

DISS. ETH NO. 29128

DYNAMIC MODELS OF ESCHERICHIA COLI METABOLISM:
AN ACCURATE DESCRIPTION OF OUR PATHETIC THINKING.

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES
(Dr. sc. ETH Zurich)

presented by
MICHIEL ANTON PIETER KARRENBELT

M.Sc., Wageningen University

born on 28.07.1987

accepted on the recommendation of

Prof. Dr. Uwe Sauer
Prof. Dr. Bas Teusink
Prof. Dr. Jörg Stelling
Prof. Dr. Manfred Claassen

2024

Abstract

The introduction of this thesis consists of two parts. In the first of these we will introduce systems biology as a field of study by placing it in the historical context of the overarching discipline of biology. A quick stroll through this history offers us a much needed and often absent perspective on the field as part of the larger body of scientific philosophy that it is part of. One might argue such positioning is paramount in any systems science whose existence predicates on the essentiality of contextualization. In this excursion we will swiftly move from the Greeks to the 20th century, touching on historical figures and events to give an impression of how the philosophy of biological sciences developed throughout. Then follows an examination of the scientific method, the development of which took center stage in the scientific debate during the 20th century, where we review different models of scientific inquiry and introduce the question of how to distinguish science from non-science, known as the demarcation problem. The section concludes with the introduction of the question of what systems biology is assumed to be, according to self-proclaimed founders of the field and their predecessors, and what promises accompanied its advent.

The second section of the introduction concerns itself with cellular metabolism. Here the relevant background on different aspects relating to the specific system studied in this work is provided to the reader. First a general perspective on central carbon metabolism and the different regulatory mechanisms that control it is laid out. Finally, an overview of the study of regulation of central carbon metabolism in the model species *Escherichia coli* is presented in which we introduce the current state of the field by summarizing contemporary work. This section concludes with an outline of the experimental and computational challenges that scientists in the field are faced with.

In chapter 2 we focus on the prediction of allosteric regulation in *Escherichia coli* metabolism by single reaction modeling. This model structure identification approach uses the available metabolomics, fluxomics and proteomics steady-state data over different series of nutrient-limited growth conditions, and then asks whether a generalized reversible enzyme kinetics rate law without regulation is sufficient to describe the observations, or whether the inclusion of allosteric regulation significantly improves this ability. In this study we take a reductionistic approach in that we study single reactions in isolation. Herein we sacrifice the ability to probe into how a regulatory interaction might affect emergent behavior such as network dynamics in favor of scalability that the simplistic nature of this approach offers. We systematically assess the regulatory potential of 284 metabolites as allosteric interactors of 84 metabolic reactions. In order to lower the number of false positives among the predicted top ranking interactions we include additional lines of evidence, such as the reported presence of the regulatory interaction in another organism or its detection in physical interaction studies. We then selected the top ranking interactions for follow-up study using enzyme assays in order to test whether the predicted regulators modulate enzyme activity *in vitro*. In these validation experiments we find evidence for the existence of 11 novel metabolite-protein interactions with potential physiological relevance.

In chapter 3 we concern ourselves with the prediction of allosteric regulation in the tricarboxylic acid cycle of *Escherichia coli* by using a system of coupled differential equations. This approach combines steady-state data on the metabolome, fluxome and proteome with that of thermodynamic and kinetic parameter estimates to derive model priors over the initial conditions. As in chapter 2, we opt for generalized reversible enzyme kinetics rate laws, but in contrast to it, we study the system as an interconnected network of metabolic reactions. Two additional noteworthy distinctions are the fact that we work with absolute quantification data on all the aforementioned -omics levels, and that we use time series data of the observed metabolome dynamics after a carbon source perturbation, rather than a series of steady states. Using this reaction network we exploit existing dependencies, such as those among thermodynamic parameters, but also those between

kinetic parameters, biochemical species and their thermodynamics as described by the Haldane relationship. Our primary goal is to answer the question of whether, based on the available data, we can derive a model that can explain the transient dynamics observed after a carbon source switch, and if not, to systematically identify key regulatory mechanisms that help coordinate this dynamic adaptation. Unexpectedly however, we encountered formidable challenges during this endeavor trying to reproduce the work of predecessors. For that reason the majority of our effort was spent addressing this issue in order to obtain a framework that produces models and predictions that are reproducible, reliable and reusable. Venturing back to the tricarboxylic acid cycle, we find that part of the observed dynamics cannot be sufficiently explained. By assembling an ensemble of models with different regulatory interaction topologies we assess which protein-metabolite interactions help to increase the explanatory potential of the model in describing the observed dynamics most, and derive a set of predictions from these results.

Finally, we summarize the key findings of this thesis. The first achievement is that we predicted and found *in vitro* evidence for the existence of up to 11 metabolite-protein interactions from our single reaction modeling study using steady-state data and *in vitro* enzyme assays. However, the absence of baseline expectations poses a problem that highlights inadequacies in scientific rigor. The second achievement is that we have produced a modeling pipeline that generates models and predictions that are reproducible, models that are both reliable, in that the automated construction process minimizes the risk of errors, and reusable, since components are properly annotated, and robust, in that the predictions obtained are less variable than those in previous work. Finally, we show that the transient metabolome dynamics that we observe cannot be sufficiently explained by combining the available data sources, and generate predictions on which allosteric interactions might exist that help shape the response. We conclude with a discussion on the current state of the field, the scientific merit of the endeavor, and bring the whole to fruition with a series of recommendations for future experiments.

Sommario

L'introduzione di questa tesi si compone di due parti. Nella prima di queste, introdurremo la biologia dei sistemi come campo di studio, posizionandola nel contesto storico della disciplina biologica generale. Una rapida passeggiata attraverso questa storia ci offre una prospettiva spesso necessaria e spesso assente sul campo come parte del più ampio corpo di filosofia scientifica di cui fa parte. Si potrebbe sostenere che tale posizionamento è fondamentale in ogni scienza dei sistemi il cui esistenza si basa sull'essenzialità della contestualizzazione. In questo excursus, passeremo rapidamente dai Greci al XX secolo, toccando figure e eventi storici per dare un'idea di come si sia sviluppata la filosofia delle scienze biologiche. Segue quindi un'esaminazione del metodo scientifico, lo sviluppo del quale ha avuto un ruolo centrale nel dibattito scientifico durante il XX secolo, in cui esamineremo diversi modelli di ricerca scientifica e introdurremo la questione di come distinguere la scienza dalla non-scienza, nota come problema della demarcazione. La sezione si conclude con l'introduzione della domanda su cosa si presume sia la biologia dei sistemi, secondo i fondatori autoproclamati del campo e i loro predecessori, e quali promesse accompagnino la sua nascita.

La seconda sezione dell'introduzione si occupa del metabolismo cellulare. Qui viene fornito al lettore il contesto rilevante su diversi aspetti relativi al sistema specifico studiato in questo lavoro. Prima di tutto, viene delineata una prospettiva generale sul metabolismo centrale del carbonio e sui diversi meccanismi di regolazione che lo controllano. Infine, viene presentata un'overview dello studio della regolazione del metabolismo centrale del carbonio nel ceppo modello *Escherichia coli*, in cui introduciamo lo stato attuale del campo riassumendo il lavoro contemporaneo. Questa sezione si conclude con un'analisi delle sfide sperimentali e computazionali che i ricercatori in questo campo devono affrontare.

Nel capitolo 2 ci concentriamo sulla previsione della regolazione alosterica nel metabolismo di *Escherichia coli* mediante la modellazione di reazioni singole. Questo approccio di identificazione della struttura del modello utilizza dati stazionari di metabolomica, fluxomica e proteomica in diverse serie di condizioni di crescita a limitazione di nutrienti, e si chiede se una legge generale di cinetica enzimatica reversibile senza regolazione sia sufficiente per descrivere le osservazioni, o se l'inclusione della regolazione alosterica migliora significativamente questa capacità. In questo studio adottiamo un approccio riduzionistico nello studio delle singole reazioni in isolamento. In questo sacrificiamo la capacità di indagare come un'interazione regolatoria possa influenzare comportamenti emergenti come la dinamica di rete a favore della scalabilità che la natura semplicistica di questo approccio offre. Valutiamo sistematicamente il potenziale regolatorio di 284 metaboliti come interattori alosterici di 84 reazioni metaboliche. Per ridurre il numero di falsi positivi tra le interazioni previste in alto includiamo ulteriori linee di evidenza, come la presenza segnalata dell'interazione regolatoria in un altro organismo o la sua rilevazione in studi di interazione fisica. Selezioniamo poi le interazioni in alto per uno studio di verifica utilizzando saggi enzimatici per verificare se i regolatori previsti modulano l'attività enzimatica *in vitro*. In questi esperimenti di validazione troviamo prove dell'esistenza di 11 nuove interazioni metabolite-proteina con potenziale rilevanza fisiologica.

Nel capitolo 3 ci occupiamo della previsione della regolazione alosterica nel ciclo dell'acido tricarbossilico di *Escherichia coli* utilizzando un sistema di equazioni differenziali accoppiate. Questo approccio combina dati stazionari sul metaboloma, il fluxoma e il proteoma con stime dei parametri termodinamici e cinetici per derivare ipotesi del modello sulle condizioni iniziali. Come nel capitolo 2, optiamo per leggi generali di cinetica enzimatica reversibile, ma a differenza di esso, studiamo il sistema come una rete interconnessa di reazioni metaboliche. Due distinzioni ulteriori degne di nota sono il fatto che lavoriamo con dati di quantificazione assoluta su tutti i livelli di -omica sopra menzionati, e che utilizziamo dati di serie temporali della dinamica osservata del metaboloma dopo

una perturbazione della fonte di carbonio, anziché una serie di stati stazionari. Utilizzando questa rete di reazioni sfruttiamo dipendenze esistenti, come quelle tra i parametri termodinamici, ma anche quelle tra i parametri cinetici, le specie biochimiche e la loro termodinamica come descritto dalla relazione di Haldane. Il nostro obiettivo principale è rispondere alla domanda se, basandoci sui dati disponibili, possiamo derivare un modello che possa spiegare la dinamica transitoria osservata dopo un cambio della fonte di carbonio e, se no, identificare sistematicamente i meccanismi regolatori chiave che aiutano a coordinare questo adattamento dinamico. Inaspettatamente, tuttavia, ci siamo imbattuti in sfide formidabili durante questo sforzo cercando di riprodurre il lavoro dei predecessori. Per questo motivo la maggior parte del nostro sforzo è stata dedicata a risolvere questo problema al fine di ottenere una struttura che produca modelli e previsioni riproducibili, affidabili e riutilizzabili. Ritornando al ciclo dell'acido tricarbossilico, troviamo che parte della dinamica osservata non può essere spiegata in modo sufficiente. Assemblando un insieme di modelli con diverse topologie di interazioni regolatorie, valutiamo

Infine, riassumiamo le principali scoperte di questa tesi. Il primo risultato è che abbiamo previsto e trovato evidenze *in vitro* dell'esistenza di fino a 11 interazioni metabolita-proteina dal nostro studio di modellazione delle reazioni singole utilizzando dati di stato stazionario e saggi enzimatici *in vitro*. Tuttavia, l'assenza di aspettative di base pone un problema che mette in evidenza le lacune nella rigerosità scientifica. Il secondo risultato è che abbiamo creato una pipeline di modellazione che genera modelli e previsioni riproducibili, modelli che sono affidabili, nel senso che il processo di costruzione automatizzato minimizza il rischio di errori, e riutilizzabili, dal momento che i componenti sono adeguatamente annotati, e robusti, nel senso che le previsioni ottenute sono meno variabili rispetto a quelle in lavori precedenti. Infine, mostriamo che le dinamiche transitorie del metaboloma che osserviamo non possono essere spiegate in modo sufficiente combinando le fonti di dati disponibili e generiamo previsioni sulle interazioni alosteriche che potrebbero contribuire a plasmare la risposta. Concludiamo con una discussione sullo stato attuale del campo, il merito scientifico dell'impresa e portiamo tutto a compimento con una serie di raccomandazioni per future ricerche.

Dedication

I dedicate this dissertation to fellow and future students in pursuit of, or considering the pursuit of, a doctorate in the field of systems biology.

If you wish to obtain the degree as a means to signal high value, I encourage you to adopt the attitude and mindset of a sales consultant. Know that the use of standards and reproducibility of research is optional, so embrace a businesslike pragmatism and don't linger in the former. By the time others have to deal with your work, they will either adopt the same mindset, or not, and if not you can rest assured that their inability to reproduce your work is not something you will be held responsible or accountable for; it can be attributed to the inaptitude of the newcomer that you by now outrank. Your truth shall be that the best dissertation is a finished dissertation.

If you wish to do science for the sake of science, then I encourage you to challenge the established practices first and foremost, and to use only those that you can fundamentally understand and have mastered through endless iteration. You will find that many of the most elementary of questions have been insufficiently or not at all addressed and that most of what is perceived-to-be knowledge is feeble or false. Do not let yourself be nauseated by the credulous folly of your colleagues that adore lustrous stories; we are only human, all too human. Your truth shall be that convictions are more dangerous enemies of truth than lies.

I write this to you, dear Lisbeth, only in order to counter the most usual proofs of believing people, who invoke the evidence of their inner experiences and deduce from it the infallibility of their faith. Every true faith is indeed infallible; it performs what the believing person hopes to find in it, but it does not offer the least support for the establishing of an objective truth. Here the ways of men divide. If you want to achieve peace of mind and happiness, then have faith; if you want to be a disciple of truth, then search.¹

¹Friedrich Nietzsche, Selected Letters (Hackett Publishing, 1996).

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree, nor was any part of it submitted for publication in a scientific journal at any time during the duration of my doctoral studies. Except where stated otherwise by reference or acknowledgment, the work presented is entirely my own.

Acknowledgments

I want to express my sincere gratitude to the following individuals for their invaluable contributions and unwavering support throughout my journey in the field of molecular systems biology:

Elad Noor, for the sighs of discomfiture preceding every answer in response to any question ever asked to him, his help in addressing an innumerable number of challenges I faced, both scientifically as well as in communicating concepts and ideas.

Marieke Buffing, for her meticulous use of controls and rigorous testing of changes to the experimental procedure, practices which I hope more people in the lab would adopt, and for introducing me to the mass spectrometer that I've come to know and love so much, the endless source of joy its maintenance brought.

Christoph Gruber and Tomek Diederer, for engaging discussions on the state of systems biology and academia.

Dimitris Christodoulou, Markus Basan and Ruben Mars for their camaraderie and shared office banter.

Tobias Furher, for the time he allotted helping me with my experimental work and for the enjoyable moments we shared during dinners, beers and various other extracurricular activities.

Philipp Warmer, for leading the way in overcoming challenges and inspiring my path forward.

Ohad Golan, for his grounded approach and self-assured demeanor.

Pau Perez, for his openness and willingness to share stories of his personal scientific skirmish.

Maren Diether, for her valuable assistance with the experimental work conducted by MSc. student, Xavier, whom I was supervising.

Xavier Hernandez, for his outstanding work and his pleasant and collaborative approach throughout our research collaboration.

Brendan Ryback, for sharing the insights of his study of the fast filtration perturbation experiment.

Sarah Cherkaoui, for her kindness and positive spirit, which added a pleasant atmosphere to our interactions.

Alexis Delabriere, for his expertise in the computational analysis of mass spectrometry data.

Alaa Othman, for providing honest feedback on my mass spectrometry data and offering critical insights and unveiling remark on previously published data from an experiment I was emulating.

Evgeniya Schastnaya for my first and only Tinder experience, one we shared in Innsbruck.

Andrei Dmitrenko, for his pragmatic approach and collaboration on machine learning projects during our time in the lab.

Mauro Masiero, for all the socially awkward moments during coffee break and many other events.

Sammy Pontrelli, for his excitement about extremophilic organisms, a fascination that I share.

Duncan Holbrook-Smith and Karin Ortmayr, for many extracurricular ventures after-hours.

Karin Meier, for her kindness and positive presence during our time working together and for the memorable skiing adventures we shared.

And last but not least of all Uwe Sauer, for providing me the opportunity to pursue my scientific interest and his patience during the final phase of my thesis work.

I extend my deepest appreciation to each of you for your contributions, camaraderie, and support, which have enriched my journey throughout this PhD.

Contents

1	Introduction	19
1.1	The history and philosophy of systems biology	20
1.1.1	From Aristotle to Galenus	20
1.1.2	The Romans and Middle Ages	21
1.1.3	The Renaissance	21
1.1.4	Mechanistic physiology	21
1.1.5	The scientific revolution	23
1.1.6	Classification of nature	23
1.1.7	Teleomechanics	24
1.1.8	German <i>Naturphilosophie</i>	27
1.1.9	Reductionistic Physiology	29
1.1.10	Holistic Physiology	32
1.2	The contemporary framework of systems biology	35
1.2.1	What is systems biology, anyway?	35
1.2.2	Critique of the "whole is greater than the sum of its parts" notion	36
1.2.3	The emergence of systems biology from bioinformatics	36
1.2.4	Defining systems biology	37
1.2.5	The role of models in systems biology	39
1.2.6	Modes of explanation	40
1.3	Cellular metabolism	44
1.3.1	Significance and molecular foundations	44
1.3.2	The discovery of cellular metabolism	44
1.3.3	Central carbon metabolism	45
1.3.4	Regulation of Metabolism	47
1.3.5	Experimental techniques for studying central carbon metabolism	51
1.3.6	Computational techniques for understanding metabolic flux	51
1.4	<i>Escherichia coli</i> metabolism	53
1.4.1	Transcriptional regulation	53
1.4.2	Allosteric regulation	54
1.4.3	Constraint-based models	55
1.4.4	Kinetic models	56
1.4.5	Experimental challenges	58
1.4.6	Computational challenges	59
1.5	Research goals	60
2	Prediction of allosteric interactions by single reaction modeling	72
2.1	Introduction	73
2.1.1	The regulation of cellular metabolism	73
2.1.2	The identification of allosteric regulation	73
2.2	Materials and Methods	75
2.2.1	<i>In vitro</i> experimentation	75
2.2.2	<i>In silico</i> experimentation	76
2.3	Results	83
2.3.1	Integration of heterogeneous multiomics data	83
2.3.2	Analysis of flux control	83
2.3.3	Systematic analysis of <i>in silico</i> screening results	86
2.3.4	Validation by <i>in vitro</i> enzyme assays	86

2.4	Discussion	92
3	Prediction of allosteric interactions in the tricarboxylic acid cycle using a system of coupled differential equations	98
3.1	Introduction	99
3.1.1	Metabolic regulation	99
3.1.2	Allosteric regulation in <i>E. coli</i> CCM	101
3.1.3	Research objectives	102
3.2	Materials and Methods	104
3.2.1	<i>In vivo</i> experimentation	104
3.2.2	<i>In silico</i> experimentation	106
3.2.3	Model simulation	108
3.3	Results	113
3.3.1	Experimental data	113
3.3.2	Computational modeling	115
3.4	Discussion	126
3.4.1	Differences in methodological approaches	127
3.4.2	Suggestions for future work	134
4	Conclusions and outlook	146
4.1	Conclusions	147
4.2	Outlook	150
4.3	Personal reflections	151
4.3.1	A fractured foundation	151
4.3.2	A dichotomy of motivations	151
A	Appendix	154
A.1	Supplement Chapter 2	155
A.2	Supplement Chapter 3	155
A.2.1	Rate Equations	174
A.2.2	Differential Equations	178

List of Figures

- 1.1 **Metabolic Regulation.** Enzyme abundances can be controlled by transcriptional regulation. This involves transcription of the DNA to RNA, translation of RNA into proteins, as well as protein folding and complex formation, before such regulation is effective. Typical time-scales of these processes are on the order of 1 minute for transcription, 1 minute for translation, and 1 millisecond to 1 minute for protein folding and complex formation. Even though in prokaryotes genes can start to be translated while they're still being transcribed, the process is still estimated to take at least a minute. On the other hand, enzyme activity can also be regulated directly by metabolites. For example, if substrate pools increase, this will increase the thermodynamic driving force of a reaction rate. However, metabolites may also modulate enzyme activity directly. If they are structurally similar to the cognate substrate they may do so by non-covalently binding at the active site, thereby acting as competitive inhibitors. Alternatively, they may do so by non-covalently binding outside the catalytic site, instead binding to a so-called allosteric site, and inducing a conformational change. This latter process is referred to as allosteric regulation, and such metabolites are called allosteric effectors or regulators. Finally, we also consider post-translational modifications, which are covalent modifications, such as phosphorylation, acetylation, methylation and glycosylation, that can also directly affect enzyme activity by altering the protein's conformation. The typical time-scale of the process of ligand-induced conformational change is on the order of a millisecond, whereas that of metabolite turnover is on the order of a second. . . . 48
- 1.2 **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 and green metabolites are activators of the indicated reactions. Adopted from Reznik *et al.* [184]. 55
- 2.1 **Reaction fitting analysis.** (A) Schematic representation the approach in which we compare the measured flux and the predicted flux based on a generalized reversible Michaelis-Menten rate law. The enzyme depicted in red and the metabolites, substrates (S_1 and S_2), products (P) and putative allosteric regulator (A or I) depicted in green. (B) For transketolase (TKT2) the measured (red) and base model predicted fluxes (blue dashed) across nine conditions match well. (C) There exist two isoforms (TktA and TktB) that catalyze the transketolase reaction, which converts E4P and Xu5P to F6P and G3P. (D) For aspartate carbomoyltransferase (ASPCT) the measured (red) and base model predicted fluxes (blue dashed) across nine conditions do not align. A model including allosteric activation by succinate yielded a much better fit to the data (blue solid). (E) A holoenzyme composed of two catalytic trimers (PyrB) and three regulatory dimers (PyrI) catalyzes the aspartate carbomoyltransferase reaction that converts Asp and CbP to CbP-Asp and phosphate. Abbreviations: E4P (erythrose-4-phosphate), Xu5P (xylulose-t-phosphate), F6P (fructose-6-phosphate), G3P (glyceraldehyde-3-phosphate), Asp (aspartate), CbP (carbamoyl-phosphate), CbP-ASP (N-carbamoyl-aspartate), Pi (phosphate), Succ (Succinate). 84

2.2	Confidence boosting of predictions obtained after reaction fitting. (A) The normalized area under the curve (AUC) of receiver operating characteristic (ROC) curves indicates the overlap between the interactions predicted by the fitting procedure with those of other sources of information. Interactions in the NMR and LiP-MS studies, as well as those obtained from literature, are binary: either detected or not. In the case of elasticity coefficients we used a cut off of $ \epsilon > 0.5$. (B) AUCs of the reaction fitting procedure of single regulators and (C) AUCs of the reaction fitting procedure of pairwise combinations of regulators.	85
2.3	Likelihood improvement upon addition of pairwise allosteric regulators across the 84 reactions. The likelihood of the unregulated reaction (red) is shown, as well as the best likelihood upon inclusion of single regulators (orange) and pairwise combinations of regulators (yellow). For each reaction, nine adjacent smaller bars correspond to the nine conditions. Only regulators that improve the akaike information criterion (AIC) with respect to the unregulated model are considered. Reaction abbreviations correspond to BiGG [30] identifiers.	87
2.4	Allosteric regulation of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DDPA). A. DDPA catalyzes the conversion of phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P) to 3-deoxy-D-arabino-heptulosonate 7-phosphate (DHAP) and phosphate (Pi). The three isoenzymes that catalyze this reaction are AroG, AroF and AroH. B. The rate law equation of the reaction with the term for allosteric regulation given in red. C. On top two of the best performing predicted pairwise interactions and below two best performing predicted single interactions, with the measured (red), base model predicted (blue dashed) and regulated model predicted (blue solid) fluxes. D. The FIA-TOF enzyme assay results of 4mM phenylalanine and novel inhibitors of AroG (4mM PEP, 3.4mM indolepyruvate, 4mM 4-hydroxyphenylpyruvate, and 4mM hippurate)) show significant effects, based on a one-tailed T-test. E. Absorbance spectroscopy enzyme assay (1/4 diluted substrates and inhibitors, and 1/2 diluted enzyme with respect to FIA-TOF enzyme assays) and measuring the decrease in absorbance at 232nm by PEP. Significance was determined using a one-tailed T-test: $p < 0.05$ (*); $p < 0.01$ (**); $p < 0.001$ (***)	88
2.5	Allosteric regulation of phosphofructokinase (PFK) and fructose-bisphosphate aldolase (FBA). A. PFK phosphorylates fructose-6-phosphate (F6P) to form fructose-bisphosphate (FBP). Two known isoforms are PfkA and PfkB. Two known inhibitors are citrate and FBP [6]. B. <i>in vitro</i> FIA-TOF enzyme assay results for PfkA show significant inhibition by controls (citrate and fbp) and indolepyruvate. C. <i>in vitro</i> FIA-TOF enzyme assay results for FbaA. The controls here are PEP and 3PG, a reported activator and inhibitor, respectively. Conflicting previous reports we detect an activatory effect for 3pg on FbaA. D. FBA catalyzes the subsequent reaction in glycolysis involving the cleavage reaction of FBP to produce DHAP and G3P. Two known isoforms are FbaA and FbaB. Known regulation includes the activation by PEP [4] and inhibition by 3PG [61] Significance was determined using a one-tailed T-test: $p < 0.05$ (*); $p < 0.01$ (**); $p < 0.001$ (***) . Error bars indicate the standard error of the mean across triplicates.	90
2.6	Allosteric regulation of phosphoserine phosphatase (PSP_L) and adenylosuccinate synthetase (ADSS). (A) Phosphoserine phosphatase catalyzes the dephosphorylation of phosphoserine, which constitutes the last step in serine biosynthesis. The only known isoenzymes is SerB and the only known allosteric regulator is L-serine, which is also the product [52]. (B) <i>in vitro</i> FIA-TOF enzyme assay results for SerB show significant inhibition by the control (L-serine) and the newly predicted inhibitor hypoxantine. (C) <i>in vitro</i> FIA-TOF enzyme assay results for PurA. Fumarate and IMP are known inhibitors used as controls. For all three newly predicted inhibitors, AMP, acetyl-glutamate and pantothenate, we see a significant inhibitory effect. Note that none of these enzyme assays was calibrated. (D) Adenylosuccinate synthetase catalyzes the first committed step toward the <i>de novo</i> synthesis of AMP from IMP. The only known isoenzyme is PurA. The known allosteric inhibitors are fumarate and IMP [20, 7]. Significance was determined using a one-tailed T-test: $p < 0.05$ (*); $p < 0.01$ (**); $p < 0.001$ (***)	91

3.1	Tricarboxylic acid cycle of <i>E. coli</i>. Model boundary species are extracellular: pyruvate (PEP), acetate (Acetate), succinate (Succinate), phosphoenolpyruvate (PEP), adenosine triphosphate (ATP), adenosine diphosphate (ADP), nicotinamide adenine dinucleotide (NAD), reduced nicotinamide adenine dinucleotide (NADH), nicotinamide adenine dinucleotide phosphate (NADP), reduced nicotinamide adenine dinucleotide phosphate (NADPH), L-glutamate (Glu), and D-lactate (Lac). The graphic is created with the use of Escher [62]. Note: Protons (H ⁺) and water (H ₂ O) are ignored in reaction kinetics.	110
3.2	Sampled initial metabolite concentrations. The sampling range of the generic prior for metabolites without data is shown in grey (Range). The empirically determined actual sampling range is shown in black (Sampled). Metabolites for which there was data available was constrained to three standard deviations from the mean as reported by Link <i>et al.</i> [75]. For comparison the data obtained by Gerosa <i>et al.</i> and Kochanowski <i>et al.</i> [37, 63] is shown in blue.	118
3.3	Best one percent of predictions without allosteric regulation. On top of the simulated time series depicted for each of the metabolite the geometric mean and standard deviation of the log residuals is provided.	122
3.4	Top ranking allosteric interaction predictions. The results of all single and pairwise interaction models was combined in order to derive a ranking for each of the individual interactions tested. Abbreviations: isocitrate dehydrogenase (ICDHr), phosphoenolpyruvate carboxylase (PPC), phosphoenolpyruvate carboxykinase (PPCK), sedoheptulose 7-phosphate (s7p), 6-phospho-D-gluconate (6pg), citrate (cit), isocitrate (icit), L-tyrosine (tyr), L-phenylalanine (phe), D-xylulose 5-phosphate (xu5p), cis-aconitate (acon).	123
3.5	Best one percent of predictions with sedoheptulose 7-phosphate (s7p) inhibiting isocitrate dehydrogenase (ICDHr). On top of the simulated time series depicted for each of the metabolite the geometric mean and standard deviation of the log residuals is provided.	124
A.1	Calibration curves to correct the effect of the regulation presence in the FIA-TOF. A. Duplicates of calibration curves of the FbaA products glyceraldehyde-3-phosphate and dihydroxyacetone-phosphate (1:1). B. Duplicates of calibration curves of the PfkA product fructose-1,6-bisphosphate. C. Duplicates of calibration curves of the SerB product serine. The ratio of the slope between curves with and without the regulator present is used as the correction factor.	155
A.2	A model without allosteric regulation can explain a lot of the fluxes in 62% of the reactions. The Pearson correlation coefficient indicates correspondence of observed and predicted fluxes in nine different conditions. Reactions for which $R^2 \geq 0.35$ are shown in green, those below in red.	157
A.3	correlations among the metabolites across nine conditions. In our reaction fitting procedure high correlations among metabolites that are tested as putative regulators will result in similar prediction scores, and hence reflect chances of false positive results of predictions are made solely using this data to differentiate candidate regulators.	158
A.4	Most fluxes and protein concentrations tend to depend linearly on the growth rate. A. Correlation between protein concentrations of the 84 analyzed reactions and growth rate. B. Correlation of fluxes with growth rate. These linear dependencies are used to inter/extrapolate proteomic and fluxomic values and compare data at the same growth rate.	158
A.5	Correlation between ¹³C flux analysis and CCM fluxes determine by FBA. Each figure plots the two flux sources against the growth rate across the eight glucose-limitation (blue) and eight glutamate-limitation (red) conditions. The correspondence of FBA-derived (circles) and ¹³ C flux analysis (diamonds) is high for the growth rate and the 26 measured CCM fluxes.	159

A.6	Likelihood improvement upon addition of single allosteric regulators across the 84 reactions. The likelihood of the unregulated reaction (red) is shown, as well as the best likelihood upon inclusion of single regulators (orange). For each reaction, nine adjacent smaller bars correspond to the nine conditions. Only regulators that improve the akaike information criterion (AIC) with respect to the unregulated model are considered. Reaction abbreviations correspond to BiGG [6] identifiers.	160
A.7	AroG absorbance enzyme assay with 1mmol of 4-hydroxyphenylpyruvate and 250μmol erythrose-4-phosphate. The absorbance indicates consumption of 4-hydroxyphenylpyruvate suggesting that it acts as a competitive substrate. The lines correspond to three biological replicates with an average slope of $-4.765e4$ units s^{-1} , which is significantly different from the slope observed in the absence of enzyme $-6.276e5$ units s^{-1} (One-tailed t-test p-value < 0.001).	161
A.8	Central metabolism of Escherichia coli. Biochemical pathways that constitute central metabolism from the genome scale metabolic model iJO1366 [12]. The graphic is created with the use of Escher [7].	162
A.9	Central metabolism of Escherichia coli. Biochemical pathways that constitute central metabolism from the core model [11]. The graphic is created with the use of Escher [7].	163
A.11	Flux projection of steady-state pyruvate data. Data from [3] were projected on the metabolic model depicted in Figure 3.1 using known as minimization of metabolic adjustment (MOMA) [17]. On the left hand side a plot of projects versus reported flux data. Residuals with respect to the original data are shown on the right hand side.	164
A.10	Steady-state fluxes during growth on pyruvate projected onto the TCA model demarcated in this study. These steady-state fluxes are part of the initial conditions of the system prior to switching to glucose-rich medium. Data are from [3].	164
A.12	Catalytic constants of the TCA cycle. Data are obtained from the BRENDA Enzyme Database [16]. The data covers 11 catalytic constants and for the majority of these there is only a single literature reported value available.	164
A.13	Michaelis-Menten constants of the TCA cycle. Data are obtained from the BRENDA Enzyme Database [16]. The data covers 45 dissociation constants and for the majority of these we obtain multiple values from the published literature.	165
A.14	Equilibrium constants of the TCA cycle. Estimates were obtained using the eQuilibrator biochemical thermodynamics calculator [1]. Reactions with a negative Gibbs energy of formation are thermodynamically favorable and indicated in green. Those with a positive Gibbs free energy are thermodynamically unfavorable.	165
A.15	Metabolite concentrations of the TCA cycle during steady-state growth on pyruvate. Data shown are from [9].	165
A.16	Cell volume of <i>E. coli</i> grown on pyruvate. Data were obtained from [18] and the kernel density estimate was plotted to highlight the issue of computing mean and standard deviation in linear scale, and not supplying the raw data.	166
A.17	Sampled enzyme concentrations. Absolute enzyme count data were obtained from [15] and cell volume data from [18]. The result of reporting only the mean and standard deviation on linear scale results in skewed sampling results, with unrealistically high enzyme concentrations.	166
A.18	Analysis of best peaks from the pyruvate-glucose-pyruvate shift experiment. Samples were measured as biological triplicates from independent experiments, each of which was injected twice from the same well on the same 96-well plate to assess technical variability.	182
A.19	Pyruvate-glucose-pyruvate switch. The black dots and error bars indicate the mean and standard deviations of the measurements, the red dashed line is a linear interpolation of the data. Data were obtained from [9].	184

-
- A.20 **Acetate-glucose-acetate switch.** Among the 11 time points with 3 replicates each, there were 24 out of 65 metabolites for which more than half of time points could be integrated from the mass spectrometry data. For those metabolite present in the model, this covers only pyruvate, ADP and AMP. We could integrate the majority of metabolites to some extent despite the loss of half the data points. For those in the model, this includes, for example, PEP, citrate and alpha-ketoglutarate. For others such as Succinate, SucCoA, Fumerate and Malate, there was only one replicate that could be integrated. Squares, circles and triangles indicate individual biological replicates, yellow lines are non-normalized times series, blue lines are ¹³C-normalized time series, dashed lines are relative quantification data, solid lines are absolute quantification data computed from the calibration curves, and dotted-lines are linear interpolation of the initial time point and the time point of the control sample where the cells were perfused by not perturbed. 185
- A.21 **Pyruvate-glucose-pyruvate switch.** Among the 11 time points with 3 replicates each, there were 35 out of 65 metabolites for which more than half of time points could be integrated from the mass spectrometry data. For those metabolite present in the model, this covers aconitate, ADP, AMP, malate and PEP. A more detailed analysis of variance in the data with respect to tehcnical replicates, biological replicates, and the effect of ¹³C normalization, can be found in Figure A.18. Squares, circles and triangles indicate individual biological replicates, yellow lines are non-normalized times series, blue lines are ¹³C-normalized time series, dashed lines are relative quantification data, solid lines are absolute quantification data computed from the calibration curves, and dotted-lines are linear interpolation of the initial time point and the time point of the control sample where the cells were perfused by not perturbed. 186
- A.22 **Succinate-glucose-succinate switch.** Among the 11 time points with 3 replicates each, there were 35 out of 65 metabolites for which more than half of time points could be integrated from the mass spectrometry data. For those metabolite present in the model, this covers aconitate, ADP, AMP, (Iso)citrate, PEP and succinate. Squares, circles and triangles indicate individual biological replicates, yellow lines are non-normalized times series, blue lines are ¹³C-normalized time series, dashed lines are relative quantification data, solid lines are absolute quantification data computed from the calibration curves, and dotted-lines are linear interpolation of the initial time point and the time point of the control sample where the cells were perfused by not perturbed. 187

List of Tables

1.1	Well-known transcription factors (TF) involved in regulating <i>E. coli</i> metabolism. .	54
2.1	Multi-omics data used in this study. Nine conditions, four from a carbon titration and five from a nitrogen titration, from three experiments conducted in two different studies were combined. The metabolomics data and fluxomics data are obtained from Kochanowski <i>et al.</i> (2021) [33], whereas the proteomics data are obtained from Hui <i>et al.</i> (2015) [27]. The growth rates are given in units of h^{-1} . The induction agent used in all of the carbon source titration experiments is 3-MBA, and in all nitrogen titrations IPTG was used, both of which are given in units of μM . Strain NQ381 achieves variable lactose uptake by regulating LacY expression [63]. Strain NQ393 has a glutamate dehydrogenase (<i>ghdA</i>) deletion and an IPTG inducible glutamate synthase (<i>gltBD</i>) [27, 55]. Strain NQ1243 has a 3-methylbenzyl alcohol (3MBA) titratable <i>ptsG</i> expression system to for controlled glucose uptake [5]. Strain NCM3722 is a glutamine prototrophic strain derived from <i>E. coli</i> K-12 [60].	77
2.2	Predicted regulatory interactions. Five reactions, their respective isolated isoenzyme and the predicted regulatory effectors tested in enzyme assay validation experiments. In green those for which we find support in <i>in vitro</i> enzyme assays, in orange those for which we find experimental evidence for the opposite mode of action as predicted. The astrix indicates that a calibration curve was used to correct for matrix effects. 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DDPA), phosphofructokinase (PFK), fructose-bisphosphate aldolase (FBA), phosphoserine phosphatase (PEP_L) and adenylosuccinate synthetase (ADSS).	86
3.1	Condition-dependent cell volume of <i>E. coli</i> . The units are: h^{-1} growth rate, μm cell length, μm cell width, fL single cell volume, 10^8 cells $\text{mL}^{-1}\text{OD}^{-1}$ OD-specific cell concentration, and $\mu\text{L mL}^{-1}\text{OD}^{-1}$ OD-specific total cell volume. Adapted from [121].	107
3.2	Model data coverage	115
3.3	Overview of numbers relating to the models and allosteric search space of previous and current studies. The glycolytic gluconeogenic switch in glycolysis was studied in Link <i>et al.</i> (2013), the effect of reactive oxygen species on the penthose phosphate pathway (PPP) in Christodoulou <i>et al.</i> (2018), and the dynamics resulting from a pyruvate to glucose switch on tricarboxylic acid (TCA) cycle metabolism in this study.	134
A.1	Reaction data coverage	156
A.2	<i>In vitro</i> enzyme assays	157
A.3	Genes present in the <i>E. coli</i> TCA cycle model. These gene identifiers were used to map absolute protein quantification data from Schmidt <i>et al.</i> [15] to the model.	167
A.4	Gene reaction rules of the <i>E. coli</i> TCA cycle model. Gene-reaction-rules represent how genes, via the proteins they encode, are associated with metabolic reactions. These rules are represented using Boolean expressions: 'or' signifies that either (set of) gene can catalyze the reaction, while 'and' mandates the presence of both (sets of) genes.	168

A.5	Reactions in the <i>E. coli</i> TCA cycle model. These reaction identifiers were used to retrieve data from the BRENDA database [16]. Reactions are BiGG identifiers [14]. Other identifier databases: metanetx [2], ec-code [19], kegg [5]. The following annotation are omitted here for clarity: SBO, biocyc, reactome, rhea.	169
A.6	Reaction formulae of the <i>E. coli</i> TCA cycle model. This is the model used for flux balance analysis from which the kinetic model is derived. Reactions prefixed with 'EX_' are exchange reactions used for flux balance analysis. Metabolites are suffixed with a compartment identifier: '_c' for cytosol and '_e' for extracellular.	170
A.7	Metabolites in the <i>E. coli</i> TCA cycle model. These metabolite identifiers were used to retrieve data from the BRENDA database [16]. Metabolites are BiGG identifiers [14]. Other identifier databases: seed [4], kegg [5] and metanetx [2]. The following annotation are omitted here for clarity: reactome, chebi, hmdb, biocyc, lipidmaps.	171
A.8	Metabolite formulae and charge of the <i>E. coli</i> TCA cycle model. Metabolites are BiGG identifiers [14]. Boundary condition indicates whether or not the metabolite was considered as such in the kinetic model. Rule assigned denotes whether a species were modeled using a linear interpolated rule and experimentally determined time series data.	172
A.9	Small Molecule Regulatory Network Table of <i>E. coli</i>. This table presents data extracted from Reznik et al.'s supplement [13], featuring columns detailing biochemical reactions (Reaction), associated metabolites (Metabolite), Enzyme Commission numbers (E.C. number), Kyoto Encyclopedia of Genes and Genomes IDs (KEGG ID), mode of action (Mode), and the regulatory mechanisms (Mechanism). Only those reactions and metabolite that are part of our study are presented, using their respective BiGG identifiers.	173
A.10	Generior prior on states and parameters. These values were taken from Lubitz <i>et al.</i> [10]. Symbols: k_{catf} : forward catalytic constant, k_{catr} : reverse catalytic constant, K_m : Michaelis-Menten constant, K_i : inhibition constant, K_a : activation constant, K_{eq} : equilibrium constant, u : enzyme concentration. c : metabolite concentration.	179
A.11	Enzyme counts on pyruvate for reactions in the <i>E. coli</i> TCA cycle model. Enzyme counts for each reaction were estimated using the absolute protein quantification data from Schmidt <i>et al.</i> [15] in conjunction with the gene-reaction-rules present in the model (See Supplementary Table A.4) [12]. These counts were aggregated on a per-reaction basis using these boolean rules: AND indicates both proteins are needed, hence we took the minimum, whereas OR indicates both can be used, and hence summation was performed, in order to approximate to total number of enzymes that could catalyze a given reaction.	180
A.12	Metabolite concentrations on pyruvate for species in the <i>E. coli</i> TCA cycle model. Metabolite concentrations were estimated by combining and averaging the data from Gerosa <i>et al.</i> and Kochanowski <i>et al.</i> [3, 8].	180
A.13	Steady-state fluxes on pyruvate for reactions in the <i>E. coli</i> TCA cycle model. Fluxes data were obtained from Gerosa <i>et al.</i> [3].	181
A.14	Standard Gibbs free energy for reactions in the <i>E. coli</i> TCA cycle model. Standard Gibbs free energy change (ΔG°) values and their corresponding standard deviations. These values were acquired using Equilibrator [1], where we assumed a constant temperature of 37°C, an intracellular pH of 7.4, and an ionic strength of 0.25M [21, 20]. We sampled the covariance matrix, which is not shown here for convenience. Units are kJ/mol.	181
A.15	Forward catalytic rate constants for reactions of the <i>E. coli</i> TCA cycle model. Data were obtained from the BRENDA database [16].	182
A.16	Michaelis-Menten constants for reaction-metabolite pairs in the <i>E. coli</i> TCA cycle model. Data were obtained from the BRENDA database [16].	183

Chapter 1

Introduction

M.A.P. Karrenbelt wrote the chapter.

1.1 The history and philosophy of systems biology

Systems biology is often considered to have sprung as a field of science at the beginning of 21st century. As such, the advent of the field coincides with the completion of the human genome project, a time in which the notion that biology could not be comprehensively understood from static sequencing data alone became more pervasive. As a logical consequence, people started to devote more attention to the behavior and dynamic properties of biological systems that result from the interactions of its constituent parts. Having spent much time working from a methodological reductionist framework in which molecules were studied in isolation, an antagonistic perspective known as holism came to be readily adopted in the study of complex molecular interaction networks that ensued.

Attempts to define systems biology often yield nebulous descriptions that provide little meaningful insight into its distinguishing features as a scientific discipline. That systems biology is based on the understanding that "the whole is greater than the sum of the parts"¹ is a platitude that one will inevitably encounter and one that characterizes the esoteric nature of the so-called holistic approach adopted by scientists in the field. We will scrutinize this notion in an attempt to define what systems biology is, and will come to see that even among scientists in the field there exist no consensus. However, before we attempt to define what systems biology is, we will look into its historical and philosophical underpinnings. A stroll through past events will provide us an understanding of the context of the field in terms of preceding and contemporary paradigms that are, in a way, the ancestors that gave birth to the discipline.

1.1.1 From Aristotle to Galenus

The Greek philosopher and naturalist Aristotle was the first of whom we know that he involved himself with the systematic study of the phenomenon of life. Contrasting the vision of Plato, who held that the earthly reality was a false world, a defective copy of an unchanging eternal and perfect world, Aristotle held the believe that the transient elements in the sublunary deserved as much attention as the – then believed to be – eternal and unchanging celestial bodies. Aristotle was not only a pioneer in his choice of living beings as subject of research, but also on a methodological level he brought about innovation in science. Virtually all Greek minds regarded the human mind as the primary source of knowledge of nature, for only the spirit could penetrate into the eternally unchangeable world. Aristotle, on the other hand, placed great importance on sensory perception and thought that the data this yielded should take precedence over the outcomes of rational reasoning. Furthermore, he saw the function (the goal) as the main cause for the existence of a (part of an) organism [171].

Although Aristotle himself appears to have conducted few experiments, Alexandrian researchers that succeeded him – such as Theophrastus around the end of the 4th century B.C. as the center of Greek culture shifted from Athens to the Nile Delta – were among the first that systematically conducted *in vivo* experiments by performing vivisection. This Aristotelian style of research came to an end with Aelius Galenus – around the end of 2nd century – who was the last great life scientist of this tradition, and who is famous for his use of direct observation, dissection and vivisection in order to study physiological processes [170, 171].

Aristotle's pioneering engagement with systematic study of life and biological phenomena laid foundational principles that resonate in modern systems biology. His emphasis on empirical observation challenged conventional philosophical paradigms and fostered an evidence-based approach, aligning with systems biology's reliance on data-driven analysis. Galen, in turn, established a legacy of empirical investigation through methods like vivisection and direct observation, prefiguring the systemic approach embraced by contemporary systems biology. These historical underpinnings set the stage for understanding complex biological systems as interconnected networks, a hallmark of both Aristotle's and Galen's philosophies and modern systems biology.

¹The quoted phrase here is often falsely attributed to Aristotle, however it originates from Gestalt psychology [226]. In Aristotle's *Metaphysics* we find the following: πάντων γὰρ ὅσα πλείω μέρη ἔχει καὶ μὴ ἔστιν ὅσον σωρὸς τὸ πᾶν ἀλλ' ἔστι τι τὸ ὅλον παρὰ τὰ μέρη. which more accurately translates to "For however many things have a plurality of parts and are not merely a complete aggregate but instead some kind of a whole beyond its parts" [186]

1.1.2 The Romans and Middle Ages

The Roman natural sciences, and in particular biology, were of a considerably lower quality than that of the Greeks. The Romans were mainly interested in the practical use, and instead of observing nature themselves, as their predecessors had done, the Roman researchers limited themselves to summarizing what others had written before them. They generally worked quite superficially. Without too much critical reflection, they collected all kinds of data from very diverse sources. They hardly made any attempt to link the separate facts or to theorize. All in all, Roman science has added very little to the existing knowledge of living nature. Also in the Middle Ages, biology remained what it had become in Roman times, namely a science based on literature research [133]. Their treatises were essentially nothing more than collections of excerpts from the works of Aristotle, Galenus, and others who were regarded as recognized authorities. They derived additional information from all kinds of literature, ranging from poetry to the bible. In general, there was little need to verify data. As a result, medieval biology appears to be a peculiar mixture of fact and fantasy, which is well exemplified in the work of Conrad Gessner, specifically his *Tierbuch* and *Fischbuch* [175, 74].

As biology underwent limited growth during the Roman and Middle Ages, marked by a utilitarian approach based on literature research, a paradigm shift emerged with the Renaissance. This era saw a resurgence of empirical inquiry and a departure from tradition, embracing methods that laid the groundwork for the modern scientific approach. This transition from passive compilation to active observation set the stage for a renewed exploration of the natural world.

1.1.3 The Renaissance

Nearly two thousand years after Aristotle's arguments for the importance of sensory perception, empiricism again became the methodical foundation of biology during the Renaissance. A first indication of this can be found in the drawings made by Hans Weiditz in the treatise of the German doctor Otto Brunfels, who broke with the habit of copying existing images and started using plants as examples for his drawings [38, 80]. Brunfels was so in awe of the writers from antiquity that he apologized for describing a plant not previously described by them. Almost a decade later Hieronymus Bock published a book on botany in which he would only describe plants that he himself had observed [22, 146]. It is worth noting that he had no qualms about rejecting the findings of his predecessors when they were inconsistent with his own observations and conclusions. Bock was one of the first researchers in biology to break free from this rule of tradition. As in the Middle Ages, treatises on plants and animals were of a highly utilitarian nature. It was not until the end of the sixteenth century that this started to change, as we can see in botanical books by Andrea Cesalpino and Adam Zaluziansky, both of whom held the ideal to transform botany into a independent field of science that would concentrate solely on acquiring knowledge of the essential structure and properties of plants, rather than their practical applications in medicine [47, 83]. Knowledge of forms of plants and animal species was considerably expanded by the critical empirical approach. The situation was completely different with regard to research into the functions of living organisms. Here, theories of Aristotle and Galenus prevailed and no substantial innovation took place. In this animistic view of nature it was assumed that the activities in plants and animals were produced and guided by immaterial forces, often referred to as the soul [93]. With these vitalist theories all kinds of life phenomena could be explained quite easily, but the empirical basis of such theories was slim, and criticism of the speculative character of the explanations offered would soon follow.

1.1.4 Mechanistic physiology

In the sixteenth century, vitalist explanations also prevailed in other natural sciences. In the early seventeenth century scientists who reasoned from a mechanistic conception of nature started attacking this notion. The French philosopher, mathematician and scientist René Descartes is the most famous one among these. His criticism of vitalist theories found its philosophical justification in the what is called substance dualism. The thinking and consciousness of man were, in his view, the only phenomena in which the soul played a part; the rest of nature consisted of nothing but matter. He considered all processes that took place to be explainable on the basis of mechanical interactions between elementary particles. The matter from which animals and plants are formed was, according to Descartes, inert in that it has no source of motion of its own. At the beginning of

time, matter had brought with it a certain amount of motion. By collision, the elementary particles constantly transferred that motion to each other. His theory was both mechanistic and reductionist: in order to understand the phenomena observed at macro-scale, one had to understand the particles and their interactions at micro-scale. He formulated the notion that complex systems can be reduced to their parts, which can each be examined in isolation, and subsequently reassembled in order to understand their behavior. What was hitherto considered impossible, Descartes regarded as an irrefutable truth: the soul is not necessary to understand the phenomenon of life [66, 32].

