRN could be a useful resource to answer other system-level questions about metabolism which were not in the scope of this work, for instance about its stability, robustness or bifurcation points. Furthermore, the dense network of interactions could facilitate topological studies and the search for design principles, using methods that have become common practice in analyzing transcriptional networks [49], [50].

A tractable path forward for decoding the control of flux by the SMRN is to integrate data on changes in *in vivo* metabolite abundance across conditions/perturbations. As we showed in Figure 5, doing so creates a map of the condition-dependent regulatory capacity (*i.e.* the elasticity) of small molecules (Figure 5 and Supplementary Figure 3), and prioritizes interactions which appear particularly relevant in a subset of conditions. For canonical interactions with available data on binding affinities, this kind of analysis can be used as a first step in understanding the importance of a regulatory interaction, perhaps by focusing on those interactions with a particularly large change in elasticity between two conditions. For instance, we found that certain metabolites (*e.g.* fructose bisphosphate) may become substantially stronger inhibitors in a subset of conditions. Doing so will likely add to the rich composition of transcriptional, post-translational, and small molecule regulatory interactions which we know to control metabolic flux.

Author contributions

E.R., D.C. and E.N. designed the analysis, performed the analysis and wrote the paper. J.G, E.B. performed/ contributed to the analysis. U.S. and D.S. supervised the study. All authors read and approved the final manuscript.

Acknowledgements

Methods

Assembly of the SMRN

Raw data were obtained from BRENDA and BioCyc databases. Scripts for parsing the obtained data, filtering unwanted values (i.e. for mutants or negative results), and mapping between BRENDA ligand IDs, KEGG identifiers and BiGG metabolite IDs were written in Python and can all be found on GitHub (https://github.com/eladnoor/small molecule-regulation, available under an MIT license). Likewise, all computations and analyses done for this paper can be found in the same repository.

Genome-scale metabolic model

All calculations involving the metabolic network of *E. coli* (i.e. distance calculations, thermodynamics of regulated and unregulated reactions, etc.) were done using the iJO1366 metabolic model downloaded from the BiGG website (http://bigg.ucsd.edu/models/iJO1366) [45] and with the help of the cobrapy toolbox (https://opencobra.github.io/cobrapy/) [46]

Distance calculations in the SMRN

First, we removed all the rows corresponding to co-factors from the genome-scale stoichiometric matrix of *E. coli* (the full list of co-factors is provided in the Supplementary Text). Then, the matrix was converted into an undirected bipartite graph where nodes were either metabolites or reactions. An edge was added between every reaction and all of its substrates and products. Then, the distance between each metabolite and enzyme was calculated by first finding the shortest path between the two on the bipartite graph, and counting the number of enzyme nodes along that path (i.e. excluding the metabolite nodes and the target enzyme itself). For example, the distance between an enzyme and one of its substrates is 0.

Cross-species analysis of small molecule regulation

All data regarding activation or inhibition was extracted from the BRENDA database. The R package taxize [47] was used to recover taxonomic information using the species name provided in the BRENDA database. Activating/inhibiting interactions with the same ligand ID (regulating metabolite) and EC number (target enzyme) were aggregated, and the number of such unique entries for each taxonomic group was calculated. Additionally, the number of unique literature references supporting each interaction was recorded, in order to preclude cases where a single interaction reported across multiple species was supported by a small number of independent sources.

Abbreviations

G6P: glucose-6-phosphate; 6PGL: 6-phosphogluconolactone; 6PGC: 6-phosphogluconate; Ru5P: ribulose-5-phosphate; KDG: 2-keto-3-deoxygluconate; R5P: ribose-5-phosphate; S7P: sedoheptulose-7-phosphate; E4P: erythrose-4-phosphate; X5P: xylulose-5-phosphate; F6P:

fructose-6-phosphate; FDP: fructose-1,6-bisphophate; DHAP:dihydroxacetone-phosphate; G3P: glyceraldehyde-3-phosphate; BPG: 1,3-bisphosphogylcerate; 3PG: 3-phosphoglycerate; 2PG: 2-phosphogylcerate; PEP: phosphoenolpyruvate; PYR: pyruvate; ACCOA: acetyl-CoA; OAA: oxaloacetate; CIT: citrate; ACON: aconitate; ICIT: isocitrate; GLX: glyoxylate; AKG: alpha-ketoglutarate; SUCCOA: succinyl-CoA; SUCC: succinate; FUM: fumarate; MAL: malate; ACTP: acetyl-phosphate; ACE: acetate, PI: phosphate, PPI: diphosphate, NADPH: Nicotinamide adenine dinucleotide phosphate - reduced.

Figures

Figure 1: Framework for reconstruction and analysis of the small molecule regulatory network (SMRN)

The BRENDA and BioCyc databases were mined for each reaction taking place in *E. coli*. The identified entries (including data on EC numbers, enzyme names, activating or inhibiting small molecule – enzyme interactions, metabolite names, K_m and K_I values) were stored and then matched by EC number to reactions in the most recent genome-scale reconstruction of *E. coli*, iJO1366. This dataset was searched and analyzed for regulatory small molecules, yielding a comprehensive Small Molecule Regulatory Network (SMRN). The SMRN was used as the primary resource for the remainder of the analysis. We analyzed the topological properties of the SMRN, evaluated the similarities and differences in the kinetic properties of reactions and interactions, and used published metabolite concentration data in order to evaluate the functional role of inhibitory small molecule/enzyme interactions.

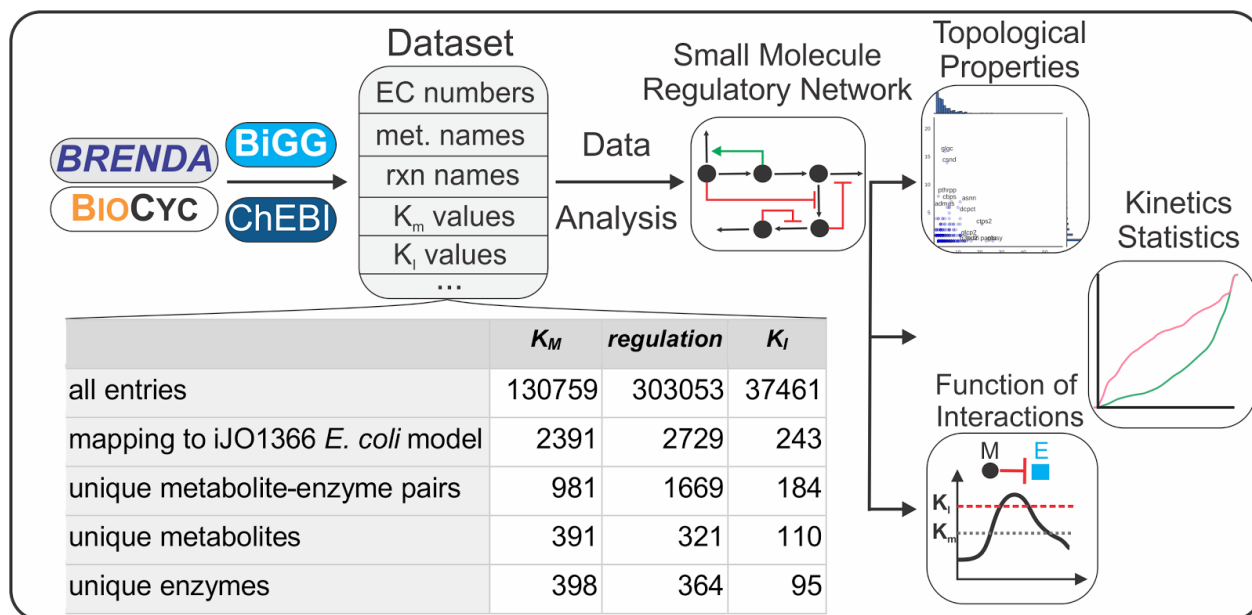|	K_M	regulation	K_I
all entries	130759	303053	37461
mapping to iJO1366 *E. coli* model	2391	2729	243
unique metabolite-enzyme pairs	981	1669	184
unique metabolites	391	321	110
unique enzymes	398	364	95