It is worth noting that although this approach gained widespread support, the majority of his colleagues held more moderate views. Giovanni Alfonso Borelli serves as a good example. In the preface to his work *De motu animalium*[28] he stated that it was natural to regard the soul as the ultimate primary cause of motion. While Borelli assumed the existence of a soul as the cause of life-phenomena, for him this starting point functioned exclusively as a philosophical concept. He did not include the soul in the field of research of the physiologist. As a researcher, Borelli worked only with the direct so-called secondary causes which were purely physical and chemical in nature. As a result, his theories were no less mechanistic than those of Descartes.

Mechanistic reasoning has strongly influenced the production of biological knowledge. It was a more or less obligatory framework for the interpretation of physiological observations and largely determined which theories were and were not accepted by the scientific community. Illustrative of this influence are the developments in embryology by the seventeenth century biologists. Initially they worked with the epigenesis theory, a theory that was mainly based on observations of incubated chicken eggs that were opened at successive time intervals. In doing so, one first saw the heart appear and then gradually the other organs. From this he drew the logical conclusion that the formation of the embryo was a process in which the various organs came into existence in the course of time.

1.1.4.1 Preformation theory

In the second half of the seventeenth century the rivaling preformation theory was formulated. The proponents of this theory denied that the embryological development was characterized by difference and growth as the epigeneticists claimed. Jan Swammerdam, one of the founders of preformation theory, had done extensive research into the reproduction of insects. In butterflies he discovered parts of the adult animal that were already present in the caterpillar that was about to pupate. Although this observation related only to part of the developmental process, Swammerdam stated that the adult animal must also be found in the ovum. According to their theory, the organism was already complete, but present in miniature form in an egg or sperm cell [29]. The answer to the question of the origin of this preformed creature was that it had existed since the dawn of time; the first individual of each species had contained its entire progeny. Neither he nor other researchers of the time had the means to empirically substantiate this conclusion and therefore Swammerdam indicated that further research was needed to definitely demonstrate the correctness of the theory.

Compared to the epigenesis theory, the preformation theory had a weak empirical basis. Moreover, there were all kinds of phenomena that the theory could not account for. The existence of monstrosities and the fact that the offspring could show characteristics of both the father and the mother could not be satisfactorily explained by the preformation theory. Given these shortcomings, one would expect the preformationists to have lost out in their battle with the epigeneticists. The opposite is the case; the preformation theory soon became the dominant theory of embryology. Most seventeenth-century researchers did not think it was necessary to keep a reservation with Swammerdam's theory and unreservedly supported it [29]. Clearly, the theory did not derive its appeal from the power of its observations. The decisive factor was its compliance with the mechanistic principles. The gradual formation and development of organs, as the epigenesis theory posited, required the assumption of a guiding principle that coordinated development, according to most researchers. The theory thus seemed to necessitate a vitalist conception of life's processes. The preformation theory, on the other hand, fit in perfectly with the dominant mechanistic philosophy of nature. The growth of already present and preformed parts could be represented mechanistically without much difficulty. Although the requirements of empiricism were of paramount importance to the seventeenth-century biologists, in this case they were still subordinate to the prevailing view of nature. Notwithstanding, this mechanistic philosophy has been of great significance, the fruits of which will be harvested in the nineteenth century.

1.1.5 The scientific revolution

The sixteenth and seventeenth centuries are the era of the scientific revolution for the history of science. Natural science broke away from the Greek tradition and the foundations were laid for modern scientific practice. Rationalistic thinking, in which the reasoning mind had primacy, gave way to empiricism that gave priority to perception. The experiment, considered artificial by the Greeks, became a valid source of knowledge. The mechanistic worldview began to be adopted, whereby the world was no longer seen as a living organism but as a machine. In physics and astronomy this development, through the work of men like Galileo, Kepler, Huygens, and Newton, led to the most remarkable results.

The first major change in biology dates from the sixteenth century and concerned the reintroduction of empiricism: people began to collect observations themselves again and no longer relied solely on surviving knowledge. In the second half of the seventeenth century, the mechanistic view of nature and the principle of uniformity became the most important presuppositions in the study of living nature. However, the break with the past was not as great as in the physical sciences. From the classical heritage, teleology and essentialism continued to play an important role in certain parts of biology [93]. In addition, a fundamental characteristic of the changes that occurred in the physical sciences, namely the mathematization of the acquired knowledge of nature, is almost absent in the research of living nature. Until the nineteenth century there was a lack of a universal theory such as Newton formulated it for physics. Biology would only get a theory of comparable scope with Darwin's theory of evolution. The thread connecting the era of scientific revolution with later periods is formed by the basic ideas that developed in the sixteenth and nineteenth century: the empirical research method, the mechanistic model, the principle of uniformity and the idea of the existence of a natural order in the plant and animal world.

1.1.6 Classification of nature

In the first edition of the *Encyclopaedia Britannica* (1771), natural history was defined as the science that gives classifications and complete descriptions of the products of nature. Carolus Linnaeus (1707-1778) and Georges-Louis Leclerc de Buffon (1707-1788) were the leading representatives of the classifying and descriptive-explanatory schools of eighteenth-century natural history, respectively.

1.1.6.1 Linnaeus's system

Linnaeus' system was founded on essentialism as we have already encountered in Cesalpino, to whom he wholeheartedly acknowledged his tributary debt. Following Cesalpino and stimulated by the discovery of the sexuality of plants, Linnaeus regarded the reproductive organs as the suppliers of the essential criteria of botanical systematics. Here he was strengthened by the firm conviction that God had decreed that plants should be distinguished by fructification, as he called the parts involved in reproduction. All other structures flowed logically from the essence and were seen as of secondary importance. In the definition of a species, genus or other group one should, according to Linnaeus, limit oneself to the essential characteristic. Description of characteristics derived from, for example, root, leaf or stem was not only superfluous, but even appears to Linnaeus as contrary to the interests of the system. The essence was a predetermined classification principle. The science thus created resembled an exercise in traditional logic rather than empirical research of nature. That was the intention. In his theoretical reflections, Linnaeus was somewhat contemptuous of purely empirical research and argued that the scientific status of systematics was determined first and foremost by its rational-deductive character [206]. Linnaeus has succeeded in his aim of making systematics a recognized and mature discipline. His action resulted in the great majority of eighteenth-century natural history focusing on this branch of science. People accepted Linnaeus' methods and aims, but showed little receptivity to the theoretical framework in which he had placed them.

1.1.6.2 Buffon's perspective

Linnaeus's work was strongly criticized by Buffon. Buffon's critical attitude towards the systematists was fueled by the conviction that they were more concerned with reasoning based on *a priori* assumptions than with unbiased observations of phenomena. As a self-declared supporter of the empirical ideal of science, he was against such an approach; true knowledge of nature could

never be obtained with this rationalistic method. He credited Linnaeus for having put an end to the chaos in nomenclature and that classifications generally provided a convenient overview. To Buffon, however, the system was clearly no more than a useful tool, he did not see it as an essential adjunct to natural history. In his view, the organization of nature was characterized by gradual transitions and by the absence of created divisions. Unlike systematists, he did not define species by form features; for Buffon, a species consisted of individuals capable of reproducing and producing fertile offspring. Groupings other than species, such as genera, orders, and classes, which systematists thought they could distinguish, existed, in his view, exclusively in their minds. Remarkable is Buffon's emphasis on the importance of anatomy to natural history. While natural historians had until then focused almost exclusively on the external structures, Buffon argued that these were only of secondary importance. He wanted to arrive at general theories based on empirical details. The ultimate goal was to understand the more fundamental principles behind natural historical phenomena. For Buffon, those principles lay in the processes responsible for the origin and maintenance of life forms.

The majority of eighteenth-century natural historians found the answer to the question of the origin of the life forms in the bible. Linnaeus summed up the prevailing view in a famous aphorism: "There are as many species as the infinite being created diverse forms in the beginning, which, following the laws of generation, produced many others, but always similar to them: therefore there are as many species as we have different structures before us today" [136]. Buffon came up with a radically different vision. He suggested that the emergence of new levels could be explained entirely from the properties of matter. For this he needed a different understanding of matter than that which had been used in the life sciences since Descartes. The inert particles of matter, widely regarded as the elementary building blocks of natural objects, Buffon replaced for the plant and animal world with what he called "organic molecules" [42]. He assigned this type of atomic particle its own force, the *force penetrantes* [41], with which they could, without external influence, bring about the characteristic processes of life. He further assumed that life had begun in a primordial sea, and suggests that some chemical process gives rise to organic molecules in this primordial sea, which join together to form living beings. Although he was not an evolutionist, Buffon's genesis theory gave a strong dynamic picture of nature. The essence of his theory is well captured in his own words: *le grand ouvrier de la nature est le temps* [42].

1.1.7 Teleomechanics

Just as Newton had established the laws of the planetary system, so also for living nature a system of laws describing its functioning had to be formulated. An indication of this need for a comprehensive 'science of living things' is provided by the introduction of the term biology by the Gottfried Reinhold Treviranus (1776-1837), a physician and comparative anatomist from Bremen. Treviranus outlined the research program that was to give biology the status of a full-fledged Newtonian science. He was an exponent of the so-called teleomechanistic tradition in natural research, which originated in Göttingen. Unlike most universities in the eighteenth century, scientific research at the University of Göttingen occupied an important place. The professors were expected not only to be engaged in teaching, as was customary elsewhere, but also in scientific research. By the end of the century, a research program had been developed that united the mainstreams of eighteenth-century thinking about living nature and synthesized it in a unique way. The founders of this program were the comparative anatomist Johann Friedrich Blumenbach (1752-1840) and the philosopher Immanuel Kant (1724-1804). They initially worked independently of each other – Blumenbach in Göttingen and Kant in the Prussian Königsberg – but after becoming acquainted with each other's writings, an interaction was established. Kant laid the theoretical foundation of the teleomechanistic tradition, while Blumenbach and his followers focused on the biological elaboration hereof.

1.1.7.1 Kant's teleological distinction

Kant's contribution is mainly contained in his *Kritik der Urteilskraft* (1790) [112]. In this work he attempted to formulate the philosophical foundations for the study of living nature, as he had done for inanimate nature in an earlier work, the *Kritik der reinen Vernunft* (1781) [111]. According to Kant, there was a fundamental difference between the organic and the inorganic world. Mechanistic explanatory models used in inorganic natural research were inadequate for the study of living things. The core of the difference lay in the character of the causal relations of

animate nature. In inanimate nature there are linear chains of cause and effect: phenomenon A causes phenomenon B, B then causes C, C leads to D, and so on. In animate nature, on the other hand, we find chains of the form A causes B, B causes C, and then C causes A. Here the clear distinction between cause and effect suddenly disappears: A causes C (via B), but is also itself caused by C. A is therefore both cause and effect of C. Consider, for example, the relationship between food intake and muscular labor. Digestion is a cause of the muscle's ability to work, because the required energy comes from digestion. But in its turn, muscular labor is a cause of digestion, for the work of the muscles is indispensable in the absorption and processing of food – each one is thus both cause and effect of the other.

In Kant's view, this kind of causal relationship was characteristic of living beings. Every part of the organism was dependent on all the other parts and thus on the whole animal or plant. It was due to this interaction that the organism formed an effectively functioning whole. In other words: because the life processes proceeded in mutual dependence, they seemed to be guided by an intention, a goal, namely the adequate exercise of the bodily functions. This teleological aspect was absent in inanimate nature. For the human mind, only explanatory models that proceed according to the linear cause-effect scheme are comprehensible. The full understanding of teleological causation is beyond our intellectual capabilities. For the mind can understand how teleological processes work, but not how they arise, how they are produced by nature. After all, if A is the cause of C, A does explain the occurrence of C, but the existence of A itself requires the existence of C to be presupposed, so it remains unclear how the process ever started. Since the mind cannot comprehend the origin of the process, teleological causation, and hence the functional organization of living beings, can only be accepted as a starting point. In fact, we may not even say, as Aristotle and many after him did, that the organism is purposeful. It presents itself to us as purposeful, it behaves as if it is driven by intentions, but we can never know for sure whether it is 'really' so. Research of living nature, says Kant, will always have to work with the restriction that the functional organization is a given that we simply have to accept before the empirical investigation of the organism can begin. This clarifies the term teleomechanics: the study of living nature had to be based on mechanistic explanatory models, but the purposive character of life had to be accepted as an axiom [126, 127].

1.1.7.2 Blumenbach and the *Bildungstrieb*

Johann Friedrich Blumenbach studied medicine in Jena and Göttingen and became professor of anatomy in Göttingen in 1776. He expressed the apparent purposefulness of living beings in the term *Bildungstrieb* [18], which was the "formative" principle responsible for development, efficient functioning, and reproduction. He developed the concept independently of Kant well before the appearance of his *Kritik der Urteilskraft*, their train of thought, in both cases inspired by Buffon's writings, showed unmistakable parallels. The *Bildungstrieb* should not be seen as a mysterious force like Buffon's *moule intérieur* [64], nor could it exist independently of matter, as the soul could, but neither could it be reduced to matter. It was the expression of matter's potential for organization that was characteristic of living beings. According to Blumenbach, the *Bildungstrieb* could be compared to Newton's gravitational force. As with gravity, this force's 'essence' remains hidden from us, while we can study its effects and trace the laws to which those effects conform.

The research program that Blumenbach developed on the basis of this principle had two objectives. First, the natural system of organisms had to be traced. Second, by studying the effects of the *Bildungstrieb*, he wanted to trace the laws to which the functional organization of living beings obeyed. In so many words he intended to do for the organic world what Newton had done for the inorganic world. In his program for systems analysis of nature, Blumenbach joined Buffon's critique of Linnaeus. He shared the view that the higher systematic categories were purely theoretical constructions. On the basis of a few arbitrarily chosen external characteristics it was impossible to establish a natural division. Natural kinship rested on similarity in organization, and thus in *Bildungstrieb*. Buffon had assumed the reproduction criterion: animals that can produce mutually fertile offspring belong to the same species. Given that the species was the unit of organization, races could be understood as variants of the organizational plan of the species. Blumenbach discussed an example of this in a treatise on the races of man, a work for which he made his name: *De generis humani varietate nativa* (1795) [20]. According to Blumenbach, the origin of the different human races was the result of small changes in the *Bildungstrieb*, brought about by differences in geographical conditions and in diet and living habits. He saw race formation as a process of degeneration. The white Caucasian race constituted the original type or *Urbild* [19].

The other varieties were degenerate variants of this, created under the influence of less favorable environmental factors.

On this point Blumenbach differed from Kant, who also raised the question of race. Kant also saw the species as a unity of organization, but he regarded the *Urbild*, which he called *Stammgattung* [112], not as a concretely existing form, but as the totality of adaptability contained within the organizational plan of the species. The human races were not degenerate forms of one original race, but environmental variants of the same organizational potential. This difference in interpretation is significant, for only in Blumenbach's version is there any transformation from one form into another. In Kant, nothing new actually ever arises: the *Stammgattung* responds to the environment with a preexisting variant of the organizational plan. Kant and Blumenbach also shed light on the historical dimension of the species question. The presence of extinct animals in older geological strata, for which evidence became increasingly convincing towards the end of the eighteenth century, could be explained without much difficulty on the basis of teleomechanistic assumptions. In the course of the earth's history, different forms of the *Urbild* could have developed under the influence of changing geological conditions, but the possibilities for variation of the *Bildungstrieb* were not unlimited. Sooner or later, an organism could end up in an environment where the *Bildungstrieb* could not provide an adequate response; extinction was then the result. As an embryologist, he defended the epigenetic point of view, supported by most other teleomechanists: the organism is not preformed in the egg, but develops gradually, under the influence of the *Bildungstrieb* [126, 127].

1.1.7.3 Kiemeyer's development

The basic principles of the teleomechanistic program formulated by Kant and Blumenbach were further developed by Blumenbach's pupil Carl Friedrich von Kiemeyer (1765-1844). He studied medicine at the Karlsschule in Stuttgart and then attended the education of Blumenbach in Göttingen. In Göttingen he was appointed professor at the University of Tübingen in 1796, first in chemistry and later also in botany and pharmacy.

Kiemeyer developed a detailed comparative-anatomical and physiological research plan that was supposed to reveal the natural connections between organisms. The higher systematic categories, above the genus level, also received attention. Like Blumenbach, Kiemeyer promoted the study of the relationships between organisms and their environment. For example, he wanted to come up with a *Psychologie der Thiere*, which would, among other things, have to map out how organisms succeeded in finding the right food and a suitable habitat. Extrapolating Blumenbach's ideas about the species to the higher systematic units, Kiemeyer argued that the richness of forms in nature originated in a relatively small number of hypothetical *Grundformen* [53]. Within these, transformation was possible: under the influence of environmental factors, new species could develop from the existing ones through the occurrence of modifications in the *Bildungstrieb*. However, the *Grundformen* were strictly separated from each other and there was therefore no question of one continuous development series.

A new element introduced by Kiemeyer was the comparative study of the developmental history of individual organisms, the comparative embryology. Common patterns in the embryological development of different species could provide insight into the relationships between different organizational forms. In this context, Kiemeyer posited a proposition that can be regarded as a prototype of a hypothesis that would profoundly influence nineteenth century anatomy and embryology. He argued that ontogeny – the development of the individual – was an abbreviated recapitulation of phylogeny – the development of the species.

Kiemeyer's contribution to the teleomechanistic program was mainly of a theoretical nature. He has published few empirical research results in support of the new ideas he introduced. His field of work included comparative anatomy and plant chemistry, the forerunner of organic chemistry. He attached great importance to chemical research, for the material basis of the functional organization of living beings lay in their chemical constitution. However, the research of organic compounds was just beginning and for the time being the ideal of a chemical analysis of the functional organization of living beings was unattainable.

An example of the influence exerted by Kiemeyer's theoretical considerations can be found in the work of the aforementioned Treviranus. It was he who introduced the concept of biology in Germany in his six-part *Biologie, oder Philosophie der lebenden Natur* (1802-1822) [215]. All the essential ingredients of the Göttingian program as drawn up by Kant, Blumenbach and Kiemeyer can be found in Treviranus's work. It contained little new empirical data but brought together the

biological knowledge of the time and placed it within the framework of the teleomechanistic view of living nature. The work was an even broader attempt to achieve what Kiehmeyer had also set out to achieve: to elevate the study of living nature to a Newtonian science by unraveling the laws of functional organization [53].

1.1.8 German *Naturphilosophie*

The development of the teleomechanistic program was embraced by other scholars of the Götting school, leading to deeper explorations in the study of living organisms. Notably, Alexander von Humboldt, Karl Ernst von Baer, and Johannes Peter Müller made significant contributions to this evolving perspective.

1.1.8.1 Alexander von Humboldt: dynamic functional organization

Alexander von Humboldt (1769-1859) was a botanist, physiologist and geographer who studied under Blumenbach for a time. He is best known for the journey through South and Central America that he made in the years 1799-1804. The vast amount of data he collected on this journey formed the basis for much of his later scientific work. Comparative anatomical studies of plants led von Humboldt to the conclusion that all plant species and genera could be reduced to a small number of types, analogous to Kiehmeyer's *Grundformen* [221]. But more than the form aspect, von Humboldt was interested in the dynamics of the functional organization. He collected quantitative data on the distribution of plant species across geographic zones and conducted experimental research on animals to determine the significance of environmental stimuli [151].

Building upon the foundational principles of teleomechanism, the Göttingen school attracted the attention of other like-minded scholars, such as Alexander von Humboldt. His botanical and physiological studies laid the groundwork for a deeper exploration of functional dynamics within living organisms. This focus on dynamics later converged with the embryological investigations of Karl Ernst von Baer, resulting in an evolving perspective on ontogeny and organismal relationships.

1.1.8.2 Karl Ernst von Baer: embryological contributions

Karl Ernst von Baer (1792-1876) was the most important of the researchers who focused their attention on embryological aspects of the Göttingen programme. Von Baer was from Estonia and studied medicine in Dorpat. He then moved west and studied zoology and comparative anatomy in Vienna, Würzburg and Berlin. From 1817 to 1834 he worked as a zoologist and anatomist at the University of Königsberg. In 1834 he left for Saint Petersburg, where he was appointed a member of the Academy of Sciences. Embryology was involved in Kiehmeyer's research because, in his view, ontogenetic development provided starting points for determining an animal's place in the natural system. His most pioneering publications came about in his Königsberg period. In 1827 he published his famous *Über die Bildung des Eies der Säugethiere und des Menschen* [10], in which he demonstrated the existence of the mammalian egg. Up to that point the egg stage was known only in the lower vertebrate classes, so von Baer could now state that the embryonic development of all vertebrates began with the egg. A year later he presented his findings on that development process in *Über die Entwicklungsgeschichte der Thiere* [220]. This work clearly shows von Baer's teleological framework. He presented the development of the embryo as a goal-directed process, which is, as it were, driven by the intended end result. In fact, according to von Baer, the complete organism logically preceded the parts; the whole 'caused' the development of the parts.

Von Baer's study supported Blumenbach's view that the future organism was not present in miniature form in the fertilized egg. According to von Baer, the egg initially consisted of a more or less homogeneous substance, from which increasingly complex structures slowly developed. His epigenetic view nevertheless also had a preformistic element. He assigned no more than a stimulating function to the male sperm and believed that the assembly of chemical compounds that represented the organizational plan of the new organism was completely present in the egg before fertilization. He established that the development of the fertilized egg always started with the formation of two clearly visible layers of tissue, the germ layers. The two germ layers split in two, and each of the four layers thus formed served as the substrate for a specific group of organs. Which organs originated from which germ layer and at what time this happened depended on the group to which the animal studied belonged. From these observations, von Baer derived a powerful tool for determining relationships between animals – the purpose that comparative embryology was

intended to serve in Kiehmeyer's teleomechanistic program. In addition to anatomical similarity between organs, the criterion for relatedness could now also be that those organs originated in the same germ layer and from there went through the same development for a shorter or longer period of time. The significance of this kinship criterion for the development of embryology can hardly be overestimated. Generations of researchers have regarded the search for what we now call homologies as their main task. Comparative embryological research was completely dominated by it until well into the nineteenth century. By the way, we still know the germ layer theory, in the formulation that Robert Remak gave it in 1855 [181]: von Baer called the upper germ layer the ectoderm, the lower one the entoderm, and he combined the two intermediate layers into the mesoderm.

An equally far-reaching influence had a second theory that von Baer launched in his *Entwicklungsgeschichte* [220]. He divided the animal kingdom into four groups, the *vertebrata*, *mollusca*, *radiata* and *articulata*. Within the four groups, the organizational plan of all animals could be traced back to one and the same basic plan or type. This was reflected in the fact that the embryonic development of the animals within a group showed remarkable parallels. This similarity, according to von Baer, only related to early stages of development, it was only the embryos of the different groups that resembled each other. Moreover, at some point each embryo began to show characteristics specific to the group to which it belonged, and then the similarity with other embryos ceased. Other researchers would extend the theory to the entire animal kingdom, arriving at the position that ontogeny is an abbreviated and accelerated recapitulation of phylogeny, the developmental history of the entire animal kingdom. For von Baer, the theory only provided clues about the way in which type manifested itself ontogenetically. Like Blumenbach and Kiehmeyer, von Baer assumed that the species were mutable and that transformation could occur within each type. But, following Kant, he set limits to that possibility, limits determined by the organizational plan. Transformation was based on the expression of the possibilities for adaptation contained in this plan, so in that sense nothing new ever came into existence [164, 89].

1.1.8.3 Johannes Peter Müller: physiological and comparative anatomy

A final example of the influence of the Göttinger tradition on the study of living nature is provided by the work of Johannes Peter Müller (1801-1858), one of the most important physiologists and comparative anatomists of the nineteenth-century. Müller studied medicine in Bonn and Berlin and was appointed professor of comparative anatomy and physiology in Bonn in 1826. In 1833 he acquired the chair of anatomy and physiology in Berlin, where he remained until the end of his career. In the 1820s Müller mainly conducted physiological research, in which he advocated for the application of experimental methods. This experimental approach was not completely new, but German physiology was still mainly a field of observation and description. It often came down to trying to give a functional interpretation of anatomical structures on the basis of observations.

Müller, on the other hand, promoted experimentation as an aid to arrive at such functional statements, in addition to observation. He studied subjects such as human sensory perception, sensorimotor transmission in the spinal cord and reflex movement. His sensory research led him, among other things, to the formulation of his 'law of specific sensory energies'. According to Müller, the nature of the stimulation was unimportant for the qualitative response of a sense: the sense always responded in a characteristic way, according to its own 'specific energy'. The eye, for example, responded exclusively to light sensations, whether it was optically, chemically, or mechanically stimulated. This is where Kant's theory of knowledge comes in, which says that we never get to know an object in reality as it 'really' is; we come to know it through our perception of it. How that sensation corresponds to 'the object in itself' (the *Ding an sich*) cannot be ascertained. The essence of the object always remains hidden from us; we can only base our knowledge on the way in which the object stimulates our senses.

The question of the function of an organ and the significance of that function for the organism as a whole was central to his research. The answer to this question was sought in (comparative) anatomical and embryological data. The experiment was a tool for Müller to obtain such data. However, with the rise of organic chemistry in the first half of the nineteenth century, physiologists increasingly began to investigate mechanisms of action. They wanted to find out not only what the function of an organ was, but also how the organ functioned. This causal question required knowledge of physics and chemistry and a more advanced experimental approach than had been customary until then. This direction in physiological research was taken by Müller's students, but Müller himself felt insufficiently at home in the physical and chemical field. Although his

experimental approach had brought about an important innovation in physiological research, he also struggled with his teleological conception of life, which made him extremely skeptical of the rigorous experimental approach of his students. In Müller's belief, the functioning of an organ could never be fully understood in physical and chemical terms. The knowledge obtained from experiments was subject to limitations. The organism was a purposeful functioning unit and it reacted as if it were controlled by a *Lebenskraft* [154] – a force similar to Blumenbach's *Bildungstrieb* – which itself could not be further analyzed. The nature of the response to an experimental stimulus, Müller reasoned, was determined not only by that stimulus, but also by the "specific energies" of the organism. just as it happened with the senses. The *Lebenskraft* played a guiding role in this, which is why experiments could never lead to a complete causal explanation [153]. A number of Müller's students did not share this skepticism.

1.1.9 Reductionistic Physiology

Anatomy and physiology were equivalent parts of biology to the teleomechanists: investigation of form and function were both necessary for an understanding of the organism as an efficient unit. But the experimental dimension was lacking in the first generation of teleomechanists; it only took shape in the work of Johannes Müller and his school. The transition to an experimental approach involved more than just a change of method. Müller's objections to unbridled experimentalism reflected his teleomechanistic point of view, and in rejecting these objections his students at the same time distanced themselves from Müller's view of living nature. The breakthrough of experimental physiology in Germany was thus accompanied by the dismantling of teleomechanics.

1.1.9.1 Theodor Schwann: founder of cell theory

One of his students who initiated this development was Theodor Ambrose Hubert Schwann (1810-1882), the founder of cell theory. In 1839 he was appointed professor of anatomy and physiology in Leuven and in 1848 he exchanged Leuven for Liège, where he continued to work as an ordinarius until 1879. As a physiologist, he made a name for himself with his discovery of the enzyme pepsin and with his thesis that alcoholic fermentation is the work of a living organism. Schwann was at the cradle of quantitative experimental physiology. Already in his first physiological studies he tried to record his observations quantitatively. He distanced himself from his teleomechanistic *Lebenskraft*. The same laws applied to living beings as to inanimate nature, Schwann argued. In the study of living beings, the same methods of research could be used as in physics and chemistry, and the aim should be to explain the activities of the organism in physicochemical terms. The core of Schwann's point of view can be found in his most famous work, the *Mikroskopische Untersuchungen* from 1839, in which he developed his cell theory [194, 128].

1.1.9.2 Cell theory and cellular discoveries

Briefly some backgrounds of the research into the cell. The first observations of cells were made by the seventeenth-century microscopists. Robert Hooke, in his *Micrographia Illustrata* (1665) [102], described the dead cell walls in cork and compared the small *cellulae* (chambers) he observed to a honeycomb. Antoni van Leeuwenhoek also observed cells and depicted them [85]. For Van Leeuwenhoek, both living and dead matter were made up of small spheres, the globules, and for him cells were nothing but globules. The emphasis in their descriptions was on the cell wall, hence the name *cellulae*. But the early microscopists had no clear idea of what a cell is. It is also often difficult to say what exactly they saw – real cells or optical illusions created by the spherical and chromatic aberration of their simple microscopes. Certainly the significance of the cell as an elementary building block of the organism was not clear to them.

In the eighteenth century, the idea of the cell as a building block did not play a role either. The idea did take hold that plants and animals are built from the same elementary units. However, those units were not cells, but fibers. The physician Herman Boerhaave and a number of contemporaries believed that tissues and organs were made up of orderly arranged fibres [86]. Little progress was made in microscopic research in the eighteenth century. Van Leeuwenhoek's observations with his microscopes turned out to be inimitable, not least because Van Leeuwenhoek had carefully kept his refined lens-grinding technique secret.

1.1.9.3 Evolution of microscopy techniques

An improvement in microscopic technique, which started in the 1820s, slowly put an end to the confusion. The chromatic aberration was largely controlled and an ever-increasing resolution was achieved. In the 1830s, researchers gradually got the improved instruments at their disposal and reliable observations of both plant and animal material increased rapidly. Robert Brown stated in 1831 that all plant cells have a nucleus [37]. Evangelista Purkinje saw the cell nucleus in 1830 and gave descriptions of nerve cells and brain cells in 1837-1838 [178]. He also pointed out general similarities between plant and animal cells. Felix Dujardin described the cell fluid in unicellular organisms in 1835 [71] and Purkinje gave it the name protoplasm in 1839.

These and similar observations culminated in the cell theory of Matthias Schleiden and Theodor Schwann. The botanist Schleiden (1804-1881), also a student of Johannes Müller, concluded in the late 1830s that the cell is the elementary building block of the plant. Stimulated by Schleiden's work, Schwann went in search of a comparable structure in animals. Due to the lack of a cell wall and the great diversity of the cells, it is much more difficult to recognize the cellular structure of the tissues in animals than in plants. However, careful microscopic examination of numerous animal tissues led Schwann to conclude that Schleiden's thesis also applied to animals. This is how the cell theory took shape in Schwann's *Mikroskopische Untersuchungen* from 1839 [194]: all living beings, both plants and animals, are made up of the same elementary units, the cells. Although cell division had already been observed by several microscopists in the 1830s, its significance was still unknown to Schleiden and Schwann. Observations by Carl Nägeli and Robert Remak, among others, made it plausible in the 1840s and 1850s that division is the cell's normal multiplication mechanism. Rudolph Virchow summarized these studies in 1855 in his famous statement *omnis cellula e cellula*: every cell (emerges) from a cell.

Theorists could only speculate about what went on inside the cell. Its fundamental role in metabolism and reproduction could not be demonstrated until later, and the development of cell physiology research that would eventually take place was largely a twentieth-century affair. This did not prevent Schwann from developing a new vision of the functioning of living beings based on cell theory. This brings us back to Schwann's role in the development of physiology after Johannes Müller. At the end of his *Mikroskopische Untersuchungen*, Schwann presented what he himself called a new 'theory of the organism'. For Schwann, cell theory was not just a theory of the unity of structure of the plant and animal organism. In his eyes it also had a physiological dimension: the cell was also the elementary unit of function. He had two arguments for this statement. The first was that a cell could exist on its own. The second argument he derived from the way cells were formed and developed. The process started with the formation of nuclei in the cytotblastema. These nuclei formed by a kind of crystallization of substances from the ground substance of the cytotblastema. After reaching a certain size in this way, a membrane formed around the core. The further development of the cell was now based on imbibition, the active uptake of substances through the membrane, and intussusception, the insertion of substances between the membrane and the nucleus.

He compared cell formation to the formation of crystals in inorganic nature. The implication was that the formation of living and inanimate structures was not essentially different. Although crystals grew by accretion, growth from the outside, while organisms grew by imbibition and intussusception, Schwann did not think this difference was fundamental. This is evidenced by his aphoristic statement that organisms are nothing but "daß die Organismen nichts sind als die Formen, unter denen imbibitionsfähige Substanzen kristallisieren" [61]. The formation of the organism was a physico-chemical process, according to Schwann, for which the same necessary laws applied as in inorganic nature, and in which the same blind forces were at work; not purposeful forces: the assumption of a *Lebenskraft* was superfluous.

Schwann's description of cell formation denounced yet another facet of teleomechanistic thinking, namely the priority of the whole over the parts. In von Baer's embryology, for example, it was assumed that the development of the embryo was controlled by the intended end result. The complete organism, in a sense, 'caused' its own development from the ovum; the whole logically preceded the parts. The cell theory, according to Schwann, suggested just the opposite. Living beings were made up of structurally and functionally independent parts that together determined the organism as a whole. The formation of those elementary parts themselves relied on physico-chemical processes. And so the parts determined the whole, not the other way around [128].

1.1.9.4 Justus von Liebig: chemical analysis in physiology

Schwann's cell theory expanded the horizons of understanding life at the microscopic level, paralleling the shift in thought initiated by Justus von Liebig's advocacy for chemical analysis in physiology. Liebig's exploration of organic chemistry's role in life processes challenged the notion of vital forces, providing a chemically-based perspective that resonated with Schwann's cellular revelations. As Schwann illuminated the microscopic realm, Liebig, a contemporary of his, delved into the molecular intricacies. In the forties of the nineteenth century, Liebig emerged as a leading proponent of applying chemical methods of analysis to physiology. Remarkably, he also found room for the concept of vital force, albeit in its original teleomechanistic sense, which he deemed indispensable for comprehending the complexities of life phenomena.

Justus von Liebig's journey into the realm of science began as a chemist, having studied at the universities of Bonn and Erlangen, and later apprenticed with the eminent chemist Joseph-Louis Gay-Lussac in Paris. In 1824, he assumed a professorship of chemistry in Gießen, embarking on a pioneering career that would shape the understanding of organic compounds [33]. This was an era when the elemental composition of organic substances – carbon, hydrogen, oxygen, and nitrogen – was unravelled by Lavoisier and his contemporaries in the late eighteenth century [213]. As organic chemistry rapidly advanced, the pursuit shifted towards quantifying the proportions of these elements in various compounds.

Liebig's description of the breathing process was grafted onto Lavoisier's work. Breathing, he had said, was slow combustion. When carbon and hydrogen from the nutrients came into contact with oxygen, they were oxidized to carbon dioxide and water. The heat released thereby compensated for the heat loss of the organism to its environment. Thus, the ability of animals to maintain their body temperature at a high, constant level relied on a chemical process. Liebig's vision of the breathing process was that an animal's heat production was entirely based on the combustion of sugars and fats. The carbon dioxide and water vapor excreted through the lungs by an organism at rest came entirely from the oxidation of sugars and fats. The nitrogenous nutrients, the proteins, were responsible for the constructive metabolic processes. They did not play a role in heat management but ensured the construction and maintenance of the blood and tissues. The background to these ideas was Liebig's belief that proteins were present in a ready-made form in food and hardly needed to undergo any changes in order to be absorbed into the tissues and blood. This ruled out the participation of differently composed substances, such as sugars and fats, in tissue building, Liebig reasoned. Because in that case the elements that contributed these substances would eventually have to be removed again and there was no reason to assume such a pointless exercise [128, 213].

1.1.9.5 Herman von Helmholtz: reductionist physiology

We find the same attitude, strongly oriented towards physics, in Herman von Helmholtz (1821-1894). Helmholtz studied medicine in Berlin and trained in physiology under Johannes Müller. He mainly focused on sensory physiology and in particular on the physical aspects of this. Eventually he would switch completely to physics: in 1871 he was given the chair of physics in Berlin. Before that he had been appointed extraordinarius in physiology in Berlin in 1849, professor of anatomy and physiology in Bonn in 1855, and professor of physiology in Heidelberg in 1858. During the years that Helmholtz worked in Müller's laboratory, he undertook a number of deliberate attempts to undercut the 'vitalism' of his generation of teachers. Liebig's *Organische Chemie in ihrer Anwendung auf Physiologie und Pathologie* of 1842 [132], which described physiological processes in terms of chemical transformations, put him on the trail.

To put this to the test, Helmholtz designed a series of ingenious electrophysiological experiments with frog legs. The result of this was, firstly, that the chemical changes that a contracting frog's leg underwent were entirely due to material conversion processes in the muscle tissue. Second, according to Helmholtz, the experiments showed that the nervous system – often regarded by vitalists as the main or even the sole seat of the vital force – played no direct role in these chemical transformations. Helmholtz succeeded in quantitatively measuring the heat generation of contracting muscles. These experiments with frog legs were part of a broader investigation that led Helmholtz in 1847 to confirm the law of conservation of 'force' ('energy', in modern terminology): all forms of energy in nature – mechanical, chemical, electrical and so on – are interconvertible and each conversion begins and ends with forms of energy that are exactly equivalent [95]. No new energy is generated, no energy is lost. Helmholtz's experiments are also typical illustrations

of reductionist physiology. Anatomy played no explanatory role, the methods used were the same as in chemistry and physics, instruments formed an essential part of the experimental design, and attempts were made to quantify the relationships described [119].

1.1.10 Holistic Physiology

Jacques Loeb (1859-1924) has long been known as a typical follower of the German reductionist school of Helmholtz. This view is mainly based on a collection of essays that Loeb published in 1912 under the title *The Mechanistic Conception of Life* [139] and which can indeed be read as a reductionist creed. But research into Loeb's life and work shows that this is only part of the story. Until 1912, Loeb had sharply opposed the German school. The collection of essays marked the beginning of a new period of research in which he radically broke with the principles that had dominated his work until then. Loeb had studied medicine in Berlin and Munich before doing doctoral research with the brain physiologist Friedrich Goltz in Strasbourg. Goltz was a student of Hermann von Helmholtz, but he opposed his reductionist approach to life phenomena. He rejected the then widespread localization theory, which held that every brain function could be located in a specific anatomical sub-area of the brain. For example, in the brain there would be a speech center, a touch center, a smell center and so on. Goltz believed that brain functions were a complex set of dynamic, constantly interacting processes. In his opinion, localization theory was insufficient to understand the dynamics and interplay of brain functions [6].

In his dissertation, which he completed in 1884, Loeb adopted this perspective from his teacher. In the following years he was strongly influenced by the philosopher and physicist Ernst Mach. Mach's ideas, he believed, provided a philosophical justification for Goltz and himself's approach to brain function. A core of Mach's philosophy was that theoretical scientific concepts, such as 'electron' and 'gene', were no more than brief descriptions of observational data. It was not permissible to attribute a real existence to the entities for which those concepts stood. The scientist had to deal only with observable and measurable phenomena, and any statement about 'underlying' entities by which those phenomena would be caused, belonged, according to Mach, to metaphysics [140]. For example, the concept of 'gene' sensibly summarized a number of hereditary phenomena, but that did not entitle the researcher to conclude that genes really existed.

Loeb decided to look at movement responses in animals and soon came with remarkable results. In certain butterfly species that deposit their eggs close to the ground on brushwood, he saw that the larvae moved upwards immediately after leaving the egg. This is how they reached their food, the young leaves at the ends of the branches. Loeb determined experimentally that the behavior resulted from positive heliotropy, an involuntary orientation toward a strong light source. The compulsion of the reaction was dramatically demonstrated when the larvae were placed on a bush with a lamp attached at the base of the trunk. The animals then oriented downwards instead of upwards, with the result that they starved to death. In another experiment, Loeb managed to produce a two-headed *Tubularia polyp* by permanently keeping the larva in a floating position in the water, thus avoiding contact with the bottom. This contact was apparently necessary for the differentiation of a foot (stereotropism), and in the absence of that stimulus a second head developed at the bottom [6].

Results like this led Loeb to develop a new physiological research program in the years 1888-1891. Scientific activity, Loeb argued, was characterized by practicality, by observation, measurement, and weighing of observable phenomena. It was useless to look for the imperceptible causes of these phenomena, for in doing so the scientist fell into metaphysical speculations. Thus, not causal analysis was the goal, but description in terms of reproducible measurement results. This led Loeb, following Ernst Mach, to reject the distinction between science and technology. To do science was to take action, and it made no fundamental difference whether natural responses of the organism were observed or artificial ones provoked for some technical purpose. The technical point of view was the heart of Loeb's program. His research gave him confidence that full control over the phenomena of life could be obtained through technical manipulation. This also meant that the researcher would have far-reaching possibilities to change the behavior of organisms in a direction desired by him. Man could ultimately make living nature subservient to his own interests; Loeb was aiming for a kind of biological engineering.

Loeb's greatest success came in 1899: artificial parthenogenesis. He discovered that unfertilized sea urchin eggs in seawater of certain salt concentrations spontaneously began to divide and develop into larvae [138]. So even a mechanism as basic as reproduction could be controlled with

experimental methods, Loeb concluded. A simple saline solution could substitute for the role normally played by the sperm. Loeb obtained a similar result with frog eggs, and he was not afraid to predict that eventually the reproduction of mammals, including humans, could also be controlled by physico-chemical techniques. His imaginative results and his suggestion that humanity would take its future fate into its own hands also brought him fame among the general public. From that moment on, he had no lack of resources to continue his research.

Around 1910 we are dealing with a completely different Loeb, namely as an archreductionist. The turnaround had several causes. Over time, Loeb grew tired of the role of public figure that had been thrust upon him after his initial successes. He was more and more inclined to abandon his controversial biotechnological vision and retreat to purely scientific terrain. He now saw it as his task to solve the riddles of life in a 'purely scientific' way. Instead of controlling the phenomena of life, causal analysis now came to be at the forefront, and this led to a break with Mach's philosophy. Loeb eventually joined the previously despised reductionist tradition that sought analysis down to the lowest, physico-chemical level [6, 169].

1.1.10.1 Lawrence Joseph Henderson: biochemical buffer systems

Lawrence Joseph Henderson (1878-1942) studied medicine at Harvard and physical chemistry with the colloid chemist Franz Hofmeister in Strasbourg. In 1905 he was appointed to teach biochemistry at Harvard University. Henderson had already become interested in the biological implications of Svante Arrhenius's theory of electrolytic dissociation during his student days. At Harvard University, he started research into the buffer system of the blood, the system that ensures that the pH, the acidity of the blood, remains within relatively narrow limits despite fluctuations in the concentrations of acid and base [98]. Faced with the great chemical complexity of the blood, Henderson first decided to study simplified physiological buffers containing only some of the elements found in the blood. In order to be able to compare different buffers, he drew up an equation that enabled him to express the buffer effect quantitatively. His research taught him that physiological buffers are much more efficient at maintaining the correct acidity than the simple buffers used by chemists. In addition, he established that combinations of buffers, such as those found in the blood, greatly increase the buffering capacity. Henderson's most striking result was that the buffers in the living organism performed their work in conjunction with other chemico-physiological systems and achieved their greatest effectiveness within that context. For example, he discovered a buffer system that was only moderately effective *in vitro*, but that operated within the organism in conjunction with the breathing gases, achieving much greater efficiency.

After these preliminary studies, Henderson attempted to map the blood pH system in its entirety [98]. He was not so much concerned with the individual chemical reactions as with the interactions between the various subsystems and with the organization of the system as a whole. The main problem Henderson faced was that a change in one of the variables affected all the others. He eventually found the solution by arranging his experimental data in a so-called Cartesian nomogram, in which the total effect of each change in one of the variables can be read. *In-vitro* studies were insufficient to map the complexity of life phenomena. Physicochemical reactions formed the basis of the functioning of organisms, but a reductionist approach failed to understand the higher-level interactions and the mutual coordination of the subsystems that enabled the efficient functioning of the individual as a whole. The blood's buffering system, Henderson argued, was one of the regulatory systems that maintained that internal environment. The system no doubt rested entirely on physico-chemical mechanisms, but much remained hidden from the reductionist analyst who looked no further than these mechanisms [5].

1.1.10.2 Walter Bradford Cannon: homeostasis and dynamic equilibrium

The ability of the organism to self-regulate was also central to the work of Henderson's colleague, the physician and physiologist Walter Bradford Cannon (1871-1945). Cannon, too, had studied at Harvard and remained here after graduation, first as a zoology teacher and, from 1906, as a professor of physiology. He was particularly interested in the role of the nervous and endocrine systems in the regulation of bodily processes. He studied part of the autonomic nervous system, the sympathetic system, which controls the body's regulatory mechanisms. Cannon made this controlling role visible by removing parts of the sympathetic system in a cat [44]. Certain regulatory mechanisms, for example to maintain body temperature, became inoperative as a result. The cat could continue to live with this without many problems, provided that the conditions in which the animal found

itself, the temperature in this case, were kept almost constant. Any slightly larger change caused problems, not because the regulatory mechanisms were missing, but because the 'order' for these mechanisms to come into effect was not forthcoming. Cannon proposed the term homeostasis to denote the regulatory processes that kept the body's internal environment in dynamic equilibrium, a term still in use today. Like Henderson, he believed that homeostasis was established on the basis of purely physical and chemical processes, yet also denied that the phenomenon could be fully understood at that lowest level of analysis. Understanding the system was only possible by focusing on the dynamics of the whole [5].

1.2 The contemporary framework of systems biology

As we navigate through the historical currents that have shaped our understanding of life, we arrive at a pivotal juncture where the prevailing notion asserts that true comprehension of a system demands a focus on its entirety. Having journeyed through the annals of biology's evolution, it is now only fitting, if not imperative, to inquire: What exactly is systems biology and how does it weave into this narrative?

In the realm of biology, diversity flourishes across various levels and scales – from the minute nuances of nucleotides and proteins to the orchestration of cells, tissues, organs, and entire ecosystems. However, what defines biology at its core? Enter Francis Harry Compton Crick (1916-2004), who boldly stated that modern biology's essence lay in explicating everything through the language of physics and chemistry, embodying a strong reductionist stance [82]. In this viewpoint, biological systems emerge as mere amalgamations of molecules and their interactions, with the behavior of organisms emerging from the underlying physical properties of their molecular constituents.

Yet, biology embarks on a transformative journey that sets it apart from the mere realms of chemistry and physics. While the arrangement of organic bases in DNA can be fashioned in a myriad of ways that adhere to the tenets of physics and chemistry, the true intrigue surfaces within the confines of the cell. It is within these microscopic landscapes that specific nucleotide sequences are meticulously tailored to function as a code – an initiation of the concept of information processing. This pivotal shift towards a systems perspective marks the departure of biology from its reductionist foundations.

A quintessential characteristic of life resides in the intricate fabric of information processing that occurs across multiple levels. As we accumulate evolutionary knowledge over time and transmit genetic instructions from DNA to various intra- and intercellular processes, a complex web of information unfolds. This intricate web of information processing speaks to the multi-level operation that lies at the heart of biology's complexity, an enigmatic phenomenon often referred to as emergence.

1.2.1 What is systems biology, anyway?

Emerging from this landscape, systems biology takes center stage. It's the discipline that envisions organisms as interwoven entities functioning across diverse spatiotemporal scales. In essence, it represents the fusion of biology, mathematics, technology, and computer science. It interlaces the various threads of theoretical biology, biochemistry, molecular biology, evolution, ecology, systems theory, network science, game theory, pattern formation, and nonlinear dynamics. At its core, systems biology seeks to unravel organisms as holistic systems, composed of dynamic cellular components like genes, proteins, and metabolites, and to fathom the intricate interactions birthing their function and behavior.

Within this domain, models assume a central role. These models, intricate constructs woven from a blend of physical laws, observational insights, and well-informed conjectures, serve as the cornerstone of our understanding. While the number of modeling formalisms is plentiful, ranging from cellular automata to differential equations [141], we can group them into two categories: forward and reverse modeling. In reverse modeling, data is dissected to unearth correlations suggesting causal relationships. This methodology gained prominence in the wake of data deluges brought about by omics technologies. Conversely, forward modeling commences with postulated causal relationships, constructing mathematical models from these conceptual foundations. Often, it is this forward trajectory that unveils profound insights into the intricate behaviors exhibited by complex systems. These models transcend mere mathematical abstractions. Instead, they materialize as logical engines, crafted to deduce the consequences of our assumptions or knowledge about a system. In molecular systems biology, they function as crucibles where we test the validity of our assumptions against physical laws and observational evidence, potentially reshaping our understanding of the underlying processes and mechanisms of life.

So, what is systems biology, anyway? According to Wikipedia at the date of writing this, the birth of systems biology as a distinct field began around the year 2000, with the establishment of the Institute for Systems Biology (ISB) in Seattle [55]. As we inquire into the very nature of systems biology, we find the ISB addressing the fundamental question, "What is systems biology?", on its official website [208]. We're confronted with their definition of systems biology, a definition replete with grandiose terms and lofty aspirations. However, upon closer scrutiny, it becomes evident

that this definition is not immune to the trappings of cliché and vagueness. Let's deconstruct this definition and explore why it might fall short of providing the clarity and substance we seek.

1.2.2 Critique of the "whole is greater than the sum of its parts" notion

The phrase "Systems biology is based on the understanding that the whole is greater than the sum of the parts" is one of the central tenets of the contemporary definition of systems biology. At first glance, it might appear to hold profound insight into the very essence of systems biology. Yet, upon closer inspection, we find it lacking in concrete elaboration. This assertion fails to expound on the mechanisms that give rise to this emergent property of systems. It stands as an assertion without the underpinnings of detailed explanation or empirical grounding.