Figure 2: Overview of small molecule interactions from BRENDA

Statistics of the computationally reconstructed *E. coli* SMRN, covering (a) the proportions of 1039 *E. coli* metabolites which are inhibitors, activators, or both, and (b) proportion of 669 *E. coli* enzymatic reactions which are inhibited, activated, or both. (c)-(d) Scatter plots depicting the number of activating and inhibiting interactions each metabolite (c) and each reaction (d) participates in.
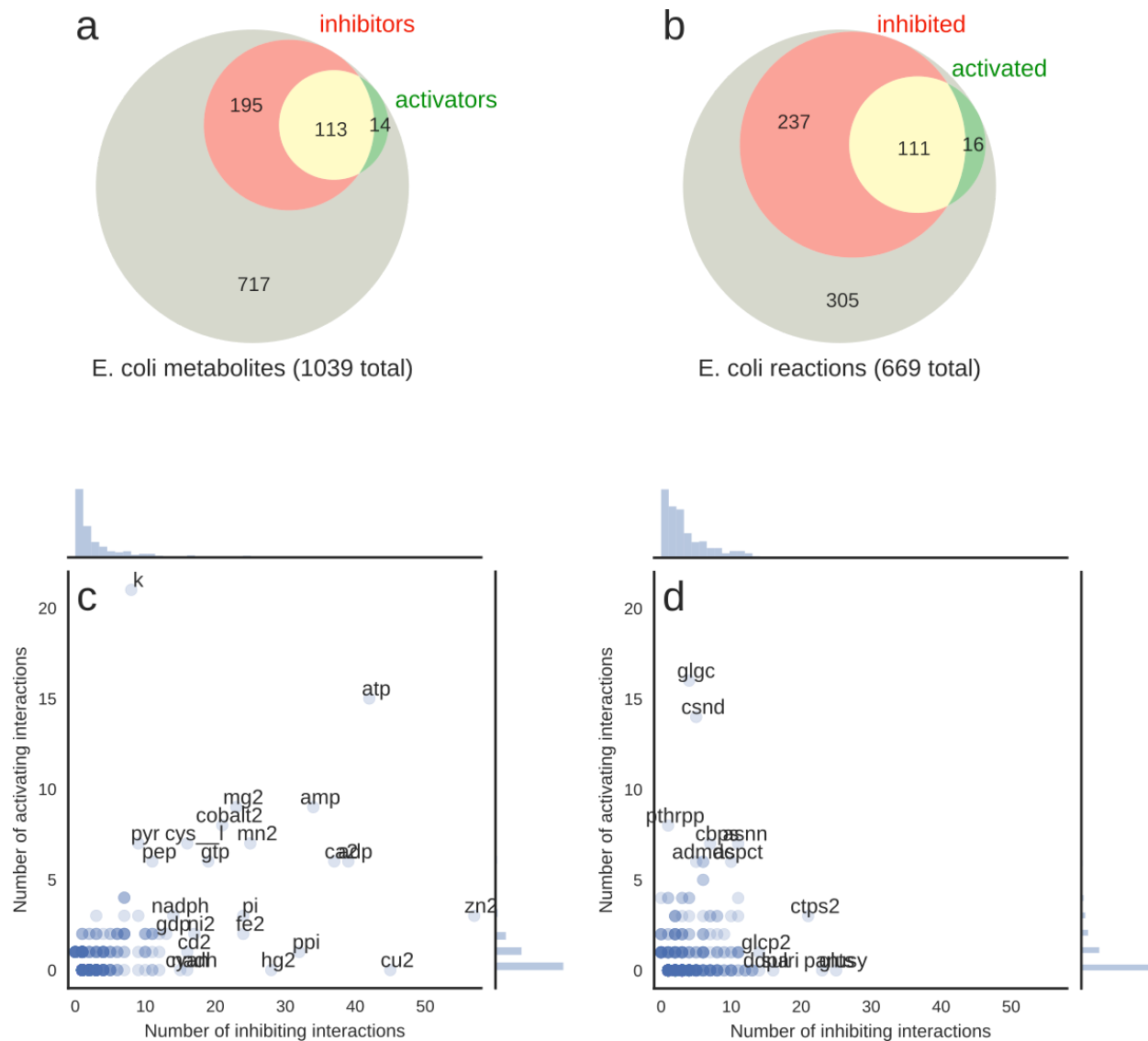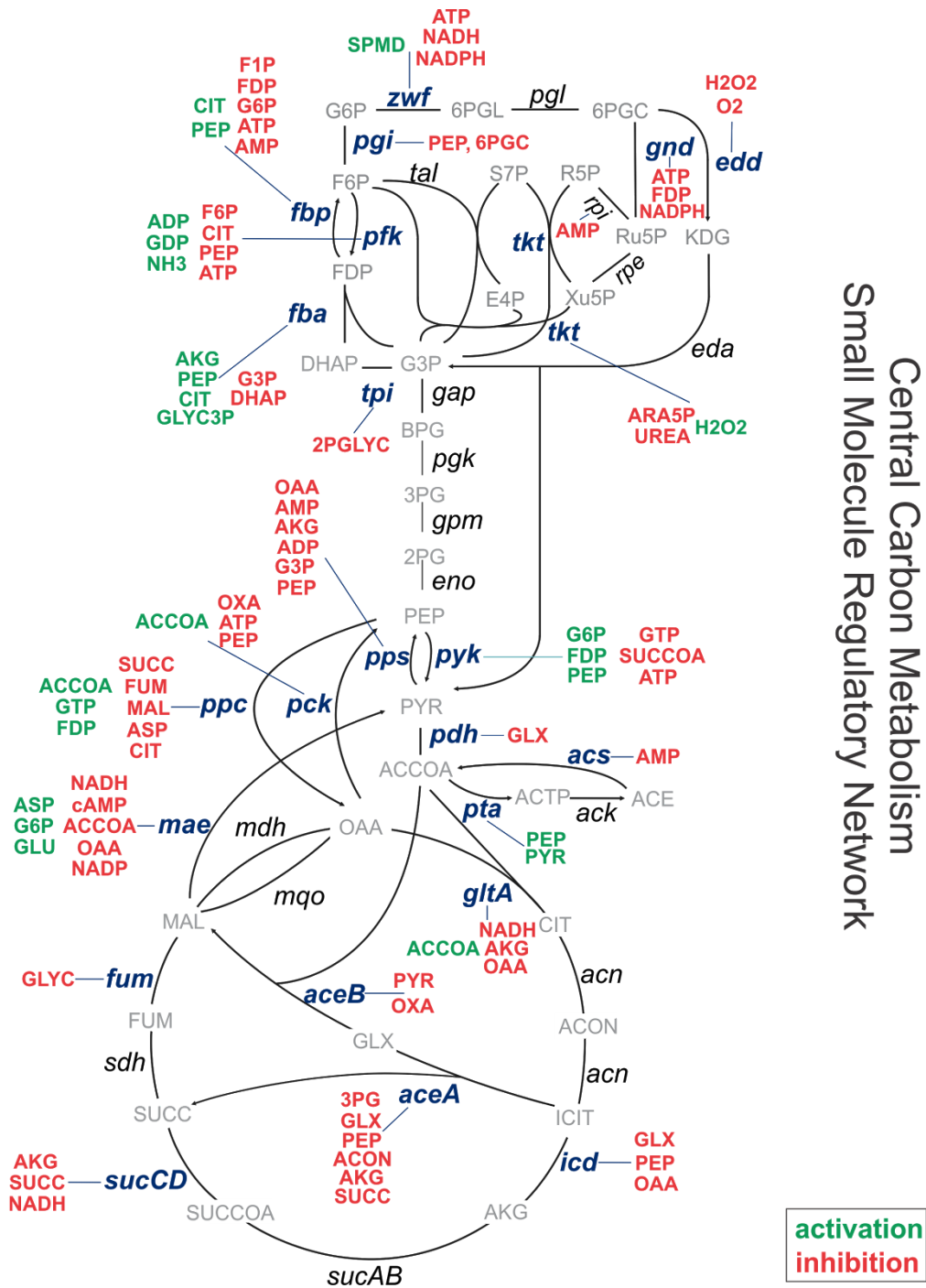
Figure 3: Small Molecule Regulatory Network of *E. coli* Central Carbon Metabolism
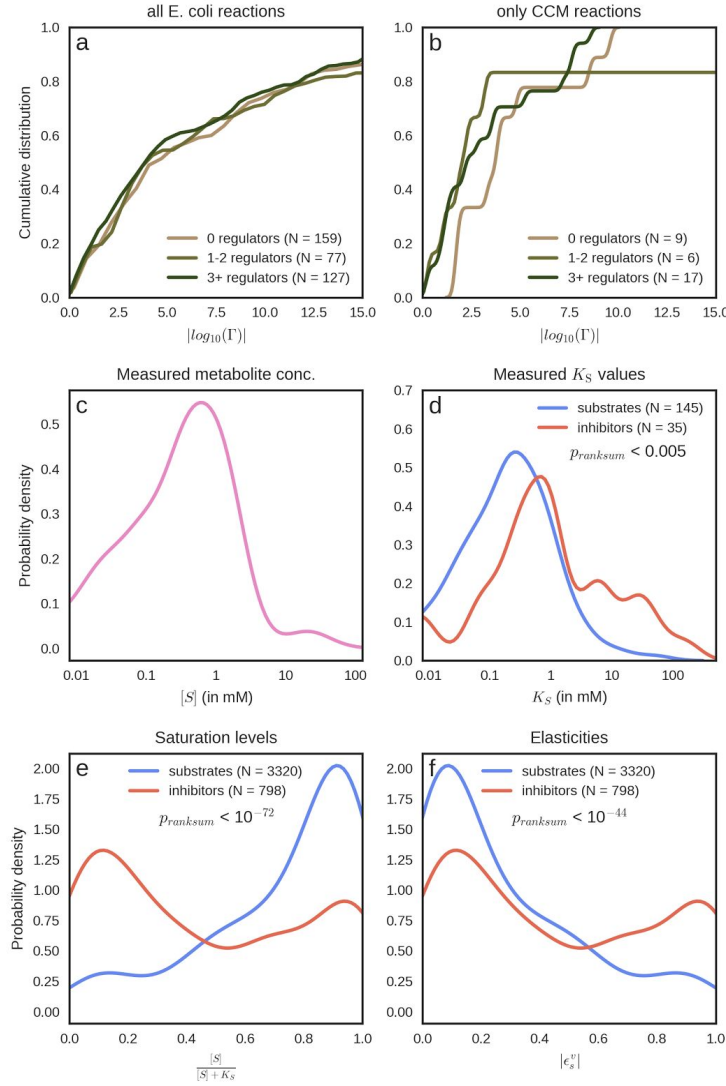