The phrase "The whole is greater than the sum of its parts" has become a widespread notion used to encapsulate the concept of emergence, where complex systems exhibit properties that cannot be directly deduced from the properties of individual components. It's a principle that permeates various fields, from philosophy to sociology. It's an idea well-tread and widely accepted, not exclusive to the domain of systems biology. In fact, its frequent invocation in diverse contexts has somewhat diluted its impact and relegated it to the status of a platitude rather than a thought-provoking insight. While the concept may carry theoretical significance, its ambiguity not only makes formulating testable hypotheses or predictions challenging but also renders it elusive as it fails to provide an operational framework for scientists working in the field of biological research.

The concept of emergence can also lead to an unnecessary division between levels of organization, creating a false dichotomy between reductionist and holistic perspectives. In reality, there is a continuum of complexity and organization, and attributing properties solely to emergence can oversimplify the intricate relationships that exist across different scales. By focusing on emergent properties, one risks ignoring the underlying principles and interactions that contribute to the emergence itself.

Furthermore, emergence may not be a necessary concept to explain complex phenomena. Reductionist approaches, when appropriately applied, can account for the emergence of properties through the interactions and behaviors of individual components. Chaos theory, for instance, demonstrates that seemingly unpredictable and emergent behavior can arise from deterministic, nonlinear interactions at lower levels.

In systems biology, the notion of emergence takes on particular significance, given the field's focus on unraveling the mechanisms underlying emergent properties and interactions in biological systems. The field does not merely rest on the repetition of this well-worn phrase but aims to go beyond by delving into the specifics of how and why the whole is greater than the sum of its parts. Systems biology seeks to elucidate the intricate processes and behaviors that give rise to emergent phenomena, offering a deeper understanding of the mechanisms driving the complex interplay within biological systems. The critique of this platitude is not a dismissal of its validity but rather a call for a more rigorous and detailed exploration of its implications. It prompts us to move beyond surface-level acknowledgment and to engage in the intricate investigation of the interconnectedness, feedback loops, and dynamic behaviors that characterize complex biological systems. Systems biology aspires to be the vanguard of this exploration, dedicating itself to uncovering the underlying principles that truly manifest the phenomenon, thus transforming the cliché into a scientifically grounded and actionable insight. As we deconstruct and reassess the conventional definition of systems biology, we find ourselves standing at the crossroads of its inception and the broader historical context. It is clear that while the Institute for Systems Biology's (ISB) grand assertion of the "whole being greater than the sum of its parts" captures an essential concept, it lacks the precision and depth necessary to guide the field with tangible methodologies.

With this exploration in mind, we transition to the next subsection, delving deeper into the historical roots of systems biology and its emergence from the field of bioinformatics.

1.2.3 The emergence of systems biology from bioinformatics

The roots of systems biology trace back to the realm of bioinformatics, a field dedicated to deciphering the intricate information processes within living organisms. Coined by Hesper and Hogeweg in 1970, bioinformatics encompasses the study of informatic processes in biotic systems, revealing profound insights into the complexities of life. They defined it as "the study of informatic processes in biotic systems" [99]. Clearly, they recognized the intrinsic significance of information processing

as a defining property of life. They saw that information accumulation during evolution, information transmission from DNA to intra- and intercellular processes, and the interpretation of such information at multiple levels, were pivotal processes within living systems. This understanding led them to propose that information processing could serve as a powerful metaphor for comprehending the intricate workings of biological entities. This information-centric perspective echoed the essence of molecular biology's focus on genetic information. Concepts like the "genetic code" and the "central dogma" highlighted the pivotal role of information in biological systems. The notion of "transfer of genetic information" encapsulated the essence of molecular biology before the era of sequence data. This thematic emphasis on information has persisted over time, periodically reemerging in various contexts, such as the identification of distinct research fields and the exploration of processes like sensing the environment and dynamic modifications of molecules.

The theoretical underpinnings of biology were gaining further traction with works like Stuart Kaufman's exploration of random Boolean networks [114] and their application to transcriptional regulation networks [115]. Kaufman's work introduced the concept of large-scale regulatory networks and envisioned cell types as attractors within multidimensional dynamical systems. This foundational concept of viewing biological systems as dynamic entities laid the groundwork for later developments in systems biology. Within the scope of bioinformatics, two distinctive paradigms emerged:

Static bioinformatics: This approach involves analyzing real-world data obtained from various biological contexts. Static bioinformatics focuses on recognizing patterns embedded within these data through thorough analyses. By deciphering these patterns, researchers infer and hypothesize about the underlying biological mechanisms that might have given rise to the observed data. For instance, large-scale phylogenetic studies uncover evolutionary relationships among species, while Genome-Wide Association Studies (GWAS) pinpoint genes associated with specific diseases. Static bioinformatics relies on data-driven insights to unveil the mechanisms shaping biological systems.

Dynamic bioinformatics: Contrasting with static bioinformatics, the dynamic approach engages with the inherent processes that generate patterns within biological systems. In dynamic bioinformatics, researchers construct formal models that represent basic biological assumptions derived from observations. These models capture the processes and interactions underlying the system's behavior, allowing researchers to study emergent patterns or outcomes. For instance, by applying dynamical evolutionary modeling, scientists explore prevalent phenomena like whole-genome duplications in extant species. This approach enables researchers to simulate and examine the impact of various assumptions on the behavior of the system, leading to insights into complex biological processes.

The dynamic approach in bioinformatics, particularly the exploration of processes and interactions, has paved the way for the birth of systems biology. Systems biology capitalizes on the principles of dynamic bioinformatics to delve into the holistic understanding of organisms as intricate networks of interconnected components. By integrating data-driven insights with modeling techniques, systems biology transcends the reductionist perspective and captures the complex interactions that define life's intricacies. The journey from bioinformatics to systems biology reflects the evolution of our quest to unravel the mysteries of life, shifting from patterns in data to the dynamic interactions that give rise to these patterns. This pivotal shift lays the foundation for a more comprehensive and precise definition of systems biology.

1.2.4 Defining systems biology

To define a research field, it's insightful to pinpoint its most ambitious ultimate objective, the question that, when resolved, would epitomize its research endeavors. For biology, this overarching question is "What is Life?"². How does this aspiration transpose into the ultimate aim of Systems Biology?

Boogerd *et al.* have characterized systems biology as a biological approach that can sidestep the consideration of evolution [27, 26]. Given the widespread adherence to Dobzhansky's maxim "Nothing in biology makes sense except in the light of evolution" [211], this assertion appears almost heretical. Biology's scientific philosophy is deeply rooted in historical exploration of phenomena, setting it apart from non-historical sciences like physics and chemistry. Boogerd *et al.* acknowledge this, suggesting that the absence of evolutionary perspectives is a temporary drawback, indicating

²Title of a seminal paper by Erwin Schrödinger[192]

that systems biology will eventually align with mainstream biology and its evolutionary interpretations. Contrarily, Kirschner argues that the primary objective of systems biology is to understand how biology creates variation, a concept distinct from physical sciences and foundational to all of life [118]. Breitling proposes that systems biology is the research endeavor that provides the scientific foundation for successful synthetic biology [30]. While he acknowledges that an evolutionary outlook initially aids in comprehending general design principles, such as the balance between robustness and evolvability in cellular and developmental circuitry, according to him the ultimate demonstration of understanding life's organizational principles necessitates the ability to engineer entirely novel, unevolved life forms.

While a number of scholars have proposed various perspectives on systems biology, ranging from sidestepping evolutionary considerations to the foundational aspect of generating biological variation or underpinning synthetic biology, a common thread among these definitions seems elusive; instead these definitions offer a somewhat fragmented view. Taking into account the diverse viewpoints and the challenges in identifying a unified perspective, let's delve deeper to provide a more comprehensive definition of systems biology. This understanding hinges not just on singular perspectives but on capturing the vast complexities and nuances inherent to biological systems:

Systems biology is an integrative scientific approach that investigates how interactions between processes operating at different organizational levels and on varying timescales give rise to the behaviour and functions of living organisms.

This complexity reveals itself in several defining properties:

- **Locality of interactions:** Unlike the standardized particles in physics, entities in biology are distinct. This leads to distinctive local interactions in biological systems, which is often accompanied by spatial pattern formation.
- **Multiple levels of organization:** Organisms operate on multiple organizational levels that cohesively interrelate and exert mutual influences. Not only do parts constitute the whole, there also exists feedback from the higher level entities on lower level ones.
- **Diverse timescales:** Biological processes unfold over varying timescales, including ecology, evolution, and regulation. There needs to be an awareness of the inherent risk and potential fallaciousness associated with assuming a clear separation of timescales in systems biology.
- **Evolutionary signatures:** All living organisms are products of evolution and thus carry its imprints. Notably, neutrality plays a pivotal role, and evolution itself can undergo transformative shifts, a phenomenon that underscores the adaptability of biological systems.

To account for individual difference, researchers in systems biology may employ localized models, such as cellular automata, or adopt an individual-based modeling framework. A frequently used example illustrating how local or individual rules can give rise to complex behavior is the phenomenon of self-organization, which is observed in studies of population dynamics, tissue morphogenesis, epidemiology, and microbial consortia [183, 58, 90, 97, 180]. Multiple levels of organization and selection emerge as fundamental aspects of biological systems. For instance, the early death of RNA replicators may seem counterintuitive at a micro-level, yet they evolve due to influences from macro-level entities, such as propagating Darwinian waves [129, 209, 188]. Furthermore, understanding the interaction between different timescales is crucial. For instance, the separation of evolutionary and ecological timescales can prove fallacious, as demonstrated by research on time-dependent fitness in animal populations and host-parasitoid systems [158, 24]. Lastly, the exploration of evolutionary signatures further enriches our comprehension of biological systems. This is exemplified by the existence of pseudogenes, constraints on molecular interactions due to co-evolution, and the identification of phylogenetic signals in gene expression patterns [130, 231, 182]. Based on these insights we may extrapolate Dobzhansky's famous aphorism to formulate two additional guiding principles:

- Nothing in biology makes sense, except in light of self-organization, and
- Nothing in biology makes sense, except in light of multilevel evolution.

In the following sections, we will delve deeper into several of these concepts, exploring their implications and significance in the context of systems biology.

1.2.5 The role of models in systems biology

The intuitive work flow in experimental biology is: observation of an interesting phenomenon, formulation of a hypothesis, and testing of this hypothesis. For example, if we observe an organism that exclusively lives in a particular geospatial location, we hypothesize that there must be a reason for this. Perhaps the environmental conditions at this location are uniquely suitable for it, due to the presence of specific nutrients or the absence of predators. In forward modeling we take a different approach: we construct a model with minimal assumptions to see whether these assumptions are sufficient to reproduce the observed behavior. If this is not the case, we explore the validity of our assumptions and explore the parameter space to see what we might be missing. A model that describes the organism in its environment, however, may show that geometric features of the landscape and random dispersal may be sufficient to explain the phenomenon. Where the intuitive approach immediately tries to seek an explanation for the observed pattern, modeling allows one to discover that no additional explanation is needed.

To meaningfully describe biological systems, we must appreciate the power of modeling and understand the strengths and weaknesses of different modeling approaches. Models need to encapsulate the necessary complexity and no more than that, presenting a conundrum: what is the right level of complexity to address the question at hand? There is no straightforward way to answer this question and we need to explore different models and modeling approaches. Therefore, it is important to know what the limitations of the model are in order to understand what can and cannot be learned from them. For example, ordinary differential equations may not be able to explain the existence of subpopulations, whereas a spatial or stochastic modeling approach might [12, 201]. Similarly, while growth rate optimization in constraint-based models might be a useful objective for understanding flux distributions in microbes, it may be unsuitable for understanding those in more complex organisms, such as animals. Importantly, not only do we need to take into consideration constraints on the system, there exist feedback mechanisms within and between different levels of organization. In such a complex system, the rules at microscale, be it the interactions of individual molecules or organisms, gives rise to patterns at mesoscale whereof the behavior cannot be predicted *a priori* from our knowledge of the behavior at a local level – that is to say not without simulation. This also implies that not all biological behavior can be understood in terms of fitness optimization, instead, observations may often be considered as side-effects of higher-level feedback mechanisms. Even if there is a fitness benefit, we need to ask to whom this fitness benefit is conveyed.

A striking example of this is evident in the world of RNA-like replicators. Research has illuminated a seemingly counterintuitive phenomenon: the evolution of early death in these replicators. At first glance, early death might appear as a disadvantage. After all, an earlier death implies a reduced window of opportunity for an individual replicator to reproduce. However, in a dense competitive environment with mutant populations, the early demise of individual replicators can free up space, facilitating the growth and propagation of neighboring replicators. This can lead to the formation of beneficial structures, like faster-rotating spiral waves, which in turn can give an edge to the population in competition against mutants. Therefore, while early death does not favor the individual replicator, it indirectly enhances the overall fitness of the population, providing a competitive advantage in certain contexts [129, 188]. In the end, if multiple modeling approaches provide evidence for the emergence of a particular behavior, then it is more likely to be true; multiple points-of-view are needed in order to truly understand a biological process.

Finally, we consider what purpose a model serves. Models can provide us a baseline expectation. For example, we may use a metabolic network model to provide us insight into which biochemical pathways could be used to synthesize a specific precursor. Models can provide us a means to explore novel ideas and new areas of research. For example, the RNA genotype-phenotype mapping paradigm provided a template which guided the discovery of evolutionary properties of systems. The Lotka-Volterra equations, describing predator-prey interactions, are another good example; although the model is clearly incorrect – e.g. prey don't die – it has taught us a lot about population dynamics [223]. Models can serve as a proof of principle. For example, the group selection model in the study evolution was used to show that a particular type of behavior is, at least theoretically, possible [229]. Models can help to debug ideas and assumptions by forcing us to make them explicit. For example, classical models of the *E. coli* Lac operon, based on 50 years of experimental work, showed that there was bistability and hysteresis in the system [167]. The model is relatively complex and many of its experimentally determined parameter value estimates varied several orders of magnitude. When other researchers later evolved the model, they obtained

a model without bistability [217]. When the work was due to be published, however, a paper came out that showed experimental evidence for the existence of the bistability that the original model predicted [198]. Years later, it was discovered that this bistability resulted from the use of an artificial inducer, IPTG, which, unlike lactose, could not be metabolized [179]. Debugging of both experiment and model was necessary here in order to obtain this finding, which dispelled the dominant theory at the time – that of the existence of bistability in the Lac operon system. Models can also be predictive. For a model to be predictive we need both the correct interaction structure as well as its parameter values and initial conditions. There are different flavors and purposes of models, but the simplest answer to why we use them is that they provide us an understanding: starting from known or assumed interactions at a micro-scale, consequences of complex behaviour at meso-scale can be derived, which may not be obvious and be counter-intuitive when considering local interactions alone.

1.2.6 Modes of explanation

Beyond the mere constructs of models lies the deeper quest of 'why' – the inherent human drive to explain phenomena. Exploring the multifaceted nature of explaining biological phenomena, we begin with a whimsical Dutch saying that perfectly captures the essence of curiosity: 'Waarom, waarom, zijn de bananen krom?'. Translated to English, it means: Why, oh why, are bananas bent? This saying reminds us that there are different ways to approach explanations, each offering a unique perspective on the underlying principles of complex systems.

1.2.6.1 Tautological

The first mode of explanation involves the use of tautology, a form of reasoning based on self-evident truths. In this context, we might consider the bent shape of a banana as a defining characteristic – essentially, a yellow fruit is a banana, and if it's not bent, it's not a banana. However, while this approach appears straightforward, it tends to fall short of offering deep insights. It's akin to stating that survival of the fittest is fundamental for evolution – undeniably true, yet lacking in profound illumination. Yet, despite its apparent simplicity, this perspective doesn't necessarily downplay its significance. To illustrate this point, consider the concept of survival of the fittest. Although it's often assumed to be a universal truth tied to explicit fitness, it's not always the case. A remarkable example of this lies in the intricate interplay of a host-parasitoid system, as explored by Boerlijst and colleagues. This system presents an intriguing scenario where even low-fitness parasitoids can emerge as dominant players in the long term, under the condition that they arise in the central regions of spiral waves [23, 24]. The exploration of this host-parasitoid system reveals an unexpected twist. In the realm of ordinary differential equation (ODE) models, increasing the death rate of a replicator is typically considered a disadvantageous trait. However, when we venture into the realm of cellular automata (CA) models and allow for mutations in the death rate of individuals, a fascinating phenomenon unfolds. This phenomenon defies initial intuitions and necessitates a shift in perspective. In this CA model, where growth and interactions are governed by local rules, the process of replication depends on the availability of adjacent empty spaces. Replicators interact and assist each other in their replication process. Spiral waves, a key feature of this system, exhibit intriguing dynamics where faster-rotating spirals can overtake and replace slower-rotating ones. This principle of spirals competing for space is established through theoretical insights from excitable media studies [124, 24]. Further insights into this host-parasitoid system unveil the pivotal role of the core region within the spiral waves. All offspring originate from this central region, which leads to a profound consequence. While a higher death rate might seem detrimental for an individual replicator – offering less time for replication – it turns out that a higher death rate can, in fact, evolve within the population. The intricate interplay becomes clearer when considering the dynamics within the spirals. Faster-rotating spirals exhibit a faster succession of replicator generations – A followed by B, then C, and so forth. The speed at which these replicators replicate and vacate space becomes a determining factor in the competition among spirals. A higher death rate, paradoxically, enables replicators to more swiftly vacate space, resulting in a faster-rotating spiral wave. This phenomenon transcends the conventional trade-off mechanisms and introduces a new level of selection dynamics.

In summary, the exploration of this host-parasitoid system offers a thought-provoking perspective on the intricate interplay between fitness, replication dynamics, and the emergence of dominant traits. It challenges our conventional understanding of fitness optimization and illustrates how

seemingly counterintuitive traits can emerge under specific contexts. This example underscores the richness and complexity of biological systems, where even apparently straightforward concepts like survival of the fittest can unfold into intricate narratives of adaptation and evolution. Moreover, tautologies can convey profound insights. The notion that survival of the fittest is equivalent to competitive exclusion is, at its core, a tautology. Recognizing that the different terms stem from the same underlying concept can shed light on their meaning and implications, and understanding where these different terms come from is insightful.

1.2.6.2 Generic property

Another mode of explanation involves considering prevalence. When bananas are almost always bent, a straight banana becomes the exception. In this light, our assumptions or stereotypes might perceive a straight banana as a deviation from the norm. This concept aligns with the notion that studying specific cases, such as model organisms, can shed light on general properties. However, investigating these 'deviations' can also lead us to uncover the inherent characteristics of systems, or prompt us to question whether "why" is indeed the right question to ask. To illustrate this point, let's delve into the fascinating example of B-cell nodule formation in lymph nodes. In lymph nodes, B cells congregate in clusters amidst T cells, forming distinctive B cell nodules. Both cell types enter the lymph node randomly, which raises two pivotal questions:

1. How are these nodules formed?
2. Why did the system evolve this pattern?

Addressing these inquiries, Hogeweg crafted a cellular automaton (CA) model in 1989 [101]. In this model, the 2D grid serves as a cross-section of the lymph node, where each grid point represents the presence or absence of a T cell or a B cell. The evolution of each grid point is governed by processes of birth and death, as well as influx and efflux of cells. As T cells and B cells traverse the lymph node, their movements contribute to the overall dynamics, characterized by localized influx and efflux. The model assumes a stochastic influx of T and B cells, with T cells autonomously proliferating and B cells proliferating in proximity to T helper cells [163]. Remarkably, this seemingly straightforward model yields a remarkable observation: B cell nodules emerge, featuring a higher density of B cells at the edges of nodules, where they are spurred to divide by T helper cells. This phenomenon mirrors real-life observations within lymph nodes. This illustration underscores a profound insight: patterns need not necessarily originate from a purposeful mechanism or serve a functional explanation. In fact, this model incorporated no specific mechanism to generate the pattern, nor did it attribute any explicit advantage or function to the observed arrangement. Curiously, in this instance, the clumped B cell configuration is actually disadvantageous to the system. These clustered patterns can impede the interactions between B cells and T cells, which are pivotal for proliferation. Paradoxically, while optimal proliferation might require a well-mixed system, local proliferation dynamics render achieving such homogeneity challenging. This example offers a pivotal revelation: patterns need not solely result from optimization mechanisms or specific causal chains. Frequently, patterns serve as the default expectation, particularly in the context of biology and cellular automaton models. The prevalence of non-uniform distribution resulting from local interactions emerges as a fundamental aspect, challenging the assumption that patterns must always emerge from deliberate optimization or functional intent. So, why do lymph nodes exhibit their distinct structure? The answer lies in the fact that this pattern reflects the inherent default state – suboptimal yet prevalent – while achieving randomness is, in reality, a formidable task.

1.2.6.3 Imposing value or benefit

Imposing a certain value or benefit presents another avenue for explanation. For example, one might posit that specific phenomena, such as the curvature of bananas, hold optimality for certain purposes. However, such assertions require precise definitions of terms like "optimal" to carry meaningful weight. Furthermore, the question "optimal for whom?" underscores the intricate and multi-level nature of biological systems. To illustrate this concept, consider Rodney Brooks' robots [35]. He designed state-of-the-art robots capable of planning optimal paths through rooms. However, these robots often spent too much time planning and reconsidering due to changing environments. Brooks' later work emphasized a more adaptable approach, where robots continuously

adjusted their behavior based on real-time feedback from the environment [34]. This example underscores that sometimes, instead of rigidly adhering to preconceived notions of optimality, adaptable behaviors driven by immediate environmental cues lead to more effective outcomes. Another example comes from self-organization in animal groups. Flocking birds, schooling fish, and migrating herds demonstrate complex behaviors emerging from simple individual rules in inert environments. These behaviors aren't necessarily optimized but result from interactions with their surroundings [183, 58]. This further exemplifies that behavior can stem from simple rules and cues, rather than explicit optimization strategies.

Many behaviors that are hailed as optimized could stem from what is known as TODO-based behavior – simple rules guiding behaviors based on local environmental cues [58, 97]. A fundamental principle that sheds light on this concept is the "do what there is to do" principle, which underscores that behavior is steered by local information. Organisms respond to cues in their environment, triggering specific behaviors when necessary. Additionally, the "do based on what is done" principle emphasizes that an organism's perception of its environment evolves as a result of its actions. As organisms engage with their surroundings, a shared memory collectively shapes their behaviors. This framework reveals that flexible behavior often emerges from rigid rules. The concept of automatic adaptation becomes apparent, where organisms don't engage in behaviors lacking the required environmental cues to trigger its enactment. The fallacy of assuming optimality is a critical consideration in understanding biological systems. Many features observed in nature are not necessarily optimized, yet it is a common error to assume that these features evolved to be optimal. Instead, they might exist due to side-effects, constraints, or from neutrality. The wide variations seen in certain features could stem from the absence of significant fitness advantages or costs associated with those variations. Hence, the notion of optimality in biology must be approached with caution, taking into account the complex interplay between organisms and their environments, the influence of local cues, and the possibility of non-optimal but functional configurations emerging due to various factors.

1.2.6.4 Epiphenomenon

Side-effects often provide alternative explanations for observed phenomena, offering valuable insights into the underlying mechanisms at play. In this mode of explanation, certain features or behaviors emerge not as direct products of evolution but as byproducts or consequences of other processes. Banana curvature, for instance, might arise as a side-effect of growth patterns influenced by gravity or bunching, shifting the focus from evolution to physical forces during growth. This perspective suggests that understanding the cause of a phenomenon might not necessarily reveal an adaptive purpose. Consider the case of bananas: explaining their bentness could lead us to identify the gene or genes responsible for this trait. Yet, diving deeper into the genetic basis might uncover an infinite regression of explanations, as we inquire about the origins of those genes and the factors that influenced their development. This endless quest could potentially trace back to fundamental cosmic events, like the Big Bang, rendering such explanations impractical for understanding the immediate biological context.

Epiphenomena can manifest in various ways across biological systems. A striking example is the phenomenon of social differentiation observed in bumblebee colonies. This differentiation is dependent on nest structure and growth rates, with certain bees becoming elite workers that lay eggs. While this might initially seem puzzling from an evolutionary optimization standpoint, a closer examination reveals that these properties are not heritable and, instead, serve as integral components of a socially regulated clock that governs the bumblebee colony's life-cycle. This sheds light on how what appears as an evolved trait might, in fact, be a side-effect of complex social interactions and environmental factors [227, 8]. Similarly, other examples from nature highlight the concept of side-effects. Grouping behavior among animals might not necessarily arise from an innate desire to form groups but can be an epiphenomenon resulting from distinct foraging behaviors or priorities [97]. This insight challenges assumptions about the purposes of certain behaviors and illustrates how behaviors that appear optimized might, in fact, be unintended consequences of individual actions interacting with their environment. Another compelling illustration of how side-effects can lead to unexpected outcomes in evolutionary processes is provided by Hillis' work on sorting algorithms. Hillis aimed to evolve efficient algorithms for parallel sorting. While the primary goal was a fast sorting algorithm, he selected algorithms based on their ability to correctly sort a certain number of problems. Interestingly, this selection criterion led to the emergence of fast sorting algorithms as a side-effect [100].

In summary, considering epiphenomena as explanations opens doors to understanding the intricate interplay between evolving organisms and their environments. While evolution may optimize specific traits, it can also produce side-effects that contribute to the diversity and complexity of biological systems. By recognizing these non-adaptive consequences, we gain a more comprehensive perspective on the forces shaping the natural world, challenging assumptions of optimization and offering new avenues for exploration and understanding.

To conclude, we might undertake the task of constructing a detailed model to decipher the unique curvature of bananas. By integrating all available data, our objective is to shed light on the myriad factors influencing this distinct growth pattern. Central to this quest is the realization that even ostensibly trivial questions can produce multifaceted responses, contingent on the perspective and level of analysis employed. It is judicious to address any question or behavior through multiple lenses for a holistic comprehension. Central to our philosophy is the application of analogous principles across varied systems to unveil common behaviors, accentuating their robustness to shifts in model structure and foundational theories.

In the broader arena of scientific modeling, this philosophy finds echoes. While models founded on the basic laws of physics might be hailed as unbiased representations of reality, their counterparts in biology aren't as clear-cut. The equations that we use, such as for example mass action kinetics, are not physics or chemistry, but phenomenology. Although biology is grounded in physics, its models don't have the same predictive prowess as those sprung from the fundamental laws of physics. In biology we need to make assumptions. The pharmacologist James Black understood this well. In his 1988 Noble prize lecture he stated: "Models in analytical pharmacology are not meant to be descriptions, pathetic descriptions, of nature; they are designed to be accurate descriptions of our pathetic thinking about nature" and that "they are meant to expose assumptions, define expectations and help us to devise new tests" [17].

The same philosophy stands true in biology. The justification for any model therefore lies in its use to test the pathetic assumptions it is was build on. From this perspective, it may seem attractive to focus on the assumptions to see what surprising behavior might emerge if we tweak or add more of these. The scientific publication record, however, is testimony of the fact that the results of these endeavors are often mundane. Instead, we should focus on the question we wish to address, assess whether our assumptions are sufficient, and ask to what extent they are consequential to the result or conclusion the model provides. Specifically in biology, the assumptions that we make may have significant consequences for the results we obtain. Ultimately, the choice of modeling framework is subjective and the certainty of conclusions derived from the model are, and should always be considered, in relation to its assumptions. They are, echoing James Black's resonant phrase, "accurate descriptions of our pathetic thinking" [92].

1.3 Cellular metabolism

Cellular metabolism encompasses the suite of chemical reactions within cells, falling broadly into two categories: catabolism, which degrades molecules to release energy and smaller components; and anabolism, which assembles these components into larger structures, consuming energy in the process.

1.3.1 Significance and molecular foundations

To appreciate the vastness of cellular metabolism, we must understand the multifaceted roles it plays in cellular functioning and broader biological contexts:

Energy production and maintenance: At its core, cellular metabolism is responsible for the production of energy needed for cellular activities. Through processes like glycolysis, the tricarboxylic acid cycle (TCA), and oxidative phosphorylation, cells generate adenosine triphosphate (ATP), the universal energy currency. ATP fuels cellular processes, such as active transport, muscle contraction, and DNA replication, ensuring the continuity of life.

Biomolecule synthesis: Beyond energy production, metabolism plays a pivotal role in synthesizing the building blocks required for the construction of complex biomolecules. Carbohydrates, lipids, proteins, and nucleic acids are intricately woven into the fabric of living organisms, forming the basis for cellular structures, enzymes, hormones, and genetic material.

Regulation of homeostasis: Cellular metabolism intricately intertwines with the imperative task of maintaining internal equilibrium, commonly known as homeostasis. Metabolic pathways regulate the concentrations of various molecules within cells, tissues, and organs. For instance, the levels of glucose and other metabolites are tightly controlled to ensure adequate energy supply and prevent detrimental imbalances.

Adaptation to environmental changes: Cells must continuously adapt to changes in their environment to survive and thrive. Metabolic pathways allow cells to adjust their activities based on nutrient availability, oxygen levels, and other environmental cues. This adaptability is essential for cells to respond to challenges, such as nutrient scarcity or exposure to toxins.

Interplay with signaling pathways: Metabolism and cellular signaling pathways are interconnected, influencing each other's activities. Metabolites, the intermediates and products of metabolic reactions, often serve as signaling molecules that modulate cellular processes. This intricate cross-talk enables cells to integrate information from their surroundings and adjust their metabolic responses accordingly.

Health and disease: Dysregulation of cellular metabolism is implicated in numerous diseases, including metabolic disorders, cancer, and neurodegenerative conditions. Understanding the molecular foundations of metabolism provides insights into disease mechanisms and offers opportunities for therapeutic interventions.

Evolutionary conservation: Many metabolic pathways are conserved across diverse species, reflecting their essential role in sustaining life. Evolution has fine-tuned these pathways over millions of years, resulting in the remarkable diversity of metabolic strategies among organisms.

In the sections ahead, we'll trace the history of cellular metabolism, highlighting its pioneers and origins. We'll then examine its molecular pathways, cycles, and regulations, offering insights into cellular function and the complexity of metabolism.

1.3.2 The discovery of cellular metabolism

The discovery of cellular metabolism can be traced back to the early 19th century and coincides with the discovery of photosynthesis. It cannot be pinned to a specific date, as it involved a gradual process to which many people contributed. The discovery started with Jean Baptiste van Helmont, who performed an experiment in which he grew a willow tree in a pot of soil and measured the mass of the tree and the soil over a period of five years. He found that the mass of the tree increased by significantly, by 74 kg, but the mass of the soil remained relatively constant, losing only 57g. In the posthumous publication of this work, in 1648, he attributed this difference to the uptake of water [96]. More than a century later, in 1774, Joseph Priestley discovered that plants produce and absorb gas [177]. Although he didn't realize it at the time, his experiments proved that oxygen is present in the air. Five years later, in 1779, Jan Ingenhousz performed an experiment in which he kept plants submerged in jars and observed bubble formation on the surface of the leaves, but only when the jars were placed in sunlight [106]. From this he concluded that plants

use light to produce oxygen. It would take almost two more decades until, in 1796, Jean Senebier demonstrated that plants absorb carbon dioxide and release oxygen with the help of sunlight [196]. However, it was not until 1845, after Julius Robert von Mayer published his seminal work on the conservation of energy [148], that von Mayer enunciated the idea that the movement of molecules in living organisms is the source of their energy [147]. It was around this time that the first studies on chemical reactions in cells commenced, such as those conducted by Jean-Baptiste Dumas and Jean-Baptiste Boussingault [72], that the chemical theory of metabolism was proposed by Justus von Liebig [131], defining metabolism as a concept and ushering in a new field of research.

In the late 19th and early 20th centuries, scientists such as Louis Pasteur, Eduard Buchner, Otto Warburg, and Leonor Michaelis made further discoveries that paved the way for the development of the modern study of cellular metabolism. Following up on the work of Theodor Schwann's cell theory, Louis Pasteur conducted a series of experiments that showed that micro-organisms are responsible for the fermentation of sugars. In 1858 he showed that lactic acid bacteria convert sugar into lactic acid, while yeast convert it into alcohol [168], and that the rate of these fermentation processes diminishes in the presence of oxygen, something which is nowadays referred to as the Pasteur effect. In subsequent fermentation study conducted in 1897, Eduard Buchner showed that fermentation can occur in the absence of living cells, which provided the first evidence for the existence of enzymes [39]. In the 1924, Otto Heinrich Warburg discovered the respiratory enzyme, cytochrome, and its involvement in electron transfer during the conversion of glucose and oxygen into energy, which helped to establish the basic principles of cellular respiration and energy production [224]. In furtherance of understanding cellular respiration, one of his students, Hans Adolf Krebs, would come to discover a cyclic series of biochemical reactions that is used for the production of energy from carbohydrates and fatty acids [123]. This cycle, which is known as the tricarboxylic acid cycle or Krebs cycle, is part of central carbon metabolism and will take center stage in the third chapter of this thesis.

1.3.3 Central carbon metabolism

At the core of metabolism lay pathways that are universal to life. This system of biochemical pathways, referred to as central carbon metabolism (CCM), is a hub at the intersection of catabolic and anabolic processes. Central carbon metabolism comprises several interconnected metabolic subsystems, such as glycolysis, the pentose phosphate pathway (PPP) and the tricarboxylic acid cycle (TCA).

Glycolysis is a fundamental metabolic pathway that occurs within cells and is conserved across diverse organisms. It plays a crucial role in energy production and provides a source of pyruvate that can enter various downstream metabolic processes. The process initiates with the phosphorylation of glucose, trapping it within the cell and priming it for subsequent breakdown. Glycolysis comprises two main phases: an energy-investment phase and an energy-payoff phase. In the energy-investment phase, ATP is consumed to convert glucose into glucose-6-phosphate, which is then further metabolized. During the energy-payoff phase, three-carbon molecules generated from glucose are processed to yield ATP and NADH, valuable energy-rich molecules for cellular functions. The final product of glycolysis is pyruvate, a versatile metabolite that can be channeled into various metabolic pathways.

Gluconeogenesis is the reverse of glycolysis and is essential for maintaining glucose levels when glucose availability is limited. This process occurs in specific tissues and cells, ensuring a constant supply of glucose for critical functions. While glycolysis breaks down glucose into pyruvate, gluconeogenesis synthesizes glucose from precursors like pyruvate, lactate, and select amino acids. Converting pyruvate to glucose requires the reversal of specific glycolytic steps. Unique enzymes facilitate several of these reactions, which are regulated to prevent wasteful cycling between glycolysis and gluconeogenesis. This process consumes energy in the form of ATP and GTP and is tightly controlled to meet the energy demands of the cell or organism. Collectively, glycolysis and gluconeogenesis assume a crucial role in maintaining energy equilibrium and optimizing nutrient utilization across diverse biological systems.

The pentose phosphate pathway (PPP), also known as the hexose monophosphate shunt, is a metabolic pathway that operates parallel to glycolysis. It serves multiple essential functions

within cells by generating reducing equivalents, NADPH, and producing pentose sugars that contribute to various biosynthetic processes. The PPP consists of two interconnected phases: the oxidative phase and the non-oxidative phase. In the oxidative phase, glucose-6-phosphate is oxidized to form 6-phosphogluconate, generating NADPH in the process. NADPH plays a critical role in an array of biochemical reactions, including the synthesis of nucleotides and fatty acids, as well as in redox regulation and defense against oxidative stress. The non-oxidative phase involves the interconversion of sugars and is pivotal for generating ribose-5-phosphate, a key precursor for nucleotide synthesis. This phase provides cells with the building blocks necessary for DNA, RNA, and ATP synthesis. The PPP's dual role in producing NADPH and ribose-5-phosphate highlights its significance in both maintaining redox balance and supporting the synthesis of crucial cellular components. Moreover, NADPH generated by the PPP helps protect cells from oxidative damage. It fuels various antioxidant systems, including the regeneration of glutathione, a potent antioxidant that scavenges reactive oxygen species generated during cellular respiration. By maintaining cellular redox equilibrium, the PPP contributes to cell viability and overall health.

The tricarboxylic acid cycle (TCA cycle), also known as the citric acid cycle or Krebs cycle, is a fundamental metabolic pathway that operates in both prokaryotic and eukaryotic cells, playing a central role in energy generation and the synthesis of key biomolecules. In prokaryotes, like *E. coli*, the TCA cycle occurs in the cytoplasm, the cell's main compartment, whereas the cycle operates in mitochondria in eukaryotic cells, which are specialized organelles responsible for energy production. The cycle begins with the condensation of acetyl-CoA, derived from various carbon sources, with oxaloacetate to form citrate. This initiates a series of enzymatic reactions that ultimately lead to the production of reducing equivalents such as NADH and FADH₂. The electrons carried by NADH and FADH₂ are crucial for the electron transport chain (ETC), which is embedded in the prokaryotic cell membrane. This chain facilitates the transfer of electrons along a series of protein complexes, ultimately leading to the establishment of an electrochemical gradient across the membrane. This gradient drives ATP synthesis, similarly to the process in eukaryotes. In addition to its role in energy production, the TCA cycle is vital for biosynthesis. Intermediates of the cycle can be extracted to serve as precursors for the synthesis of various biomolecules, including amino acids and fatty acids. This versatile aspect of the TCA cycle contributes to the overall metabolic flexibility of prokaryotic and eukaryotic cells. Overall, the TCA cycle is a central hub of metabolic activity in prokaryotes, contributing to energy generation, the synthesis of essential molecules, and maintaining metabolic balance. Its universality across diverse organisms underscores its fundamental significance in cellular metabolism.

Anaplerotic reactions are crucial cellular mechanisms that replenish intermediates within metabolic pathways, enabling cells to maintain balanced fluxes in response to varying metabolic demands. These reactions play a vital role in central carbon metabolism, ensuring the availability of essential metabolites for various cellular processes. Central carbon metabolism features several key anaplerotic reactions:

- **Phosphoenolpyruvate carboxylase (PPC)**: This enzyme facilitates the conversion of phosphoenolpyruvate (PEP) to oxaloacetate (OAA) in the presence of CO₂. Replenishing OAA is pivotal for the continuation of the TCA cycle and other biosynthetic pathways.
- **Phosphoenolpyruvate carboxykinase (PEPCK)**: Functioning in the reverse direction of PEPCK, this enzyme transforms oxaloacetate back into phosphoenolpyruvate, consuming ATP. This reaction contributes to maintaining a balance between energy production and biosynthetic processes.
- **Malic enzyme (ME)**: The malic enzyme plays a critical role by converting malate to pyruvate using NAD(P)H. This reaction assists in pyruvate regeneration, contributing to energy equilibrium and redox homeostasis.
- **Glyoxylate shunt**: As a bypass of the TCA cycle, the glyoxylate shunt is particularly relevant in organisms utilizing acetate or fatty acids as carbon sources. The isocitrate lyase-mediated conversion of isocitrate to succinate and glyoxylate leads to glyoxylate condensing with acetyl-CoA to produce malate. This malate then feeds into the gluconeogenic pathway, allowing for the conversion of fatty acids into carbohydrates. This adaptation enables cells to grow using acetate and fatty acids as sole carbon sources.

The glyoxylate shunt not only showcases the cell's metabolic flexibility in generating essential molecules from diverse carbon substrates but also highlights the intricate coordination of pathways for sustainable growth and survival. Anaplerotic reactions contribute to metabolic balance by dynamically adjusting the availability of key intermediates, reflecting the sophisticated strategies cells employ to optimize their metabolic responses under different conditions. These reactions underscore the complexity and adaptability of cellular metabolism, ensuring the harmonious functioning of diverse biochemical processes.

1.3.3.1 Interplay with other pathways

Central carbon metabolism serves as a critical hub in the intricate web of cellular activities. It interacts with essential pathways like amino acid biosynthesis, lipid metabolism, and energy production, highlighting its vital role in cellular functionality. For instance, central carbon intermediates contribute to amino acid synthesis, ensuring the creation of necessary proteins. This process not only aids protein formation but also optimizes resource use by harnessing energy and reducing equivalents. The relationship with lipid metabolism is similarly important, as central carbon molecules are essential for generating fatty acids and lipids used in membranes, energy storage, and signaling. Moreover, central carbon metabolism is key for energy production, converting nutrients into ATP through glycolysis, the TCA cycle, and oxidative phosphorylation. This energy powers cellular processes like transport and biosynthesis. The coordinated interplay of these pathways maintains metabolic balance and resource allocation across diverse conditions.

1.3.4 Regulation of Metabolism

The precise orchestration of enzymatic reaction rates, also known as fluxes, lies at the heart of cellular metabolism. The efficient generation of ATP, the regeneration of cofactors, and the synthesis of essential biomolecules, including amino acids and fatty acids, necessitate a delicate balance of metabolic fluxes. Achieving such balance is crucial for maintaining cellular homeostasis and adapting to changing environments. In order to achieve this, regulatory mechanisms have evolved, which play a pivotal role in allowing cells to dynamically adjust their metabolic states on both short and long time scales. The magnitude of metabolic fluxes is determined by multiple factors, including the availability of substrates and products, as well as the abundance and activity of enzymes involved. Two primary avenues of regulating reaction rates exist: modulation of enzyme abundance and modulation of enzyme activity. Modulation of enzyme abundance is often referred to as hierarchical regulation, whereas modulation of enzyme activity is often referred to as metabolic regulation (Figure 1.1).

1.3.4.1 Hierarchical Regulation

Hierarchical regulation plays a pivotal role in shaping the cellular landscape, influencing enzyme abundance which, in turn, affects the capacity of metabolic pathways. The concentration of enzymes within the cell is determined by a dynamic interplay of processes including transcriptional control, post-transcriptional regulation, translation, and protein degradation. These mechanisms collectively orchestrate the availability of enzymes, a key determinant of cellular function and adaptation.

Transcriptional regulation affects the process of DNA to RNA transcription and includes various mechanisms:

- **promoter-related regulation**, which involves the interaction of transcription factors with specific DNA sequences [145], RNA polymerase recruitment to the promoter region [191], and mediator complex activation [203].
- **enhancer/silencer-mediated regulation**, such as enhancers and silencers of gene expression [205], cis- and trans-regulatory elements [4], and transcriptional activators and repressors [4, 15].
- **epigenetic regulation**, such as DNA methylation and histone modifications, which control gene expression by modifying chromatin structure [173], and chromatin remodeling complexes that alter the accessibility of DNA for transcription [52].

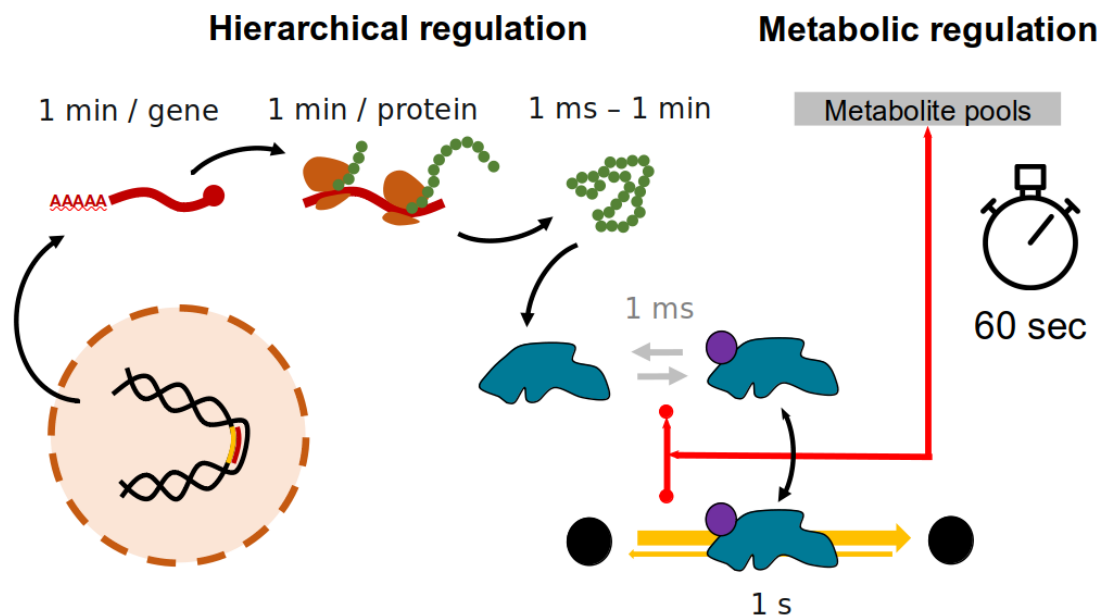


Figure 1.1: Metabolic Regulation. Enzyme abundances can be controlled by transcriptional regulation. This involves transcription of the DNA to RNA, translation of RNA into proteins, as well as protein folding and complex formation, before such regulation is effective. Typical time-scales of these processes are on the order of 1 minute for transcription, 1 minute for translation, and 1 millisecond to 1 minute for protein folding and complex formation. Even though in prokaryotes genes can start to be translated while they're still being transcribed, the process is still estimated to take at least a minute. On the other hand, enzyme activity can also be regulated directly by metabolites. For example, if substrate pools increase, this will increase the thermodynamic driving force of a reaction rate. However, metabolites may also modulate enzyme activity directly. If they are structurally similar to the cognate substrate they may do so by non-covalently binding at the active site, thereby acting as competitive inhibitors. Alternatively, they may do so by non-covalently binding outside the catalytic site, instead binding to a so-called allosteric site, and inducing a conformational change. This latter process is referred to as allosteric regulation, and such metabolites are called allosteric effectors or regulators. Finally, we also consider post-translational modifications, which are covalent modifications, such as phosphorylation, acetylation, methylation and glycosylation, that can also directly affect enzyme activity by altering the protein's conformation. The typical time-scale of the process of ligand-induced conformational change is on the order of a millisecond, whereas that of metabolite turnover is on the order of a second.

Post-transcriptional regulation operates on mRNA molecules after transcription, and encompasses a wide range of mechanisms:

- **alternative processing**, such as alternative splicing of pre-mRNA into different isoforms [207] and alternative polyadenylation to the 3' end of mRNA [214].
- **sequence modification**, such as the enzymatic conversion of adenosine to inosine via RNA editing [88, 160], or RNA modifications such as methylation or pseudouridylation [187].
- **interference and degradation**, such as RNA interference by microRNAs (miRNAs) and small interfering RNAs (siRNAs) [2], regulation of RNA stability via specific sequences in the 5' and 3' untranslated regions (UTRs) [228], and RNA-binding proteins masking the ribosome binding site or other regions [204].
- **localization and structural changes**, such as transport to specific subcellular location before translation [137] and the binding of small molecules to riboswitches which induce a conformational change [197].

Translational regulation focuses on controlling the process of protein synthesis itself, and includes the following mechanisms:

- **initiation factors and ribosome availability**, which control the initiation of translation and overall translation rate and efficiency [122].
- **translational activators and repressors**, which enhance translation by recruiting ribosomes and initiation factors, or inhibit it by blocking ribosome binding or recruitment, respectively [157].
- **codon usage and tRNA abundance** impact the translation process, where efficient translation depends on the availability of tRNAs that can recognize and deliver the appropriate amino acid for each codon [104].
- **metabolite sensing**, as a means to control translation initiation, enables the cell to respond to metabolic changes by adjusting protein synthesis [156].
- **protein product feedback regulation**, can act as a negative regulator to control its own translation, contributing to homeostasis through feedback inhibition [157].

Protein degradation includes passive degradation by instability, characterized by a half-life, as well as active mechanisms:

- **ubiquitin-proteasome system and tagging**, includes degradation via ubiquitin tagging and proteasome-mediated breakdown [155], as well as degradation based on specific N-terminal and C-terminal amino acid sequences [103, 125].
- **lysosomal and vesicular degradation**, includes the sequestration of cellular components in autophagosomes [65], degradation in lysosomes by lysozymes [25], and exosome-mediated degradation [218].
- **cellular quality control and signaling**, includes endoplasmic reticulum-associated degradation of misfolded proteins [185], regulated intramembrane proteolysis of membrane proteins to release functional fragments [36], and apoptotic proteolysis [159].

These mechanisms form a hierarchical cascade that operates on longer time scales, intertwining various layers of regulation to ensure fine-tuned control over enzyme abundance, which is essential for long-term adaptation of cellular metabolism. We note that protein degradation via the ubiquitin-proteasome system is a post-translational modification, and that lysosomal degradation, as well as regulated intramembrane proteolysis, can be induced via post-translational modifications [152, 149].

1.3.4.2 Metabolic Regulation

Metabolic regulation primarily influences enzyme activity, fine-tuning the catalytic efficiency of individual enzymes in response to immediate cellular demands. Enzyme activity can be rapidly influenced by protein-metabolite interactions, including covalent and non-covalent modifications.

Orthosteric regulation, also known as competitive inhibition, is a regulatory mechanism that involves the non-covalent binding of a molecule to the active site of an enzyme, blocking the cognate substrate from binding. This results in the inhibition of enzyme activity and subsequent metabolic reactions.

Allosteric regulation involves the non-covalent binding of regulatory molecules at specific sites other than the active site of an enzyme, resulting in conformational changes that impact enzyme activity. Allosteric regulation can be further categorized into:

- **Non-competitive regulation:** The regulatory molecule binds to an allosteric site, inducing a conformational change that either reduces or increases enzyme activity in a manner that is not dependent on substrate binding.
- **Uncompetitive regulation:** The regulatory molecule binds to the enzyme-substrate complex, inhibiting or activating its function and by altering either substrate or product binding affinity.

Post-translational regulation involves covalent modifications of enzymes and other proteins after translation. These modifications influence enzyme activity, stability, and subcellular localization. Examples of post-translational regulatory mechanisms include phosphorylation, acetylation and methylation and glycosylation, among many others.