Depiction of the small molecule regulatory interactions in the central carbon metabolism of *E. coli*. Red metabolites are inhibitors, green metabolites are activators of the indicated reactions.

Figure 4: Thermodynamics, saturation, and elasticity of small molecule regulatory interactions in *E. coli*

(a) -(b) Comparison of reversibility indices (Γ) [30] between regulated and unregulated reactions. $|\log_{10}(\Gamma)|$ reflects how much freedom the reactants' concentrations need for reversing the flux. For instance, a reaction whose $|\log_{10}(\Gamma)| > 3$ would require a concentration range of at least $1{:}10^3$


(e.g. 10 uM to 10 mM) in order to reverse its direction. Physiological constraints typically limit this range to 10^3-10^4. (a) shows the cumulative distribution of $|\log_{10}(\Gamma)|$ for all reactions in the *E. coli* model for which an equilibrium constant could be computed using component contributions [29]. This means that about 40%-50% of reactions are reversible, with virtually no regard to whether they are regulated or not. The difference between the distribution of $|\log_{10}(\Gamma)|$ for regulated and unregulated reactions is not significant (Mann-Whitney U test p-value = 0.5). In (b), we focus only on reactions in central carbon metabolism, where the difference is not significant either (p-value = 0.2). (c) The distribution of measured metabolite concentrations, where the peak value is slightly below ~1 mM. (d) The histograms of K_M and K_I values are significantly different (Mann-Whitney U test p-value < 0.005). For a more detailed comparison on a single-metabolite basis, see Supplementary Figure S6. (e) Conversion of measured binding constants to saturation levels using measurements of metabolite concentrations across 13 conditions further highlights the difference between substrates (K_M values) and inhibitors (K_I

values) (Mann-Whitney U test p-value < 10^{-72}). When comparing the distributions of scaled elasticities (f) we find that inhibitors have significantly higher values (p-value < 10^{-44}), and seem to have a bimodal distribution which is split exactly at 0.5. Note that the absolute elasticity value for inhibitors is exactly equal to the saturation, therefore the blue histogram is the same in (e) and (f). For substrates, however, the elasticity is equal to 1 minus the saturation, so the red histogram in (f) is the mirror image of the one in (e).

Figure 5: Functional role of small molecule-enzyme interactions in central metabolism

(a) The opposing relationship between elasticity and concentration for a prototypical substrate or inhibitor. In general, substrates have high elasticity at low concentrations, while inhibitors have high elasticity at high concentrations. (b) A heatmap of the median values of each metabolite's elasticity values across all enzymes that utilize it as a substrate (left-hand side) and across all enzymes inhibited by it (right-hand side). The different columns correspond to different growth conditions (batch growth on minimal media with single carbon sources, samples in mid-exponential phase). Saturation levels were calculated using the formula $[S] / ([S] + K_S)$, where K_S is either the Michaelis-Menten coefficient (K_M) or the inhibition constant (K_I) and in turn elasticities were calculated as described in main text. The numbers next to each metabolite in parentheses count the number of different K_M or K_I values, respectively, that a metabolite has in our database (for different reactions). If a metabolite has more than one K_M or K_I value (i.e. for more than one enzyme), the median of all elasticities is shown. For more details, see Supplementary Figure S5.

Figure 6: Small molecule regulation across kingdoms of life
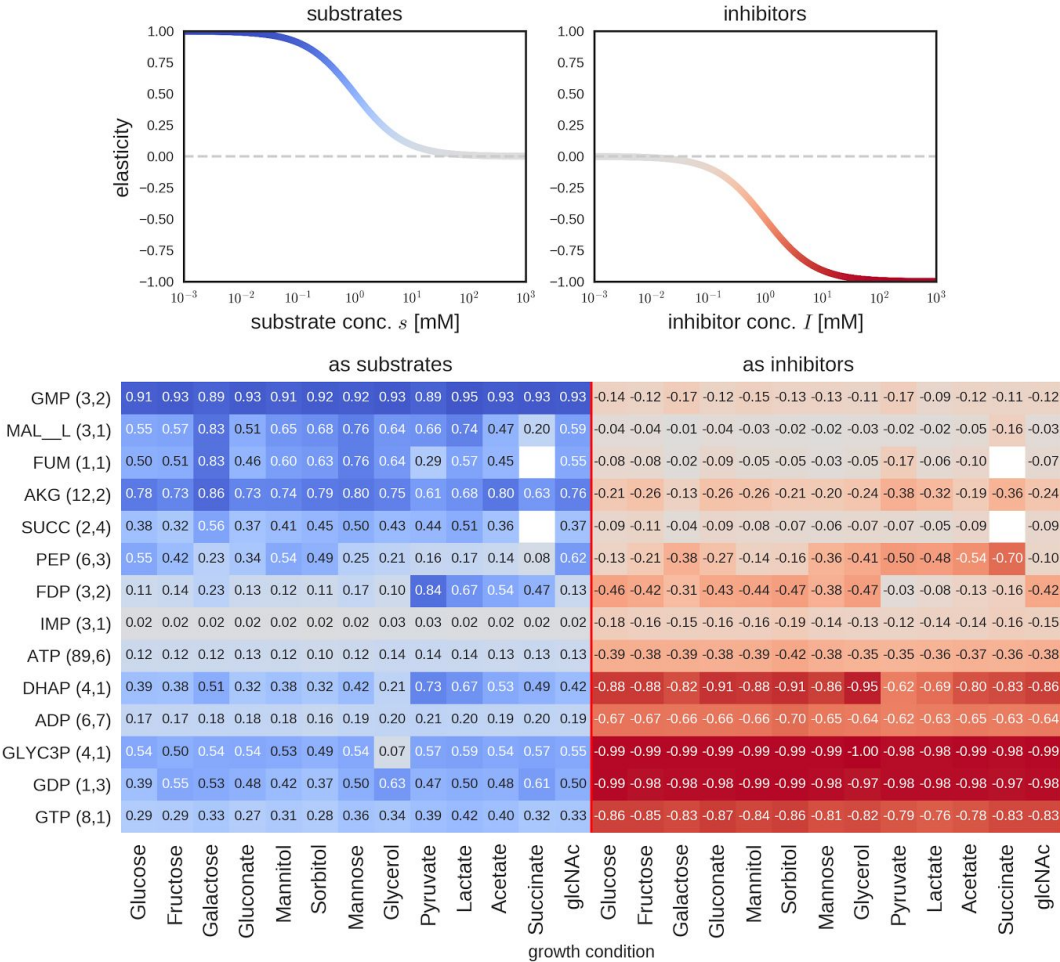
The BRENDA database was mined for all reports of small molecule regulatory interactions across all species. These interactions were aggregated by unique metabolite-reaction pairs. For each interaction evident in at least 10 different organisms and supported by evidence from at least 10 different published studies, followed by manual curation of the results. We identified the broad phylogenetic taxon within which the interaction was present. Nearly all conserved interactions are inhibitory, with a two exceptions: the activation of PFK by three metabolites, AMP, ADP, and fructose-2,6-bisphosphate, and the activation of PEP carboxylase by G6P.
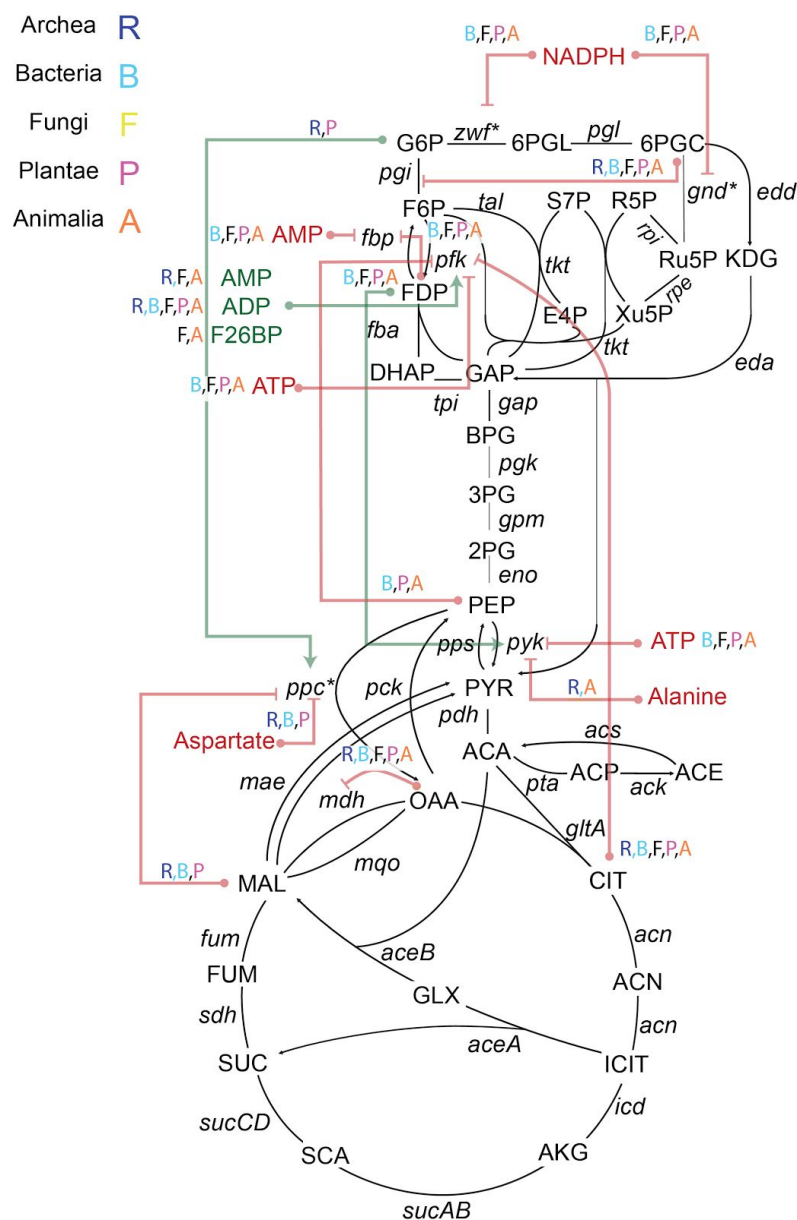
Figure 7: Combined architecture of direct small molecule and indirect transcriptional regulation via endogenous metabolites in *E coli*