These regulatory mechanisms discussed in this section collectively ensure the precise control of metabolic pathways; they operate on a similar time scale, providing rapid adaptation to environmental changes. However, the contemporary nomenclature used to classify these mechanisms may sometimes lead to a perceived differentiation across dimensions of time and space. Orthosteric and allosteric on the one hand, and post-translational modifications on the other, are often considered separate categories, seemingly differentiating them by their temporal and spatial characteristics. However, this differentiation blurs when considering that both competitive inhibition and allosteric regulation occur post-translationally. Moreover, many post-translational modifications occur outside the active site, whereas others may occur at the active site, effectively making them allosteric and orthosteric regulators, respectively [162]. Considering this, an alternative and potentially better approach might be to group these mechanisms based on the nature of their interactions; whether they are covalent or non-covalent. It seems that this approach could provide a more comprehensive and nuanced understanding of how these regulatory mechanisms influence metabolic pathways, bridging the apparent gap and false dichotomy allosteric and post-translational regulation.

The interconnectedness of hierarchical and metabolic regulation we've explored gives rise to a truly intricate web of control governing cellular metabolism. It's essential to recognize that these regulatory mechanisms often intersect and influence one another. This complexity, which is amplified by the contemporary definitions, challenges any oversimplified attempt to separate these mechanisms solely across temporal or spatial dimensions.

1.3.4.3 Cellular signaling

The intricate web of central carbon metabolism is not an isolated entity within the cell, but rather intricately intertwined with cellular signaling pathways. This interconnectedness underscores the dynamic nature of metabolism, which responds to various cellular signals and cues. Central carbon metabolism exhibits profound sensitivity to nutrient availability and growth factors. Cellular signals triggered by nutrient levels, such as glucose and amino acids, are transmitted to metabolic pathways to regulate flux in an attempt to ensure an optimal balance between energy production and biosynthesis. This allows cells to efficiently adapt their metabolism to match changing nutritional conditions, enhancing their survival and fitness. Moreover, the interplay between central carbon metabolism and cellular signaling pathways extends beyond nutrient availability. Growth factors and signaling cascades can modulate enzyme activities and expression levels, thereby influencing metabolic fluxes. These interactions highlight the intricate integration between metabolic processes and cellular signalling processes, where metabolic fluxes serve as both indicators and effectors of cellular states. In essence, central carbon metabolism not only governs energy generation and biosynthesis but also integrates cellular signaling inputs to fine-tune metabolic outcomes. This

dynamic interplay between metabolism and signaling ensures that the cell's metabolic activities are in harmony with its physiological needs, ultimately contributing to cellular function and survival.

1.3.5 Experimental techniques for studying central carbon metabolism

The exploration of central carbon metabolism has been significantly advanced by modern experimental techniques that leverage high-throughput methods to generate vast amounts of data. These methods offer insights into the dynamic behavior of metabolic networks and enable the identification of key regulatory nodes. The resulting "data deluge" poses both opportunities and challenges, as making sense of the collected data and extracting meaningful insights are complex undertakings.

Metabolomics plays a crucial role in unveiling the dynamic interplay of metabolites within central carbon metabolism. Advances in mass spectrometry and nuclear magnetic resonance spectroscopy have enabled the simultaneous quantification of numerous metabolites in a single experiment [67, 199, 7, 230, 75]. Metabolomic profiling generates comprehensive snapshots of metabolite concentrations, highlighting changes in response to environmental conditions, genetic modifications, and metabolic interventions. By integrating metabolomic data with computational tools, researchers can construct metabolic maps and predict how perturbations impact metabolic fluxes.

Stable isotope tracing techniques provide a powerful means to investigate metabolic fluxes through central carbon pathways [76, 60]. By introducing isotopically labeled substrates, such as labeled glucose or amino acids, researchers can track the fate of labeled carbon atoms as they traverse the metabolic network. Mass spectrometry and nuclear magnetic resonance spectroscopy are again instrumental in measuring isotopic enrichment in metabolites. These techniques yield dynamic flux information, allowing the identification of pathway contributions and the quantification of metabolic turnover rates.

High-throughput screening methodologies have revolutionized the study of central carbon metabolism by enabling the rapid characterization of enzyme activities, metabolite interactions, and regulatory effects [199, 234]. These methods leverage robotic platforms and microfluidics to perform thousands of experiments in parallel. For instance, enzyme assays can be miniaturized to measure reaction rates for numerous enzyme-substrate pairs simultaneously. Such approaches facilitate the identification of allosteric effectors, substrates, and inhibitors on a global scale.

"Omics" technologies integration, including genomics, transcriptomics, and proteomics, yield extensive molecular data related to central carbon metabolism. Genomic sequencing and gene expression profiling identify genes involved in metabolic pathways and regulatory networks. Transcriptomics reveals gene expression patterns, highlighting transcriptional responses to changing conditions. Proteomics quantifies protein abundances and post-translational modifications, offering insights into how the proteome responds to metabolic cues. The integration of these high-throughput techniques has generated a wealth of data, shedding light on the complexity of central carbon metabolism [81]. However, managing and interpreting this abundant information presents challenges in data integration, model construction, and hypothesis generation. Computational tools, encompassing network reconstruction algorithms and machine learning approaches, play a pivotal role in transforming raw data into analyzable formats. These techniques empower researchers to decipher the regulatory mechanisms underlying metabolic responses, contributing to a deeper exploration of cellular behavior.

1.3.6 Computational techniques for understanding metabolic flux

The exploration of central carbon metabolism is greatly facilitated by computational techniques that harness the power of data integration and modeling to unravel its intricate dynamics. These approaches bridge the gap between experimental data and mechanistic understanding, enabling researchers to simulate and analyze metabolic behaviors across various conditions.

Constraint-based models are computational frameworks that leverage stoichiometric information and known reaction constraints to predict feasible metabolic flux distributions. These models provide insights into the possible metabolic states that a cell can achieve, aiding in the

interpretation of high-throughput experimental data. Constraint-based models offer a balance between computational efficiency and biological accuracy, making them valuable tools for analyzing large-scale metabolic networks. One of the pivotal breakthroughs in the recent history of systems biology is the advent of genome-scale metabolic models (GSMMs) [166, 91]. These models encapsulate the entirety of cellular metabolic reactions, providing a comprehensive framework to study flux distributions, nutrient utilization, and regulatory interactions. The reconstruction of GSMMs for various organisms, including *Escherichia coli* [166], *Saccharomyces cerevisiae* [70], and even of human cells [200] tissues [222] and their gut microbiota [143], has enabled predictions of cellular phenotypes, growth rates, and metabolic responses to varying environmental conditions. The maturation of GSMMs has been paralleled by the emergence of constraint-based modeling techniques. Flux balance analysis (FBA) and related methods leverage stoichiometric constraints to predict metabolic flux distributions while accounting for thermodynamics and cellular objectives [165]. These approaches have offered insights into optimal metabolic states, potential drug targets, and the interplay between metabolic networks and cellular functions.

Kinetic models integrate enzyme kinetics and reaction mechanisms to simulate the dynamic behavior of metabolic pathways and even entire subsystems [48, 210]. By considering the rate equations for individual reactions, these models capture the time-dependent changes in metabolite concentrations and fluxes. While more complex and computationally intensive than constraint-based models, kinetic models offer a deeper level of detail and can reveal non-linear interactions and regulatory effects. Allosteric regulation, a cornerstone of systems biology, has been brought into focus with the development of dynamic models that incorporate regulatory interactions. Advanced computational methods, such as kinetic modeling and dynamic flux balance analysis (dFBA) [144], enable the simulation of metabolic responses to perturbations and environmental changes. This has shed light on how allosteric effectors, feedback loops, and regulatory motifs influence cellular behavior, contributing to our understanding of metabolic homeostasis and adaptability.

Hybrid and multi-scale models often integrate multiple modeling techniques, offering the ability to obtain a more comprehensive understanding of cellular physiology. This approach enables the exploration of genetic and environmental influences on metabolic responses, elucidating system dynamics across various organizational levels and time scales, and even extending to encompass whole-cell models [59, 113]. Furthermore, the power of multi-scale models lies in bridging scales and times, connecting molecular interactions with cellular behavior [62]. In addition, hybrid modeling techniques, such as the fusion of constraint-based and kinetic models, or the integration of boolean networks with constraint-based models, expand the frontier of metabolism research [233, 56, 45]. Together, the integration of omics data and hybrid modeling propels us toward a unified understanding of metabolic regulation, revealing key regulatory nodes and network motifs that orchestrate the complex cellular processes.

Machine learning techniques have gained significant traction in deciphering intricate patterns and predicting metabolic behaviors from large and diverse datasets. Leveraging the power of computational algorithms, these approaches extend from simpler regression and classification analyses to advanced deep learning models. By mining complex data structures, machine learning models excel at extracting meaningful insights that may otherwise remain hidden. They play a pivotal role in identifying correlations, classifying diverse metabolic states, and contributing to the discovery of novel regulatory interactions. Through their predictive capabilities, machine learning techniques guide the formulation of hypotheses that can inform experimental designs and elucidate intricate metabolic networks [63, 120, 57]. Moreover, these methods provide a versatile toolbox for advancing our comprehension of metabolic dynamics and their functional implications.

In essence, the recent history of systems biology in cellular metabolism has been marked by the integration of diverse approaches to decipher the orchestration of metabolic networks. The synergy between experimental techniques, computational modeling, and omics data has propelled our understanding of how metabolic pathways respond to environmental cues, adapt to changing conditions, and contribute to cellular fitness. This multidisciplinary endeavor continues to refine our knowledge of cellular metabolism, paving the way for novel insights and potential applications in biotechnology, medicine, and synthetic biology.

1.4 *Escherichia coli* metabolism

The study of *Escherichia coli* (*E. coli*) metabolism has a history that stretches over a century. In the early 1900s, scientists first began to investigate its metabolic processes, focusing on its ability to break down sugars and other organic compounds. These studies laid the foundation for the identification of key metabolic pathways in *E. coli*, such as the glycolytic and the pentose phosphate pathway. In the subsequent decades many more enzymes, catalytic reactions and metabolic pathways were identified, as well as a number of mechanisms that control metabolic processes.

The dynamic interplay between metabolic processes and gene expression in *E. coli* reveals a sophisticated regulatory network that ensures efficient growth and adaptation to varying environmental conditions. This regulatory coordination involves a series of mechanisms that help the bacterium detect and respond to nutrient availability, prioritize carbon source utilization, and balance protein synthesis with energy generation. However, our understanding of the role of these regulatory interactions in shaping responses, such as those seen in metabolism during the adaptation to a change in the environment, is still very limited.

1.4.1 Transcriptional regulation

The phenomenon of preferential carbon utilization is exemplified by the classic case of the glucose-lactose diauxic shift, where *E. coli* initially consumes glucose exclusively and then switches to lactose once glucose is depleted [105]. Glucose is identified as a favored carbon source in many organisms, often inhibiting the use of secondary sources through a mechanism called glucose repression or carbon catabolic repression (CCR) [142]. CCR is a vital regulatory mechanism affecting numerous genes in various bacterial species [87, 50]. The molecular basis of CCR has been elucidated for sugars transported by the phosphotransferase system (PTS), including glucose and mannose. This transport pathway leads to reduced cyclic AMP (cAMP) levels, which in turn affects cAMP receptor protein (CRP), a global regulator essential for activating carbon utilization promoters. PTS sugars lower CRP activity and inhibit alternative carbon systems, and recently discovered small regulatory RNAs (sRNAs) have been implicated in CCR [13, 21]. Transcriptional regulation in *E. coli* extends beyond individual pathways, embodying the concept of operons. This phenomenon, initially proposed by Jacob and Monod [107], involves grouping functionally related genes under a single promoter, streamlining their co-expression. CRP binds to specific DNA sites and activates the transcription of genes involved in energy-intensive processes, such as glucose metabolism, enabling it to prioritize glucose consumption over that of other carbon sources when it is readily available. The relative contribution of these mechanisms to CCR varies depending on the carbon source, while the metabolic and energetic state of the cell also impacts cAMP levels [232]. α -ketoacids, central carbon metabolites, can negatively influence cAMP levels under certain conditions, forming a feedback loop that regulates carbon uptake based on cell needs between anabolism and catabolism [69]. A study into the utilization of non-glucose sugars in *E. coli* showed preferential uptake of specific sugars over others, revealing a hierarchy based on the growth rates these sugars support, which is linked to the activity of the master regulator cAMP-CRP. Under certain conditions both sequential and simultaneous expression of sugar systems was observed, suggesting a multi-objective optimization strategy that balances rapid growth and future adaptation to less preferred sugars [3].

While numerous regulatory mechanisms influence metabolic transitions, determining which are active drivers of adaptation in response to environmental changes is challenging. The commonly used approach involves inferring active regulatory events by comparing steady-state endpoints of adaptations. Extending this, Gerosa *et al.* conducted a study in which they investigated the dynamics of gene expression in *E. coli* central carbon metabolism using a "pseudo-transition" analysis approach. Pseudo-transition analysis is a method that aims to identify active regulatory events driving dynamic cellular adaptations by assessing the proportionality between regulatory inputs and functional outputs across different steady-state conditions. This research focused on over 30 transcription factors by using a library of fluorescent transcriptional reporters to quantify the activity of central metabolic promoters across a spectrum of 26 environmental conditions. This investigation has uncovered two primary modes of regulation: the dominant influence of global regulation by the growth rate-dependent cellular expression machinery and specific, transcription factor-mediated regulation that is primarily localized to a subset of promoters, particularly those associated with the TCA cycle and the Embden-Meyerhoff-Parnas (EMP) pathway. Interestingly, the investigation identifies just two transcription factors, Cra and Crp, and a trio of regulatory

metabolites, cAMP, fructose-1,6-bisphosphate (FBP), and fructose-1-phosphate (F1P), as the primary drivers of a substantial portion of specific transcriptional regulation. However, around 70% of flux changes remain unexplained [84].

Interestingly, whereas glucose is the preferred carbon source of *E. coli*, under specific circumstances involving a single amino acid (arginine, glutamate, or proline) as the sole nitrogen source, *E. coli* exhibits slower growth on glucose than on other sugars. This counterintuitive behavior, resulting in a reversed diauxic shift, is attributed to a metabolic imbalance characterized by elevated TCA-cycle metabolite levels, particularly α -ketoglutarate, and low levels of the regulatory molecule cAMP [31]. This highlights the existence of complex interactions between various metabolic subsystems, emphasizing the need to study their intricate interplay for a comprehensive understanding. Table 1.1) provides an overview of several of the best-known transcription factors known to modulate metabolism in *E. coli*.

Table 1.1: Well-known transcription factors (TF) involved in regulating *E. coli* metabolism.

TF	Role
CRP	Responds to cellular cAMP levels, prioritizing glucose consumption and energy-intensive processes
Cra	Modulates central carbon metabolism in response to preferred carbon sources
LacI	Regulates lactose metabolism & Represses transcription of lac operon genes in the absence of lactose
ArcA	Influences gene expression during anaerobic growth, affecting electron acceptors and metabolic balance
FNR	Regulates the switch between aerobic and anaerobic metabolism based on oxygen levels
CrpP	Shares functional similarities with CRP, affecting carbon source utilization and energy metabolism
IclR	Regulates genes related to glyoxylate shunt enzymes and influences acetate utilization
MalT	Activates maltose transport and metabolism genes in response to maltose availability
LexA	Involved in the cellular response to DNA damage or inhibition of DNA replication
CysB	Regulates genes related to sulfur metabolism and cysteine biosynthesis
NtrC	Activates genes involved in nitrogen assimilation and metabolism in response to nitrogen availability
GadE	Involved in the maintenance of pH homeostasis, including the principal acid resistance system
FadR	Regulates genes related to fatty acid metabolism and fatty acid degradation pathways
NarL	Part of the two-component regulatory system for response to nitrate and nitrite availability
PhoB	Controls the phosphate regulon, impacting genes involved in phosphate transport and metabolism

1.4.2 Allosteric regulation

Despite significant accumulated experimental data on the role of small molecules in enzyme regulation, genome-scale understanding of such interactions remains limited. Allosteric regulation and competitive inhibition by small molecules are crucial for homeostasis and rapid adaptations to environmental changes. While feedback inhibition conserves resources in pathways like amino acid biosynthesis, feedforward activation allows for swift glucose import after periods of carbon starvation. In a recent study by Reznik *et al.* the authors reconstructed the small molecule regulatory network (SMRN) of *E. coli* by aggregating data from diverse sources, such as the Braunschweig Enzyme Database (BRENDA) and BioCyc database [190, 46]. The resulting network reveals widespread enzyme regulation by endogenous metabolites and can be overlaid on a genome-scale metabolic model, enabling a direct comparison of metabolic topology and regulatory patterns. Integrating the SMRN with experimental metabolite concentrations and binding affinities offered insights into the dual roles of small molecules as substrates or inhibitors and their context-dependent impact on metabolic flux regulation [184].

The reconstructed *E. coli* small molecule regulatory network (SMRN) comprises 1,669 interactions involving 321 unique endogenous metabolites and 364 unique enzymes (EC numbers). Around 83% of these interactions are inhibitory, and approximately half of the unique EC numbers in the *E. coli* genome-scale model are regulated by at least one native metabolite. Similarly, 320 distinct native metabolites were found to regulate at least one enzyme. Several metabolites and EC numbers are prominently involved in regulatory interactions. Notably, ATP, AMP, ADP, PI, PPI, NADPH, GTP, cysteine, pyruvate, and phosphoenolpyruvate (PEP) are frequent regulators. The authors also note that small molecule regulators can affect enzymes beyond their native substrates or products, potentially enabling distant signaling between metabolic pathways [184].

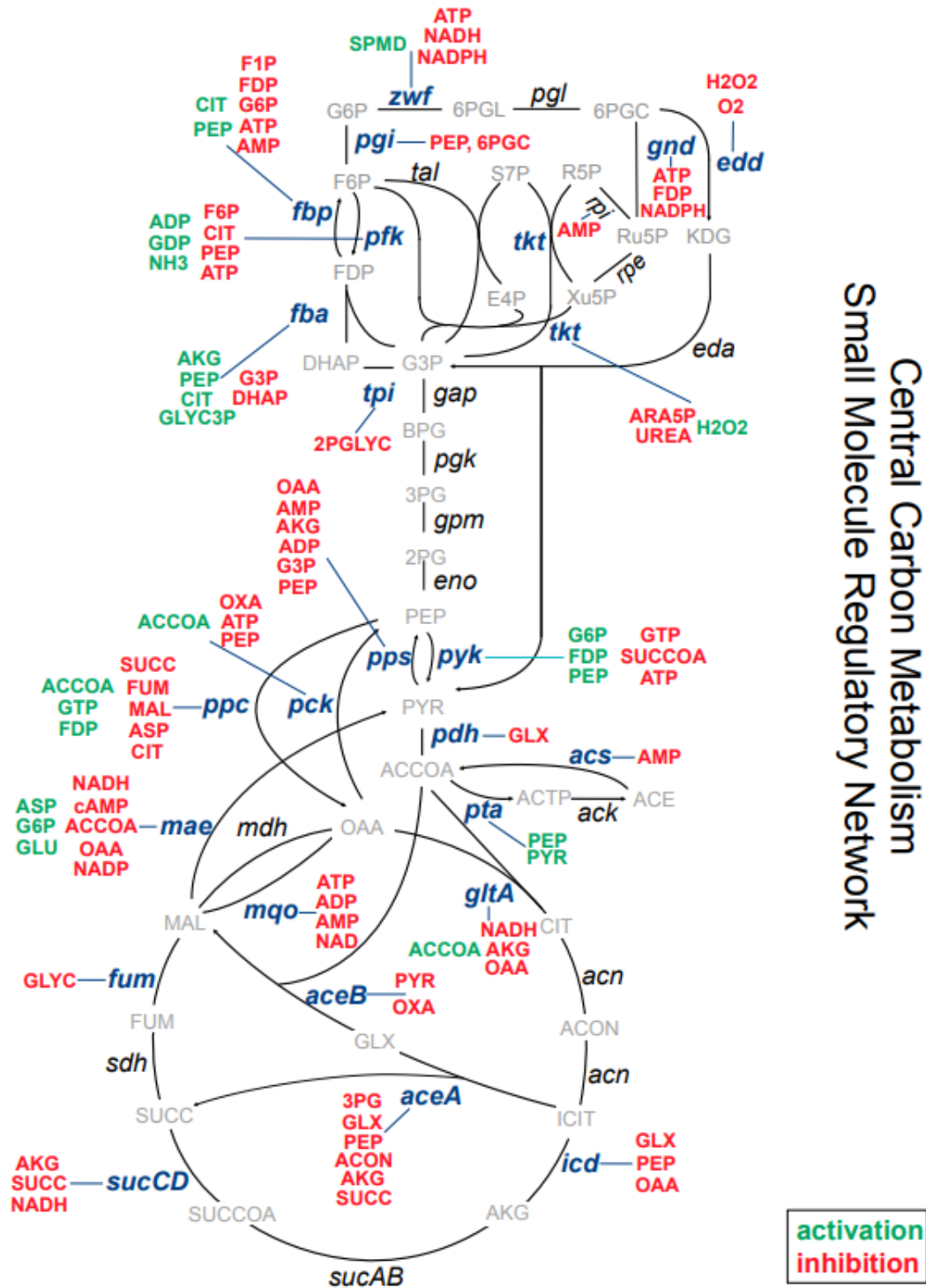


Figure 1.2: Small molecule regulatory network of *E. coli* central carbon metabolism. Depiction of the small molecule regulatory interactions in the central carbon metabolism of *E. coli*. Red metabolites are inhibitors and green metabolites are activators of the indicated reactions. Adopted from Reznik *et al.* [184].

1.4.3 Constraint-based models

The genome-scale reconstruction iJO1366, an achievement described by Orth *et al.*, revolutionized our understanding of *E. coli*'s metabolism [166]. This comprehensive reconstruction encapsulates 1366 genes, 2251 metabolic reactions, and 1136 metabolites, reflecting intricate cellular compart-

ments and newly discovered pathways. The framework, a product of experimental screenings and gap-filling strategies, empowers flux balance analysis to predict growth rates, substrate uptake rates, and product secretion rates. Moving beyond the foundational genome-scale metabolic models, researchers have employed computational techniques that enable a deeper understanding of how metabolic networks function. One such technique is Flux Balance Analysis (FBA) [165], a mathematical approach that utilizes the information encoded within these large-scale reconstructions to predict cellular behaviors under varying conditions. FBA seeks to maximize or minimize a specific objective function while ensuring that the network's reactions satisfy mass balance constraints. Typically, the objective function is chosen to represent a cellular goal, often the maximization of biomass production rate. By optimizing the allocation of fluxes through reactions, FBA provides a snapshot of how the network's reactions collaborate to achieve the chosen objective. It also considers additional constraints, such as reaction reversibility, thermodynamics, and capacity limitations, to reflect the physiological realities of the cell. FBA has provided a multitude of important insights into the metabolism of *E. coli* and other microorganisms. Some of the most significant results and findings obtained using FBA in *E. coli* include:

- **Optimal growth predictions:** predicting optimal growth conditions and rates under various nutrient scenarios, providing insights into growth regulation [165].
- **Metabolic pathway discovery:** identification of novel pathways and alternative routes for metabolite synthesis, expanding our understanding of *E. coli*'s metabolic capabilities [73].
- **Essential reactions and genes:** pinpointing essential reactions and genes by simulating knockouts, aiding in the identification of critical metabolic nodes and potential drug targets [77].
- **Metabolic engineering:** guiding metabolic engineering efforts, aiding in strain design for improved yields, selecting gene targets for modification, and predicting genetic modifications' effects [43].
- **Stress response insights:** revealing how *E. coli* adapts its metabolic network under stress, offering insights into microbial behavior in changing environments [54].
- **Trade-offs and constraints:** exposing trade-offs and constraints within the metabolic network, illuminating resource allocation's impact on cellular fitness [193].
- **Pathway analysis:** assessing flux distributions across pathways, providing a systems-level view of nutrient utilization and metabolite production [189].
- **Carbon source utilization:** predicting the preferred order of carbon source utilization and the metabolic shifts in response to different nutrient availabilities [78].
- **Phenotypic predictions:** predicting diverse phenotypes beyond growth, such as metabolite production, by constraining fluxes through specific reactions [219].
- **Comparative analysis:** enabling comparative studies of metabolic capabilities across strains or species, revealing evolutionary adaptations and differences [212].

While constraint-based genome-scale models and FBA analysis have significantly enhanced our comprehension of *E. coli* metabolism, they do have limitations. If we wish to venture beyond static steady state analysis and into the dynamic domain of transient responses, kinetic models are a more suitable alternative. Kinetic models take into account the rate equations of individual reactions, enzyme kinetics, and the time-dependent interactions between different components of the metabolic network. This allows them to capture the nuances of cellular behavior under varying environments and stimuli, providing insights into how the metabolism adapts and responds in real time.

1.4.4 Kinetic models

One of the key advantages of kinetic models in the study of *E. coli* metabolism is their ability to capture transient responses and complex regulatory mechanisms. These models can simulate how metabolite concentrations change over time, taking into account factors such as enzymatic

saturation, substrate availability, and allosteric regulation. Kinetic models are particularly useful for studying processes involving fast dynamics, such as signaling cascades and transient metabolic shifts in response to perturbations. In the context of *E. coli* metabolism, kinetic models have yielded valuable insights. Notably, the work of Chassagnole *et al.* presented a dynamic model of the central carbon metabolism of *E. coli*. The study incorporated enzyme kinetics and mass balance equations to describe the dynamics of glycolysis, pentose phosphate pathway, and the TCA cycle. This approach enabled the simulation of metabolite concentrations, fluxes, and intracellular redox balances over time. By fitting the model to experimental data, the authors demonstrated its capability to predict dynamic responses and capture transient metabolic behaviors [48]. Furthermore, kinetic models have shed light on the effects of post-translational modifications, enzyme cooperativity, and feedback regulation in *E. coli* metabolism. These models allow researchers to simulate how changes in enzyme activity and metabolite levels impact metabolic fluxes and cellular behavior. By integrating experimental data on enzyme kinetics and metabolite concentrations, kinetic models can be parameterized to match observed metabolic behaviors as closely as possible. Some of the most significant results and findings obtained using kinetic modeling in *E. coli* include:

- **Detailed enzyme kinetics:** providing a deeper understanding of enzyme kinetics, allowing researchers to quantify reaction rates, enzyme activities, and substrate affinities for various metabolic pathways [134, 195].
- **Metabolic control analysis (MCA):** enabling the application of MCA, a powerful tool for quantifying the relative control that enzymes have over metabolic fluxes. This analysis helps identify key regulatory points in the network and assess their impact on overall metabolism [79].
- **Uncovering regulatory mechanisms:** facilitating the exploration of regulatory mechanisms, including feedback loops, feedforward activation, and allosteric regulation. By integrating experimental data with kinetic parameters, researchers can better understand how regulatory interactions shape metabolic behavior [210, 51].
- **Stress responses and adaptation:** allowing the simulation of cellular responses to various stresses, such as nutrient limitations or environmental changes. These models help uncover how metabolic networks adapt to maintain essential functions and achieve homeostasis [109, 49].
- **Substrate preference and competition:** helping to explain the preference for specific substrates under varying conditions. They reveal how enzyme affinities and kinetic parameters influence substrate utilization and competition between alternative pathways [11].
- **Prediction of dynamic responses:** enabling the prediction of dynamic responses to changing conditions over time. This is crucial for understanding transient behavior during metabolic shifts and adaptation [135, 121].
- **Quantification of metabolite pools:** providing insights into the dynamics of metabolite concentrations, helping to elucidate the interplay between production, consumption, and transport of metabolites within the cell [225].
- **Flux redistribution under perturbations:** revealing how metabolic fluxes redistribute when enzymes are perturbed, aiding in the identification of potential drug targets or engineering strategies [117, 116].
- **Identification of rate-limiting steps:** allowing the identification of rate-limiting steps in metabolic pathways, guiding efforts to enhance pathway efficiency or productivity [9].
- **Informed Metabolic Engineering:** providing a quantitative framework for metabolic engineering. They guide the design of optimal genetic modifications for improving production of biofuels, pharmaceuticals, and other valuable compounds [161].
- **Integration with -omics data:** providing a comprehensive view of cellular processes through the integration of omics data (genomics, transcriptomics, proteomics), and thereby enabling a more accurate representation of metabolic behavior [59, 113].

While kinetic models provide a powerful tool for obtaining insights into *E. coli* metabolism, the study of metabolic networks also faces experimental and computational challenges that need to be addressed. In the following section, we will delve into the various challenges researchers encounter when studying the dynamics and complexities of *E. coli* metabolism, including the limitations of current experimental techniques, the role of genetic variability, and the intricacies of cellular regulation.

1.4.5 Experimental challenges

From cellular heterogeneity to temporal dynamics, and from quantifying metabolite concentrations to understanding the impact of post-translational modifications, the landscape of experimental challenges is vast. In order to unravel the complex interplay of biochemical reactions and regulatory mechanisms within the cell, researchers must confront a range of experimental challenges that impact how metabolic networks are characterized and understood. These challenges extend beyond the confines of the laboratory bench, encompassing factors such as cellular heterogeneity, dynamic perturbations, and the intricacies of post-translational modifications. Overcoming these challenges demands innovative techniques and approaches capable of capturing the nuanced metabolic behavior within living cells.

- **Metabolic interplay:** The interconnectedness of metabolic pathways means that changes in one pathway can affect others. The core challenge here is delineating system boundaries.
- **Time scales:** Metabolic processes in cells are highly dynamic and can vary over time scales ranging from milliseconds to hours. Capturing these rapid changes and understanding the kinetics of metabolic reactions in real-time presents a challenge.
- **Quantitative metabolomics:** Accurate quantification of metabolite concentrations is crucial for understanding metabolic fluxes and network behavior. However, metabolomics techniques often face challenges in quantifying low-abundance metabolites accurately.
- **Dynamic perturbations:** The effects of dynamic perturbations, such as sudden nutrient changes or environmental shifts, on metabolic networks are complex to unravel. Moreover, finding an appropriate perturbation may prove challenging.
- **Post-translational modifications (PTMs):** PTMs can significantly impact enzyme activity and metabolic regulation. Exploring the effects of PTMs on enzyme kinetics and metabolic behavior is a challenge due to the diversity of possible modifications.
- **Interaction with macromolecules:** Metabolic enzymes can interact with other macromolecules, such as RNA and proteins, influencing their activity.
- **Cellular heterogeneity:** *E. coli* populations can exhibit heterogeneity in their metabolic states, especially in response to changing environments. Single-cell techniques are needed to capture this variability and understand how individual cells contribute to the population-level behavior.

The majority of knowledge on allosteric interactions to date originates from decades of research in enzymology. Central to this research are *in vitro* activity assays in which enzyme kinetics are studied by observation of substrate consumption and/or product formation [14]. This has proven useful for the identification of catalytic reaction mechanisms as well as of metabolites regulating enzymatic activity. The advantage of an *in vitro* enzyme assay is that there are only a relatively small number of distinct states present in the test tube at any given time, and hence these can be more easily observed as compared to a situation that involves cells or cell extracts. Due to the smaller number of states in the system it is easier to induce a local change in any one of them and study local effects, since complex interaction structures, such as feedback loops involving additional states normally present in the cell, are absent here. When no other reactions are taking place that might affect substrate or product levels the observational data can be used to determine kinetic parameters of the enzyme. However, there are also serious limitations to this approach. Starting with the enzyme of interest, it must be isolated from the cell, something which is hard if not downright impossible for enzymes that are membrane bound or those part of protein complexes. Then, it has to be purified, because impurities can bring about confounding effects. The enzyme

also needs to remain active, meaning that the isolation and purification procedure should not affect its tertiary structure. Moreover, the *in vivo* physiochemical conditions present in the cell need to be emulated, including pH, ionic strength, redox state and the presence of the necessary cofactors [16]. In practice this means that the conditions for any enzyme of interest under study need to be optimized. As a result this approach is laborious, which, outside of the aforementioned constraints, makes systematic assessment of metabolite-protein interactions prohibitively expensive.

More recently, methods that allow for high-throughput measurement have started to be developed [1]. These solutions these methods offers are versatile. For example, grouping metabolites into subsets and studying the commonality of effects in samples where a specific metabolite is present [40], the detection of physical interactions by means of nuclear magnetic resonance spectrometry [68] and the detection of such by mass spectrometry-coupled limited-proteolysis [174] are new approaches that recently have been explored. While these studies have been conducted in attempt to systematically identify metabolite-protein interactions, the overlap in the interactions they find is low. To illustrate, of the 162 positive hits founds in the two studies of which the enzyme-metabolite pair was also tested in the other study; only 11 were predicted in both [174, 68]. While these methods aim to address the question of what interactions exist, yielding an important prerequisite and necessary conditions, it yields a static picture. In order to understand the functional role of regulatory interactions in the system, computational models are needed.

1.4.6 Computational challenges

Computational challenges in the study of regulation in *E. coli* metabolism arise due to the intricate nature of cellular regulatory mechanisms and the complexity of metabolic networks. Here are some key computational challenges researchers face when studying regulation in *E. coli* metabolism:

1. **Data integration:** Integrating heterogeneous omics data such as transcriptomics, proteomics, and metabolomics poses challenges in accurately representing the underlying regulatory dynamics and reconciling inconsistencies.
2. **System demarcation:** Defining the scope of the model is a nuanced task involving balancing model complexity, scale, and dimensionality. This requires including the necessary complexity to study the phenomena of interest while avoiding excessive detail. The challenge lies in determining the appropriate level of detail, which is not known *a priori*, often requires an iterative process of discovery, and remains uncertain *post hoc*.
3. **Structure identification:** Inference of missing edges in the network, both metabolic reactions and regulatory interactions, from experimental data is complex due to noise, incomplete information, and context-dependent relationships.
4. **Parameter estimation:** Accurate quantification of interaction strengths and kinetics is crucial. However, estimating parameters for regulatory models is challenging due to limited data and nonlinear regulatory dynamics.
5. **Model validation:** Validating regulatory models against experimental observations is intricate, considering the network complexity and data limitations. Incorporating time-series data is essential for validating kinetic models.

State-of-the-art kinetic models of *E. coli* CCM often ignore metabolic regulation, usually because of missing kinetic information and the inherent difficulty accurately measuring fast metabolic responses [202, 110, 94]. Kinetic models that try to capture this level of regulation rely on interactions previously reported in literature, and include these in an *ad hoc* manner [48, 216], which is prone to errors in the form of missing or falsely inferred interactions. Recently, kinetics models have been developed that combine transcriptional, reactant and metabolic regulation, however, these were only validated by their ability to predict steady-state levels in wild type and gene knockout mutants [172, 117, 176, 108], not by their ability to capture system dynamics. Interestingly, the latest model published to date suggests that metabolic regulation is the main control mechanism [150]. Moreover, the authors find that flux control is distributed throughout the network, thereby refuting claims made in previous work that indicated flux control is most strongly exerted by a limited number of ‘rate-limiting’ reactions. Evidently, there is no consensus on the relevant flux control mechanisms, reflecting that certain regulatory mechanisms are more relevant under specific conditions than others.

1.5 Research goals

This thesis aims to address key challenges in understanding the regulation of central carbon metabolism in *E. coli* through a systems biology perspective. By rectifying ambiguities and misconceptions surrounding the field's terminology, we provided a refined conceptualization that serves as an overarching guide for our subsequent investigations within a coherent systems biology framework. Our research objectives are structured as follows:

Prediction of allosteric interactions by single reaction modeling

Our first research goal involves the identification of allosteric regulators within *E. coli* central carbon metabolism. Leveraging steady-state data, we employ a single-reaction modeling approach to predict metabolite-protein interactions. Quantifying the predictive power of the Michaelis-Menten equation, we assess the contribution of allosteric regulation to these predictions. Experimental validation through enzyme assays provides tangible evidence for the existence of nine previously unrecognized allosteric interactions.

Prediction of allosteric interactions in the tricarboxylic acid cycle using an ensemble of systems of coupled differential equations

The second research goal centers on unraveling the dynamic implications of allosteric interactions within the tricarboxylic acid (TCA) cycle. Using systems of coupled differential equations to assess the regulatory importance of allosteric interactions in the metabolic response of *E. coli* CCM upon an environmental perturbation. We meticulously construct these models in order to minimize the bias introduced by assumptions, incorporate as much of the available data as possible, and leverage thermodynamic constraints to ensure a physically realistic portrayal of the system's dynamics. By assembling an ensemble of models featuring various regulatory interaction topologies, we evaluate the contributions of protein-metabolite interactions in fitting the observed time series data. By aggregating the predictions of this ensemble we compute a score that reflects how much each of the putative allosteric interactions has contributed to explain the observed dynamics. Our findings illuminate the intricate interplay of regulatory mechanisms governing the TCA cycle and emphasize the significance of considering dynamic responses.

By addressing these research goals, we aim to contribute to the advancement of systems biology and enhance our understanding of the intricate metabolic network of *E. coli*.

Bibliography

- [1] Michael G Acker and Douglas S Auld. “Considerations for the design and reporting of enzyme assays in high-throughput screening applications”. In: *Perspectives in Science* 1.1-6 (2014), pp. 56–73.
- [2] Neema Agrawal et al. “RNA interference: biology, mechanism, and applications”. In: *Microbiology and molecular biology reviews* 67.4 (2003), pp. 657–685.
- [3] Guy Aidelberg et al. “Hierarchy of non-glucose sugars in Escherichia coli”. In: *BMC systems biology* 8.1 (2014), pp. 1–12.
- [4] Bruce Alberts. *Molecular biology of the cell*. Garland science, 2017.
- [5] Garland E Allen. “Mechanism, vitalism and organicism in late nineteenth and twentieth-century biology: the importance of historical context”. In: *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36.2 (2005), pp. 261–283.
- [6] Garland E. Allen. *Life Science in the Twentieth Century*. Cambridge University Press, 1978.
- [7] Saleh Alseekh et al. “Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices”. In: *Nature methods* 18.7 (2021), pp. 747–756.
- [8] Etya Amsalem et al. “The physiological and genomic bases of bumble bee social behaviour”. In: *Advances in Insect Physiology*. Vol. 48. Elsevier, 2015, pp. 37–93.
- [9] Maciek R Antoniewicz, Joanne K Kelleher, and Gregory Stephanopoulos. “Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements”. In: *Metabolic engineering* 8.4 (2006), pp. 324–337.
- [10] von Karl Ernst Baer. “Über die Bildung des Eies der Säugetiere und des Menschen”. In: *Übersetzung ins Deutsche von Benno Ottow. Leipzig* (1927).
- [11] Arren Bar-Even et al. “The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters”. In: *Biochemistry* 50.21 (2011), pp. 4402–4410.
- [12] Megan G Behringer et al. “Escherichia coli cultures maintain stable subpopulation structure during long-term evolution”. In: *Proceedings of the National Academy of Sciences* 115.20 (2018), E4642–E4650.
- [13] Chase L Beisel and Gisela Storz. “The base-pairing RNA spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in Escherichia coli”. In: *Molecular cell* 41.3 (2011), pp. 286–297.
- [14] Hans Ulrich Bergmeyer and Erich Bernt. “UV-assay with pyruvate and NADH”. In: *Methods of enzymatic analysis*. Elsevier, 1974, pp. 574–579.
- [15] Lacramioara Bintu et al. “Transcriptional regulation by the numbers: models”. In: *Current opinion in genetics & development* 15.2 (2005), pp. 116–124.
- [16] Hans Bisswanger. “Enzyme assays”. In: *Perspectives in Science* 1.1-6 (2014), pp. 41–55.
- [17] James W Black, Gertrude B Elion, and George H Hitchings. “Nobel Lectures, 1988 Physiology Or Medicine, DRUGS FROM EMASCULATED HORMONES: THE PRINCIPLES OF SYNTOPIC ANTAGONISM, December 8, 1988 by JAMES BLACK AND THE PURINE PATH TO CHEMOTHERAPY, December 8, 1988 by GERTRUDE B. ELION AND SELECTIVE INHIBITORS OF DIHYDROFOLATE REDUCTASE, December 8, 1988 by GEORGE H. HITCHINGS, JR.” In: *Nobel Lectures* 1.4 (1995).

-
- [18] JF Blumenbach. “Über den Bildungstrieb und das Zeugungsgeschäfte, Göttingen”. In: *Search in* (1781).
- [19] Johann Friedrich Blumenbach. *Handbuch der naturgeschichte*. Lechner, 1832.
- [20] Johann Friedrich Blumenbach and Joseph Banks. *De generis humani varietate nativa*. Vandenhoeck et Ruprecht, 1795.
- [21] Maksym Bobrovskyy and Carin K Vanderpool. “Regulation of bacterial metabolism by small RNAs using diverse mechanisms”. In: *Annual review of genetics* 47 (2013), pp. 209–232.
- [22] Hieronymus Bock. *Kreutterbuch: darin Unterscheidt Namen unnd Würckung der Kreutter, Stauden, Hecken und Beümen, sampt jren Früchten, so inn teütschen Landen wachsen, auch derselbigen eygentlicher und wolgegründter Gebrauch in der Artzney, fleißig dargeben;... und jetzund von newem fleißig ubersehen, gebessert und gemehret...* Rihel, 1952.
- [23] Maarten C Boerlijst and Paulien Hogeweg. “Spiral wave structure in pre-biotic evolution: hypercycles stable against parasites”. In: *Physica D: Nonlinear Phenomena* 48.1 (1991), pp. 17–28.
- [24] Maarten C Boerlijst, Marcel E Lamers, and Pauline Hogeweg. “Evolutionary consequences of spiral waves in a host—parasitoid system”. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 253.1336 (1993), pp. 15–18.
- [25] P Bohley and PO Seglen. “Proteases and proteolysis in the lysosome”. In: *Experientia* 48 (1992), pp. 151–157.
- [26] Fred C Boogerd et al. “Afterthoughts as foundations for systems biology”. In: *Systems Biology*. Elsevier, 2007, pp. 321–336.
- [27] Fred C Boogerd et al. “Towards philosophical foundations of Systems Biology: introduction”. In: *Systems Biology*. Elsevier, 2007, pp. 3–19.
- [28] Giovanni Borelli. “De motu animalium, 2 vols”. In: *Rome: A. Bernabò* (), pp. 1680–81.
- [29] Peter J Bowler. “Preformation and pre-existence in the seventeenth century: A brief analysis”. In: *Journal of the History of Biology* 4.2 (1971), pp. 221–244.
- [30] Rainer Breitling. “What is systems biology?” In: *Frontiers in physiology* 1 (2010), p. 1576.
- [31] Anat Bren et al. “Glucose becomes one of the worst carbon sources for *E. coli* on poor nitrogen sources due to suboptimal levels of cAMP”. In: *Scientific reports* 6.1 (2016), p. 24834.
- [32] Ingo Brigandt and Alan Love. “Reductionism in biology”. In: (2008).
- [33] William H Brock. *Justus von Liebig: the chemical gatekeeper*. Cambridge University Press, 2002.
- [34] Rodney A Brooks. “Intelligence without representation”. In: *Artificial intelligence* 47.1-3 (1991), pp. 139–159.
- [35] Rodney A Brooks. “Solving the find-path problem by good representation of free space”. In: *IEEE Transactions on Systems, Man, and Cybernetics* 2 (1983), pp. 190–197.
- [36] Michael S Brown et al. “Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans”. In: *Cell* 100.4 (2000), pp. 391–398.
- [37] Robert Brown. “XXVII. A brief account of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies”. In: *The philosophical magazine* 4.21 (1828), pp. 161–173.
- [38] Otto Brunfels, Hans Weiditz, et al. “Herbarvm vivæ eicones [ie eicones]”. In: ().
- [39] Eduard Buchner. “Alkoholische gährung ohne hefezellen”. In: *Berichte der deutschen chemischen Gesellschaft* 30.1 (1897), pp. 117–124.
- [40] Marieke F Buffing et al. “Capacity for instantaneous catabolism of preferred and non-preferred carbon sources in *Escherichia coli* and *Bacillus subtilis*”. In: *Scientific reports* 8.1 (2018), pp. 1–10.
- [41] George Louis Leclerc Buffon. *Histoire naturelle*. Vol. 36. 1787.
- [42] Georges Louis Leclerc Buffon. “Histoire naturelle, générale et particulière contenant les époques de la nature”. In: (*No Title*) (1778).

-
- [43] Anthony P Burgard, Priti Pharkya, and Costas D Maranas. “Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization”. In: *Biotechnology and bioengineering* 84.6 (2003), pp. 647–657.
- [44] Walter Bradford Cannon. “Homeostasis”. In: *The wisdom of the body*. Norton, Newyork (1932).
- [45] Nuno Carinhas et al. “Hybrid metabolic flux analysis: combining stoichiometric and statistical constraints to model the formation of complex recombinant products”. In: *BMC systems biology* 5.1 (2011), pp. 1–13.
- [46] Ron Caspi et al. “The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases”. In: *Nucleic acids research* 36.suppl_1 (2007), pp. D623–D631.
- [47] Andrea Cesalpino. *De plantis libri XVI*. apud Georgium Marescottum, 1983.
- [48] Christophe Chassagnole et al. “Dynamic modeling of the central carbon metabolism of *Escherichia coli*”. In: *Biotechnology and bioengineering* 79.1 (2002), pp. 53–73.
- [49] Dimitris Christodoulou et al. “Reserve flux capacity in the pentose phosphate pathway enables *Escherichia coli*’s rapid response to oxidative stress”. In: *Cell systems* 6.5 (2018), pp. 569–578.
- [50] Victor Chubukov et al. “Coordination of microbial metabolism”. In: *Nature Reviews Microbiology* 12.5 (2014), pp. 327–340.
- [51] Victor Chubukov et al. “Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis*”. In: *Molecular systems biology* 9.1 (2013), p. 709.
- [52] Cedric R Clapier and Bradley R Cairns. “The biology of chromatin remodeling complexes”. In: *Annual review of biochemistry* 78 (2009), pp. 273–304.
- [53] William Coleman. “Limits of the recapitulation theory: Carl Friedrich Kilmeyer’s critique of the presumed parallelism of earth history, ontogeny, and the present order of organisms”. In: *Isis* 64.3 (1973), pp. 341–350.
- [54] Caroline Colijn et al. “Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production”. In: *PLoS computational biology* 5.8 (2009), e1000489.
- [55] Wikipedia contributors. *Systems biology*. Accessed on August 15th, 2023. 2023. URL: https://en.wikipedia.org/wiki/Systems_biology.
- [56] Rafael S Costa et al. “Hybrid dynamic modeling of *Escherichia coli* central metabolic network combining Michaelis–Menten and approximate kinetic equations”. In: *Biosystems* 100.2 (2010), pp. 150–157.
- [57] Zak Costello and Hector Garcia Martin. “A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data”. In: *NPJ systems biology and applications* 4.1 (2018), pp. 1–14.
- [58] Iain D Couzin et al. “Effective leadership and decision-making in animal groups on the move”. In: *Nature* 433.7025 (2005), pp. 513–516.
- [59] Markus W Covert et al. “Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*”. In: *Bioinformatics* 24.18 (2008), pp. 2044–2050.
- [60] Darren J Creek et al. “Stable isotope-assisted metabolomics for network-wide metabolic pathway elucidation”. In: *Analytical chemistry* 84.20 (2012), pp. 8442–8447.
- [61] Thomas Cremer and Thomas Cremer. “Wachstum naturwissenschaftlicher Erkenntnis: Von der Zelltheorie bis zur Chromosomentheorie der Vererbung”. In: *Von der Zellenlehre zur Chromosomentheorie: Naturwissenschaftliche Erkenntnis und Theorienwechsel in der frühen Zell- und Vererbungsforschung*. Springer. 1985, pp. 27–238.
- [62] Joseph O Dada and Pedro Mendes. “Multi-scale modelling and simulation in systems biology”. In: *Integrative Biology* 3.2 (2011), pp. 86–96.
- [63] Joseph M Dale, Liviu Popescu, and Peter D Karp. “Machine learning methods for metabolic pathway prediction”. In: *BMC bioinformatics* 11.1 (2010), pp. 1–14.