A map of the reactions in central carbon metabolism that are regulated directly or indirectly by metabolite(s). On the left are reactions which are reported to have at least one metabolite - enzyme interaction. Middle diagram indicates reactions which are indirectly regulated by metabolites via transcription: in each case, the reaction is regulated by transcription factors that are recipients of metabolic signals (i.e. Cra-FDP, Crp-cAMP), as reported in [36]. Some reactions, *e.g.* those in intermediate glycolysis, are regulated exclusively transcriptionally. Map on the right overlays small-molecule and transcriptional regulation.
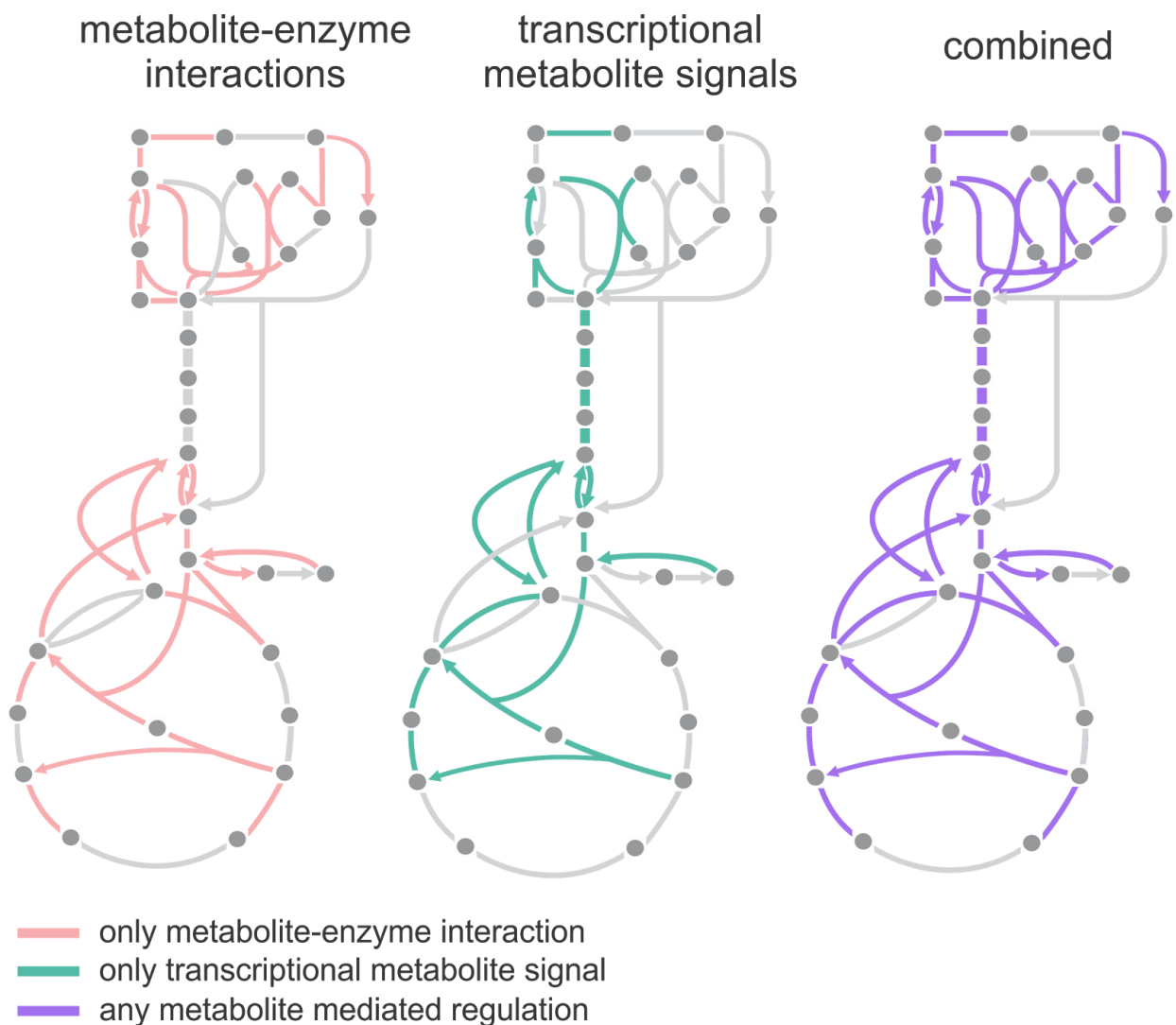
References:

[1] J. M. Berg, J. L. Tymoczko, and L. Stryer, "Biochemistry (5th edn)." WH Freeman, New York, 2002.