-
- [64] Georges Louis Le Clerc De Buffon. *Histoire naturelle generale et particuliere avec la description du Cabinet du Roy: Servant de suite à la Théorie de la Terre, & de préliminaire à l'histoire des Végétaux.... Supplément, Tome Second*. Imprimerie Royale, 1775.
- [65] Vojo Deretic. *Autophagosome and phagosome*. Springer, 2008.
- [66] René Descartes. “The philosophical works of Descartes.[2 vols.]” In: (1955).
- [67] Katja Dettmer, Pavel A Aronov, and Bruce D Hammock. “Mass spectrometry-based metabolomics”. In: *Mass spectrometry reviews* 26.1 (2007), pp. 51–78.
- [68] Maren Diether et al. “Systematic mapping of protein-metabolite interactions in central metabolism of *Escherichia coli*”. In: *Molecular systems biology* 15.8 (2019), e9008.
- [69] Christopher D Doucette et al. “ α -Ketoglutarate coordinates carbon and nitrogen utilization via enzyme I inhibition”. In: *Nature chemical biology* 7.12 (2011), pp. 894–901.
- [70] Natalie C Duarte, Markus J Herrgård, and Bernhard Ø Palsson. “Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model”. In: *Genome research* 14.7 (2004), pp. 1298–1309.
- [71] Félix Dujardin. “Observations sur les Rhizopodes et les Infusoires”. In: *L’Academic. des Sciences Paris, Comptes Rendus* 1 (1835), pp. 338–340.
- [72] Jean Baptiste A Dumas and Jean Baptiste JD Boussingault. *The chemical and physiological balance of organic nature, an essay by JB Dumas and JB Boussingault*. Saxton & Miles, 1844.
- [73] Jeremy S Edwards and Bernhard O Palsson. “The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities”. In: *Proceedings of the National Academy of Sciences* 97.10 (2000), pp. 5528–5533.
- [74] Florike Egmond. “A collection within a collection: rediscovered animal drawings from the collections of Conrad Gessner and Felix Platter”. In: *Journal of the History of Collections* 25.2 (2013), pp. 149–170.
- [75] Abdul-Hamid Emwas et al. “NMR spectroscopy for metabolomics research”. In: *Metabolites* 9.7 (2019), p. 123.
- [76] Teresa W-M Fan et al. “Stable isotope-resolved metabolomics and applications for drug development”. In: *Pharmacology & therapeutics* 133.3 (2012), pp. 366–391.
- [77] Adam M Feist and Bernhard O Palsson. “The biomass objective function”. In: *Current opinion in microbiology* 13.3 (2010), pp. 344–349.
- [78] Adam M Feist et al. “A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information”. In: *Molecular systems biology* 3.1 (2007), p. 121.
- [79] David A Fell and Herbert M Sauro. “Metabolic control and its analysis: additional relationships between elasticities and control coefficients”. In: *European Journal of Biochemistry* 148.3 (1985), pp. 555–561.
- [80] Friedrich August Flückiger. “Otto Brunfels, Fragment zur Geschichte der Botanik und Pharmacie”. In: *Archiv der Pharmazie* 212.6 (1878), pp. 493–514.
- [81] Marco Fondi and Pietro Liò. “Multi-omics and metabolic modelling pipelines: challenges and tools for systems microbiology”. In: *Microbiological research* 171 (2015), pp. 52–64.
- [82] CRICK Francis. “What Mad Pursuit: A Personal View of Scientific Discovery”. In: *United State of America* (1988).
- [83] Holger Funk. “Adam Zaluzansky’s “De sexu plantarum”(1592): an early pioneering chapter on plant sexuality”. In: *Archives of natural history* 40.2 (2013), pp. 244–256.
- [84] Luca Gerosa et al. “Pseudo-transition analysis identifies the key regulators of dynamic metabolic adaptations from steady-state data”. In: *Cell systems* 1.4 (2015), pp. 270–282.
- [85] Howard Gest. “The discovery of microorganisms by Robert Hooke and Antoni Van Leeuwenhoek, fellows of the Royal Society”. In: *Notes and records of the Royal Society of London* 58.2 (2004), pp. 187–201.
- [86] David S Goodsell. “The machinery of life”. In: (2009).

-
- [87] Boris Görke and Jörg Stülke. “Carbon catabolite repression in bacteria: many ways to make the most out of nutrients”. In: *Nature Reviews Microbiology* 6.8 (2008), pp. 613–624.
- [88] Jonatha M Gott and Ronald B Emeson. “Functions and mechanisms of RNA editing”. In: *Annual review of genetics* 34.1 (2000), pp. 499–531.
- [89] Stephen Jay Gould. *Ontogeny and phylogeny*. Harvard University Press, 1985.
- [90] Volker Grimm et al. “A standard protocol for describing individual-based and agent-based models”. In: *Ecological modelling* 198.1-2 (2006), pp. 115–126.
- [91] Changdai Gu et al. “Current status and applications of genome-scale metabolic models”. In: *Genome biology* 20 (2019), pp. 1–18.
- [92] Jeremy Gunawardena. “Models in biology: ‘accurate descriptions of our pathetic thinking’”. In: *BMC biology* 12.1 (2014), pp. 1–11.
- [93] A Rupert Hall. *The revolution in science 1500-1750*. Routledge, 2014.
- [94] Timo Hardiman et al. “Prediction of kinetic parameters from DNA-binding site sequences for modeling global transcription dynamics in *Escherichia coli*”. In: *Metabolic engineering* 12.3 (2010), pp. 196–211.
- [95] Hermann von Helmholtz. *Über die Erhaltung der Kraft: Eine physikalische Abhandlung*. Berlin: Reimer, 1847.
- [96] Johannes Baptista van Helmont. *Ortus medicinae, id est initia physicae inaudita. Progresus medicinae novus in morborum ultionem ad vitam longam, ed. Francisco Mercurio van Helmont*. Elzevirius.
- [97] Charlotte K Hemelrijk and Hanno Hildenbrandt. “Schools of fish and flocks of birds: their shape and internal structure by self-organization”. In: *Interface focus* 2.6 (2012), pp. 726–737.
- [98] Lawrence Joseph Henderson. “Blood: a study in general physiology”. In: *(No Title)* (1928).
- [99] Ben Hesper and Paulien Hogeweg. “Bioinformatica: een werkconcept”. In: *Kameleon* 1.6 (1970), pp. 28–29.
- [100] W Daniel Hillis. “Co-evolving parasites improve simulated evolution as an optimization procedure”. In: *Physica D: Nonlinear Phenomena* 42.1-3 (1990), pp. 228–234.
- [101] P Hogeweg. “Local TT cell and TB cell interactions: a cellular automaton approach”. In: *Immunology Letters* 22.2 (1989), pp. 113–122.
- [102] Robert Hooke. “Robert Hooke’s Micrographia of 1665 and 1667”. In: *JR Coll Physicians Edinb* 40 (2010), pp. 374–6.
- [103] Cheol-Sang Hwang, Anna Shemorry, and Alexander Varshavsky. “N-terminal acetylation of cellular proteins creates specific degradation signals”. In: *Science* 327.5968 (2010), pp. 973–977.
- [104] Toshimichi Ikemura. “Codon usage and tRNA content in unicellular and multicellular organisms.” In: *Molecular biology and evolution* 2.1 (1985), pp. 13–34.
- [105] Toshifumi Inada, Keiko Kimata, and Hiroji Aiba. “Mechanism responsible for glucose–lactose diauxie in *Escherichia coli*: challenge to the cAMP model”. In: *Genes to Cells* 1.3 (1996), pp. 293–301.
- [106] Jan Ingenhousz. *Experiments Upon Vegetables, Discovering Their Great Power of Purifying the Common Air in the Sun-shine, and of Injuring it in the Shade and at Night: To which is Joined, a New Method of Examining the Accurate Degree of Salubrity of the Atmosphere*. P. Elmsly, and H. Payne, 1779.
- [107] François Jacob and Jacques Monod. “Genetic regulatory mechanisms in the synthesis of proteins”. In: *Journal of molecular biology* 3.3 (1961), pp. 318–356.
- [108] Nusrat Jahan et al. “Development of an accurate kinetic model for the central carbon metabolism of *Escherichia coli*”. In: *Microbial cell factories* 15.1 (2016), pp. 1–19.
- [109] Paul A Jensen and Jason A Papin. “Functional integration of a metabolic network model and expression data without arbitrary thresholding”. In: *Bioinformatics* 27.4 (2011), pp. 541–547.

-
- [110] Tuty Asmawaty Abdul Kadir et al. “Modeling and simulation of the main metabolism in *Escherichia coli* and its several single-gene knockout mutants with experimental verification”. In: *Microbial cell factories* 9.1 (2010), pp. 1–21.
- [111] Immanuel Kant. *Immanuel Kants Kritik der reinen Vernunft*. Mayer & Müller, 1889.
- [112] Immanuel Kant. *Kritik der urteilskraft*. Vol. 39. F. Meiner, 1924.
- [113] Jonathan R Karr et al. “A whole-cell computational model predicts phenotype from genotype”. In: *Cell* 150.2 (2012), pp. 389–401.
- [114] S Kauffman and WS McCulloch. “Random nets of formal genes”. In: *Quarterly Progress Report* 34 (1967).
- [115] Stuart A Kauffman. “Metabolic stability and epigenesis in randomly constructed genetic nets”. In: *Journal of theoretical biology* 22.3 (1969), pp. 437–467.
- [116] Ali Khodayari and Costas D Maranas. “A genome-scale *Escherichia coli* kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains”. In: *Nature communications* 7.1 (2016), pp. 1–12.
- [117] Ali Khodayari et al. “A kinetic model of *Escherichia coli* core metabolism satisfying multiple sets of mutant flux data”. In: *Metabolic engineering* 25 (2014), pp. 50–62.
- [118] Marc W Kirschner. “The meaning of systems biology”. In: *Cell* 121.4 (2005), pp. 503–504.
- [119] Leo Koenigsberger. *Hermann von Helmholtz*. Clarendon press, 1906.
- [120] Maria Kogadeeva and Nicola Zamboni. “SUMOFLUX: a generalized method for targeted ¹³C metabolic flux ratio analysis”. In: *PLoS computational biology* 12.9 (2016), e1005109.
- [121] Oliver Kotte et al. “Phenotypic bistability in *Escherichia coli*’s central carbon metabolism”. In: *Molecular systems biology* 10.7 (2014), p. 736.
- [122] Marilyn Kozak. “Initiation of translation in prokaryotes and eukaryotes”. In: *Gene* 234.2 (1999), pp. 187–208.
- [123] Hans Adolf Krebs and William Arthur Johnson. “Metabolism of ketonic acids in animal tissues”. In: *Biochemical Journal* 31.4 (1937), p. 645.
- [124] VI Krinsky and KI Agladze. “Interaction of rotating waves in an active chemical medium”. In: *Physica D: Nonlinear Phenomena* 8.1-2 (1983), pp. 50–56.
- [125] Philipp F Lange and Christopher M Overall. “Protein TAILS: when termini tell tales of proteolysis and function”. In: *Current Opinion in Chemical Biology* 17.1 (2013), pp. 73–82.
- [126] Timothy Lenoir. “Kant, Blumenbach, and vital materialism in German biology”. In: *Isis* 71.1 (1980), pp. 77–108.
- [127] Timothy Lenoir. “Teleology without regrets. The transformation of physiology in Germany: 1790–1847”. In: *Studies in history and philosophy of science Part A* 12.4 (1981), pp. 293–354.
- [128] Timothy Lenoir. *The strategy of life: Teleology and mechanics in nineteenth-century German biology*. Vol. 13. University of Chicago Press, 1989.
- [129] Orgel Leslie E. “Prebiotic chemistry and the origin of the RNA world”. In: *Critical reviews in biochemistry and molecular biology* 39.2 (2004), pp. 99–123.
- [130] Wen-Hsiung Li, Takashi Gojobori, and Masatoshi Nei. “Pseudogenes as a paradigm of neutral evolution”. In: *Nature* 292.5820 (1981), pp. 237–239.
- [131] Freiherr von Liebig and Justus von Liebig. *Die organische Chemie in ihrer Anwendung auf Agricultur und Physiologie*. Vieweg, 1840.
- [132] Justus Freiherr von Liebig. *Die organische Chemie in ihrer Anwendung auf Physiologie und Pathologie*. F. Vieweg und Sohn, 1842.
- [133] David C Lindberg. *The beginnings of Western science: The European scientific tradition in philosophical, religious, and institutional context, prehistory to AD 1450*. University of Chicago Press, 2010.
- [134] Hans Lineweaver and Dean Burk. “The determination of enzyme dissociation constants”. In: *Journal of the American chemical society* 56.3 (1934), pp. 658–666.

-
- [135] Hannes Link, Karl Kochanowski, and Uwe Sauer. “Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo”. In: *Nature biotechnology* 31.4 (2013), pp. 357–361.
- [136] Carolus Linnaeus. *Systema naturae*. Vol. 1. Stockholm Laurentii Salvii, 1758.
- [137] Howard D Lipshitz and Craig A Smibert. “Mechanisms of RNA localization and translational regulation”. In: *Current opinion in genetics & development* 10.5 (2000), pp. 476–488.
- [138] Jacques Loeb. *The dynamics of living matter*. Columbia University Press, 1906.
- [139] Jacques Loeb. *The Mechanistic Conception of Life*. University of Chicago Press, 1912.
- [140] Ernst Mach. *Die Mechanik in ihrer Entwicklung: historisch-kritisch dargestellt*. Brockhaus, 1912.
- [141] Daniel Machado et al. “Modeling formalisms in systems biology”. In: *AMB express* 1.1 (2011), pp. 1–14.
- [142] Boris Magasanik. “Catabolite repression”. In: *Cold Spring Harbor symposia on quantitative biology*. Vol. 26. Cold Spring Harbor Laboratory Press. 1961, pp. 249–256.
- [143] Stefania Magnúsdóttir et al. “Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota”. In: *Nature biotechnology* 35.1 (2017), pp. 81–89.
- [144] Radhakrishnan Mahadevan, Jeremy S Edwards, and Francis J Doyle. “Dynamic flux balance analysis of diauxic growth in *Escherichia coli*”. In: *Biophysical journal* 83.3 (2002), pp. 1331–1340.
- [145] Agustino Martinez-Antonio and Julio Collado-Vides. “Identifying global regulators in transcriptional regulatory networks in bacteria”. In: *Current opinion in microbiology* 6.5 (2003), pp. 482–489.
- [146] Heinrich Marzell. “Das Buchsbaum-Bild im Kräuterbuch (1551) des Hieronymus Bock”. In: *Sudhoffs Archiv für Geschichte der Medizin und der Naturwissenschaften* H. 2 (1954), pp. 97–103.
- [147] Julius Robert von Mayer. *Die organische Bewegung in ihrem Zusammenhange mit dem Stoffwechsel*. C. Drechsler, 1845.
- [148] Robert Mayer. *Bemerkungen über die Kräfte der unbelebten Natur*. 1839.
- [149] Aonghus J McCarthy, Caroline Coleman-Vaughan, and Justin V McCarthy. “Regulated intramembrane proteolysis: emergent role in cell signalling pathways”. In: *Biochemical Society Transactions* 45.6 (2017), pp. 1185–1202.
- [150] Pierre Millard, Kieran Smallbone, and Pedro Mendes. “Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in *Escherichia coli*”. In: *PLoS computational biology* 13.2 (2017), e1005396.
- [151] Eric L Mills. *The Secular Ark. Studies in the History of Biogeography*. 1986.
- [152] Olga Moreno-Gonzalo, Carolina Villarroya-Beltri, and Francisco Sánchez-Madrid. “Post-translational modifications of exosomal proteins”. In: *Frontiers in immunology* 5 (2014), p. 383.
- [153] Johannes Müller. *Elements of physiology*. Vol. 2. Lea and Blanchard, 1843.
- [154] Johannes Müller. *Handbuch der physiologie des menschen*. Vol. 1. J. Hölscher, 1838.
- [155] Masafumi Muratani and William P Tansey. “How the ubiquitin–proteasome system controls transcription”. In: *Nature reviews Molecular cell biology* 4.3 (2003), pp. 192–201.
- [156] Ali Nahvi et al. “Genetic control by a metabolite binding mRNA”. In: *Chemistry & biology* 9.9 (2002), pp. 1043–1049.
- [157] David L Nelson and Michael M Cox. *Absolute Ultimate Guide for Lehninger Principles of Biochemistry*. 2008.
- [158] Alexander J Nicholson and Victor A Bailey. “The Balance of Animal Populations.—Part I.” In: *Proceedings of the zoological society of London*. Vol. 105. 3. Wiley Online Library. 1935, pp. 551–598.

-
- [159] DW Nicholson. “Caspase structure, proteolytic substrates, and function during apoptotic cell death”. In: *Cell Death & Differentiation* 6.11 (1999), pp. 1028–1042.
- [160] Kazuko Nishikura. “Functions and regulation of RNA editing by ADAR deaminases”. In: *Annual review of biochemistry* 79 (2010), pp. 321–349.
- [161] Elad Noor et al. “The protein cost of metabolic fluxes: prediction from enzymatic rate laws and cost minimization”. In: *PLoS computational biology* 12.11 (2016), e1005167.
- [162] Ruth Nussinov et al. “Allosteric post-translational modification codes”. In: *Trends in biochemical sciences* 37.10 (2012), pp. 447–455.
- [163] Stephen L Nutt et al. “The generation of antibody-secreting plasma cells”. In: *Nature Reviews Immunology* 15.3 (2015), pp. 160–171.
- [164] Jane Marion Oppenheimer. “Essays in the History of Embryology and Biology”. In: (1967).
- [165] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. “What is flux balance analysis?” In: *Nature biotechnology* 28.3 (2010), pp. 245–248.
- [166] Jeffrey D Orth et al. “A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011”. In: *Molecular systems biology* 7.1 (2011), p. 535.
- [167] Ertugrul M Ozbudak et al. “Multistability in the lactose utilization network of Escherichia coli”. In: *Nature* 427.6976 (2004), pp. 737–740.
- [168] Louis Pasteur. “Nouveaux faits concernant l’histoire de la fermentation alcoolique”. In: *Comptes Rendus Chimie* 47 (1858), pp. 1011–1013.
- [169] Philip J Pauly. *Controlling life: Jacques Loeb and the engineering ideal in biology*. Oxford University Press on Demand, 1987.
- [170] Lee T Percy. “Galen and Stoic Rhetoric”. In: *Greek, Roman, and Byzantine Studies* 24.3 (1983), pp. 259–272.
- [171] Pierre Pellegrin. *Aristotle’s classification of animals: biology and the conceptual unity of the Aristotelian corpus*. Univ of California Press, 1986.
- [172] Kirill Peskov, Ekaterina Mogilevskaya, and Oleg Demin. “Kinetic modelling of central carbon metabolism in Escherichia coli”. In: *The FEBS journal* 279.18 (2012), pp. 3374–3385.
- [173] Craig L Peterson and Marc-André Laniel. “Histones and histone modifications”. In: *Current Biology* 14.14 (2004), R546–R551.
- [174] Ilaria Piazza et al. “A map of protein-metabolite interactions reveals principles of chemical communication”. In: *Cell* 172.1-2 (2018), pp. 358–372.
- [175] Johannes Pommeranz. “Das Tierbuch von Conrad Gesner”. In: (2009).
- [176] Kесе Pontes Freitas Alberton et al. “Simultaneous parameters identifiability and estimation of an E. coli metabolic network model”. In: *BioMed research international* 2015 (2015).
- [177] Joseph Priestley. *Experiments and observations on different kinds of air*. Vol. 2. J. Johnson, 1776.
- [178] Jan Evangelista (Purkinje) Purkyne. *Mikroskopisch-neurologische Beobachtungen*. Berlin: Veit & Comp., 1845, pp. 281–295.
- [179] Selwyn Quan et al. “Adaptive evolution of the lactose utilization network in experimentally evolved populations of Escherichia coli”. In: *PLoS genetics* 8.1 (2012), e1002444.
- [180] Steven F Railsback and Volker Grimm. *Agent-based and individual-based modeling: a practical introduction*. Princeton university press, 2019.
- [181] Robert Remak. *Untersuchungen über die Entwicklung der Wirbelthiere*. Vol. 1. G. Reimer, 1855.
- [182] Liam J Revell, Luke J Harmon, and David C Collar. “Phylogenetic signal, evolutionary process, and rate”. In: *Systematic biology* 57.4 (2008), pp. 591–601.
- [183] Craig W Reynolds. “Flocks, herds and schools: A distributed behavioral model”. In: *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*. 1987, pp. 25–34.

-
- [184] Ed Reznik et al. “Genome-scale architecture of small molecule regulatory networks and the fundamental trade-off between regulation and enzymatic activity”. In: *Cell reports* 20.11 (2017), pp. 2666–2677.
- [185] Karin Römisch. “Endoplasmic reticulum-associated degradation”. In: *Annu. Rev. Cell Dev. Biol.* 21 (2005), pp. 435–456.
- [186] William David Ross et al. *Aristotle’s metaphysics*. Vol. 2. Clarendon Press, 1924.
- [187] Ian A Roundtree et al. “Dynamic RNA modifications in gene expression regulation”. In: *Cell* 169.7 (2017), pp. 1187–1200.
- [188] Kepa Ruiz-Mirazo, Carlos Briones, and Andres de la Escosura. “Prebiotic systems chemistry: new perspectives for the origins of life”. In: *Chemical reviews* 114.1 (2014), pp. 285–366.
- [189] Vinay Satish Kumar, Madhukar S Dasika, and Costas D Maranas. “Optimization based automated curation of metabolic reconstructions”. In: *BMC bioinformatics* 8 (2007), pp. 1–16.
- [190] Ida Schomburg, Antje Chang, and Dietmar Schomburg. “BRENDA, enzyme data and metabolic information”. In: *Nucleic acids research* 30.1 (2002), pp. 47–49.
- [191] Laura Schramm and Nouria Hernandez. “Recruitment of RNA polymerase III to its target promoters”. In: *Genes & development* 16.20 (2002), pp. 2593–2620.
- [192] Erwin Schrodinger. *What is life? The physical aspect of the living cell*. At the University Press, 1951.
- [193] Robert Schuetz et al. “Multidimensional optimality of microbial metabolism”. In: *Science* 336.6081 (2012), pp. 601–604.
- [194] Theodor Schwann. “Mikroskopische Untersuchungen über die Uebereinstimmung in der Struktur und dem Wachstum der Thiere und Pflanzen Berlin”. In: *Microscopical researches into the accordance in the structure and growth of animals and plants* (1839), p. 141.
- [195] Irwin H Segel. “Enzyme kinetics: behavior and analysis of rapid equilibrium and steady state enzyme systems”. In: (1975).
- [196] Jean Senebier. *Mémoires physico-chymiques: sur l’influence de la lumière solaire pour modifier les êtres des trois règnes de la nature, et sur tout ceux du règne végétal*. Chez Barde, Manget & Compagnie, 1788.
- [197] Alexander Serganov and Evgeny Nudler. “A decade of riboswitches”. In: *Cell* 152.1 (2013), pp. 17–24.
- [198] Yaakov Setty et al. “Detailed map of a cis-regulatory input function”. In: *Proceedings of the National Academy of Sciences* 100.13 (2003), pp. 7702–7707.
- [199] Daniel C Sévin et al. “Nontargeted in vitro metabolomics for high-throughput identification of novel enzymes in *Escherichia coli*”. In: *Nature methods* 14.2 (2017), pp. 187–194.
- [200] Martin I Sigurdsson et al. “A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1”. In: *BMC systems biology* 4 (2010), pp. 1–13.
- [201] Emrah Şimşek and Minsu Kim. “The emergence of metabolic heterogeneity and diverse growth responses in isogenic bacterial cells”. In: *The ISME journal* 12.5 (2018), pp. 1199–1209.
- [202] Vivek Kumar Singh and Indira Ghosh. “Kinetic modeling of tricarboxylic acid cycle and glyoxylate bypass in *Mycobacterium tuberculosis*, and its application to assessment of drug targets”. In: *Theoretical Biology and Medical Modelling* 3.1 (2006), pp. 1–11.
- [203] Julie Soutourina. “Transcription regulation by the Mediator complex”. In: *Nature reviews Molecular cell biology* 19.4 (2018), pp. 262–274.
- [204] AS Spirin. “On “masked” forms of messenger RNA in early embryogenesis and in other differentiating systems”. In: *Current topics in developmental biology* 1 (1966), pp. 1–38.
- [205] François Spitz and Eileen EM Furlong. “Transcription factors: from enhancer binding to developmental control”. In: *Nature reviews genetics* 13.9 (2012), pp. 613–626.
- [206] Frans Antonie Stafleu. “Linnaeus and the Linnaeans: The spreading of their ideas in systematic botany, 1735-1789”. In: (*No Title*) (1971).

-
- [207] Stefan Stamm et al. “Function of alternative splicing”. In: *Gene* 344 (2005), pp. 1–20.
- [208] Institute for Systems Biology. *What is Systems Biology?* Accessed on August 15th, 2023. 2023. URL: <https://isbscience.org/about/what-is-systems-biology/>.
- [209] Nobuto Takeuchi and Paulien Hogeweg. “Multilevel selection in models of prebiotic evolution II: a direct comparison of compartmentalization and spatial self-organization”. In: *PLoS computational biology* 5.10 (2009), e1000542.
- [210] Bas Teusink et al. “Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry”. In: *European journal of biochemistry* 267.17 (2000), pp. 5313–5329.
- [211] MARCH THEODOSIUS DOBZHANSKY. “Nothing in biology makes sense except in the light of evolution”. In: *The American Biology Teacher*, <https://www.pbs.org/wgbh/evolution/library/10/2/text/html Revised> 14 (1973).
- [212] Ines Thiele and Bernhard Ø Palsson. “A protocol for generating a high-quality genome-scale metabolic reconstruction”. In: *Nature protocols* 5.1 (2010), pp. 93–121.
- [213] Thomas Thomson. *The history of chemistry*. e-artnow, 2021.
- [214] Bin Tian and James L Manley. “Alternative polyadenylation of mRNA precursors”. In: *Nature reviews Molecular cell biology* 18.1 (2017), pp. 18–30.
- [215] Gottfried Reinhold Treviranus. *Biologie oder Philosophie der lebenden Natur für Naturforscher und Aerzte*. Vol. 4. bey JF Röwer, 1814.
- [216] Yoshihiro Usuda et al. “Dynamic modeling of Escherichia coli metabolic and regulatory systems for amino-acid production”. In: *Journal of biotechnology* 147.1 (2010), pp. 17–30.
- [217] MJA Van Hoek and P Hogeweg. “In silico evolved lac operons exhibit bistability for artificial inducers, but not for lactose”. In: *Biophysical journal* 91.8 (2006), pp. 2833–2843.
- [218] Ambro Van Hoof et al. “Exosome-mediated recognition and degradation of mRNAs lacking a termination codon”. In: *Science* 295.5563 (2002), pp. 2262–2264.
- [219] Amit Varma and Bernhard O Palsson. “Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110”. In: *Applied and environmental microbiology* 60.10 (1994), pp. 3724–3731.
- [220] Karl Ernst Von Baer. *Über Entwicklungsgeschichte der Thiere; Beobachtung und Reflexion*. Vol. 1. Bei den gebrüderm Bornträger, 1828.
- [221] Alexander Von Humboldt. *Views of nature*. HG Bohn, 1850.
- [222] Yuliang Wang, James A Eddy, and Nathan D Price. “Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE”. In: *BMC systems biology* 6.1 (2012), pp. 1–16.
- [223] Peter J Wangersky. “Lotka-Volterra population models”. In: *Annual Review of Ecology and Systematics* 9 (1978), pp. 189–218.
- [224] Otto Warburg. “Über den stoffwechsel der karzinomezellen”. In: *Biochem Z* 152 (1924), pp. 309–344.
- [225] Michael Weitzel et al. “13CFLUX2—high-performance software suite for 13C-metabolic flux analysis”. In: *Bioinformatics* 29.1 (2013), pp. 143–145.
- [226] Max Wertheimer. “Gestalt theory.” In: (1938).
- [227] Charles W Whitfield et al. “Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera*”. In: *Science* 314.5799 (2006), pp. 642–645.
- [228] Gavin S Wilkie, Kirsten S Dickson, and Nicola K Gray. “Regulation of mRNA translation by 5′- and 3′-UTR-binding factors”. In: *Trends in biochemical sciences* 28.4 (2003), pp. 182–188.
- [229] David Sloan Wilson. “A theory of group selection.” In: *Proceedings of the national academy of sciences* 72.1 (1975), pp. 143–146.
- [230] David S Wishart. “Quantitative metabolomics using NMR”. In: *TrAC trends in analytical chemistry* 27.3 (2008), pp. 228–237.

-
- [231] Mark EJ Woolhouse et al. “Biological and biomedical implications of the co-evolution of pathogens and their hosts”. In: *Nature genetics* 32.4 (2002), pp. 569–577.
- [232] Conghui You et al. “Coordination of bacterial proteome with metabolism by cyclic AMP signalling”. In: *Nature* 500.7462 (2013), pp. 301–306.
- [233] Katsuyuki Yugi et al. “Hybrid dynamic/static method for large-scale simulation of metabolism”. In: *Theoretical Biology and Medical Modelling* 2.1 (2005), pp. 1–11.
- [234] Mattia Zampieri et al. “Frontiers of high-throughput metabolomics”. In: *Current opinion in chemical biology* 36 (2017), pp. 15–23.

Chapter 2

Prediction of allosteric interactions by single reaction modeling

X. Hernandez-Alias conducted extensive computational analyses, enzyme assays, and metabolomics measurements, and authored the MSc. thesis forming the foundation of this chapter.

M.A.P. Karrenbelt structured the project, supervised the study, reviewed and validated code implementations, and provided comprehensive feedback on the thesis writing.

M. Diether provided guidance and oversight in enzyme assays and metabolite measurements.

Prof. Dr. Manfred Claassen and Prof. Dr. Uwe Sauer offered supervision and consultation during the study.

2.1 Introduction

Metabolism comprises the integrality of chemical reactions that provide cells with the energy and molecular building blocks needed to thrive in a set of highly diverse and dynamic environments. In order to successfully acclimatize to these environments cells need to adapt their metabolism to accommodate growth, reproduction and homeostatic maintenance. Metabolic adaptation consists of the tuning of rates at which metabolites are converted in a chemical reaction process, commonly referred to as metabolic flux. This conversion is often catalyzed by enzymes, and their abundance and activity can be modulated to adjust metabolic fluxes in order to meet anabolic and catabolic demands. To facilitate the balancing of fluxes in accordance with these demands regulatory mechanisms have evolved that help orchestrate the adaptation process. In order to gain a comprehensive understanding of how cells adapt to fluctuating environments knowledge of their regulatory interaction system is a prerequisite. Knowledge of these regulatory interactions can in turn be used to help optimize biotechnological production processes or therapeutic treatment programs that are often plagued by a limited understand hereof.

2.1.1 The regulation of cellular metabolism

One way to regulate enzymatic flux is via a hierarchical cascade comprised of transcriptional, post-transcriptional and translational regulation, which ultimately control enzyme abundance. The regulation of enzyme abundance is a process that occurs over longer time-scales because gene transcription, mRNA translation and protein folding all need to occur before metabolic flux is affected. Another way to regulate fluxes is through direct modulation of enzyme activity by metabolites, either by binding at the active site, or by binding outside of the active site, a process referred to as allosteric regulation. The modulation of enzyme activity by metabolites is direct and therefore far more rapid, constituting a process that occurs on the millisecond to second timescale [57]. Of these two regulatory processes, the regulation of enzyme abundance has been studied more thoroughly. This is because these processes are easier to observe, which can be attributed, among others, to the following:

1. effects have a relatively slow onset and are comparatively long-lived
2. signals at the transcript level can be amplified, e.g. using next generation sequencing technologies
3. amenability to engineering and the availability of molecular tools: overexpression libraries [31], inducible expression systems [19], tuneable degradation systems [56], and libraries of promoters, ribosomal binding sites [47] and GFP-fusion proteins [64].

Based on the time-scales of the two regulatory processes one might think that the primary role of metabolic regulation is to govern the initial response, and that subsequently hierarchical regulation takes over as the driver of long term adaptation. However, recent work suggests a more prominent role for metabolic regulation by concluding that small molecule regulation, rather than hierarchical control, is responsible for the majority of long-term metabolic flux changes [14, 15, 37, 22, 44]. With regard to *E. coli* specifically, Gerosa *et al.* (2015) suggest that 70% of the observed difference in steady state flux distributions between growth on eight different carbon sources cannot be explained by transcriptional and reactant regulation alone [18]. Based on these findings we focus on the identification of allosteric regulation in this study.

2.1.2 The identification of allosteric regulation

Different computational approaches have been established in furtherance of the aim to understand metabolic regulation. These can broadly be classified as top-down approaches, which are those starting with data and seeking to find causal explanations for the observed correlations, and bottom-up approaches, which are those that impose a pre-existing structure to the problem in the form of a model containing known or suspected causalities. To the former of these two categories most of the modern machine learning algorithms belong, and their popularity is reflected by the number of studies performed using them [25, 50, 38, 2, 24]. In these part of the existing knowledge of allosteric interactions is used to train a model that is subsequently used to predict the existence of allosteric interactions in the left out data. One limitation of this approach is the availability

of data, as the largest database of allosteric interactions currently contains 1949 allosteric proteins [26]. Another challenge is a bias in reporting positive results, hence many of these approaches deal with by use randomly selected combinations of metabolite-protein pairs as negatives. It should furthermore be noted that the work cited in this paragraph focused solely on optimization of performance metrics with respect to yet available data; newly predicted interactions on randomly selected protein-metabolite pairs serving as negatives are considered false positives in these studies and no investigation of newly predicted interactions or validation hereof was performed.

To the latter of the two categories belong approaches that use models which reflect previously acquired knowledge of the system, such as knowledge of enzyme kinetics or the metabolic network structure. One such method is allosteric regulation FBA, a constraint-based method that has been used to try to understand the contribution of allostery on flux distributions [43]. This and other constraint-based methods are centered around a steady-state assumption that facilitates the prediction of fluxes by sacrificing the ability to include information on or predict metabolite or enzyme concentrations. An approach that does inherently have this capability is kinetic modeling, which has been applied in several studies on allosteric regulation [54, 12, 40, 13]. These approaches, however, require additional parameters to present each interaction in the network. Although these methods are capable of describing complex intricacies more accurately, they scale poorly because they suffer from the curse of dimensionality since the parameter space grows exponentially with the size of the system.

In pursuit of addressing the challenge of scalability, a recent publication by Hackett *et al.* introduced a single-reaction modeling approach aimed at predicting allosteric interactions using steady-state omics data [22]. The authors ask whether the observed variation in metabolic fluxes across 25 conditions can be sufficiently explained by generalized Michaelis-Menten kinetics, or whether the inclusion of allosteric regulation improves the models' ability to explain the data. Using this approach the authors were able to retrieve several of the known allosteric regulators in *Saccharomyces cerevisiae* metabolism, as well as predict new interactions that they afterwards validated using enzyme essays.

Additionally, Kotte and Heinemann proposed a divide-and-conquer strategy to analyze underdetermined biochemical models [35]. Their approach focuses on breaking down complex models into smaller, solvable submodels, providing insights into the behavior of underdetermined systems. While their method differs from the single-reaction modeling approach of Hackett *et al.*, it demonstrates the diversity of computational approaches used in the pursuit of understanding metabolic regulation.

While we acknowledge the valuable contributions of diverse methodologies in understanding metabolic regulation, we found Hackett's approach particularly compelling due to its suitability for predicting allosteric interactions even in distally related metabolites and enzymes. Building upon the foundation established by Hackett *et al.*, our study takes a step further by uniquely extending this approach to the realm of *Escherichia coli* metabolism. Although the foundational methodology serves as a blueprint, we navigate different terrain by utilizing distinct datasets and tailored version of the computational approach, implementation from scratch, to suit our needs. The developed method integrates proteomic, fluxomic and metabolomics data in order to assess the *in vivo* relevance of allosteric regulators. We more than double coverage of the metabolites by combining targeted and non-targeted data, almost doubling the number of reactions we cover and more than doubling the number of allosteric interactions whose regulatory potential we can assess for each of these reactions. In addition, we do not limited our search to regulatory interactions previously reported in other organisms, but exhaustively search through all possible interactions. Furthermore, we increased the penalty of conditional errors in order to increase sensitivity to detect events of condition-specific regulation. Finally, we include additional information from orthogonal studies in order to increase our confidence in the final set of predictions. Experimental validation using *in vitro* enzyme essays of the top scoring interactions lead us to confirm the existence of nine newly predicted allosteric interactions that are in agreement with our predictions. We also find experimental evidence for two allosteric interactions whose effect is opposite of what we predicted.

2.2 Materials and Methods

2.2.1 *In vitro* experimentation

2.2.1.1 Protein expression and purification

E. coli strains for overexpression of AroG, PfkA, FbaA, FumA, HisG, PurF, TktA, SerB, IlvA, GlmS, PurA and ThrC proteins were obtained from the ASKA library [31]. Shake flasks containing 50mL LB medium were inoculated in a 1:100 ratio with LB precultures and protein overexpression was induced with 0.1mM isopropyl β -D-1-thiogalactopyranoside (IPTG). Cultures were incubated for 16h at 37°C while being shaken at 300RPM. Cells were harvested by centrifugation at 5000RCF and 4°C for 15min and afterwards resuspended in 5.7mL lysis buffer [B-PER lysis buffer, 500mM NaCl, 20mM imidazole, 2mg mL⁻¹ lysozyme (Fluka), 0.2mg mL⁻¹ DNase I (Roche), 1mM MgCl₂ and 4mM phenylmethanesulfonyl fluoride]. Cells were disrupted by vortexing and shaking for 10min at room temperature and cell extracts were separated from cell debris by centrifugation at 21000RCF and 4°C for 30min. His-tagged proteins were purified from cell extracts using His-GraviTrap TALON columns and the elution buffer was replaced by enzyme assay buffer [10mM Tris-HCl buffer of pH7.5 and 1mM MgCl₂] using ZEBRA desalting columns (Thermo Scientific) with a 7kDa cutoff. Final protein concentration was determined by NanoDrop A280 measurements, as well as sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

2.2.1.2 FIA-TOF enzyme assays

Enzymatic activity assays were performed at room temperature in a reaction master mix of pH7.5 containing 1mM of the required substrates in enzyme assay buffer, as well as the corresponding regulatory compounds at 4mM or the closest to that concentration possible given the solubility constraints. Different enzyme concentrations were tested to find one at which the product formation rate of the reaction exhibited linearity. Enzyme assays were started by the addition of 10 μ L of the purified enzyme solution in 190 μ L of reaction master mix. At different time intervals - 0, 30, 60, 120, 240 and 300 s - the reactions were quenched by addition of methanol of -20°C to 80% v/v. Each enzyme was analyzed in triplicate.

The product concentration of these samples was determined through flow injection analysis coupled to time-of-flight (FIA-TOF) on a platform consisting of a Hitachi L-7100 liquid chromatography pump coupled to a Gerstel MPS2 autosampler and an Agilent 6520 QTOF mass spectrometer (Agilent, Santa Clara, CA, USA) operated as previously published [17]. The isocitrate flow was 150 μ L min⁻¹ of mobile phase consisting of isopropanol:water (60:40v/v) buffered with 5mM ammonium fluoride at pH9 for negative ionization mode. For online mass axis correction, 2-propanol, taurocholic acid, and Hexakis(1H,1H,3H-tetrafluoropropoxy)-phosphazine were used. Mass spectra were recorded in profile mode from 50 to 1000m/z with a frequency of 1.4 spectra/sec for double injection using the highest resolving power (4GHz HiRes). Source temperature was set to 325°C with 5L min⁻¹ drying gas and a nebulizer pressure of 30 psi. Fragmenter, skimmer and octopole voltages were set to 175V, 65V and 750V, respectively. All steps of mass spectrometry data processing and ion annotation were performed with Matlab R2017b (The Mathworks, Natick, MA, USA) using functions embedded in Bioinformatics, Statistics, Database and Parallel Computing toolboxes as previously described [17]. Reaction velocities were determined by linear regression analysis.

For the regulators showing a significant effect calibration curves were generated of the product and regulator if interest for a particular reaction in order to be able to differentiate between regulatory effects on enzyme kinetics and FIA-TOF artifacts (Supplementary Figure A.1). These calibration curves were measured in duplicate and enzyme assay data were normalized with respect to this data to correct for confounding effects (matrix effects). The final measurement uncertainty was determined by means of propagation of errors [36], taking into consideration the variance in the data.

$$f = f(x_1, x_2), \quad x_1 \pm \delta x_1, x_2 \pm \delta x_2 \quad (2.1)$$

Since the uncertainty depends on two variables we use the following formula in which we add the relative uncertainties in quadrature.

$$\delta f = \sqrt{\left(\frac{\partial f}{\partial x_1} \delta x_1\right)^2, \left(\frac{\partial f}{\partial x_2} \delta x_2\right)^2} \quad (2.2)$$

where f denotes the normalized data, and x_1 and x_2 are the measurement data of the samples and calibration curves. Which can be rewritten to obtain the relative uncertainty as follows:

$$\frac{\delta f}{f} = \frac{\sqrt{(x_2 \delta x_1)^2 + (x_1 \delta x_2)^2}}{x_1 x_2} = \sqrt{\frac{\delta x_1^2}{x_1} + \frac{\delta x_2^2}{x_2}} \quad (2.3)$$

2.2.1.3 Photometric enzyme assays

For the reaction catalyzed by the enzyme AroG we had no chemical stock available to prepare a calibration curve. Hence, instead of using the mass spectrometer, *in vitro* enzymatic assays were performed in which we monitored the decrease of phosphoenolpyruvate by measuring the absorbance of the solution at 232 nm [46]. A 1 in 4 dilution of both the reaction master mix and the solutions containing the putative regulatory compound as were used in the FIA-TOF assay were used for this assay. As with the FIA-TOF enzyme assays different enzyme concentrations were tested to find a range in which the product formation rate of the reaction exhibited linearity. Absorbance was measured on the TECAN Infinity M200 (Tecan Trading AG, Switzerland) at intervals of 13s and time points in the range of 100 to 300s were used to determine the reaction rate by linear regression analysis. The enzyme assay was conducted in quadruplicate. Similar as done in the TIA-TOF assays, a calibration curve was used to normalize the samples in order to remove a potential bias, and error propagation was performed in order to determine the final uncertainty of the normalized measurement data.

2.2.2 *In silico* experimentation

2.2.2.1 Steady-state multiomics data

The metabolomics data consists of liquid chromatography-mass spectrometry (LC-MS) data, flow injection analysis coupled to time of flight (FIA-TOF) mass spectrometry and physiological data on different *E. coli* strains across a set of 16 steady-state conditions [33]. Eight of these conditions belong to a carbon titration where glucose uptake was controlled, whereas the other eight are part of a nitrogen titration in which glutamate uptake was controlled. Both these titrations were performed using strains that had different inducible uptake systems whose expression was controlled by chemical induction. The inducing agent for the carbon titration was 3-methylbenzyl alcohol (3-MBA), and for the nitrogen titration isopropyl- β -D-1-thiogalactopyranoside (IPTG) was used. We confirmed that the qualitative trends in both LC-MS and FIA-TOF data from this study were consistent, and combined them into a single data set. Since we only use relative changes in our modeling approach, we opted to use absolute quantification data of the LC-MS method over the relative quantification data from FIA-TOF when both were available for a particular metabolic species. The final data set covers 284 different metabolites.

The fluxomics data covers 26 fluxes in central carbon metabolism in the same 16 conditions, which were derived using ^{13}C metabolic flux analysis [33]. This data was used in a constraint-based modeling approach to obtain flux estimates for the entire metabolic network.

The proteomics data was obtained from Hui *et al.* and covers 1043 proteins across 14 conditions [27]. These conditions do not match the aforementioned conditions exactly, and only 10 conditions were used in our analysis. Five of these belong to a carbon titration where lactose uptake was controlled, and the other five are part of a nitrogen titration in which glutamate uptake was controlled. A noticeable difference is that the carbon source titration in this study is performed using lactose instead of glucose, however, the proteome response in these conditions is highly similar [33].

Finally, in the last step of data integration procedure we paired up the conditions from the studies that were closest in growth rate of these studies. We were able to match nine conditions, four for the carbon source titration and five for the nitrogen titration (See Table 2.1). We corrected for small differences in growth rates in the conditions that were paired up by exploiting the observation that metabolic fluxes and protein concentrations tend to be linearly dependent on the growth rate (Supplementary Figure A.4), and used this to adjust the fluxes and protein concentrations in accordance with the growth rate of the metabolomics experiment.

Table 2.1: Multi-omics data used in this study. Nine conditions, four from a carbon titration and five from a nitrogen titration, from three experiments conducted in two different studies were combined. The metabolomics data and fluxomics data are obtained from Kochanowski *et al.* (2021) [33], whereas the proteomics data are obtained from Hui *et al.* (2015) [27]. The growth rates are given in units of h^{-1} . The induction agent used in all of the carbon source titration experiments is 3-MBA, and in all nitrogen titrations IPTG was used, both of which are given in units of μM . Strain NQ381 achieves variable lactose uptake by regulating LacY expression [63]. Strain NQ393 has a glutamate dehydrogenase (*ghdA*) deletion and an IPTG inducible glutamate synthase (*gltBD*) [27, 55]. Strain NQ1243 has a 3-methylbenzyl alcohol (3MBA) titratable *ptsG* expression system to for controlled glucose uptake [5]. Strain NCM3722 is a glutamine prototrophic strain derived from *E. coli* K-12 [60].

		Carbon Titration				Nitrogen Titration				
Metabolomics	strain	NCM3722	NQ1234	NQ1234	NQ1234	NCM3722	NQ393	NQ393	NQ393	NQ393
	growth rate	0.90	0.85	0.70	0.57	0.91	0.85	0.74	0.50	0.45
	induction	0	400	100	0	0	100	50	40	30
Fluxomics	strain	NCM3722	NQ1234	NQ1234	NQ1234	NCM3722	NQ393	NQ393	NQ393	NQ393
	growth rate	0.95	0.91	0.74	0.62	0.92	0.88	0.61	0.51	0.44
	induction	0	400	100	0	0	100	50	40	30
Proteomics	strain	NCM3722	NQ381	NQ381	NQ381	NCM3722	NQ393	NQ393	NQ393	NQ393
	growth rate	1.04	0.87	0.67	0.58	0.97	0.89	0.72	0.60	0.46
	induction	0	500	50	25	0	100	50	40	30

2.2.2.2 Flux balance and flux variability analyses

Constraint-based modeling using the genome-scale metabolic model of *E. coli* iJO1366 [49] was performed to derive a flux distribution for the entire metabolic network. To facilitate this process the Cobra Toolbox v3.0 [23] was used using a Matlab 2017b environment. We constrained the exchange rates in accordance with the uptake and secretion rates reported in the physiological data and the intracellular fluxes from the ^{13}C metabolic flux analysis [33].

Constraints

- Major exchange rates of glucose, acetate and 2-oxoglutarate in accordance with the reported mean exchange rates, allowing for 5% error.
- Minor exchange rates of succinate, malate, citrate, phenylpyruvate, glycine, valine and glutamine in accordance with the reported mean exchange rates, allowing 10% error.
- The growth rate in accordance with the reported mean growth rate
- Flux ratios of phosphofructokinase - G6P dehydrogenase, phosphoglucoisomerase - phosphogluconate dehydratase, malate dehydrogenase - PEP carboxylase and enolase - PEP carboxykinase in accordance with ^{13}C flux analysis data
- Glyoxylate shut and malic enzyme flux set to zero in accordance with ^{13}C flux analysis data
- Glutamate dehydrogenase flux to zero in accordance with the knockout strain used in the glutamate titration
- Bypass reactions Fructose 6-phosphate aldolase and Sedoheptulose 1,7-bisphosphate D-glyceraldehyde-3-phosphate-lyase set to zero to avoid the existence of thermodynamically infeasible cycles

To obtain the final set of fluxes we performed a two-step optimization procedure. First we set ATP-production as the objective and found a flux solution that maximizes this objective within the set of given constraints. Second, we fixed the production rate found in the first step by setting it as a constraint and then minimized the sum of absolute fluxes. Since this just provides us with a single solution while we're interested in how individual fluxes may vary around this optimal solution we also performed a variability analysis to determine the effective flux bounds by minimizing and maximizing flux through individual reactions [21]. The estimated fluxes agree well with those reported in the ^{13}C metabolic flux analysis (Supplementary figure 7). In total we found 532 reactions to carry flux in at least one of the conditions that was considered well-constrained give the following cutoff criterion.

$$\left| \text{median}\left(\frac{FVA_{max} - FVA_{min}}{(FVA_{max} + FVA_{min})/2}\right) \right| > 0.3 \quad (2.4)$$

where the median refers to the median value computed across the set of conditions under consideration. The threshold value of 0.3 was adapted from Hackett *et al.*, who utilized a similar criterion in their study [22].

2.2.2.3 Reaction-by-reaction fitting

The reaction-by-reaction fitting procedure that we use in this work has been written in Python (v3.6.2) and can be found here: https://github.com/hexavier/RbRfit_Algorithm. As input the method takes omics data that quantify the either absolute or relative changes in the states of metabolites, fluxes and proteins. These entities need to be linked by specifying which metabolites are converted in specific reactions, and which enzymes catalyze these reactions. Hereto a genome-scale metabolic network may be provided, such as iJO1366 used in this study, as long as care is taken that the identifiers of reactions, enzymes and metabolites in the data sets match those in the model. For each of the reactions a model in the form of a generalized reversible Michaelis-Menten rate law is constructed [39].

$$j^P = \left(\sum_h E_h k_{kcat}^h \right) \frac{\frac{1}{\prod_i K_d^{\alpha_i}} (\prod_i A_i^{\alpha_i} - \frac{1}{K_{eq}} \prod_j B_j^{\beta_j})}{1 + \sum_k \sum_{m=1}^{\gamma_{R_k}} \left(\frac{R_k}{K_d^{R_k}} \right)^m} \quad (2.5)$$

where j^P denotes the metabolic flux, which is expressed as a function of metabolite concentrations, for which A_i , B_j and R_k denote the substrates, products and all reactants taken together and α_i , β_j and γ_i denote their stoichiometric coefficients, enzyme concentrations, for which E_h denotes the isoenzyme concentration, and a set of parameters, where K_d , K_{eq} , K_{cat} denote the dissociation, equilibrium and catalytic constant, respectively.

The available data for each of the reactions is then retrieved from the different data sets and an overview is generated that shows which information is available for each individual reaction. We obtained sufficient coverage for 84 reactions in the metabolic network (Supplementary Table A.1). All of the metabolites for which data is available are considered to be putative allosteric regulators of a reaction, and for all combinations of enzymes and putative regulators a model is constructed by augmenting the rate law with additional terms representing this regulation. To include allosteric interactions to the model the follow terms are added to the rate law expression: $\frac{A}{A+K_a}$ for activation and $\frac{1}{1+\frac{I}{K_i}}$ for inhibition, where A and I represent the concentrations of the activator and inhibitor, respectively, and K_a and K_i their respective affinity constants.

Once the mathematical model of the reaction has been formulated we try to find the parameter set that best fits the model to the flux data. Since we work with relative quantification data, meaning that no physical units are associated, model parameter lose their physical interpretability. As a consequence we cannot retrieve their values from databases or primary literature and instead they need to be inferred.