[2] Y.-F. Xu, D. Amador-Noguez, M. L. Reaves, X.-J. Feng, and J. D. Rabinowitz, "Ultrasensitive regulation of anapleurosis via allosteric activation of PEP carboxylase.," *Nat. Chem. Biol.*, vol. 8, pp. 562–568, Apr. 2012.

[3] H. Link, K. Kochanowski, and U. Sauer, "Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo.," *Nat. Biotechnol.*, vol. 31, pp. 357–361, Apr. 2013.

[4] S. R. Hackett *et al.*, "Systems-level analysis of mechanisms regulating yeast metabolic flux.," *Sci.*, vol. 354, Oct. 2016.

[5] F. A. Chandra, G. Buzi, and J. C. Doyle, "Glycolytic oscillations and limits on robust efficiency.," *Sci.*, vol. 333, pp. 187–192, Jul. 2011.

[6] F. He, E. Murabito, and H. V. Westerhoff, "Synthetic biology and regulatory networks: where metabolic systems biology meets control engineering.," *J. R. Soc. Interface*, vol. 13, Apr. 2016.

[7] Z. Chen, S. Rappert, and A.-P. Zeng, "Rational design of allosteric regulation of homoserine dehydrogenase by a nonnatural inhibitor L-lysine.," *ACS Synth. Biol.*, vol. 4, pp. 126–131, Feb. 2015.

[8] H. R. Christofk *et al.*, "The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth.," *Nature*, vol. 452, pp. 230–233, Mar. 2008.

[9] X. Li, X. Wang, and M. Snyder, "Systematic investigation of protein-small molecule interactions.," *IUBMB life*, vol. 65, pp. 2–8, Jan. 2013.

[10] M. M. Savitski *et al.*, "Tracking cancer drugs in living cells by thermal profiling of the proteome.," *Sci.*, vol. 346, Oct. 2014.

[11] F. B. M. Reinhard *et al.*, "Thermal proteome profiling monitors ligand interactions with cellular membrane proteins.," *Nat. methods*, vol. 12, pp. 1129–1131, Dec. 2015.

[12] T. Orsak, T. L. Smith, D. Eckert, J. E. Lindsley, C. R. Borges, and J. Rutter, "Revealing the allosterome: systematic identification of metabolite-protein interactions.," *Biochemistry*, vol. 51, pp. 225–232, Jan. 2012.

[13] Y. V. Nikolaev, K. Kochanowski, H. Link, U. Sauer, and F. H.-T. Allain, "Systematic Identification of Protein-Metabolite Interactions in Complex Metabolite Mixtures by Ligand-Detected Nuclear Magnetic Resonance Spectroscopy.," *Biochemistry*, vol. 55, pp. 2590–2600, May 2016.

[14] Y. Feng *et al.*, "Global analysis of protein structural changes in complex proteomes.," *Nat. Biotechnol.*, vol. 32, pp. 1036–1044, Oct. 2014.

[15] H. Link, D. Christodoulou, and U. Sauer, "Advancing metabolic models with kinetic information," *Curr. Opin. Biotechnol.*, vol. 29, pp. 8–14, Oct. 2014.

[16] O. Schueler-Furman and S. J. Wodak, "Computational approaches to investigating allostery," *Curr. Opin. Struct. Biol.*, vol. 41, Dec. 2016.

[17] Z. A. King *et al.*, "BiGG Models: A platform for integrating, standardizing and sharing genome-scale models," *Nucleic Acids Res.*, vol. 44, Jan. 2016.

[18] A. Chang *et al.*, "BRENDA in 2015: exciting developments in its 25th year of existence.," *Nucleic acids Res.*, vol. 43, no. Database issue, pp. D439–D446, Jan. 2015.

[19] G. Cochrane, I. Karsch-Mizrachi, T. Takagi, and International Nucleotide Sequence Database Collaboration, "The International Nucleotide Sequence Database Collaboration.," *Nucleic acids Res.*, vol. 44, no. D1, pp. D48–D50, Jan. 2016.

[20] J. D. Orth *et al.*, "A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011.," *Mol. Syst. Biol.*, vol. 7, Oct. 2011.

[21] I. Thiele, N. Jamshidi, R. M. T. Fleming, and B. Ø. Palsson, "Genome-scale reconstruction of Escherichia coli's transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization.," *PLoS Comput. Biol.*, vol. 5, Mar. 2009.

[22] C. S. Henry, M. DeJongh, A. A. Best, P. M. Frybarger, B. Linsay, and R. L. Stevens, "High-throughput generation, optimization and analysis of genome-scale metabolic models.," *Nat. Biotechnol.*, vol. 28, no. 9, pp. 977–982, Sep. 2010.

[23] D. E. Atkinson, "The energy charge of the adenylate pool as a regulatory parameter. Interaction with feedback modifiers.," *Biochemistry*, vol. 7, pp. 4030–4034, Nov. 1968.

[24] M. J. Page and E. Di Cera, "Role of Na+ and K+ in enzyme function.," *Physiol. Rev.*, vol. 86, pp. 1049–1092, Oct. 2006.

[25] J. D. Orth, R. M. T. Fleming, and B. Ø. Palsson, "Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide.," *EcoSal Plus*, vol. 4, Sep. 2010.

[26] V. Chubukov, L. Gerosa, K. Kochanowski, and U. Sauer, "Coordination of microbial metabolism.," *Nat. Rev. Microbiol.*, vol. 12, pp. 327–340, May 2014.