Apart from testing individual regulatory interactions, we also explore pairwise combinations of interactions to understand potential synergistic or antagonistic effects. Due to the combinatorial complexity of this analysis, we decided to limit the search space by focusing on 100 metabolites for pairwise interactions. This pragmatic approach effectively reduces the search space by two orders of magnitude. To select these 100 metabolites, we devised a simple heuristic based on two key considerations. First, we observed that metabolites that exhibit greater variability across conditions are more likely to play a role in flux regulation [41]. Second, there is a tendency for regulatory interactions observed in one organism to exist in others [53]. To quantify the selection criteria, we introduced a scoring metric:

$$Score_i = Var_i [5 \times \text{number of database entries} + 1] \quad (2.6)$$

where $Score_i$ represents the calculated score for metabolite i , Var_i represents the variability of metabolite i across condition, and the term $(5 \times \text{number of database entries} + 1)$ accounts for the importance of metabolites with existing regulatory interactions in the database. Using this scoring metric, we selected the top 100 metabolites with the highest scores for further analysis. For each of the 84 reactions, we then applied three types of models: one without any regulation, 568 models considering single interaction regulation, and 4950 models exploring pairwise interaction regulations.

2.2.2.4 Parameter estimation

If we assume that deviations between predicted and measured fluxes can be modeled as independent and identically distributed continuous random variables, we can apply Bayes' theorem to define the posterior probability distribution for the parameters:

$$Pr(\Theta|M, E, j^M) = \frac{Pr(M, E, j^M|\Theta)Pr(\Theta)}{Pr(M, E, j^M)} \quad (2.7)$$

where M , E , and j^M denote the data on the metabolites, enzymes, and fluxes, respectively, and Θ denotes the parameters. $Pr(\Theta|M, E, j^M)$ denotes the posterior distribution, which we compute using the likelihood, denoted by $Pr(M, E, j^M|\Theta)$, and the prior over the parameters, denoted as $Pr(\Theta)$, and the marginal distribution of the data, denoted as $Pr(M, E, j^M)$. Since the marginal distribution over the data is just a normalizing constant that does not depend on the parameters, we can ignore it in our parameter estimation procedure and rewrite Bayes' theorem as follows:

$$Pr(\Theta|M, E, j^M) \propto Pr(M, E, j^M|\Theta)Pr(\Theta) \quad (2.8)$$

which states that the posterior probability is proportional to the product of the likelihood and the prior probability. In order to sample the parameter space for estimation of the posterior distribution, we used a Markov Chain Monte Carlo (MCMC) algorithm for sampling that consists of the following steps:

1. Construction of an uninformative prior over the parameters. For K_d , K_a and K_i a log-uniform distribution around the median values of their respective metabolite concentrations was taken. A log-uniform distribution around the median of the reaction quotient Q_r was chosen for K_{eq} .
2. A parameter set is drawn from the prior distribution
3. In case of missing data for one of the states in the system in a given condition, this was sampled from a normal distribution around the median value based on the data available for the other conditions.
4. We estimate the catalytic constant k_{cat} by linear regression using non-negative least squares.
5. We compute the flux estimates using the generalized reversible Michaelis-Menten rate law given in (2.5)
6. We compute the likelihood of Θ . If we use the point estimates from our constraint-based modeling as the measured fluxes the likelihood is computed as follows

$$\mathcal{L}(\Theta|M, E, j^M) = \sum_{c=1}^9 \log[\phi(x = j_c^M; \mu = j_c^P \sigma^2 = \frac{\sum_{c=1}^9 (j_c^M - j_c^P)^2}{9})] \quad (2.9)$$

where ϕ denotes a Gaussian probability density function. Given that we account for uncertainty in our experimentally determined fluxes by means of using the variability estimate the equation above must be rewritten to include the uniform density between the lower and upper flux bound

$$\begin{aligned} \mathcal{L}(\Theta|M, E, j^M) = & \sum_{c=1}^9 \log\left[\frac{1}{j_c^{M+} - j_c^{M-}} \right. \\ & \left. [\phi(x = j_c^{M+}; \mu = j_c^P \sigma^2 = \frac{\sum_{c=1}^9 (\frac{j_c^{M+} + j_c^{M-}}{2} - j_c^P)^2}{9}) \right. \\ & \left. - \phi(x = j_c^{M-}; \mu = j_c^P \sigma^2 = \frac{\sum_{c=1}^9 (\frac{j_c^{M+} + j_c^{M-}}{2} - j_c^P)^2}{9})] \right] \end{aligned} \quad (2.10)$$

where J_c^{M-} and J_c^{M+} denote the lower and upper flux bound.

-
7. Determine the posterior probability of Θ , which is proportional to the likelihood. The non-uniform prior distributions, introduced in the first step, play a significant role in shaping the parameter space exploration during the MCMC procedure. Here, the likelihood computed in the previous step combines with priors for each parameter. This yields the unnormalized posterior distribution, guiding parameter acceptance or rejection. Non-uniform priors fundamentally affect acceptance probability, biasing parameter space exploration to empirical distributions.
 8. If the likelihood improves we accept the parameter set. Otherwise we accept the parameter set with a probability of

$$\frac{Pr(M, E, j^M | \Theta^{proposed})}{Pr(M, E, J^M | \Theta^{current}) + Pr(M, E, j^M | \Theta^{proposed})} \quad (2.11)$$

where $\Theta^{current}$ and newly $\Theta^{proposed}$ denote the current and proposed set of kinetic parameters, respectively.

9. We iterate over steps 2 to 8 recording the parameter set every 20 iterations to reduce artificial correlations that result from the MCMC sampling procedure until a total of 200 samples have been collected. The algorithm was executed independently 10 times from diverse initial conditions. By comparing the outcomes of the 10 distinct Markov chains obtained from each model using the multivariate potential scale reduction factor (MPSRF) [8], we confirmed the convergence of all instances to equivalent posterior distributions.

After the MCMC sampling the maximum a posteriori (*MAP*) estimate was computed

$$\hat{\Theta} = \underset{\Theta}{argmax}(Pr(\Theta | M, E, j^M)) \quad (2.12)$$

$$MAP = Pr(\hat{\Theta} | M, E, j^M) \quad (2.13)$$

To assess how well a proposed model fits the experimental data we need to assess how much of its discrepancy can be attributed to uncertainty in the data. In order to do so, we use a multivariate delta method to estimate the collective error of all species in the reaction [42]

$$Var(j_c^P) = \left(\frac{\delta v_c}{\delta s_c}\right)^T \Sigma \left(\frac{\delta v_c}{\delta s_c}\right) \quad (2.14)$$

where s is the set of all species involved in the reaction, j the pathway-level flux, v the reaction-level fluxes and Σ denotes the covariance matrix of the species. Here, $Var()$ represents the variability or uncertainty in the model fit. Pathway-level flux encapsulates the cumulative flow of metabolites through a specific metabolic pathway; it represents the overall rate of metabolite transformation within the pathway, capturing the collective behavior of all contributing reactions. In the context of assessing model fit, this equation quantifies the variability in pathway-level flux (j) by considering the relationships between reaction-level fluxes (v) and the concentrations of the involved metabolites (s). The covariance matrix (Σ) encapsulates the interdependencies among species in the reaction.

2.2.2.5 Confidence assessment

In order to assess the confidence of the regulatory interactions modeled we compare the likelihood of a model that describes the reaction kinetics without regulation with those of models that include additional regulatory interactions. Next we wish to assess the significance of these regulatory interactions, for which we use three different metrics: the likelihood ratio test statistic, the elasticity and the Akaike Information Criterion (AIC). In order to increase sensitivity of our method in order to be able to pick up regulatory interactions that are functionally relevant in only a subset of the conditions we computed a weighting factor that we used to generate our summary statistics.

$$W = \frac{\ln(x)}{\sum \ln(x)} \quad (2.15)$$

where x denotes the p-value, elasticity coefficient or AIC of the respective analyses. The inclusion of this weighting factor aligns with our aim to capture context-specific regulatory effects and

ensures that our analysis is capable of identifying condition-specific regulatory interactions within the metabolic network. This weighting factor is applied to combine the multiple metrics into a comprehensive summary score that accounts for the relative contributions of each metric, allowing us to effectively prioritize regulatory interactions based on their functional relevance across different conditions.

The likelihood ratio test is used for significance testing. It informs us whether, at a confidence level of our choice, we can accept the hypothesis of a simpler model without regulation over the alternative hypothesis of a more complex model that includes regulation.

$$\Lambda = -2 \ln \frac{L(\Theta_0)}{L(\hat{\Theta})} \quad (2.16)$$

where L denotes the maximized likelihood and Θ_0 and $\hat{\Theta}$ denote the parameter sets of the null hypothesis and alternative hypothesis, respectively. It is important to note that this test can only be used to differentiate between hypothesis of nested models, which is to say models in which the parameters of the null hypothesis are a subset of those in alternative hypothesis. This is because the test-statistic is assumed to follow a χ^2 distribution with k degrees of freedom, where k denotes the number of parameters in our model. This assumption holds for nested models because the parameters that are absent in the smaller model are by their exclusion implicitly assumed to be zero. However, when both models contain parameters absent in the other model this assumption is violated. In our case this test can thus be used to compare a model without regulation, the null hypothesis, with models that contain additional regulation as the alternative hypothesis, but not in the comparison of models containing different regulatory interactions. In the latter case we will need to resort to the use of the AIC.

Elasticity coefficients are a measure of the sensitivity of reaction rates to changes in reactant concentrations [29]. They quantify how small variations in metabolite concentrations influence reaction fluxes, providing insights into local sensitivities. Mathematically, the elasticity coefficient for a reactant in reaction flux is defined as follows:

$$\epsilon_S^v = \frac{\delta \ln v}{\delta \ln S} = \frac{\delta v}{\delta S} \frac{S}{v} \quad (2.17)$$

where ϵ_S^v denotes the elasticity coefficient, v the reaction level flux and S the reactant concentration.

Elasticity values are always in the range of 0 to 1, where 0 indicates that no potential for flux regulation, and 1 indicates strong sensitivity of the local reaction flux. It is essential to note that elasticity coefficients capture local sensitivities, making them suitable for understanding immediate impacts of changes in reactant concentrations. However, in situations where the operating point is near saturation conditions or when global regulatory effects are of interest, elasticity coefficients might not provide a comprehensive depiction. Therefore, while they offer valuable insights into local sensitivities and potential control points, additional methodologies may be necessary to account for non-linear dynamics and global regulatory behavior.

The AIC is an estimator of prediction error that can be used to decide which model fits the data best [1, 11]. More specifically, it estimates the Kullback-Leibler information loss between a model generating the data and a fitted candidate model. Two important requirements that need to be met in order to be able to use AICs to perform model selection are that the competing models were constructed on the same data and that model likelihoods have been maximized; AICs computed for models that have not been fitted should not be compared. In contrast to the likelihood ratio test, it does not make any assumptions on the distribution of differences between AICs when comparing models. It therefore offers a way to compare and perform model selection when these models are not nested [9, 10].

$$AIC = 2k - 2 \ln(\hat{L}) \quad (2.18)$$

where k denotes the number of parameters in the model and \hat{L} the maximum likelihood estimate of the model. The AIC reflects the relative loss of information when assuming a particular model represents the process that generated the information, hence a lower score indicates less information loss and a higher quality model. We use an extension the AIC that corrects for small sample sizes (AICc)

$$AICc = AIC + \frac{2k^2 + 2k}{n - k - 1} \quad (2.19)$$

We therefore only consider models containing allosteric interactions that have a lower AIC than their counter parts without allosteric regulation in our downstream analysis.

2.2.2.6 Prediction prioritization

The previous work by Hackett *et al.* (2016) has noted that the major potential difficulty of their approach is the fact that false positives inevitably arise as a result of correlations between metabolites [22]. We note that the absence of data, such as the inability to use any of the available information on parameters due to the absence of physical units, further exacerbates this. Unlike Hackett *et al.* (2016), who limit themselves to testing regulatory interactions found in other organisms previously, we don't bias our search and exhaustively test all combinations of 84 reactions and 284 metabolites for which data is available. Since we don't limit our search to interactions previously reported in other organisms we expect to suffer from a false positives rate that is at least as large as theirs. In order to address this issue we invoke a prediction prioritization scheme that boost predictions for which additional evidence exists, consisting of the following:

- Prior information from literature reporting the existence of an allosteric interaction in another species. Similar to Hackett *et al.* (2016) we expect some level of evolutionary conservation, or even convergent evolution for that matter, of allosteric regulation. The information on previously reported allosteric regulation was taken from a study in which the authors constructed the small-molecule regulatory network across species [53].
- A physical interaction study reporting protein-metabolite interactions in central carbon metabolism of *E. coli* that were detected using nuclear magnetic resonance (NMR) spectroscopy [16].
- A physical interaction study reporting protein-metabolite interactions in central carbon metabolism of *E. coli* that were detected using a limited proteolysis-coupled mass spectrometry (LiP-MS) approach [51].
- Regulatory elasticity coefficients indicative of the potential for metabolites to exert flux control. If the absolute elasticity coefficient is over 0.5 ($|\epsilon| > 0.5$) we categorize it as positive, otherwise negative. These were computed from the data used in the fitting procedure of this study.

The evidence of each of these sources was compared to the predictions made using our reaction-by-reaction fitting approach by generating receiver operating characteristic (ROC) curves. The discrimination threshold was defined as follows

$$\Delta AIC = AIC(L(\Theta_0)) - AIC(L(\hat{\Theta})) \quad (2.20)$$

where $L(\Theta_0)$ and $L(\hat{\Theta})$ denote the maximum likelihood estimates of the null hypothesis - the model without regulation - and the alternative hypothesis - the model with regulation. With respect to each of these, if we predict:

- An interaction and it was reported to exist or be detected we consider it a true positive, otherwise a false positive.
- Absence of interaction and it was not reported or detected we consider it a true negative, otherwise a false negative.

For these ROC curves the area under the curve (AUC) is a measure of information content of each of these sources with respect to the predictions resulting from fitting individual reactions. Each of the AUC was normalized as follows:

$$W = \frac{AUC}{\sum AUC} \quad (2.21)$$

and used as a weighting factor to prioritize predictions of the fitting procedure.

$$Score = \sum_i W_i AICc_i \quad (2.22)$$

where $AICc_i$ refers to the corrected Akaike Information Criterion (AICc) value for the individual model or reaction denoted by the index i .

2.3 Results

2.3.1 Integration of heterogeneous multiomics data

In order to assess the potential of metabolites as regulators of enzymatically catalyzed reactions we leveraged data on nutrient-limited steady-state growth conditions from two studies: metabolomics and fluxomics data from [32] with the proteomics data from [27], as described in materials and methods section 2.2.2.1. The resulting data set comprises four carbon-source limited growth conditions, and five nitrogen-source limited growth conditions. For each of these, the data set contains relative changes of 284 metabolites, relative changes of 345 metabolic enzymes, and 26 absolute fluxes from central carbon metabolism, as well as uptake and secretion rates of 10 metabolites and cellular growth rates.

Since information on metabolites, enzymes and fluxes are needed in our fitting procedure we estimated the complete cellular metabolic flux distribution using a genome-scale metabolic model (iJO1366 [49]) and a flux balance analysis approach [48]. Using the available data on the 26 fluxes and the uptake, secretion and growth rates were used to constrain the solution space we perform a two-step procedure to estimate genome-scale fluxes. Under the assumption that cells maximize energy production in nutrient-limited growth conditions, we first set ATP production as the objective. In the second step, based on the assumption that cells minimize enzyme usage, the ATP production rate is fixed to the value found in the previous step and minimization of the sum of absolute fluxes is performed to approximate this resource-to-enzyme allocation (see section 2.2.2.1 for details). The result hereof indicated non-zero cellular fluxes for a total of 532 reactions in at least one of these conditions, which from here on out will be referred to as "measured" or "observed" fluxes in order to contrast them with fluxes predicted later on using our fitting procedure, which correlated well with the ^{13}C flux analysis data (Supplementary Figure A.5). Flux variability analysis [21] was performed in order to assess the uncertainty of these flux estimates.

Given that our fitting procedure requires data on metabolites, enzymes and fluxes, we end up with sufficient coverage for 84 reactions (Supplementary Table A.1). Although central carbon metabolism has the highest coverage of any metabolic subsystem, most of the reactions are part of different subsystems and overall coverage of the metabolic network is sparsely distributed.

2.3.2 Analysis of flux control

First we investigated how much of the flux changes across conditions can be explained using a generalized reversible Michaelis-Menten kinetics rate law [39] without any allosteric regulation. We find that for 62% of the models without regulation we obtain a Pearson correlation coefficient of $R^2 \geq 0.35$ when we compared measured and predicted metabolic fluxes (Supplementary Figure A.2). This indicates that a model without allosteric regulation can largely explain the observed fluxes for these reactions. For the remaining reactions, however, we are unable to obtain a reasonable fit to the data, suggesting that we might be missing allosteric regulation.

The best fit was obtained for the transketolase reaction TKT2, which constitutes a reversible reaction that connects glycolysis and the pentose phosphate pathway (Figure 2.1B). We observe flux reversal in the three conditions with the most severe glutamate limitation and the non-regulated model is capable of capturing the reversible nature of this reaction. On other end of the spectrum we find models for which we do not obtain a proper fit, such as aspartate carbamoyltransferase (ASPCT). ASPCT catalyzes the transfer of carbamoyl from carbamoyl-phosphate (CbP) to aspartate (Asp), which is the first reaction in the *de novo* synthesis of pyrimidine nucleotides. Here the search through models augmented with allosteric regulation resulted in a prediction of activation by succinate, which turns out to be a well-known activator of ASPCT in conditions of excess carbamoyl phosphate and limiting aspartate [28].

In the process of evaluating metabolite regulatory potential during the reaction fitting procedure, it became evident that relying solely on these data might introduce false positives due to correlations between these metabolites. To address this concern and reduce the likelihood of false positives in our top-ranking predictions, we strategically gave precedence to predictions backed by additional evidence. This encompassed reports confirming interactions in other organisms [53], detection in physical interaction studies employing LiP-MS [51] or NMR [16], and the utilization of elasticity coefficients computed from available data (Figure 2.2). Our rationale was clear: an interaction reported in another organism was deemed more likely to exist in *E. coli*, potentially via

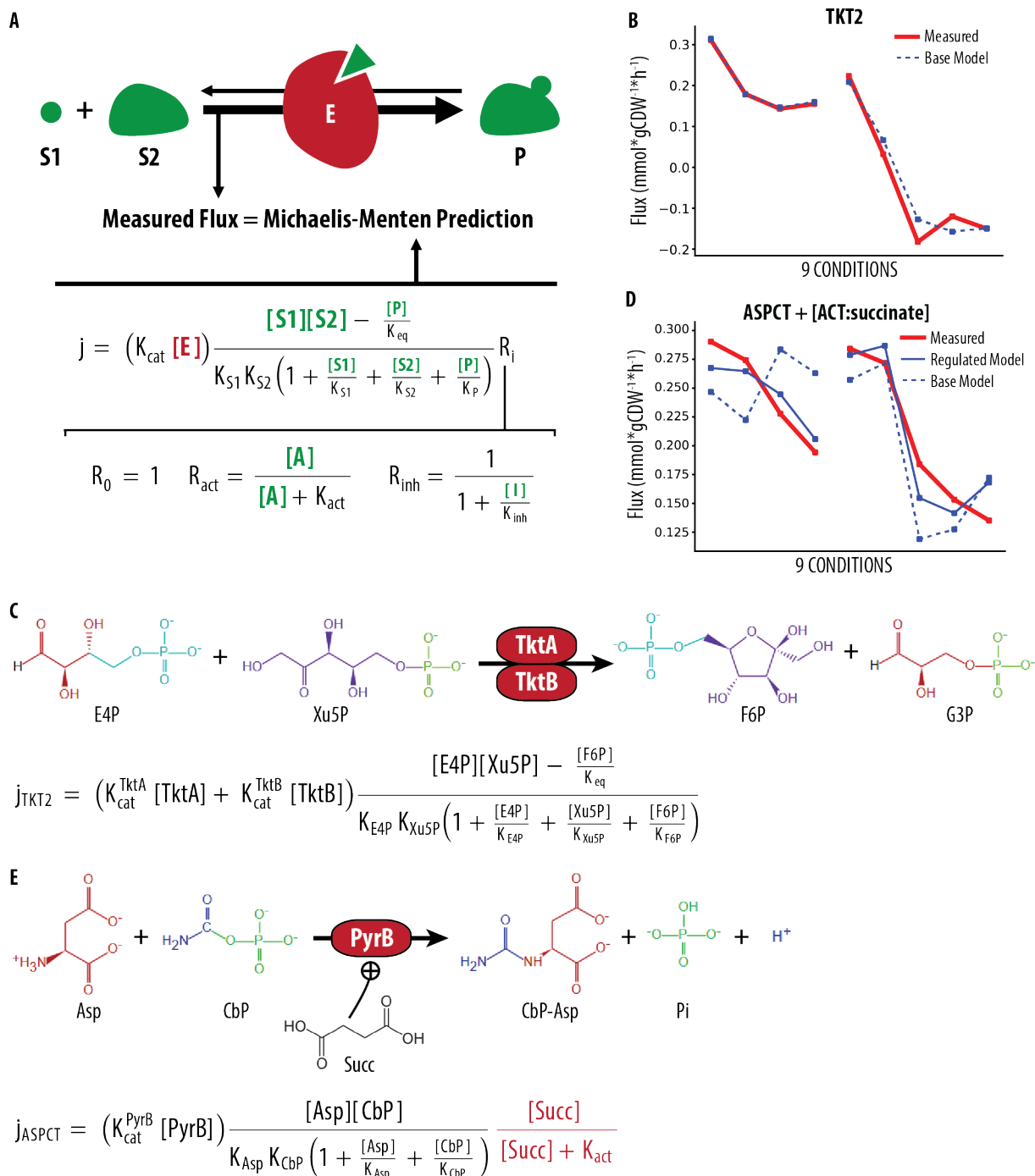


Figure 2.1: Reaction fitting analysis. (A) Schematic representation the approach in which we compare the measured flux and the predicted flux based on a generalized reversible Michaelis-Menten rate law. The enzyme depicted in red and the metabolites, substrates ($S1$ and $S2$), products (P) and putative allosteric regulator (A or I) depicted in green. (B) For transketolase (TKT2) the measured (red) and base model predicted fluxes (blue dashed) across nine conditions match well. (C) There exist two isoforms (TktA and TktB) that catalyze the transketolase reaction, which converts E4P and Xu5P to F6P and G3P. (D) For aspartate carbomoyltransferase (ASPCT) the measured (red) and base model predicted fluxes (blue dashed) across nine conditions do not align. A model including allosteric activation by succinate yielded a much better fit to the data (blue solid). (E) A holoenzyme composed of two catalytic trimers (PyrB) and three regulatory dimers (PyrI) catalyzes the aspartate carbomoyltransferase reaction that converts Asp and CbP to CbP-Asp and phosphate. Abbreviations: E4P (erythrose-4-phosphate), Xu5P (xylulose-t-phosphate), F6P (fructose-6-phosphate), G3P (glyceraldehyde-3-phosphate), Asp (aspartate), CbP (carbamoyl-phosphate), CbP-ASP (N-carbamoyl-aspartate), Pi (phosphate), Succ (Succinate).

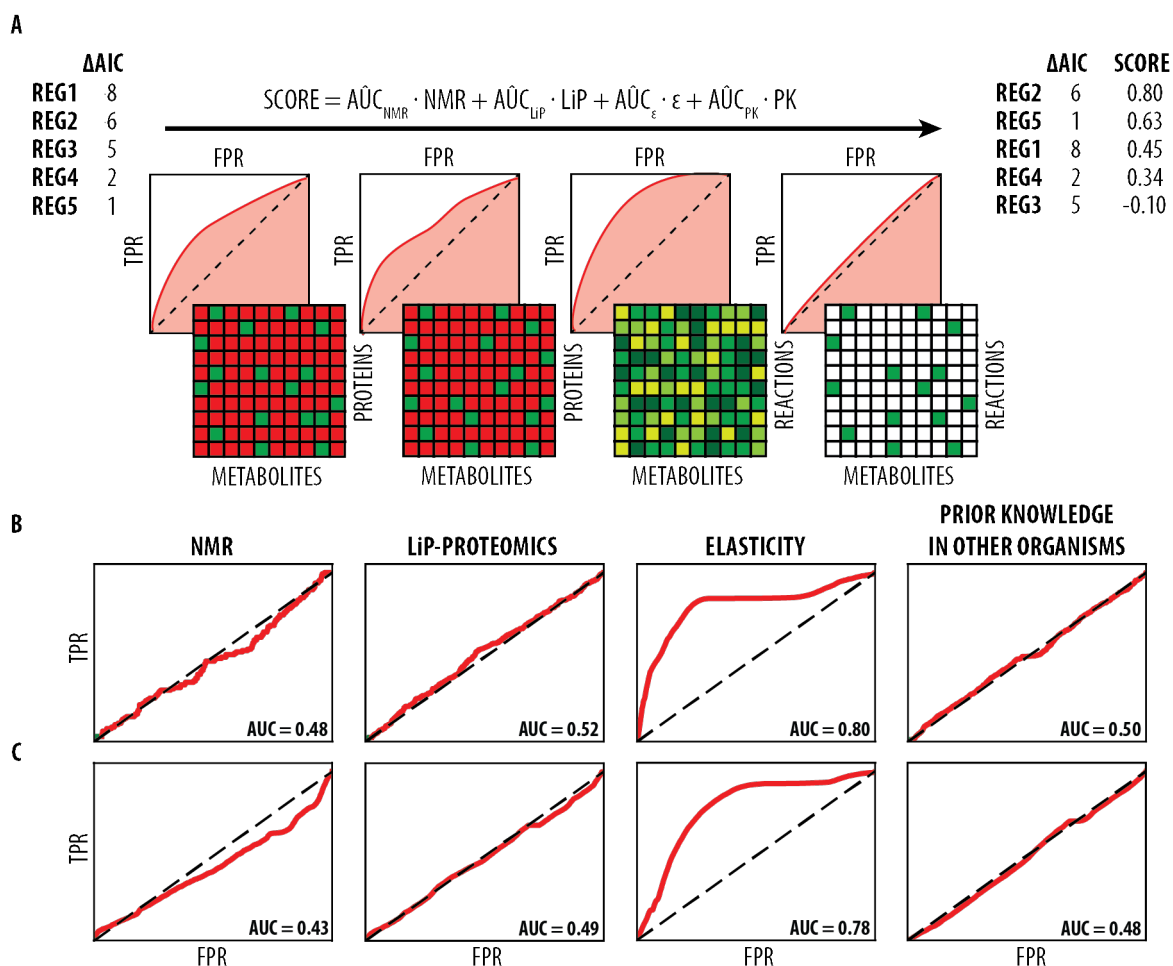


Figure 2.2: Confidence boosting of predictions obtained after reaction fitting. (A) The normalized area under the curve (AUC) of receiver operating characteristic (ROC) curves indicates the overlap between the interactions predicted by the fitting procedure with those of other sources of information. Interactions in the NMR and LiP-MS studies, as well as those obtained from literature, are binary: either detected or not. In the case of elasticity coefficients we used a cut off of $|\epsilon| > 0.5$. (B) AUCs of the reaction fitting procedure of single regulators and (C) AUCs of the reaction fitting procedure of pairwise combinations of regulators.

evolutionary conservation or convergent evolution. Furthermore, detection in physical interaction studies significantly increased the likelihood of identifying potential allosteric regulators. However, it's important to note that higher elasticity coefficients indicating a propensity for metabolites to regulate flux constituted a necessary but not sufficient condition for flux control.

To leverage this multifaceted information, we employed receiver operating characteristic (ROC) curves. These curves, while not used for direct validation, served as a means to enhance confidence in specific predictions by employing them as weighting factors. They illustrated the alignment between our fitting procedure's predictions and the various other information sources mentioned earlier. Among these sources, only the elasticity coefficients demonstrated an area under the curve (AUC) significantly greater than 0.5. This was anticipated as the same data used in the fitting procedure, coupled with the estimated parameter set, computed these coefficients. However, it's noteworthy that apart from elasticity coefficients, other sources presented AUC values close to 0.5, indicating limited discriminatory power in binary classification.

Despite the higher influence of elasticity coefficients, all information sources were utilized in generating final predictions. Unlike the approach by Hackett *et al.* [22], our predictions were not confined to allosteric interactions previously reported in other organisms. This approach ensured an unbiased search and fitting procedure, allowing for the prediction of entirely novel interactors and leveraging existing evidence for prioritization.

2.3.3 Systematic analysis of *in silico* screening results

We have sufficient data on 84 reactions, for each of which we fit one model without regulation, 568 single interaction models and 4,950 pairwise interactions models. After having obtained maximum likelihood estimates from our fitting procedure we compute the Akaike information criterion (AIC) of each model in order to be able to compare models with different allosteric regulation. As a result we find that for 57 out of 84 reactions there is at least one allosteric interactions that improves upon the model without regulation. In total 687 out of 23,856 single interaction models (Supplementary Figure A.6) and 3,611 out of 415,800 pairwise interaction models improve on their respective base model without regulation (Figure 2.3). For these interactions we found additional evidence for 105 out of 687 single interaction models and 1,006 out of 3,611 pairwise interaction models in literature or the aforementioned physical interaction studies.

2.3.4 Validation by *in vitro* enzyme assays

In order to validate our predictions we selected newly predicted allosteric interactions for experimental follow-up in an *in vitro* enzyme assay study. We selected the 12 highest ranking interactions subject to the following constraints: (1) enzymatic catalysis was not performed by a heterocomplex or (2) by membrane-bound enzymes and (3) the substrate and (4) predicted regulatory metabolite were available. We successfully managed to purify the most abundant isoenzyme known to catalyze these reactions for nine of these. Next we assessed enzymatic activity of these nine enzymes and found five that were active (Supplementary Table A.2). For these five enzymes we tested 22 newly predicted allosteric regulators (Table 2.2).

The first reaction is 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DDPA). DDPA catalyzes the conversion of phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P) to 3-

Table 2.2: Predicted regulatory interactions. Five reactions, their respective isolated isoenzyme and the predicted regulatory effectors tested in enzyme assay validation experiments. In green those for which we find support in *in vitro* enzyme assays, in orange those for which we find experimental evidence for the opposite mode of action as predicted. The asterisk indicates that a calibration curve was used to correct for matrix effects. 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DDPA), phosphofructokinase (PFK), fructose-bisphosphate aldolase (FBA), phosphoserine phosphatase (PEP_L) and adenylosuccinate synthetase (ADSS).

DDPA (AroG)		PFK (PfkA)		FBA (FbaA)		PSP_L (SerB)		ADSS (PurA)	
metabolite	prediction	metabolite	prediction	metabolite	prediction	metabolite	prediction	metabolite	prediction
indpyr	inhibition	hxan	inhibition	3pg	activation*	hxan	inhibition*	amp	inhibition
pep	inhibition	indpyr	inhibition*	ade	inhibition	gtp	inhibition	acglu	inhibition
bgly	inhibition	glu	inhibition	hxan	inhibition	gly	inhibition	pnto	inhibition
34hpp	inhibition	gly	activation	atp	inhibition*	ino	inhibition		
4pyrdx	inhibition			imp	activation*	pser	inhibition		

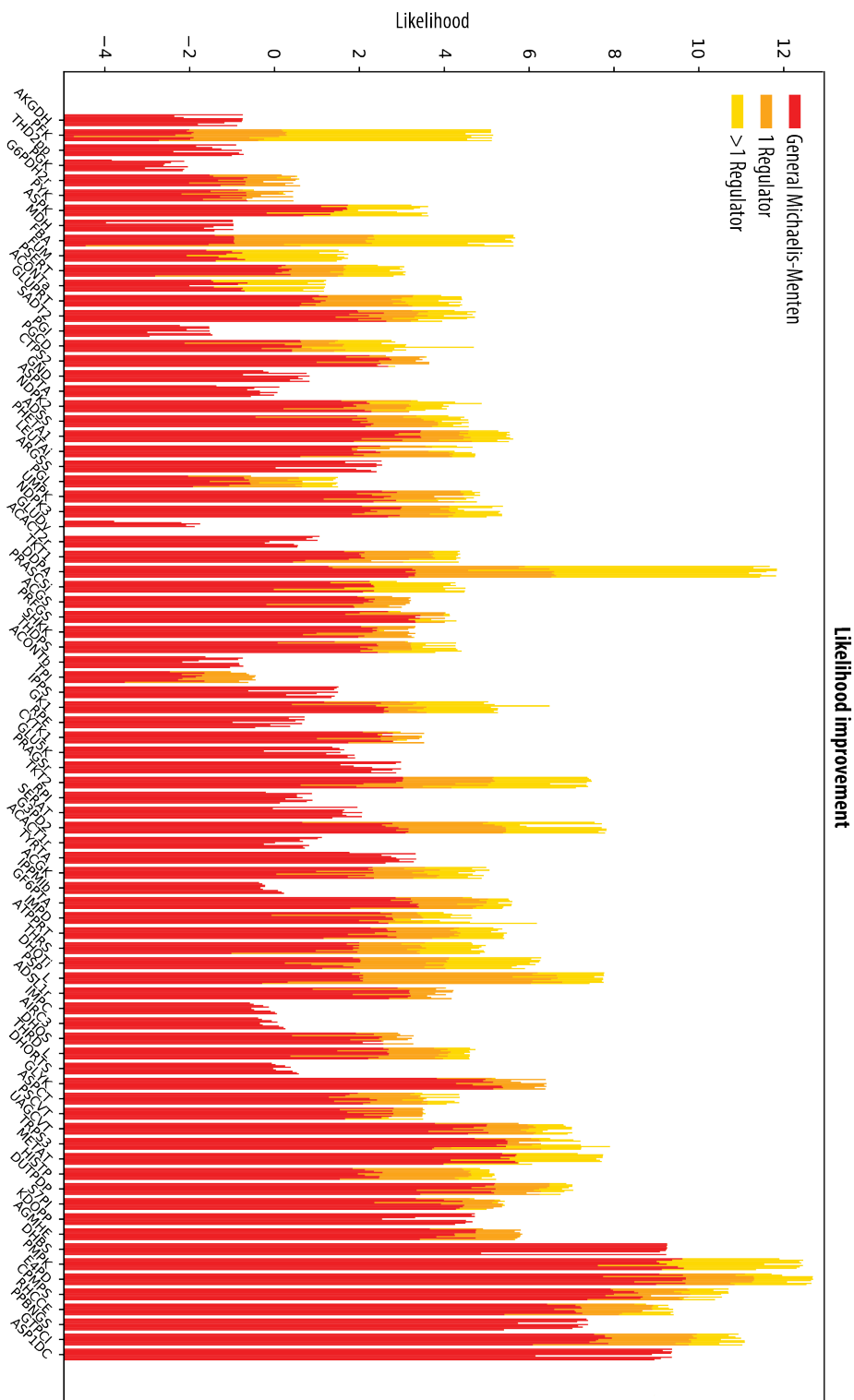


Figure 2.3: Likelihood improvement upon addition of pairwise allosteric regulators across the 84 reactions. The likelihood of the unregulated reaction (red) is shown, as well as the best likelihood upon inclusion of single regulators (orange) and pairwise combinations of regulators (yellow). For each reaction, nine adjacent smaller bars correspond to the nine conditions. Only regulators that improve the akaike information criterion (AIC) with respect to the unregulated model are considered. Reaction abbreviations correspond to BiGG [30] identifiers.

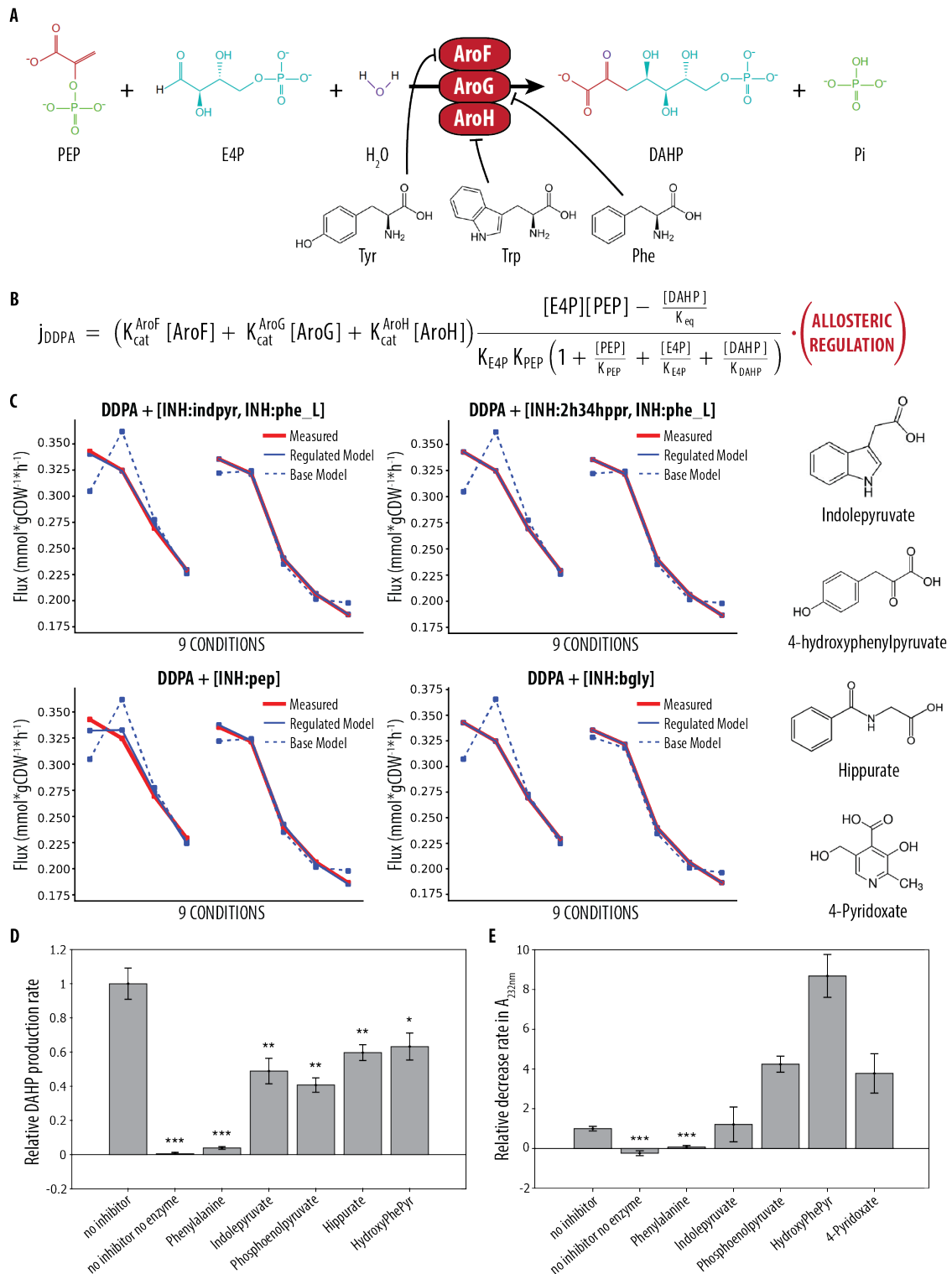


Figure 2.4: Allosteric regulation of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DDPA). **A.** DDPA catalyzes the conversion of phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P) to 3-deoxy-D-arabino-heptulosonate 7-phosphate (DHAP) and phosphate (Pi). The three isoenzymes that catalyze this reaction are AroG, AroF and AroH. **B.** The rate law equation of the reaction with the term for allosteric regulation given in red. **C.** On top two of the best performing predicted pairwise interactions and below two best performing predicted single interactions, with the measured (red), base model predicted (blue dashed) and regulated model predicted (blue solid) fluxes. **D.** The FIA-TOF enzyme assay results of 4mM phenylalanine and novel inhibitors of AroG (4mM PEP, 3.4mM indolepyruvate, 4mM 4-hydroxyphenylpyruvate, and 4mM hippurate) show significant effects, based on a one-tailed T-test. **E.** Absorbance spectroscopy enzyme assay (1/4 diluted substrates and inhibitors, and 1/2 diluted enzyme with respect to FIA-TOF enzyme assays) and measuring the decrease in absorbance at 232nm by PEP. Significance was determined using a one-tailed T-test: $p < 0.05$ (*); $p < 0.01$ (**); $p < 0.001$ (***)

deoxy-D-arabino-heptulosonate 7-phosphate (DHAP) and phosphate (Pi). This constitutes the first committed step in the synthesis of aromatic amino acids [45]. The three isoforms that are known to catalyze this reaction, AroG, AroF and AroH, are known to be regulated by pathway products phenylalanine, tyrosine and tryptophan, respectively [59]. In our analysis we also find phenylalanine and tyrosine among the top scoring regulators (Figure 2.4). Conversely, we predict phenylpyruvate to be an allosteric regulator, ranked 43th out of 628, however a previous study that of the interaction found no evidence for the interaction [46] and hence its final ranking after including the additional evidence was to 213th and we did not follow-up with enzyme essays. The newly predicted interactions are inhibition of DDPA by PEP, indolpyruvate, 4-hydroxyphenylpyruvate (4HPP) and hippurate, for each of which we detected a decrease in product formation that is indicative of inhibitory regulation (Figure 2.4D). Due to the unavailability of DHAP we could not generate a calibration curve and hence matrix effects - which refers to the effect other compounds in the sample have on the detection of the one of interest - could not be excluded. As a result this data is less reliable due to the possibility of confounding effects. We therefore decided to perform the enzyme assay again and measure PEP consumption by absorbance spectroscopy. Strikingly, none of the four newly predicted allosteric inhibitors could be detected as inhibiting AroG using this method (Figure 2.4E). Moreover, PEP and 4HPP seemed to increase the enzyme activity instead. We found that 4HPP absorbs at the same wavelength as PEP, which likely explains part of these conflicting results. Moreover, the high structural similarity of 4HPP to PEP led us to hypothesize that 4HPP might act as a competitive substrate. We confirmed that part of the decrease in absorbance is caused by consumption of 4HPP (Supplementary Figure A.7).

The next reactions which we followed up on are phosphofructokinase (PFK) and fructose-bisphosphate aldolase (FBA), which catalyze two consecutive reactions in glycolysis (Figure 2.5). PFK catalyzes the conversion of fructose-6-phosphate (F6P) and ATP to fructose-1,6-bisphosphate (FBP) and there are two known isozymes, PfkA and PfkB. The known regulators, FBP, ATP, ADP, PEP and citrate [6, 34], collectively make up 11% of the predicted pairwise interactions. Outside of these, of the four highest ranking predictions we detected significant inhibition by indolepyruvate, which was subsequently calibrated to exclude matrix effects. FBA catalyzes the conversion of FBP to glyceraldehyde 3-phosphate (G3P) and dihydroacetone-phosphate (DHAP). PEP is a known allosteric activator [4] and 3-phosphoglycerate (3PG) a known inhibitor [61]. Interestingly, we find the opposite effect for 3PG. Of the newly predicted interactions we find that IMP and ATP act as allosteric activators.

The fourth reaction involves the dephosphorylation of 3-phosphoserine (PSER), which is catalyzed by phosphoserine phosphatase (PEP_L) and of which SerB is the only isoenzyme, to produce L-serine (SER) (Figure 2.6). The only known allosteric inhibitor of this enzyme is L-serine [52]. Among the five metabolites tested we found hypoxanthine to significantly inhibit enzyme activity.

Finally, adenylosuccinate synthetase (ADSS) isoenzyme PurA, catalyzes the GTP-dependent formation of adenylosuccinate from IMP and aspartic acid. This is the first committed step toward the *de novo* synthesis of AMP from IMP. Fumerate, which is formed when adenylosuccinate is cleaved to produce AMP and fumerate, is a reported allosteric inhibitor [20, 7] that was used as a control. We tested AMP, acetyl-glutamate and pantothenate as allosteric inhibitors and found that all three inhibit the enzymes' catalytic activity.

In total we find evidence for the existence of up to 11 out of the 22 predicted interactions that we tested, for nine of which the observed effect was consistent with the predicted mode of action and two showing a significant but opposite effect. Most of these constitute cross pathway interactions, suggesting not only that cross pathway regulation might be more prevalent than currently known, but also a more important mechanism for the global as opposed to local regulation of metabolism.

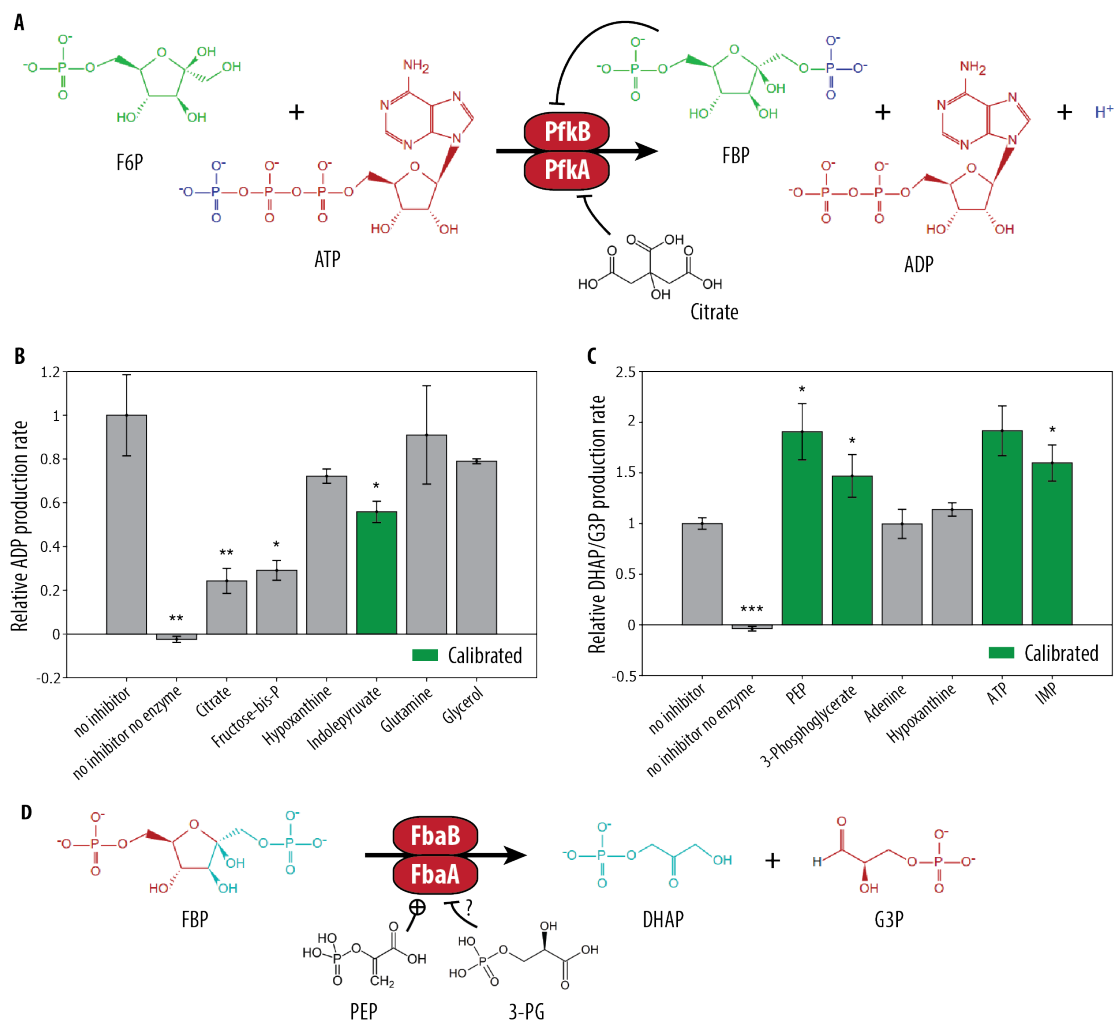


Figure 2.5: Allosteric regulation of phosphofruktokinase (PFK) and fructose-bisphosphate aldolase (FBA). **A.** PFK phosphorylates fructose-6-phosphate (F6P) to form fructose-bisphosphate (FBP). Two known isoforms are PfkA and PfkB. Two known inhibitors are citrate and fbp [6]. **B.** *in vitro* FIA-TOF enzyme assay results for PfkA show significant inhibition by controls (citrate and fbp) and indolepyruvate. **C.** *in vitro* FIA-TOF enzyme assay results for FbaA. The controls here are PEP and 3PG, a reported activator and inhibitor, respectively. Conflicting previous reports we detect an activatory effect for 3pg on FbaA. **D.** FBA catalyzes the subsequent reaction in glycolysis involving the cleavage reaction of FBP to produce DHAP and G3P. Two known isoforms are FbaA and FbaB. Known regulation includes the activation by PEP [4] and inhibition by 3PG [61] Significance was determined using a one-tailed T-test: $p < 0.05$ (*); $p < 0.01$ (**); $p < 0.001$ (***) . Error bars indicate the standard error of the mean across triplicates.