[27] L. Gerosa *et al.*, "Pseudo-transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data.," *Cell Syst.*, vol. 1, pp. 270–282, Oct. 2015.

[28] K. Kochanowski, B. Volkmer, L. Gerosa, B. R. Haverkorn van Rijsewijk, A. Schmidt, and M. Heinemann, "Functioning of a metabolic flux sensor in Escherichia coli.," *Proc. Natl. Acad. Sci. United States Am.*, vol. 110, pp. 1130–1135, Jan. 2013.

[29] E. Noor, H. S. Haraldsdóttir, R. Milo, and R. M. T. Fleming, "Consistent estimation of Gibbs energy using component contributions.," *PLoS Comput. Biol.*, vol. 9, Jul. 2013.

[30] E. Noor, A. Bar-Even, A. Flamholz, Y. Lubling, D. Davidi, and R. Milo, "An integrated open framework for thermodynamics of reactions that combines accuracy and coverage.," *Bioinforma.*, vol. 28, pp. 2037–2044, Aug. 2012.

[31] F. Daldal and D. G. Fraenkel, "Assessment of a futile cycle involving reconversion of fructose 6-phosphate to fructose 1,6-bisphosphate during gluconeogenic growth of Escherichia coli.," *J. Bacteriol.*, vol. 153, pp. 390–394, Jan. 1983.

[32] H. Link, T. Fuhrer, L. Gerosa, N. Zamboni, and U. Sauer, "Real-time metabolome profiling of the metabolic switch between starvation and growth.," *Nat. methods*, vol. 12, pp. 1091–1097, Nov. 2015.

[33] H. Akashi and T. Gojobori, "Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis.," *Proc. Natl. Acad. Sci. United States Am.*, vol. 99, pp. 3695–3700, Mar. 2002.

[34] L. Gerosa and U. Sauer, "Regulation and control of metabolic fluxes in microbes.," *Curr. Opin. Biotechnol.*, vol. 22, pp. 566–575, Aug. 2011.

[35] J. O. Park *et al.*, "Metabolite concentrations, fluxes and free energies imply efficient enzyme usage.," *Nat. Chem. Biol.*, vol. 12, pp. 482–489, Jul. 2016.

[36] K. K. G. F. B. C. V. N. Sauer, "Few regulatory metabolites coordinate expression of central metabolic genes in Escherichia coli."

[37] B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, and J. D. Rabinowitz, "Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli.," *Nat. Chem. Biol.*, vol. 5, pp. 593–599, Aug. 2009.

[38] M. Willemoës, B. Hove-Jensen, and S. Larsen, "Steady state kinetic model for the binding of substrates and allosteric effectors to Escherichia coli phosphoribosyl-diphosphate synthase.," *J. Biol. Chem.*, vol. 275, pp. 35408–35412, Nov. 2000.

[39] B. Buschmeier, W. Hengstenberg, and J. Deutscher, "Purification and properties of 1-phosphofructokinase from Escherichia coli," *FEMS Microbiol. Lett.*, vol. 29.

[40] A. W. Fenton and G. D. Reinhart, "Disentangling the web of allosteric communication in a homotetramer: heterotropic inhibition in phosphofructokinase from Escherichia coli.," *Biochemistry*, vol. 48, pp. 12323–12328, Dec. 2009.

[41] J. M. Peregrín-Alvarez, C. Sanford, and J. Parkinson, "The conservation and evolutionary modularity of metabolism.," *Genome Biol.*, vol. 10, Jun. 2009.

[42] J. M. Peregrin-Alvarez, S. Tsoka, and C. A. Ouzounis, "The phylogenetic extent of metabolic enzymes and pathways.," *Genome Res.*, vol. 13, pp. 422–427, Mar. 2003.

[43] Y.-F. Xu *et al.*, "Regulation of yeast pyruvate kinase by ultrasensitive allostery independent of phosphorylation.," *Mol. cell*, vol. 48, pp. 52–62, Oct. 2012.

[44] H. P. Morgan *et al.*, "M2 pyruvate kinase provides a mechanism for nutrient sensing and regulation of cell proliferation.," *Proc. Natl. Acad. Sci. United States Am.*, vol. 110, pp. 5881–5886, Apr. 2013.

[45] J. D. Orth and B. Palsson, "Gap-filling analysis of the iJO1366 Escherichia coli metabolic network reconstruction for discovery of metabolic functions.," *BMC Syst. Biol.*, vol. 6, May 2012.

[46] A. Ebrahim, J. A. Lerman, B. O. Palsson, and D. R. Hyduke, "COBRApy: COnstraints-Based Reconstruction and Analysis for Python," *BMC Syst. Biol.*, vol. 7, 2013.

[47] S. A. Chamberlain and E. Szöcs, "taxize: taxonomic search and retrieval in R," *F1000Research*, Oct. 2013.

[48] J. E. Lindsley and J. Rutter, "Whence cometh the allosterome?," *Proc. Natl. Acad. Sci. United States Am.*, vol. 103, pp. 10533–10535, Jul. 2006.

[49] R. Milo, S. Itzkovitz, N. Kashtan, and D. Chklovskii, "Network Motifs: Simple Building Blocks of Complex Networks," *Science*, vol. 298, pp. 11–14, 2002.

[50] J. C. Liao, R. Boscolo, Y.-L. Yang, L. M. Tran, C. Sabatti, and V. P. Roychowdhury, "Network component analysis: reconstruction of regulatory signals in biological systems.," *Proc. Natl. Acad. Sci. United States Am.*, vol. 100, pp. 15522–15527, Dec. 2003.