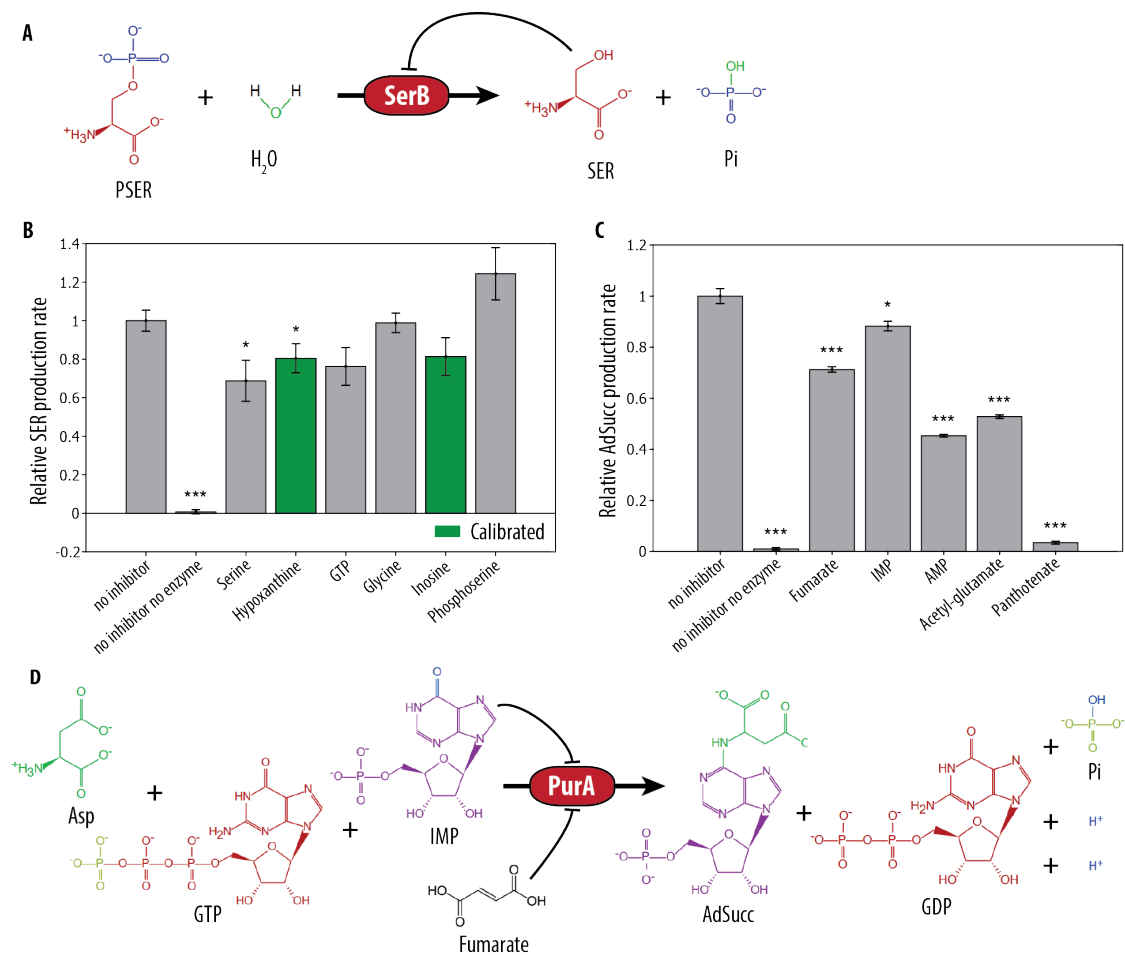


Figure 2.6: Allosteric regulation of phosphoserine phosphatase (PSP_L) and adenylosuccinate synthetase (ADSS). (A) Phosphoserine phosphatase catalyzes the dephosphorylation of phosphoserine, which constitutes the last step in serine biosynthesis. The only known isoenzymes is SerB and the only known allosteric regulator is L-serine, which is also the product [52]. (B) *in vitro* FIA-TOF enzyme assay results for SerB show significant inhibition by the control (L-serine) and the newly predicted inhibitor hypoxanthine. (C) *in vitro* FIA-TOF enzyme assay results for PurA. Fumarate and IMP are known inhibitors used as controls. For all three newly predicted inhibitors, AMP, acetyl-glutamate and pantothenate, we see a significant inhibitory effect. Note that none of these enzyme assays was calibrated. (D) Adenylosuccinate synthetase catalyzes the first committed step toward the *de novo* synthesis of AMP from IMP. The only known isoenzyme is PurA. The known allosteric inhibitors are fumarate and IMP [20, 7]. Significance was determined using a one-tailed T-test: $p < 0.05$ (*); $p < 0.01$ (**); $p < 0.001$ (***)

2.4 Discussion

Allosteric regulation is a widespread mechanism of control of protein function; effectors bind to regulatory sites distinct from the active site, usually inducing conformational changes that influence their activity. Conventional methods used to study allosteric regulation consist of *in vitro* studies of enzyme kinetics in the presence and absence of an allosteric effector. Although these experiments provide detailed information from which kinetic parameter can be inferred, it requires the isolation of the enzyme from the cell and reconstitution of an intracellular-like environment outside the cell in which they can be studied. These experiments are difficult and laborious because far from all enzymes can be isolated in an active-form, as we have also seen in this study. More recently method have been introduced that allow one to assess the physical interactions between multiple protein-metabolite pairs, or even an entire cellular extract [16, 51]. However, the low level of agreement between the detected interactions in these high-throughput methods indicates that they have strong biases or are prone to produce false positives. It also remains an open question how well the *in vitro* behavior translates to that in a native cellular context and, moreover, what their physiological relevance is to the larger cellular system it is part of.

In order to address these limitations we used a single-reaction model fitting procedure that combines the available data on the cellular metabolome, proteome and fluxome to predict the mostly likely regulatory interactions. This computational approach is scalable and uses snapshot data of the *in vivo* state of the cell, providing the ability to predict allosteric interactions that are physiologically relevant in the conditions tested. Among the top scoring candidates we tested 22 newly predicted interactions across five enzymes with the use of enzyme assays. We found evidence for the existence of half of them; nine in accordance with the mode of regulation predicted, two showing the opposite effect, and the remaining half not showing a significant effect on enzymatic activity. Negative results, meaning no effect is observed although allosteric regulation was predicted, might indicate false predictions, however it is also possible that the *in vitro* experimental conditions do not resemble the intracellular environment closely enough. The same holds for the predicted allosteric interactions whose effect is significant but opposite of what was predicted; these might simply be false positives that by chance turn out to have an opposite effect, or it might also be that conditions in the cell cause this effect, such as the presence of another modulator.

The high number of true positives is striking given that previous work using a similar approach in *S. cerevisiae* reported a high false positive rate [22]. One reason might be the data quality, and in particular the determination of the observed fluxes. The previous study relied on the inference of internal fluxes solely based on the uptake and secretion rates and growth rate of the cells. In this study isotopic labeling data was used as well, providing important information on the split of fluxes at metabolic branching points. Another contributing factor might be that the previous study limited its search through candidate interactions reported to exist in other organisms, whereas we did not and instead prioritized predictions *post hoc*. This offers us the ability to evaluate all candidates, meaning that we don't prematurely discard any good candidate. Furthermore, we boost the ranking of predictions for which we have additional evidence, such as their detection in physical interactions studies. It is also important to realize the limitation of the orthogonal approaches used in the experimental validation. In the case of FIA-TOF enzyme assays calibration is important because the compounds other than the analyte of interest can affect the ionization process and thereby taint the results (also known as the matrix effect) [58]. In the case of DDPA (AroG), where we could not calibrate the mass spectrometer due to absence of the product DHAP, we instead decided to perform a classical absorbance spectroscopy enzyme assay to measure PEP consumption. These results, with exception of the controls used, showed either no significant effect or even an opposite effect as to what we observed in the TIA-TOF enzyme assay. Notably, in the FIA-TOF assay we determined catalytic activity by observing product formation, whereas in the absorbance assay we measured the consumption of one of its substrates. Considering that in the initial conditions of the enzyme assay the substrates are present at a saturating amounts whereas the products are absent changes in product formation are more readily detected than those in substrate consumption. Lack of a calibration curve and the absence of support from the absorbance spectroscopy assay makes the confirmatory evidence for these interactions less strong. We conclude from this that there is support for those predictions that are confirmed by the FIA-TOF enzyme assay as long as these are not contradicted in the absorbance based assay. In the case of PFK (PfkA), FBA (FbaA) and PEP_L (SerB) the product of the reaction was available and hence calibration curves were used, whereas for ADSS (PurA) we did not.

Taken together we find evidence for the existence of up to 11 novel allosteric interactions. The identification of these in an organism as well studied as *E. coli* further supports the notion that there is an extensive network of allosteric regulation that we are still largely unaware of. The systematic identification of regulatory interactions is a first step towards understanding the role of allosteric regulation in metabolism. Considering the cross-pathway nature of most of the interactions found in this study, it is possible that this type of interaction is possible underestimated. However, we cannot rule out the fact that these might be spurious interactions, possibly resulting from the narrow biochemical space that small molecules occupy [3], which have no biological relevance. Determination of cellular concentrations of the metabolites and quantification of the strength of these interactions would be a next step in determining of their functional relevance. Knowing which interactions exist and what effect they have at a local level paves the way for functional analyses at a larger scale in which we study their contribution to large-scale system dynamics. This knowledge can in turn be used to develop more efficient disease treatment programs from a biomedical perspective or improve metabolic engineering strategies and biotechnological production processes in an industrial setting.

The main limitations of the approach used here are (1) that relative quantification data are used, (2) that we use data from steady-state conditions, (3) consider only generalized Michaelis-Menten kinetics and (4) that we combine data from two different studies. The first of these implies that we cannot use any information available to us in the form of kinetic or thermodynamic parameters. Ideally we would use this information, however it would require us to obtain absolute quantification data of the metabolome, proteome and fluxome. The second limitation means that we inherently limit ourselves to the detection of allosteric interactions whose effect can be observed at steady-state, and thereby exclude those relevant in coordinating transient responses. Regarding the third limitation, it is possible that different rate law approximations might be more appropriate for specific reaction, which have not been considered in this work. It could therefore be interesting to see how predictions change with different rate law approximations, specifically so for individual reaction mechanisms that have been studied in-depth and can be found in the SABIO-RK database [62]. Finally, we did not conduct our own experiments to obtain the multi-omics data and instead leveraged two previous studies used to derive our predictions. When combining data from different studies conducted by different errors, such as those introduced by the experimentalist, the instrument or the environment, are no longer systematic are therefore less easily differentiated from the biological variation of interest. Moreover and specific to our study, the carbon-limited growth condition data was derived from growth on glucose in one study [32], whereas it was derived from growth on lactose in the other [27]. Although proteome differences on these carbon sources might affect only a small subset of enzymes and growth rate limitation might be a dominant driver, this discrepancy is a shortcoming of our study we had to accept in order not to discard carbon-limited growth conditions. Such a situation that would be avoided when designing and conducting one's own experiments and the significance of this discrepancy is expected to directly affect the accuracy of our predictions.

Although knowledge of the existence of physical interactions is primary, the next step towards a more comprehensive understanding of allosteric regulation is the study of their functional relevance. In this work we took a systems identification approach in which we brought together the existing data and a generalized model of enzyme kinetics in order to see if it can explain the data for individual reactions. Bringing these individual reactions together in a model describing the larger metabolic system that they comprise would allow one to probe deeper into the complex behavior that emerges as a consequence of network dynamics. This will bring along a new set of challenges, including questions such as how to properly demarcate a subsystem of interest, ensuring that physical dependencies in the system are not violated, dealing with discrepancies in the data, dealing with a parameter space that is sparsely populated with *in vitro* data that might not translate well to an *in vivo* context, as well as more fundamental issues related to structural and practical identifiability.

Bibliography

- [1] Hirotugu Akaike. “A new look at the statistical model identification”. In: *IEEE transactions on automatic control* 19.6 (1974), pp. 716–723.
- [2] Benjamin RC Amor et al. “Prediction of allosteric sites and mediating interactions through bond-to-bond propensities”. In: *Nature communications* 7.1 (2016), pp. 1–13.
- [3] William Bains. “A trip through chemical space: why life has evolved the chemistry that it has”. In: *Evolutionary Biology: Genome Evolution, Speciation, Coevolution and Origin of Life*. Springer, 2014, pp. 371–394.
- [4] STEPHEN A Baldwin and RICHARD N Perham. “Novel kinetic and structural properties of the class-I D-fructose 1, 6-bisphosphate aldolase from *Escherichia coli* (Crookes’ strain)”. In: *Biochemical Journal* 169.3 (1978), pp. 643–652.
- [5] Markus Basan et al. “Overflow metabolism in *Escherichia coli* results from efficient proteome allocation”. In: *Nature* 528.7580 (2015), pp. 99–104.
- [6] David P Bloxham and Henry A Lardy. “7 Phosphofructokinase”. In: *The enzymes*. Vol. 8. Elsevier, 1973, pp. 239–278.
- [7] Jan M Boitz et al. “Adenylosuccinate synthetase and adenylosuccinate lyase deficiencies trigger growth and infectivity deficits in *Leishmania donovani*”. In: *Journal of Biological Chemistry* 288.13 (2013), pp. 8977–8990.
- [8] Stephen P Brooks and Andrew Gelman. “General methods for monitoring convergence of iterative simulations”. In: *Journal of computational and graphical statistics* 7.4 (1998), pp. 434–455.
- [9] Kenneth P Burnham. “Model selection and multimodel inference”. In: *A practical information-theoretic approach* (1998).
- [10] Kenneth P Burnham and David R Anderson. “Multimodel inference: understanding AIC and BIC in model selection”. In: *Sociological methods & research* 33.2 (2004), pp. 261–304.
- [11] Joseph E Cavanaugh. “Unifying the derivations for the Akaike and corrected Akaike information criteria”. In: *Statistics & Probability Letters* 33.2 (1997), pp. 201–208.
- [12] Christophe Chassagnole et al. “Dynamic modeling of the central carbon metabolism of *Escherichia coli*”. In: *Biotechnology and bioengineering* 79.1 (2002), pp. 53–73.
- [13] Dimitris Christodoulou et al. “Reserve flux capacity in the pentose phosphate pathway enables *Escherichia coli*’s rapid response to oxidative stress”. In: *Cell systems* 6.5 (2018), pp. 569–578.
- [14] Victor Chubukov et al. “Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis*”. In: *Molecular systems biology* 9.1 (2013), p. 709.
- [15] Pascale Daran-Lapujade et al. “The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels”. In: *Proceedings of the National Academy of Sciences* 104.40 (2007), pp. 15753–15758.
- [16] Maren Diether et al. “Systematic mapping of protein-metabolite interactions in central metabolism of *Escherichia coli*”. In: *Molecular systems biology* 15.8 (2019), e9008.
- [17] Tobias Fuhrer et al. “High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection–time-of-flight mass spectrometry”. In: *Analytical chemistry* 83.18 (2011), pp. 7074–7080.

-
- [18] Luca Gerosa et al. “Pseudo-transition analysis identifies the key regulators of dynamic metabolic adaptations from steady-state data”. In: *Cell systems* 1.4 (2015), pp. 270–282.
- [19] Matthew J Giacalone et al. “Toxic protein expression in Escherichia coli using a rhamnose-based tightly regulated and tunable promoter system”. In: *Biotechniques* 40.3 (2006), pp. 355–364.
- [20] Andrea Gorrell et al. “Determinants of l-Aspartate and IMP Recognition in Escherichia coli Adenylosuccinate Synthetase”. In: *Journal of Biological Chemistry* 277.11 (2002), pp. 8817–8821.
- [21] Steinn Gudmundsson and Ines Thiele. “Computationally efficient flux variability analysis”. In: *BMC bioinformatics* 11.1 (2010), pp. 1–3.
- [22] Sean R Hackett et al. “Systems-level analysis of mechanisms regulating yeast metabolic flux”. In: *Science* 354.6311 (2016).
- [23] Laurent Heirendt et al. “Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0”. In: *Nature protocols* 14.3 (2019), pp. 639–702.
- [24] Min Huang et al. “AlloFinder: a strategy for allosteric modulator discovery and allosterome analyses”. In: *Nucleic acids research* 46.W1 (2018), W451–W458.
- [25] Wenkang Huang et al. “Allosite: a method for predicting allosteric sites”. In: *Bioinformatics* 29.18 (2013), pp. 2357–2359.
- [26] Zhimin Huang et al. “ASD: a comprehensive database of allosteric proteins and modulators”. In: *Nucleic acids research* 39.suppl_1 (2011), pp. D663–D669.
- [27] Sheng Hui et al. “Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria”. In: *Molecular systems biology* 11.2 (2015), p. 784.
- [28] GARY R Jacobson and GEORGE R Stark. “Aspartate transcarbamylase of Escherichia coli. Mechanisms of inhibition and activation by dicarboxylic acids and other anions.” In: *Journal of Biological Chemistry* 250.17 (1975), pp. 6852–6860.
- [29] Henrik Kacser et al. “The control of flux”. In: *Biochemical Society Transactions* 23.2 (1995), pp. 341–366.
- [30] Zachary A King et al. “BiGG Models: A platform for integrating, standardizing and sharing genome-scale models”. In: *Nucleic acids research* 44.D1 (2016), pp. D515–D522.
- [31] Masanari Kitagawa et al. “Complete set of ORF clones of Escherichia coli ASKA library (A Complete Set of E. coli K-12 ORF Archive): Unique Resources for Biological Research”. In: *DNA research* 12.5 (2005), pp. 291–299.
- [32] Karl Kochanowski et al. “Few regulatory metabolites coordinate expression of central metabolic genes in Escherichia coli”. In: *Molecular Systems Biology* 13.1 (2017), p. 903.
- [33] Karl Kochanowski et al. “Global coordination of metabolic pathways in Escherichia coli by active and passive regulation”. In: *Molecular systems biology* 17.4 (2021), e10064.
- [34] Denise Kotlarz and Henri Buc. “[11] Phosphofructokinases from Escherichia coli”. In: *Methods in enzymology* 90 (1982), pp. 60–70.
- [35] Oliver Kotte and Matthias Heinemann. “A divide-and-conquer approach to analyze under-determined biochemical models”. In: *Bioinformatics* 25.4 (2009), pp. 519–525.
- [36] Harry H Ku et al. “Notes on the use of propagation of error formulas”. In: *Journal of Research of the National Bureau of Standards* 70.4 (1966), pp. 263–273.
- [37] Benno H ter Kuile and Hans V Westerhoff. “Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway”. In: *FEBS letters* 500.3 (2001), pp. 169–171.
- [38] Shuai Li et al. “Alloscore: a method for predicting allosteric ligand–protein interactions”. In: *Bioinformatics* 32.10 (2016), pp. 1574–1576.
- [39] Wolfram Liebermeister and Edda Klipp. “Bringing metabolic networks to life: convenience rate law and thermodynamic constraints”. In: *Theoretical Biology and Medical Modelling* 3.1 (2006), pp. 1–13.

-
- [40] Hannes Link, Karl Kochanowski, and Uwe Sauer. “Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo”. In: *Nature biotechnology* 31.4 (2013), pp. 357–361.
- [41] Athanasios Litsios et al. “Metabolic-flux dependent regulation of microbial physiology”. In: *Current Opinion in Microbiology* 42 (2018), pp. 71–78.
- [42] Michael Lynch, Bruce Walsh, et al. “Genetics and analysis of quantitative traits”. In: (1998).
- [43] Daniel Machado, Markus J Herrgård, and Isabel Rocha. “Modeling the contribution of allosteric regulation for flux control in the central carbon metabolism of *E. coli*”. In: *Frontiers in bioengineering and biotechnology* 3 (2015), p. 154.
- [44] Pierre Millard, Kieran Smallbone, and Pedro Mendes. “Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in *Escherichia coli*”. In: *PLoS computational biology* 13.2 (2017), e1005396.
- [45] David L Nelson and Michael M Cox. *Absolute Ultimate Guide for Lehninger Principles of Biochemistry*. 2008.
- [46] Yaroslav V Nikolaev et al. “Systematic identification of protein–metabolite interactions in complex metabolite mixtures by ligand-detected nuclear magnetic resonance spectroscopy”. In: *Biochemistry* 55.18 (2016), pp. 2590–2600.
- [47] Sabine Oesterle et al. “Efficient engineering of chromosomal ribosome binding site libraries in mismatch repair proficient *Escherichia coli*”. In: *Scientific reports* 7.1 (2017), pp. 1–10.
- [48] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. “What is flux balance analysis?” In: *Nature biotechnology* 28.3 (2010), pp. 245–248.
- [49] Jeffrey D Orth et al. “A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011”. In: *Molecular systems biology* 7.1 (2011), p. 535.
- [50] Alejandro Panjkovich and Xavier Daura. “PARS: a web server for the prediction of protein allosteric and regulatory sites”. In: *Bioinformatics* 30.9 (2014), pp. 1314–1315.
- [51] Ilaria Piazza et al. “A map of protein-metabolite interactions reveals principles of chemical communication”. In: *Cell* 172.1-2 (2018), pp. 358–372.
- [52] Lewis I Pizer. “The pathway and control of serine biosynthesis in *Escherichia coli*”. In: *Journal of Biological Chemistry* 238.12 (1963), pp. 3934–3944.
- [53] Ed Reznik et al. “Genome-scale architecture of small molecule regulatory networks and the fundamental trade-off between regulation and enzymatic activity”. In: *Cell reports* 20.11 (2017), pp. 2666–2677.
- [54] Manfred Rizzi et al. “In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. Mathematical model”. In: *Biotechnology and bioengineering* 55.4 (1997), pp. 592–608.
- [55] Severin Schink et al. “Analysis of proteome adaptation reveals a key role of the bacterial envelope in starvation survival”. In: *Molecular Systems Biology* 18.12 (2022), e11160.
- [56] Karthik Sekar et al. “N-terminal-based targeted, inducible protein degradation in *Escherichia coli*”. In: *PLoS One* 11.2 (2016), e0149746.
- [57] Maya Shamir et al. “SnapShot: timescales in cell biology”. In: *Cell* 164.6 (2016), pp. 1302–1302.
- [58] Alexander G Shard et al. “The matrix effect in organic secondary ion mass spectrometry”. In: *International Journal of Mass Spectrometry* 377 (2015), pp. 599–609.
- [59] Louis C Smith et al. “The control of 3-deoxy-D-arabino-heptulosonic acid 7-phosphate synthesis by phenylalanine and tyrosine”. In: *Journal of Biological Chemistry* 237.11 (1962), pp. 3566–3570.
- [60] Eric Soupene et al. “Physiological studies of *Escherichia coli* strain MG1655: growth defects and apparent cross-regulation of gene expression”. In: *Journal of bacteriology* 185.18 (2003), pp. 5611–5626.
- [61] Benjamin S Szwegold, Kamil Ugurbil, and Truman R Brown. “Properties of fructose-1, 6-bisphosphate aldolase from *Escherichia coli*: an NMR analysis”. In: *Archives of biochemistry and biophysics* 317.1 (1995), pp. 244–252.

-
- [62] Ulrike Wittig et al. “SABIO-RK—database for biochemical reaction kinetics”. In: *Nucleic acids research* 40.D1 (2012), pp. D790–D796.
- [63] Conghui You et al. “Coordination of bacterial proteome with metabolism by cyclic AMP signalling”. In: *Nature* 500.7462 (2013), pp. 301–306.
- [64] Alon Zaslaver et al. “A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*”. In: *Nature methods* 3.8 (2006), pp. 623–628.

Chapter 3

Prediction of allosteric interactions in the tricarboxylic acid cycle using a system of coupled differential equations

M.A.P. Karrenbelt contributed to the project design, conducted all physiology and metabolite experiments, carried out data analysis and wrote the chapter.

M.F. Buffing contributed to experimental method development.

T. Fuhrer contributed to experimental method development.

Dr. E. Noor contributed to the computational method development and supervised the study.

Prof. Dr. Uwe Sauer contributed to the project design and supervised the study.

3.1 Introduction

Enzymatic reaction rates, or fluxes, constitute the functional output of metabolism. At the core of metabolism lay pathways that are universal to life. This system of biochemical pathways, referred to as central carbon metabolism, is a hub at the intersection of catabolic and anabolic processes. Because of its central role in ATP generation, cofactor regeneration, amino acid, fatty acid and nitrogen base biosynthesis, it is pivotal that metabolic fluxes are distributed appropriately such that energy production and precursor biosynthesis are balanced in accordance with cellular demand. Therefore, flux control is of the highest importance for bacteria such as *Escherichia coli* if they are to achieve homeostasis and successfully adapt to new environments. In order to achieve this, regulatory circuitries have evolved that enable the cell to adopt to a large variety of states on a physiological time scale.

3.1.1 Metabolic regulation

Regulatory circuitries can be grouped into those that affect enzyme abundance and those that affect enzyme activity. Enzyme activity can be regulated by covalent and non-covalent metabolite interactions, whereas enzyme abundance is controlled through protein synthesis, degradation and dilution. Metabolic flux in the cell is therefore not only the product of chemical reaction kinetics driven by thermodynamics and physiochemical determinants hereof, such as temperature, pH and ionic strength [38], but also of the layers of regulatory circuitry that control the abundance and activity of enzymatic catalysts in the network.

3.1.1.1 Temporal dynamics of metabolic regulation

Since metabolic turnover in the cell occurs on a time scale ranging from milliseconds to seconds, locally induced changes in metabolite abundance quickly propagate through the metabolic network. As a consequence, modulation of enzymatic activity by metabolites, even by those more distally related in the metabolic network, constitutes a mechanism that operates on roughly the same timescale [81, 110]. In contrast, protein levels are controlled by processes of synthesis, degradation and dilution. Adjustment through modulation synthesis rates involves a graded response that involves additional steps such as gene transcription, mRNA translation and protein folding. These range from riboswitches to rho-dependent termination at the transcriptional level [126, 31, 119], from antisense RNA to nonsense-mediated decay at the RNA level [47, 102], and from ubiquitination to protosomal degradation at the protein level [97, 26]. Control of enzyme abundance by degradation constitutes an active regulatory mechanism that occurs on a time scale ranging from minutes to tens of minutes. It is invoked when, for example, an enzyme gets damaged as a result of oxidative stress, at which stage it gets targeted for degradation by the ubiquitin-proteasome system [91]. Dilution, on the other hand, constitutes a passive mechanism that occurs on a timescale ranging from tens of minutes to hours, depending largely on the duration of the cell cycle [95]. It is important to note that this results in a process whose first effects at the metabolic level can be observed after approximately a minute and it might take several hours before fluxes are fully adjusted and a new steady-state is attained [81, 110]. Hence in the event of a perturbation we can distinguish two phases: an initial phase in which the response is driven by thermodynamics and in which regulatory control is exerted by metabolites that directly modulate enzyme activity, and a more gradual long term response that includes the aforementioned processes that ultimately control enzyme abundance.

3.1.1.2 Importance of metabolic regulation

Understanding how cells control metabolic fluxes is of extensive interest in the study and treatment of numerous diseases that are associated with metabolic dysregulation, including diabetes [129], obesity [69], Alzheimer's [128] and many types of cancers [93]. From a biotechnological perspective the understanding of metabolic regulation is needed both for the development of new, as well as the optimization of existing, industrial production processes [82, 40, 44]. However, our understanding of these regulatory processes is far from sufficient and therefore limits our ability to successfully treat metabolic maladies or engineer strains with the desired production characteristics. One of the reasons for this is the fact that we don't possess a complete picture of the regulatory interaction network. Moreover, even if we were to possess a complete picture of both the biochemical pathway

and regulatory interaction topology, a static description of nodes and edges alone is often insufficient to predict how behavior of the system changes over time.

3.1.1.3 Understanding metabolic regulation

Chassagnole *et al.* (2002) introduce a dynamic model of glycolysis and the pentose-phosphate pathway in *E. coli* [16]. Validated with transient metabolite data, the model explores central carbon metabolism as a source of precursors. It notably links sugar transporter PTS kinetics to central carbon metabolism, emphasizing a strong nonlinear feedback/feedforward link between PTS and glycolytic pools of PEP and Pyr. The model successfully replicates experimental metabolite dynamics, including oscillations. Additionally, it's applied to metabolic control analysis, revealing shared flux control between glucose uptake system PTS and enzymes degrading PTS inhibitors G6P and pyruvate.

Kadir *et al.* (2010) focused on simulating the main metabolic pathways in *E. coli*, including glycolysis, TCA cycle, pentose phosphate pathway, and anaplerotic pathways [56]. Using enzymatic reaction models, they fine-tuned parameters based on *in vivo* metabolite concentrations. The specific ATP, CO₂, and NADPH production rates were computed and utilized for estimating growth rate, cell yield, and oxidative pentose phosphate pathway flux. The study incorporated batch and continuous cultivations, aligning changing metabolite concentrations with experimental data. Gene knockout effects were investigated for pathways like Ppc, Pck, and Pyk, revealing intricate metabolic responses. The importance of the anaplerotic route of Ppc and Pyk knockout-induced regulatory shifts was highlighted. The simulation results showed good predictability for gene knockout effects, emphasizing the model's usefulness for understanding metabolic changes due to specific gene perturbations.

Peskov *et al.* (2012) developed a more detailed kinetic model for *E. coli*'s central carbon metabolism [94]. Their model provided insight into both steady-state and dynamic behaviors of the system. By focusing on steady-state analysis, they investigated the distribution of fluxes in *E. coli*'s central carbon metabolism during glucose-limited aerobic growth. They discovered that glycolysis dominates glucose assimilation, while the PMP and Entner-Doudoroff pathways contribute to <35% of glucose consumption. The study delved into metabolic issues like the looping of the TCA cycle, revealing a possible Mdh/Mqo futile cycle for precise intracellular malate regulation. They explored the flux distribution between enzyme isoforms, demonstrating that kinetic models outperform stoichiometric models in distinguishing these fluxes. Additionally, the study provided insights into pyruvate kinase isozymes, revealing that both PykA and PykF operate during glucose-limited growth. They observed the dominance of specific enzyme isozymes in phosphofructokinase and fructose biphosphatase under varying glucose concentrations. Furthermore, the researchers studied pykA knockout mutants and refined their model to describe both steady-state flux distributions and metabolomics data. They hypothesized a novel regulatory mechanism, Gnd inhibition by PEP, which improved the agreement between model predictions and experimental metabolomics data.

Khodayari *et al.* (2014) employed an ensemble modeling strategy to construct a kinetic model of *E. coli*'s core metabolism [59]. This method involves obtaining a steady-state flux distribution using available data. Within the EM framework, reactions are deconstructed into elementary steps with explicit forward and reverse fluxes. Thermodynamic constraints and normalized metabolite concentrations guide reaction reversibility. Enzyme conservation and kinetic parameters are determined based on reversibilities and enzyme fractions. An ensemble of kinetic parameters is generated by sampling reversibilities and enzyme fractions. The system's differential equations are solved for each parameter set to achieve steady-state, and model fitness is improved using an optimizer to update parameters. Their resulting kinetic models are matched against diverse data, validated against metabolomics, kinetic constants, and cross-validations. The model successfully captured 78% of reactions within experimental ranges, and cross-validation demonstrated reliable predictions for mutants adjacent to those used in training. Furthermore, metabolite concentrations were accurately predicted for 68% of metabolites. Comparative analysis of kinetic parameters showed strong agreement with literature data for 35% of Km and 77% of kcat values.

Jahan *et al.* (2016) developed a kinetic model replicating flux data from knockout mutants in glucose-based aerobic batch cultures [54]. Validation showed good performance at higher growth rates, albeit underestimating TCA cycle fluxes in some mutants. Sensitivity analysis identified crucial enzymes for growth, particularly related to glucose uptake and irreversible reactions. Furthermore, the effect of gene regulatory and allosteric interactions was studied:

-
- With respect to gene regulation, the absence of Crp-cAMP interaction reduced TCA cycle and glyoxylate shunt fluxes. Conversely, Cra-FBP interaction absence increased gluconeogenesis, TCA cycle, and glyoxylate shunt activity. Removing PdhR-PYR interaction notably hindered glyoxylate shunt but minimally affected other fluxes.
 - With respect to allosteric regulations: PEP inhibition removal increased Pfk synthesis ratio and flux through neighboring reactions, TCA cycle, and glyoxylate shunt. FBP activation removal raised Pyk activity, enhancing TCA cycle and glyoxylate shunt flux. Malic enzyme flux decreased by removing AcCoA and cAMP inhibition. Fbp removal increased flux through TCA cycle, anaplerotic reaction, and glyoxylate shunt; Ppc activation removal increased gluconeogenic flux.

Crucially, these findings underscore the inherent complexity of biological systems and the intricacies of metabolic networks. While the current study’s insights provide a solid foundation for understanding the effects of genetic and allosteric perturbations on glucose-based growth, extrapolating to alternate carbon sources demands a cautious approach. Our ability to make informed predictions based on metabolic network analysis is invaluable, yet the dynamic interplay of diverse factors within living systems introduces a level of unpredictability that transcends our current grasp. The remarkable complexity of cellular metabolism reminds us that even with sophisticated models, we navigate a terrain where our comprehension, while advancing, remains humbly partial.

3.1.2 Allosteric regulation in *E. coli* CCM

A recent study on regulation coefficients in *Escherichia coli* (*E. coli*) has shown that thermodynamics and transcriptional regulation alone cannot explain the majority of the difference in metabolic fluxes observed between different steady-states [37]. In another study on *E. coli*, in which metabolic control analysis was performed using a dynamic model used to fit steady-state data, the authors conclude that metabolite-enzyme interactions, as opposed to gene regulation, explains the majority of experimental observations [80]. The importance of direct regulation of enzymes by metabolites has also been highlighted in the yeast *Saccharomyces cerevisiae*. One such study on steady-state data concludes that metabolite concentrations have more than double the physiological impact of enzymes [41]. Another using a dynamic modeling approach showed that the inclusion of allosteric regulation was crucial in obtaining a good correspondence between the model and steady-state as well as dynamic data [32]. These results suggest a prominent role for allosteric regulation in shaping dynamic responses in, and the homeostatic regulation of, metabolism.

3.1.2.1 Previous studies on the role of allosteric regulation

Two previous studies from our lab have scrutinized the role of allosteric regulation in glycolysis [75] and the pentose phosphate pathway [17] of *E. coli*. In the study of glycolysis a perturbation resulting in flux reversal was performed and ten interactions were predicted to be relevant, six of which were already known at the time and one that was newly confirmed, whereas the study on the pentose phosphate pathway lead to new insight at the functional level that contradicted contemporary presuppositions regarding the *prima causa* of flux rearrangement in central carbon metabolism after exposure to reactive oxygen species. Both of these studies used a dynamic modeling-based approach in which a systematic analysis of regulatory topologies was performed in order to predict the existence of allosteric interactions based on their ability to improve the correspondence between model simulation and the observed metabolite dynamics. These results show that such an approach can be used to identify unknown allosteric regulation as well as shed light on their functional relevance in context of the larger system that they inhabit.

3.1.2.2 Allosteric regulation in the tricarboxylic acid cycle

Based on our current knowledge of allosteric interactions in *E. coli* metabolism it appears that these are less prevalent in the tricarboxylic acid (TCA) cycle than in glycolysis and the pentose phosphate pathway. However, a systematic analysis of the relevance of allosteric interactions in shaping metabolite dynamics in this subsystem using a dynamic modeling-based approach has not yet been performed. Interestingly, almost all of the reported allosteric interactions in the TCA cycle, and the majority of those affecting anaplerotic pathways, were found to be of an inhibitory nature

in *E. coli* [100]. However, determining the functionality of these inhibitory interactions solely based on their inhibitory nature can be misleading, as their impact depends on the relative change of the input signal. To illustrate, an interaction that inhibits when the signal increases may actually lead to stimulation when the signal decreases, thus accelerating the process it influences. Moreover, many reported allosteric regulators might not be universally valid and could have condition-specific relevance. It hence remains an open question what role allosteric interactions in the TCA cycle of *E. coli* play during metabolic adaptation.

In this study we investigate their role in tuning transitory dynamics that result from switching from a less preferred to a preferred carbon source. Furthermore, a comparison will be made with methods used in the previous studies on glycolysis and the pentose phosphate pathway. Concerning experiments, the intricacies of the experimental procedures are examined and results of this study will be contrasted with those obtained in previous work. With regard to modeling, model dependence on assumptions and guesswork will be discussed. Finally, standards, practices and their consequences for reproducibility, reliability and reuseability are evaluated to come with suggestions on how to improve on the *status quo* moving forward.

The method we use to assess the functional relevance of allosteric interactions consists of relating observational time series data from a fast filtration experiment to simulated data from models containing different sets of allosteric interactions. In the experimental procedure we sample cells growing in mid-exponential phase, the initial pseudo steady-state condition, place them on a filter and perfuse them with medium containing a different carbon source for short periods of time, and then immediately quench and extract them. These cell extracts are subjected to liquid chromatography-coupled mass spectrometry in order to quantify intracellular metabolite concentrations. The modeling approach consists of constructing an ensemble of dynamic models that comprises a base model without regulation, models containing a single regulatory interaction and models containing combinations of pairwise regulatory interactions.

The limitation to pairwise interactions in our method is driven by the complexity and computational challenges of modeling higher-order interactions within the metabolic network. Focusing on pairwise interactions strikes a balance between computational feasibility and the potential for capturing relevant regulatory insights. This approach enables systematic exploration of the regulatory landscape and offers predictive insights through an ensemble of dynamic models. While none of these models individually can provide an accurate representation of the complex metabolic network, they act as "weak learners" that, when combined, have the potential to offer predictive insights beyond any individual model. This approach allows us to encompass a broad range of regulatory scenarios, capturing uncertainties and potential interactions. Notably, this ensemble approach is unbiased in that it does not include previously reported allosteric interactions by default. This unbiased approach allows us to assess the regulatory relevance of allosteric interactions in specific conditions, as well as the ability to predict allosteric interactions not previously reported. Any newly predicted interactions can be subjected to experimental validation using *in vitro* enzyme essays.

3.1.3 Research objectives

While previous studies have provided insights into the role of allosteric regulation in *E. coli* CCM during the initial response after a perturbation in glycolysis and the PPP, such a study has not yet been conducted for the TCA cycle and anaplerotic pathways. To address these gaps, this study aims to achieve the following objectives:

1. **Perform an unbiased assessment:** By excluding previously reported allosteric regulators from our model by default, we aim to perform an unbiased estimation of allosteric interactions within the TCA cycle of *E. coli*. This approach will allow us to discern between condition-specific regulatory interactions and those with broader functional significance.
2. **Investigate condition-dependent relevance:** Our investigation seeks to understand the condition-specific relevance of allosteric interactions during metabolic adaptation. By focusing on the switch from a less preferred to a preferred carbon source, we aim to uncover how these interactions impact transitory dynamics and contribute to the cell's ability to adapt to changing environmental conditions.
3. **Extend understanding to the TCA cycle regulation:** Building on insights gained from previous studies in glycolysis and the pentose phosphate pathway, we aim to extend

our understanding of allosteric regulation to the TCA cycle. This essential subsystem of metabolism has unique characteristics that warrant a dedicated investigation into the role of allosteric interactions.

4. **Evaluate methodological improvements:** This study will critically evaluate the methodology employed in the dynamic modeling-based approach. We will discuss the assumptions, uncertainties, and limitations associated with the model construction. Additionally, we will explore how improvements in the modeling approach can contribute to the accuracy and reliability of the predictions.
5. **Enhance reproducibility and reliability:** In the pursuit of scientific rigor, we will assess the current standards and practices in the field, with a specific focus on their implications for reproducibility, reliability, and reusability. Through this evaluation, we aim to propose suggestions for enhancing the overall quality of research in metabolic regulation.

By accomplishing these objectives, we seek to advance our understanding of allosteric regulation in the TCA cycle, provide insights into condition-specific regulatory networks, and contribute to the development of more accurate and reliable methodologies for studying metabolic regulation.

3.2 Materials and Methods

3.2.1 *In vivo* experimentation

3.2.1.1 Media

Lysogeny Broth (LB) 5g tryptone, 2.5g yeast extract and 2.5g NaCl was dissolved in 500mL deionized water. For plating medium 7.5g granulated agar was added. Subsequently the medium was autoclaved.

M9 minimal medium

- Carbon source solutions
The final concentration of the carbon source stock solutions was 0.5M for glucose, 1.0M for pyruvate, 0.5M for succinate, and 1.5M for acetate. The carbon source stock solutions were pH neutralized, using either a 37% HCl or a 10M NaOH solution, and filter sterilized.
- Base salts solution
37.6g $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, 15g KH_2PO_4 , 2.5g NaCl, 7.5g $(\text{NH}_4)_2\text{SO}_4$, was dissolved in 1.0L of deionized water.
- Trace elements solution
0.18g $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 0.12g $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$, 0.12g $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ and 0.18g $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ was dissolved in 1.0L of deionized water.
- Thiamine-HCl solution
25mg thiamine-HCl was dissolved in 50mL of deionized water.
- CaCl₂ solution
1.47g $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ was dissolved in 50mL of deionized water to yield a 0.1M solution.
- MgSO₄ solution
24.6g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ was dissolved in 50mL of deionized water to yield a 1.0M solution.
- FeCl₃ solution
1.35g $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$ was dissolved in 50mL of deionized water to yield a 0.1M solution.

The M9 minimal medium was prepared by mixing 200mL of the base salts solution, 700mL ddH₂O, 10mL trace element solution, 1.0mL of 0.1M CaCl₂, 1.0mL of 1.0M MgSO₄, 0.6mL of 0.1M FeCl₃. This solution was stored at 4°C for up to two weeks. Prior to growth experiments 2mL of the thiamine-HCl solution, and 25mL carbon source solution - with the exception of succinate where 37.5mL was added - was added, and deionized water was used to fill up to 1.0L.

3.2.1.2 Strain and cultivation

The wild-type strain *E. coli* BW 25113 was used in all experiments. For cultivation, frozen glycerol stocks were used to inoculate agar-containing LB plates and incubated at 37°C overnight, and subsequently stored at 4°C for up to a week. Single colonies were then used to inoculate liquid LB medium, and grown for 6 hours at 37°C and 250RPM. These LB pre-cultures were then used to inoculate a dilution series of M9 pre-cultures containing either 25mL of a 1.0M pyruvate stock solution, 37.5mL of a 0.5M succinate stock solution, or 25mL of a 1.5M acetate stock solution per litre, and grown overnight at 37°C and 250RPM. A preculture with an optical density at 600nm (OD₆₀₀) between 0.5 and 1.0 was chosen to inoculate 500mL shake flasks containing 100mL of M9 minimal medium with the same respective carbon source as used in the M9 preculture, and cultivated at 37°C and 250RPM until an OD₆₀₀ between 0.3 and 0.4 was reached while monitoring the OD₆₀₀ in order to determine the growth rate. The culture was then transferred to a 500mL beaker glass and placed on a magnetic stirrer plate inside a 37°C chamber until an OD₆₀₀ of 0.5 was reached.

3.2.1.3 Carbon source switches

Aliquots of 2mL of *E. coli* cultures were drawn using a 5mL pipet, and dispensed on a 0.45µm pore size Hydrophilic PVDF filter (Millipore) on top of a Buchner funnel to perform vacuum filtration. While on this filter the cells were perfused with a fresh M9 medium containing the same carbon source as was used during cultivation - either pyruvate, succinate or acetate - for 10s using a 50mL syringe. Immediately after this the perfusion media was switched to a fresh M9 medium containing glucose for a duration of 30s, using another 50mL syringe, and back again to the first perfusion media for another 30s. Perfusion solutions were kept at 37°C. This process was repeated and stopped at different stages to obtain samples of different time points during this perturbation: 10s, 15s, 20s, 25s, 40s, 45s, 50s, 55s and 70s. A control, which consisted of perfusing the cells for 70s with the fresh M9 medium containing the same carbon source as was used during cultivation, was performed to assess the effect of the experimental procedure without the carbon source switch. Every sample was immediately quenched and extracted by rapidly transferring the filter from the Buchner funnel to a 6-well plate with wells containing 4mL of a 2:2:1 acetonitrile:methanol:water extraction solution and was kept at -20°C. For the purpose of normalization 100µL of ¹³C-internal standard was added to the extraction solution directly after a sample was taken. This ¹³C internal standard is a metabolite extract of *E. coli* grown on [U-¹³C]glucose that had been previously prepared. At the end of the entire perfusion experiment an OD600 measurement was performed to correct for biomass growth during the experiment, and to ensure that an OD600 of 1.0 was not surpassed, indicating cells were still in mid-exponential phase. Fast filtration perturbation experiments were performed in triplicate starting from different single colonies of the LB agar plate for each of the different carbon source shift experiments.

3.2.1.4 Metabolic sample preparation

After the samples were placed in the extraction solution they were covered with a lid and kept at -20°C for 2 hours to complete the extraction. The extracts were then transferred to 15mL falcon tubes for overnight storage at -20°C. The next day extracts were centrifuged at 14,000RCF at 4°C for 20min to remove cell debris. Supernatants were dried at 0.12mbar to complete dryness in a SpeedVac composed of an Alpha 2-4 LD plus cooling trap, a RVC 2-33 rotational vacuum concentrator and a RC-5 vacuum chemical hybrid pump (Christ, Osterode am Harz, Germany). Dry metabolite extracts were stored at -80°C until further analysis. Dried metabolite samples were resuspended in 120µL deionized water, plated over 3 different 96-well plates each containing 30µL of the sample; one for direct measurement and two were stored for backup at -80°C.

3.2.1.5 Metabolite calibration

For absolute metabolite quantification, a mix of more than 100 metabolites from *E. coli* metabolism was used which had been previously prepared. This mix contained metabolites at a known equimolar concentration, from which a two-fold dilution series was prepared ranging from 100mM to 49µM using 1mL Eppendorf tubes. For the purpose of normalization 100µL of ¹³C-internal standard was added to the extraction solution directly after a sample was taken. 100µL of ¹³C internal standard was added to each of the samples of the dilution series, which was prepared fresh on the same day right before the fast filtration perturbation. These underwent the exact same procedure as the samples: transfer to 6-well plate with wells containing 4mL of a 2:2:1 acetonitrile:methanol:water extraction solution, kept at -20°C, transferred to 15mL falcon tubes, and dried simultaneously in the same SpeedVac as the samples.

3.2.1.6 Metabolite measurements

10µL of each of the wells was injected into a Waters Acquity UPLC with a Waters T3 column (150mm × 2.1mm × 1.8mm; Waters Corporation, Milford, MA) coupled to a Thermo TSQ Quantum Ultra triple quadrupole instrument (Thermo Fisher Scientific, Waltham, MA) with electrospray ionization. The calibration curve was measured once before each of the samples of biological replicates of a carbon source shift experiment, and once after (technical replicate).

3.2.1.7 Raw data to signal intensities

The .raw files were converted using in-house software on a server, resulting in .xml files being returned. These files are stored on an internal database accessible through <https://baobab.ethz.ch/>. Peak integration was performed manually with the in-house software “msSoftware”. ^{13}C -traces and metabolite concentrations were determined as previously described [13].

3.2.1.8 Background signal

A background signal obtained from deionized water measurements was subtracted from the raw intensities measured in the samples, for both ^{12}C and ^{13}C signals. Values below zero were removed.

3.2.1.9 Calibration curve

Calibration curves were analyzed by fitting a linear model to the log-converted intensities and concentrations. Covariances were estimated for the sake of error propagation in the inference of concentrations in the samples. Both ^{12}C and $^{12}\text{C}/^{13}\text{C}$ normalized calibration models were generated.

3.2.1.10 Sample analysis

Sample analysis started with a correction for biomass by normalizing for the inferred OD600 at sampling time by using a linear model fit to log-transformed OD600 data for each of the biological replicates. Relative changes could be directly calculated from these data, which was done by z -scoring the OD600-corrected signal intensities of the samples. For absolute metabolite quantification, the sample concentrations were inferred using the calibration models. In order to go from sample concentration to intracellular concentration, we used a conversion factor obtained by multiplying the following three elements:

1. The resuspension over sample volume (120 μL per 2mL sample)
2. An OD600 to gram cell dry weight (gCDW) factor of $0.433 \pm 0.023 \text{ gCDW OD600}^{-1} \text{ L}^{-1}$
3. a gram cell dry weight to intracellular volume factor of 2.3mL gCDW^{-1} .

The latter two conversion factors were previously determined [37]. A detailed description on the determination of intracellular metabolite concentrations can also be found in published work [9]. The inferred intracellular metabolite concentrations are in units of mM.

3.2.2 *In silico* experimentation

3.2.2.1 Annotation data

In this work we have adopted the use of BiGG database standard as primary identifiers for compounds and reactions. We have additionally annotated the compounds using a multitude of other well-established database identifiers (See Supplementary Tables ?? and ??). We have opted for BiGG identifiers as these are unique, whereas for the other databases often more than one entry exists. Moreover, they are easily human readable. For metabolites these annotations include: BiGG [105], CHEBI [23], KEGG [57], Biocyc [15], SEED [90], HMDB [127], MetaNetX [36], Reactome [19] and LipidMaps [33]. For reactions these include: BiGG [105], EC-code [4], KEGG [57], Biocyc [15], MetaNetX [36], Reactome [19] and RHEA [3]. Annotation of these identifiers was obtained programmatically, validated to be compliant with the Minimal Information Requested In the Annotation of biochemical Models (MIRIAM) standard [68]. To access the tools and scripts used for this annotation, please refer to the GitLab repository https://gitlab.ethz.ch/karrenbelt/sbml_tools.

3.2.2.2 Initial condition data

All sources used concern studies on *E. coli* strain BW25113 during mid-exponential growth.

Table 3.1: Condition-dependent cell volume of *E. coli*. The units are: h^{-1} growth rate, μm cell length, μm cell width, fL single cell volume, 10^8 cells $\text{mL}^{-1}\text{OD}^{-1}$ OD-specific cell concentration, and $\mu\text{L mL}^{-1}\text{OD}^{-1}$ OD-specific total cell volume. Adapted from [121].

growth condition	growth rate	cell length	cell width	single cell volume	OD-specific cell concentration	OD-specific total cell volume
acetate	0.29±0.02	2.3±0.6	1.2±0.1	2.4±1.3	16.8±1.7	4
fumarate	0.47±0.03	2.4±0.6	1.1±0.1	2.4±1.2	17.0±1.7	4.1
galactose	0.17±0.02	2.0±0.5	1.1±0.1	1.9±1.2	19.9±2.0	3.8
glucose	0.60±0.05	3.0±0.7	1.4±0.2	3.2±1.2	11.1±1.1	3.6
glucosamine	0.39±0.03	2.7±0.7	1.3±0.1	2.9±1.3	12.2±1.2	3.5
glycerol	0.47±0.03	2.3±0.6	1.2±0.1	2.3±1.3	19.6±2.0	4.5
pyruvate	0.40±0.03	2.2±0.6	1.0±0.1	2.1±1.2	21.0±2.1	4.5
succinate	0.49±0.02	2.4±0.6	1.1±0.2	2.4±1.3	16.7±1.7	4.1

Addressing the cell volume data is crucial, as its substantial error propagates significantly into enzyme concentration estimates. It’s important to note that the absence of raw data and the provision of summary statistics in linear scale introduce sampling bias. This bias arises from the challenge of converting arithmetic mean and standard deviation into geometric terms, which is impossible without the raw data. Consequently, when sampling from a normal distribution, there is a tendency to draw unrealistically small cell volumes. This, in turn, leads to the generation of unrealistically large enzyme concentrations, impacting the overall accuracy of our model. As the most quantities in biology are lognormally distributed, we decided to use such a distribution for sampling of the cell size:

$$f(x|\mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} \quad (3.1)$$

where μ is the geometric mean, σ the geometric standard deviation, and x the random variable. We approximated the geometric mean and standard deviation from the arithmetic data as $\mu_{\log} = \ln(\mu)$ and $\sigma_{\log} = \frac{\sigma}{\mu}$. As this approximation suffers from potential outliers during sampling, we used a truncated lognormal distribution:

$$f(x|\mu, \sigma, a, b) = \frac{1}{x\sigma\sqrt{2\pi}(\Phi(z_b) - \Phi(z_a))} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} \quad (3.2)$$

where $\Phi(z)$ is the cumulative distribution function, $z_a = \frac{\ln(a)-\mu}{\sigma}$ and $z_b = \frac{\ln(b)-\mu}{\sigma}$, and a and b denote the lower and upper bound, respectively. The smallest and largest bacterial cells have a diameter of 0.2 μm and 750 μm , respectively [108]. Assuming a perfect spherical shape, the formula $V = \frac{4}{3}\pi r^3$ was used to obtain a lower and upper bound of 3.35×10^{-17} and 1.77×10^{-6} , respectively.

Enzyme concentrations were obtained from the absolute protein quantification study by Schmidt *et al.* [106]. We have aggregated these using the gene-reaction-rules present in the model (See Supplementary Table A.4). These counts were aggregated on a per-reaction basis using these boolean rules: **and**-logic indicates both proteins are needed, hence we took the minimum, whereas **or**-logic indicates both can be used, and hence summation was performed, in order to approximate to total number of enzymes that could catalyze a given reaction (See Supplementary Table A.11). It’s important to acknowledge that considering the number of subunits is a significant aspect that should be considered. However, due to time constraints and the complexity of assessing its significance, we have regrettably not been able to address this in our current work. We note that enzyme concentrations were assumed to be constant throughout the duration of our experiment, hence also served merely as scaling factors in our simulations. Enzyme counts were also sampled using the truncated lognormal distribution, and subsequently converted to cellular concentration using concentration (M) = $\frac{\text{enzyme counts}}{\text{cell volume (L)} \times \text{Avogadro's Number}}$.

Metabolite concentration data was aggregated from the works of Gerosa *et al.* and Kochanowski *et al.* [37, 64]. We averaged these, relying on the uncertainties package [71] for error propagation (See Supplementary Table A.12).

Flux data were obtained from Gerosa *et al.* [37] (See Supplementary Table A.13).

3.2.2.3 Parameter data

Kinetic parameter data were retrieved from the BRENDA database [107]. Only data from wildtype *E. coli* were used, as kinetic properties of mutants often deviate significantly. For the derivation of priors, we computed geometric means and standard deviations from the available data for kinetic parameters, including forward catalytic rate constants (k_{catf}) and Michaelis-Menten constants

(K_m) (See Supplementary Tables A.15 and A.16). In cases where specific parameter values or their geometric standard deviations were missing, we supplemented the data using a generic prior distribution (See Supplementary Table A.10). The kinetic parameters are also sampled from a truncated lognormal distribution.

The Gibbs free energy (ΔG°) values used in this work were obtained along with their associated covariance matrix. These values were acquired using Equilibrator [34], where we assumed a constant temperature of 37°C, an intracellular pH of 7.4, and an ionic strength of 0.25M [131, 125] (See Supplementary Table A.14). We sampled the Gibbs free energy values and their associated covariance matrix from a multivariate normal distribution. The distribution was parameterized by the mean vector, denoted as $\boldsymbol{\mu}$, and the covariance matrix, denoted as $\boldsymbol{\Sigma}$. The sampling process followed the multivariate normal distribution equation:

$$f(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^k \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad (3.3)$$

where \mathbf{x} represents a vector of Gibbs free energy values, $\boldsymbol{\mu}$ is the mean vector, $\boldsymbol{\Sigma}$ is the covariance matrix, and k is the dimensionality of the vector \mathbf{x} . This sampling process allowed us to capture the uncertainty and correlations among Gibbs free energy values when estimating kinetic parameters and performing subsequent analyses. After obtaining the sample of Gibbs free energy values and their associated covariance matrix, we converted these values into equilibrium constants using the relation $\Delta G^\circ = -RT \ln(K)$, where R is the gas constant (8.314 J mol⁻¹ K⁻¹), T is the absolute temperature, and K is the equilibrium constant.

3.2.2.4 Time series data

The time series data that was obtained as described in the *in vivo* experiments section (see Figures A.20, A.21 and A.22), published time series data from [75] on the pyruvate to glucose shift was used (see Figure A.19).

3.2.3 Model simulation

In order to assess the functional relevance of allosteric interactions we construct an ensemble of models that consists of a base model without regulation, models containing a single regulatory interaction, and a set of models containing all combinations of pairwise regulatory interactions. Next, we initialize each model at pseudo steady-state, run the simulations, and estimate the prediction error based on a trade-off between goodness-of-fit and model complexity that is reflective of model quality. Finally, we aggregate the results from our modeling procedure to derive a score for each of the individual regulatory interactions based on the prediction error and the frequency of a the interaction occurring in models that improve upon the base model.

3.2.3.1 Systems demarcation

Our first aim was to collect the available data from previous studies in order to determine how to demarcate our subsystem of interest from the rest of metabolism and other cellular subsystems in both space and time. Systems demarcation is a nontrivial process and somewhat of an art that requires knowledge of both the system and data at hand, as model complexity needs to be balanced in accordance with data availability and the expected signal-to-noise ratio on data yet-to-be collected. Hence, systems demarcation is performed by means of heuristic reasoning rather than through the use of an algorithmic procedure and associated are implicit assumptions that often escape our attention. It is a part of the modeling process that most often omitted, yet it should not be and hence is explicated upon here. Decisive considerations include the following:

3.2.3.2 The modeling framework

From a menagerie of modeling frameworks, including finite-state machines, cellular automata, Petri nets, boolean network, constraint-based, stochastic and ordinary, partial and delay differential equations, our first task consists of choosing a suitable candidate. Each of these frameworks has specific advantages and technical foibles, discussion of which deserves its own treatise [5, 39, 24], essential however is to select a paradigm that is suitable for addressing the research question at hand. The research question we pose concerns the role of allosteric interactions in the TCA

cycle during metabolic adaptation. One way to approach this question is through steady-state analysis, however, previous such analyses have indicated that the majority of flux changes remained unexplained [37]. A limitation of this approach is that dynamic behaviour cannot be simulated. As the process of metabolic adaptation implies change, we opt for a differential equation framework to study how metabolite-protein interactions shape the response of the TCA cycle intermediates upon a perturbation. In order to be able to describe dynamics, however, we will need to introduce parametrized rate laws, which significantly increases the complexity of the model. There is a trade-off to be made; one where we sacrifice scale for the ability to simulate dynamic behaviour of the network.

Furthermore, we consider stochastic effects to be negligible for the following reasons. First off, since metabolic changes can be observed on a timescale ranging from milliseconds to seconds, whereas changes in enzyme abundance take at least a minute, we chose to focus on solely the first minute of the response after a perturbation. Such time-scale separation allows us to assume enzyme levels remain constant. Secondly, we note that cellular metabolite concentrations are generally much higher than protein concentrations [9, 76, 106]. Given that the cellular metabolite concentrations of interest to us are expected to be in the $1\mu\text{M}$ to 10mM range [9, 75, 37] we assume stochastic effects to be negligible even on a single cell level. Regardless, single cell metabolomics is not feasible with contemporary technology, hence our cellular metabolite concentration estimates are derived from large cell populations in the range of 10^8 to 10^{10} cells. The law of large numbers dictates that regression to the mean will occur, which is to say that stochastic effects at the single cell level average out. We specifically chose ordinary differential equations, and hence neglect spatial gradients and implicitly assume that our system is well-mixed. This entails that we are not able to model phenomena such as the diffusion of gasses [45, 27], membrane transport [124, 67], or the effects of molecular crowding [30, 89] accurately. Although we cannot accurately model membrane transport dynamics, this phenomena is not of primary interest to us. We furthermore note that, unlike eukaryotes, prokaryotes such as *E. coli* do not possess mitochondria that separate TCA cycle metabolism from the rest of the cell, lending additional support for our choice. Nonetheless, it would be a stretch to think that cells are well-mixed bags of chemicals without gradients. In fact we know they're not, as there exist intracellular pH gradients [120], protein gradients [60], and affinity gradients that can localize metals and minerals through sequestration [29, 6] and compartmentalize reactions to metabolosomes to reduce toxicity [116, 114]. Even in an initially well-mixed (bio)chemical reaction-diffusion systems spatial gradients can emerge spontaneously, which has been observed in cells [111] and cell-free systems [7, 130]. We do not include a description of spatial gradients simply because it requires far more knowledge on spatial organization in the cell than we currently possess, and greatly increases model complexity.

3.2.3.3 Existing metabolic pathway definitions

Our TCA cycle model should minimally encompass the entirety of metabolites and reactions commonly considered to be part of the TCA cycle [123]. It also includes closely associated pathways such as the glyoxylate shunt [66] and anaplerotic reactions [65], as these are closely associated and likely play a significant role in the flux adaptation process. Finally, one must also consider other reactions in which metabolites are known to be consumed or produced [61]; specifically co-factors such as ATP and NAD might play an important role, however these are known to be involved in many other reactions.

We decided to use the *E. coli* genome scale metabolic model [88] (see Supplementary Figure A.8) and core model [87] A.9 as starting templates, from which we demarcated the subsystem depicted in Figure 3.1.

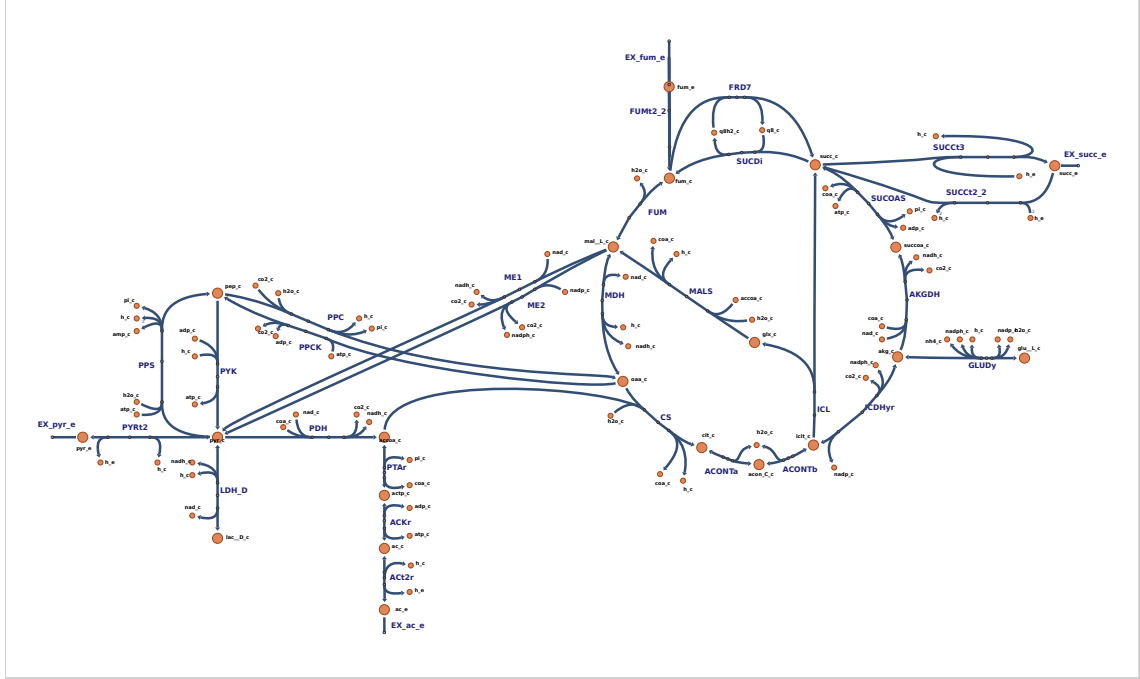


Figure 3.1: Tricarboxylic acid cycle of *E. coli*. Model boundary species are extracellular: pyruvate (PEP), acetate (Acetate), succinate (Succinate), phosphoenolpyruvate (PEP), adenosine triphosphate (ATP), adenosine diphosphate (ADP), nicotinamide adenine dinucleotide (NAD), reduced nicotinamide adenine dinucleotide (NADH), nicotinamide adenine dinucleotide phosphate (NADP), reduced nicotinamide adenine dinucleotide phosphate (NADPH), L-glutamate (Glu), and D-lactate (Lac). The graphic is created with the use of Escher [62]. Note: Protons (H+) and water (H₂O) are ignored in reaction kinetics.

3.2.3.4 Model construction and simulation

The models were constructed using the Systems Biology Markup Language (SBML) [48] and a custom pipeline written in python v3.6. Steady-state flux analysis was performed using the cobra toolbox [46]. To constrain the space for sampling of the initial metabolite concentrations, we derived linear inequality constraints from the equilibrium constants and measured metabolite concentrations, and computed the primal in order to assess whether a feasible solution existed [85]. For hit-and-run sampling we used an implementation of the algorithm available here: <https://github.com/fontclos/hitandrunk>. To simulate system dynamics, we use a generalized reversible modular rate laws to describe the reactions kinetics [72, 73]. This modular rate law is defined as follows:

$$v = E_r R_{reg} \frac{T}{D + D_{reg}} \quad (3.4)$$

where v denotes metabolic flux, E_r the enzyme concentration, T the thermodynamic term, D the denominator, and R_{reg} and D_{reg} , regulatory terms. The thermodynamic term T can be expressed in the form of simple reversible mass-action kinetics

$$T = k_f \prod \alpha_i^{n_i h_i} - k_r \prod \pi_i^{n_i h_i} \quad (3.5)$$

where k_f and k_r are forward and reverse catalytic rate constants, α and π are relative concentrations for the substrate and product respectively ($\alpha = \frac{[S]}{K_m[S]}$ and $\pi = \frac{[P]}{K_m[P]}$), n_i the stoichiometry of species and h_i the degree of cooperativity. In this work we made the simplifying assumption that cooperativity is insignificant and can be neglected for several key reasons:

- Many other reactions in the TCA cycle involve relatively simple substrate-product conversions, where the mechanisms are well understood and cooperativity effects are less likely to play a significant role. We note that there have been studies suggesting cooperativity in isocitrate dehydrogenase (ICDHr) in the yeast *S. cerevisiae* [92, 74]. Furthermore, there is evidence of negative cooperativity in the isomerization of isocitrate catalyzed by aconitases in *E. coli* [118].

- Our modeling approach prioritizes clarity and ease of interpretation, avoiding unnecessary complexity introduced by cooperativity-related parameters and equations.
- Due to limited available data and knowledge about specific enzyme cooperativity in this context, introducing such effects could lead to unreliable results.
- Lastly, our model’s primary aim is to capture essential behaviors and trends of the TCA cycle, and neglecting cooperativity aligns with this focus.

Using the Briggs-Haldane relationship [12]

$$K_{eq} = \frac{V_f K_P}{V_r K_S} \quad (3.6)$$

where V_f and V_r are the maximal forward and reverse rates, S and P denote substrate and product concentrations, we can eliminate one of the catalytic rate constants by introducing the equilibrium constant and expressing the thermodynamic term using the Haldane form

$$T = \prod \alpha_i^{n_i} \left(1 - \frac{\Gamma}{K_{eq}}\right) \quad (3.7)$$

where Γ denotes the mass-action ratio ($\Gamma = \frac{\prod_i \alpha_i^{n_i}}{\prod_i \pi_i^{n_i}}$) and K_{eq} the equilibrium constant. The denominator term D comes in a variety of different flavors [73]. In this work we opted for the use of convenience kinetics [72]

$$D = \prod_i (1 + \alpha_i)^{n_i} + \prod_i (1 + \pi_i)^{n_i} - 1 \quad (3.8)$$

which is also known as the common modular form and is equivalent to the reversible Michaelis-Menten equation [104]. Allosteric regulation in the generalized rate law is described using R_{reg} , which assumes independent regulator binding, which is to say that regulators can bind to all enzymes states.

$$D_{reg} = \prod (\rho_A + [1 - \rho_A] \frac{\mu}{1 + \mu})^\omega \prod (\rho_I + [1 - \rho_I] \frac{1}{1 + \mu})^\omega \quad (3.9)$$

where ρ_A and ρ_I are the relative basal rates, varying from 0 to 1, for the activator and inhibitor, μ is expressed as the ratio of the concentration of the allosteric effector to the corresponding dissociation constant, and ω a matrix containing the regulation numbers. The ρ values determine whether there is partial or complete activation or inhibition. In this work we assumed ρ to be zero and explore only complete allosteric regulation. This choice parallels our simplification approach for cooperativity and is motivated by a commitment to model clarity, data limitations, and the desire to capture primary behaviors. Hence this term was reduced to

$$D_{reg} = \prod \left(\frac{\mu}{1 + \mu}\right)^\omega \prod \left(\frac{1}{1 + \mu}\right)^\omega \quad (3.10)$$

Finally, the R_{reg} term represents non-allosteric regulation, such as competitive inhibition, a mechanism left unexplored in this work [73].

Symbolic solutions for equations, such as for the solving for catalytic constants, were performed using sympy [79] and JiTCODE [8]. Simulation of the differential equation model was performed using libRoadRunner [112]. Numba [70] was used to speed up the remaining numerical computation where necessary and possible. The prediction errors were computed as the sum of squares of the log residuals.

$$E[y] = \sum_{i=1}^n (\ln(\hat{y}_i) - \ln(y_i))^2 \quad (3.11)$$

where E denotes the error over the metabolites, ranging from i to n , \hat{y} the prediction and y the data. Measurement uncertainties were not explicitly incorporated in the evaluation of goodness-of-fit. Instead, we relied on the log residuals of means as our assessment metric. To be able to compare the information content of models with additional parameters we used the Akaike information criterion (AIC) [2].

$$AIC = 2k - 2 \ln(\hat{L}) \quad (3.12)$$

where k denotes the number of parameters in the model and \hat{L} the maximum likelihood estimate of the model. Note that the least squares estimation is equivalent to maximum likelihood estimation under the assumption of Gaussian errors [122]. The AIC reflects the relative loss of information when assuming a particular model represents the process that generated the information, hence a lower score indicates less information loss and a higher quality model. We use an extension the AIC that corrects for small sample sizes (AICc)

$$AICc = AIC + \frac{2k^2 + 2k}{n - k - 1} \quad (3.13)$$

where n denotes the sample size.

3.3 Results

Earlier studies on flux regulation in central carbon metabolism have proposed that non-covalent metabolite interactions play a prominent role [32, 37, 41, 80]. Previous research, most notably on the part of Link *et al.* and Christodoulou *et al.* [75, 17], describes the use of an ensemble modeling approach to determine which allosteric metabolite-protein interactions allow the model to fit the data better. In the present study, we employ this approach to investigate the significance of allosteric noncovalent metabolite-protein interactions in governing metabolic flux adaptation within the TCA cycle. Our goal is to assess whether these allosteric interactions contribute to the dynamic regulatory behavior observed during metabolic transitions. To achieve this, we first evaluate the performance of a model without regulatory mechanisms in describing the measured metabolite dynamics. Subsequently, we create an ensemble of topologically distinct models, created by augmenting the base model with either single or pairwise allosteric interactions. The ensemble modeling approach may be regarded as an unbiased estimator of the importance of allosteric regulatory interactions, in that we do not include any of the previously reported allosteric interactions. In order to assess the importance of individual allosteric interactions, we aggregate the results across models and simulations in order to compute a score for each of them, representing their contribution to reducing the error of the fits. These models can be regarded to act as a set of "weak learners" [35], each with its own contribution to predictive insight. By focusing on the predictive capacity of the ensemble as a whole, we aim to decipher the relative relevance of regulatory interactions without getting lost in the details of individual fits.

3.3.1 Experimental data

In this study, we extend the ensemble modeling approach pioneered by Link *et al.* and Christodoulou *et al.* [75, 17] to unravel the significance of allosteric noncovalent metabolite-protein interactions in governing metabolic flux adaptation within the TCA cycle. By incorporating single and pairwise allosteric interactions into the model, we aim to discern their impact on the model's ability to recapitulate experimental time series observations upon a perturbation. Since allosteric regulation involves direct protein-metabolite interactions, the effect of a change in concentration of an effector almost instantaneously propagates to the target enzyme. The effect of transcriptional regulation on metabolism, on the other hand, takes at least a minute to set in as it involves a cascade of processes that ultimately effect enzyme levels [110]. In order to capture the effects of allosteric regulators, as well as to be able to exclude the effects of transcriptional regulation, we focused on dynamics observed within the first minute after such a perturbation.

We selected perturbations involving shifts in carbon sources, specifically transitioning from a medium with a less favored carbon source to one containing glucose, which serves as *E. coli*'s preferred carbon source. "Preferred carbon source" in this context refers to a carbon source that is consumed as a priority, as indicated by previous research [1]. Notably, when transitioning to this preferred carbon source, *E. coli* has been observed to exhibit instantaneous glucose metabolism [75].

The choice of initial conditions regarding carbon sources was guided by the documented steady-state fluxes observed in *E. coli* during mid-exponential growth when using these sources as the sole carbon source [37]. Our selection of succinate and acetate as initial conditions is grounded in the remarkable contrasts in steady-state flux distribution as compared to glucose. More specifically, there are stark differences anaerobic fluxes between growth on succinate and glucose, while on acetate the glyoxylate shunt is active, a pathway that remains quiescent during growth on glucose [37]. Hence, cells need to facilitate these changes as part of their adaptive process when transitioning between these carbon sources. Consequently, cells are compelled to facilitate these metabolic adjustments as part of their adaptive response during the transition between these diverse carbon sources. Pyruvate was additionally considered in this analysis, as its role in the metabolic transition from pyruvate to glucose had been investigated previously during the switch from gluconeogenic to glycolytic modes. Not only will this serve as a control, significant changes in TCA cycle intermediates were also observed during this transition.

We cultivated *E. coli* to mid-exponential phase on a carbon source of secondary preference: either acetate, succinate or pyruvate. We then placed samples of this culture on a filter on which we perfused the cells for 10 seconds with medium containing the original carbon source of secondary preference, then for 30 seconds with medium containing glucose, and finally another 30 seconds with medium containing the original carbon source. During this time, and at specific time intervals

during the perturbation, filters with cells were quenched and extracted for metabolite concentration quantification via mass spectrometry (see section 3.2.1.3).

Our attempts to estimate metabolite concentrations from mass spectrometry data encountered significant challenges. Many crucial metabolites were difficult to reliably detect, leading to gaps in their time series. Moreover, we observed considerable variance in the estimated metabolite concentrations across both technical and biological replicates. The best results obtained after numerous iterations are presented in the appendix of this thesis (See Figure A.20, A.21, and A.22).

In our attempt to replicate the pyruvate-glucose-pyruvate switch experiment conducted by Link *et al.*, we observed that, among the 14 metabolites present in our model that Link *et al.* could accurately quantify, we encountered significant difficulty in reliably detecting signals for 10 of these metabolites. To elaborate, when we mention "reliably," we are referring to a situation in which, out of the three replicates, each encompassing 11 time points, we encountered the unfortunate outcome of losing signals for more than half of these time points. Specifically, PEP shows an initial drop upon the switch to glucose, whereafter its concentration quickly rises back to the initial state level, and drops again after the switch back to glucose, which is qualitatively similar to what Link *et al.* observed. The observed dynamics of FBP, though not present in our model, are characterized by a significant spike upon the switch to glucose, and an equally significant drop upon switching back to glucose. Looking into the other metabolites that are present in our model and detected by Link *et al.*, we find no further similarity. More precisely, for the metabolites citrate, isocitrate, alpha-ketoglutarate, succinyl-CoA, succinate, fumarate, malate and oxaloacetate, all of which show coefficients of variation under 10% across the entire time series in Link *et al.*, we do not detect a signal in more than half of the samples, with the exception of malate. Regarding co-factors, AMP, ADP and ATP, as well as NADH and NADPH, were consistently detected, but exhibited coefficients of variation greater than 100%.

Upon assessing data dissimilarity, our dataset revealed an average coefficient of variation (CV) of 39% among the top 25% (15 signals) of metabolites, markedly surpassing Link *et al.*'s benchmark of 12% across their entire dataset of 57 metabolites. This substantial discrepancy highlights the challenge in achieving data consistency essential for elucidating the intricate dynamics of metabolic flux adaptation. The high CV in our dataset impedes our ability to distinguish between various model fits and assess the relevance of allosteric interactions during metabolic transitions based on aggregated results. Consequently, we relied on previously published data concerning a pyruvate-glucose-pyruvate shift [75], where dynamics of TCA cycle intermediates were recorded but not utilized (see Figure A.19).

During discussions with a senior scientist regarding our data, concerns were raised about its reliability. Specifically, the senior scientist highlighted the high variance observed and suggested caution regarding its interpretation. This input was critical as it emphasized the challenges we faced in obtaining consistent and reliable data, corroborating our own observations.

The empirical validation in our field presents inherent challenges due to the weakness in predictions and the absence of a definitive ground truth. This limitation significantly impacts model selection and directly influences the aims of our study. The inability to establish a reliable ground truth contributes to the difficulty in interpreting and validating empirical data, particularly concerning the intricate dynamics of metabolic flux adaptation. Consequently, assessing whether data quality is sufficient becomes an arduous task without a definitive benchmark or ground truth for comparison.

The dataset limitations, such as undetectable signals for key metabolites and high coefficients of variation for consistently detected co-factors, significantly compromised the reliability of our dataset, aligning with the discussion in [77, 10] regarding the impact of data uncertainties on *in silico* studies. More specifically, limited availability of accurate metabolite concentration data poses a significant hurdle in discriminating between the fits of topologically different models, and as a consequence our ability to differentiate between the relevance of allosteric interactions during metabolic transitions derived from aggregating these results.

Moreover, adhering to the GIGO Principle (Garbage In, Garbage Out), the quality of output is determined by the quality of the input data. It emphasizes the critical role of data quality in computational processes, highlighting that using unreliable or inconsistent data will yield unreliable or inconsistent results. Therefore, based on the high variance, lack of detection for crucial metabolites, and the caution raised by the senior scientist, we made a decision not to move forward using this dataset. Therefore, we resorted to the use of data previously published on a pyruvate-

glucose-pyruvate shift [75], in which dynamics of TCA cycle intermediates had been recorded but were not used (see Figure A.19).

3.3.2 Computational modeling

3.3.2.1 Integration of heterogenous data

Our first challenge was to find a way to integrate data from different sources for model initialization. Data on initial conditions of metabolite concentrations, metabolic fluxes, enzyme counts and cell volume estimates were collected from recent publications [121, 75, 37, 106] (Supplementary Tables A.12, A.11 and A.13). Kinetic parameter data was obtained from the BRENDA database and equilibrium constants obtained with the use of a biochemical thermodynamics calculator named eQuilibrator [107, 85] (Supplementary Tables A.15, A.16 and A.14). An overview of coverage of our model by the data is provided in Table 3.2, the specifics are provided in section 3.2.2.

Table 3.2: Model data coverage

	Data	Total	Fraction
States			
cell volume	1	1	1.00
metabolites	14	30	0.47
enzymes	22	23	0.97
fluxes	14	23	0.61
Parameters			
K_{eq}	23	23	1.00
k_{cat_f}	8	23	0.35
k_{cat_r}	2	23	0.09
K_m	50	92	0.54

Among the states in our biological system we may differentiate cell volume, metabolites, enzymes and fluxes. Cell volume data is necessary since enzyme quantification data consists of counts per cell, necessitating conversion into concentrations for use in our rate law equations. With respect to enzymes we have data on 22 of the reactions out of the 23 in our system, with PPCK being the notable exception, whereas steady-state fluxes are available for 14 out of 23 reactions, and with metabolites covering 14 out of 30 species in the model.

With respect to parameters, we differentiate thermodynamic and enzyme kinetic parameters. Data on thermodynamic parameters, specifically the Gibbs free energies of formation, are available for all reaction in the model (see Figure A.14). We found data for 10 catalytic constants, which represents approximately a quarter of the catalytic constants present in the model and for seven out of these only a single value is reported in literature (see Figure A.12). With respect to the Michaelis-Menten constants we found data for 50 of them, which represents more than half of the dissociation constants in the model (see Figure A.13). When no specific data was available, we opted to use a generic prior that was derived from the entire collection of available data of different studies reported in databases [76] (Supplementary Table A.10).

The overview reveals a stark contrast: while all the equilibrium constants are known, most of the kinetic parameters remain unknown. This discrepancy is further compounded by the consideration of estimate uncertainty and the heterogeneity of data sources. Specifically, the equilibrium constants are jointly estimated for the entire network. The availability of data from which these are derived, and the fact that thermodynamic parameters don't vary between *in vivo* and *in vitro* conditions makes them a valuable source of information.

In contrast, the situation with kinetic parameters is more complex. These parameters, which underlie enzyme kinetics, are determined independently through biochemical enzyme assays. These assays have been conducted in independent studies, diverse geographical locations, and over the course of several decades. Moreover, the experimental conditions used for these assays vary widely, as optimal conditions, such as temperature, pH and the presence of cofactors, can vary considerably. Although there appears to be a tendency to move towards *in vitro* conditions that more closely mimic those found *in vivo*, achieving a faithful emulation of *in vivo* conditions remains challenging.

Hence, it is questionable whether test tube conditions are representative of behavior in the cell under physiological conditions, a source of uncertainty that affects enzyme kinetic parameters but not chemical equilibrium constants [20, 78]. Therefore, we opted to formulate the rate law in Haldane form, effectively exchanging an unknown catalytic constant for a known thermodynamic constant (see equations 3.5, 3.6, 3.7).

Previously reported metabolite-protein interactions are also available. However, we did not include these during model development because of two main reasons. First, including them will greatly increase model complexity. Specifically, with respect to the enzymatic reactions in our model of the *E. coli* TCA cycle and anaplerotic reactions, a total of 46 inhibitory and 13 activatory allosteric interactions have been reported [100]. Including all of these would significantly increase the number of parameters in our model. Moreover, not only is it likely that this set contains false positives, for example due to the use of non-physiological conditions, not all of them may be relevant in the conditions we're testing. Secondly, by not including this information *a priori* we reserve this data for validation *post hoc*.

3.3.2.2 Model initialization procedure

Here, we provide a high-level overview of the steps involved in generating samples and initializing the model:

Algorithm 1 Model Initialization

Equilibrium constants:

- Sample $\Delta rG'^{\circ}$ from a multivariate normal distribution
- Convert $\Delta rG'^{\circ}$ to equilibrium constants K_{eq}

Fluxes:

- Generate a flux solution using constrained-based optimization
- Introduce parameter to address futile cycling

Metabolites:

- Use equilibrium constants and known concentrations as constraints
- Sample metabolites using constrained-based hit-and-run sampling

Cell volume:

- Sample cell volume from a truncated lognormal distribution

Enzymes:

- Sample enzymes with data from a truncated lognormal distribution
- Use cell volume to calculate enzyme concentrations

Kinetic constants:

- Sample forward catalytic rate constants from a truncated lognormal distribution
- Sample Michaelis-Menten constants from a truncated lognormal distribution

Consistency adjustment:

- Adjust enzyme concentrations to match the steady-state flux solution
-

Each step in the initialization process contributes to building a comprehensive starting point for our computational model.

Equilibrium constants: We estimate the equilibrium constants for our system, which encompass the Gibbs free energy of formation ($\Delta fG'^{\circ}$) each of the metabolites and the Gibbs free energy of reactions ($\Delta rG'^{\circ}$) describing the transformation of one set of metabolites into another. As these Gibbs free energies are calculated for the network as a whole, we explored the resulting covariance matrix using a multivariate sampling approach (Section 3.2.2.3). The multivariate normal sampling of $\Delta rG'^{\circ}$ followed by conversion to equilibrium constants K_{eq} provides the foundation for thermodynamic considerations, parameter which we will not only use in the rate law equations, but also in conjunction with the flux estimate in the process of sampling initial metabolite concentrations, as we detail below.

Fluxes: To initialize the model at a predefined steady-state, we first obtained a steady-state flux solution for the model using constraint-based analysis approach. Specifically, our goal was to match the fluxes reported in a previous publication that compared the steady-state growth of *E. coli* on various carbon sources [37]. However, due to structural dissimilarities between the model used

in that study and ours, direct mapping of fluxes was not feasible. To address this limitation, we employed a constraint-based modeling approach known as Minimization of Metabolic Adjustment (MOMA) [109]. MOMA is commonly used to predict flux distributions in knock-out mutants by minimizing the distance to fluxes observed in the wild type. In our case, we utilized MOMA to derive a flux solution that closely resembles the flux distribution in the published data.

Although some differences exist between the reported fluxes and the solution we obtained, such as for phosphoenolpyruvate carboxykinase (PPCK) (7%), α -ketoglutarate dehydrogenase (AKGDH) (1%), and the exchange reactions involving pyruvate (9%), acetate (1%), and succinate (1%) (see Figure A.10 and Figure A.11), we regard these disparities as inconsequential for our modeling procedure. Given the inherent uncertainty in the original estimates, as well as the variation in other states and parameters within our model, we assumed that these relatively small deviations do not significantly impact our final model predictions.

Lastly, we introduced an additional parameter to address futile cycling at the PPS / PYK node. In the constraint-based model, both reactions are irreversible, and only one of them carries flux in the obtained steady-state flux solution. For growth on pyruvate, the flux-carrying reaction is PPS, whereas the flux through PYK is zero. However, the solution represents the net flux through both reactions. Since the extent of futile cycling remains uncertain, we chose to sample the ratio from a uniform range between 0.01 and 0.99. We then increased both fluxes by an amount equal to this ratio multiplied by the net flux to maintain the original steady-state conditions. This approach ensures that the preserved net flux captures the interplay between PPS and PYK under these conditions.

Initial metabolite concentrations We leverage the equilibrium constants within our system [34] and integrate physiological and steady-state flux data [37] to define the space of possible initial metabolite concentrations using linear inequality constraints (Section 3.2.2.3). For metabolites without available data, we employed an uninformative generic prior [76] (Supplementary Table A.10). To determine the thermodynamically feasible initial metabolite concentrations, we employ linear programming. The objective is to solve the following optimization problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T \mathbf{x} \\ \text{subject to} \quad & \mathbf{A}_{\text{ub}} \mathbf{x} \leq \mathbf{b}_{\text{ub}}, \\ & \mathbf{A}_{\text{eq}} \mathbf{x} = \mathbf{b}_{\text{eq}}, \\ & \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}, \end{aligned}$$

where:

- \mathbf{x} is the vector of metabolite concentrations being determined.
- \mathbf{c} is the objective vector, which we set to maximize the flux through a specific row named "mdf."
- \mathbf{A}_{ub} is the coefficient matrix for the upper bound constraints, which includes an additional "mdf" row.
- \mathbf{b}_{ub} is the right-hand side vector for the upper bound constraints, combining sampled equilibrium constants with metabolite concentration bounds.
- \mathbf{A}_{eq} is the coefficient matrix for equality constraints (not applicable here).
- \mathbf{b}_{eq} is the right-hand side vector for equality constraints (not applicable here).
- \mathbf{l} and \mathbf{u} are lower and upper bounds, respectively, for the concentrations of each metabolite.

\mathbf{A}_{ub} is a matrix embodying conditions to be satisfied, formed by multiplying the stoichiometric matrix (\mathbf{S}) with the flux directionality vector ($\text{sign}(\mathbf{v})$). \mathbf{b}_{ub} contains upper bounds, encapsulating log-transformed equilibrium constants, with a small value $\epsilon = 1.0^{-6}$ subtracted to ensure strict inequalities during optimization rather than inclusive ones during optimization. This practice is a standard procedure in optimization to address potential numerical inaccuracies, ensuring that constraints remain satisfied even in the presence of small rounding errors or other numerical intricacies. The bounds l and u were determined based on the available metabolite data and set

to encompass 3 standard deviations from the mean concentration (Supplementary Table A.12). This was achieved by calculating the relative standard error, which is the ratio of the standard deviation to the mean concentration. For metabolites without available data the upper and lower bound were set of 1.0^{-1} and 1.0^{-6} M. Prior to incorporating these bounds into the optimization process, they were also converted to the logarithmic scale, aligning with the logarithmic nature of the linear programming approach.

The optimization process is driven by the objective function, which involves specifying an objective vector denoted as c . The pivotal step in this linear programming formulation is the inclusion of the "mdf" row in the objective vector c . Within this vector, all elements are initially set to zero, except for the entry corresponding to the "mdf" row, which is assigned a value of 1. This strategic inclusion highlights the optimization's focus on maximizing the flux associated with the molecular driving force, a thermodynamic parameter critical in determining reaction directionality based on changes in free energy [85]. Additionally, uniform bounds of -1000 and $+1000$ are applied to the elements of the objective vector. It is worth noting that both the objective c and its associated bounds are arbitrary and devoid of specific physical units. Instead, they function as guiding parameters for the optimization process and are not intended to represent physical quantities.

With these matrices and vectors in place, the linear programming problem aims to find the concentration vector \mathbf{x} that minimizes the objective function while satisfying the defined inequalities. By including the "mdf" row in the objective vector \mathbf{c} and imposing upper bound constraints through \mathbf{A}_{ub} and \mathbf{b}_{ub} , we systematically determine initial metabolite concentrations that fulfill mass balance, thermodynamic, and driving force constraints. Solving this linear programming problem provides the vector x , representing the optimal initial metabolite concentrations that satisfy mass balance and thermodynamic constraints while maximizing the flux through the "mdf" row. This approach systematically identifies a feasible range of initial metabolite concentrations that align with available data and thermodynamic principles.

To systematically explore this constrained space, we utilized the hit-and-run sampling method [43] for uniform sampling. The resulting distribution of sampled metabolite concentrations, illustrated in Figure 3.2, serves as the foundation for initializing our model. The figure provides a visual representation of the sampled metabolite concentrations and their respective ranges. The grey area signifies the range of the generic prior for metabolites lacking data. In contrast, the black region represents the empirically determined actual sampling range. Metabolites with available data were further constrained based on three standard deviations from the mean, as reported by Link *et al.* [75]. Additionally, data from [37] and [63] are depicted in blue for reference. It's important to note that while Figure 3.2 offers a visual portrayal of the sampling ranges, the polytope's actual size is influenced by joint sampling of metabolites in high-dimensional space, a dimensionality that cannot be fully captured in a two-dimensional figure.

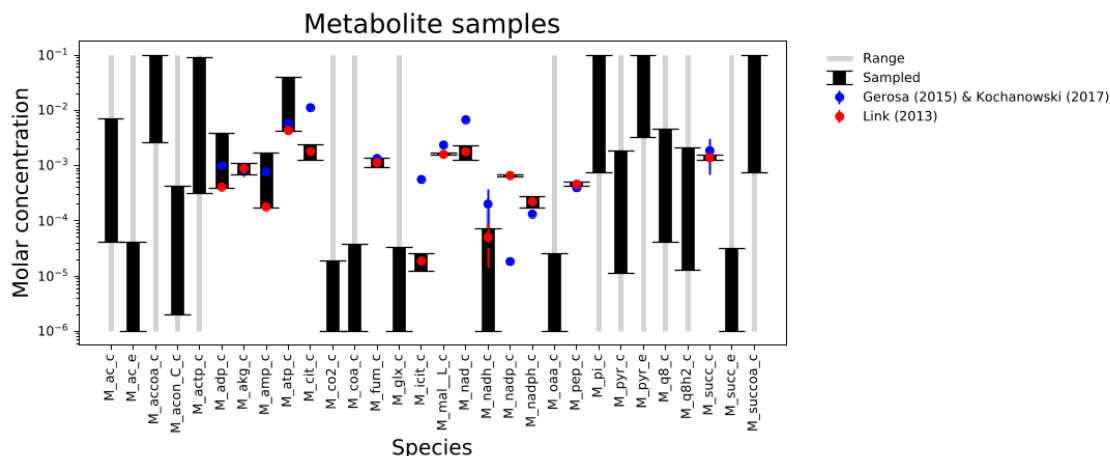


Figure 3.2: Sampled initial metabolite concentrations. The sampling range of the generic prior for metabolites without data is shown in grey (Range). The empirically determined actual sampling range is shown in black (Sampled). Metabolites for which there was data available was constrained to three standard deviations from the mean as reported by Link *et al.* [75]. For comparison the data obtained by Gerosa *et al.* and Kochanowski *et al.* [37, 63] is shown in blue.

The figure presents a visual representation of the sampled initial metabolite concentrations, shedding light on the consistency between the obtained samples and established experimental data. The black region, which represents the actual sampled range of metabolite concentrations, showcases the empirically determined sampling range as a result of constraints imposed by the combination of equilibrium constants and known concentrations. A discernible observation emerges from this figure: the disparity between our sampled concentrations and the mean and standard deviation values derived from the studies conducted by Gerosa *et al.* and Kochanowski *et al.*, symbolized by the blue data points [37, 63]. This discordance stems from a deliberate pivot towards integrating previously published time series data by Link *et al.* [75], due to unexpected challenges encountered in our own experimental time series data acquisition. The reason we opted to primarily use the data from Link *et al.* for model initialization, rather than combining data from various studies, is because these metabolic concentrations also serve as the initial data point for our time series data fitting. This decision was made to ensure consistency in the model’s initialization and subsequent analysis, as well as to align the model with the specific time-dependent behavior observed in Link *et al.*’s measurements.

Cell volume: To account for the variability in cell volume, we sample the cell volume from a truncated lognormal distribution. Cell volume sampling captures the inherent variability in cellular sizes, which is necessary for calculating enzyme concentrations from the absolute quantification data counts (Section 3.2.2.2).

Enzymes: To initialize enzyme concentrations, we utilize a truncated lognormal distribution, leveraging available data to constrain the sampling space, and subsequently compute the cellular concentration of these. Enzymes are constant in our model, effectively serving as a scaling factor, and due to error propagation of the cell volume have significant uncertainty associated.

Kinetic constants: The kinetic parameters of enzymes under consideration here include the forward catalytic rate constants (k_{cat}^f) and Michaelis-Menten constants (k_m). To initialize these kinetic constants, we sample from a truncated lognormal distribution using the available data (Supplementary Tables A.15 and A.16).

Consistency adjustment: In our model initialization procedure, a notable instance of redundancy arises as multiple variables interact to ensure the consistency of our calculations. Specifically, we address the relationship between the enzyme concentration (E), catalytic rate constant (k_{cat}^f), and the flux (v) through a reaction, given the known values of the other parameters in the generalized Michaelis-Menten equation (Equation 3.4). We are confronted with the requirement that the right-hand side of the equation must match the predefined flux (v). This dependence introduces a redundancy, as we are using both the sampled enzyme concentration and catalytic constant to compute the flux, while already having the target flux value. To reconcile this redundancy, we adjust the sampled enzyme concentration until the equation’s right-hand side aligns with the desired flux. The choice to adjust the enzyme concentration stems from a careful consideration of the intrinsic uncertainties inherent in enzyme quantification. Notably, the enzyme concentration estimation is significantly influenced by the uncertainty in cell size measurements. This uncertainty ripples through the estimation process, creating a considerable error in the computed enzyme concentration. Furthermore, our estimation of enzyme concentration is derived from gene reaction rules, which presents another layer of complexity. This method offers a rather coarse approximation, overlooking the stoichiometric coefficients of enzyme subunits, thereby limiting its accuracy and precision. Consequently, the decision to adjust the enzyme concentration offered a relatively simple and straightforward approach.

3.3.2.3 Experimental time series data as direct model input

In scenarios where we lack an explicit description of glycolysis, a pertinent question arises regarding how to describe extracellular perturbation events such as the switch to glucose. To tackle this, we chose to model the perturbation by assigning the dynamics of certain metabolites, namely phosphoenolpyruvate (PEP), lactate, and L-glutamate, to align with available experimental measurements. PEP is specifically difficult to model, as it is involved in glucose uptake via the glucose phosphotransferase system.

Furthermore, cofactors are also difficult to model, since they are involved in numerous other reactions, which thus affect their pools, though these are not part of the model. Nonetheless, cofactors likely play a crucial role in when it comes to the dynamics of the TCA cycle; not only are they essential cofactors of almost every of the individual reactions in the TCA cycle, they have also been reported to act as allosteric regulators. To illustrate, with respect to the enzymatic reactions in our model, no less than 13 allosteric interactions involving co-factors have been reported [100]. To capture their behavior during perturbation events, we integrated measurement data for cofactors—AMP, ADP, ATP, NAD^+ , NADH, NADP^+ , and NADPH—into the model as direct inputs. This approach accounts for cofactor fluctuations during metabolic transitions and their influence on the TCA cycle’s regulatory mechanisms.

The assignment of metabolite dynamics was achieved through linear interpolation and rule-based assignments using the SBML framework. This approach ensured that the model’s behavior accurately captures the experimental dynamics of these metabolites. Notably, for these assigned metabolites, we did not simulate their dynamics or explicitly account for uncertainty. Instead, we employed rule-based assignments to directly set the values of these metabolites based on provided mathematical expressions.

Finally, we have also chosen to utilize time series measurement data for metabolites to explore their potential role as allosteric regulators in our model’s enzymatic reactions. This approach allows us to assess the influence of metabolites, including those not explicitly accounted for in the system of differential equations. *In brevi*, dynamics of 9 out of 30 metabolites in the model were assigned instead of simulated, and all 57 metabolites from central carbon metabolism for which measurement data was available were tested as putative allosteric regulators (see Figure A.19).

3.3.2.4 Kinetic Parameter Optimization

Following the initialization of the model to its predefined steady state, the subsequent step involves evaluating the model's ability to replicate the observed system dynamics during a carbon source perturbation. To achieve this, we employ a constraint-based optimization approach. In this procedure, we optimize the Michaelis-Menten constants within the model. These constants are bounded by either their literature-reported values or, in the absence of specific data, by an uninformative prior distribution (Supplementary Tables A.16 and A.10).

Importantly, during each iteration of the optimization process, we ensure that the enzyme concentrations are adjusted to maintain the original and initial steady state. This preservation of the steady state serves as a critical constraint in our parameter optimization. Moreover, to maintain the physical feasibility of enzyme concentrations, if a computed enzyme concentration ever falls below zero during this procedure, our cost function returns infinity. This effectively constrains enzyme concentrations to remain within the realm of physical reality, ensuring that our optimization explores biologically meaningful parameter regions. Furthermore, we explore the parameter space in a logarithmic scale. This is done because parameters might vary over several orders of magnitude, and searching for optimal values in a linear scale could lead to numerical instabilities, slow convergence, or issues with precision.

We utilized a scoring function to assess the goodness-of-fit between the model predictions and the observed data. Specifically, we computed the sum of squares of log residuals to quantify the discrepancies between the model-predicted values (\hat{y}) and the experimental measurements (y). For the case of absolute quantification data, the log residuals are computed as per Equation 3.11. In the case of relative quantification data, we first normalized these by their initial time point, providing us fold changes.

Initially, we undertook parameter optimization using basin-hopping, a technique frequently applied to nonlinear multimodal optimization problems in structural biology [86]. However, we encountered challenges, including slow convergence rates and fluctuating minima. These issues prompted us to investigate alternative optimization algorithms. Our exploration revealed that a differential evolution algorithm [115] exhibited faster convergence and consistently led to lower minima, thus overcoming the limitations of the previous approach.

Given the complexity of the system and the large number of simulations involved, we opted not to perform sensitivity analysis at this juncture. This decision was informed by two factors. First, local sensitivity analysis using partial derivatives, while providing an avenue to explore sensitivities, lacks informativeness and is inappropriate for nonlinear systems [103]. Second, while global sensitivity analysis is better suited for nonlinear systems, it poses challenges similar to finding a suitable global optimization method, as both require extensive exploration of the parameter space.

Returning to our initial exploration, we aimed to assess the potential of a kinetic model devoid of allosteric regulation in capturing dynamics (refer to Figure 3.3). Through 10,000 optimizations, each initialized with diverse starting conditions drawn from the prior distribution, we aimed to evaluate the model's capability to explain metabolite dynamics.

While we acknowledge that we cannot definitively rule out the existence of a solution that perfectly matches the data, our outcomes suggest that the model struggles to comprehensively capture metabolite dynamics. Notably, the most substantial deviations were observed for metabolites at the junction of glycolysis and the TCA cycle, such as acetyl-CoA, citrate, aconitate, and isocitrate. Additionally, the dynamics of fumarate and malate exhibit another limitation, as the simulated response to glucose contradicts the observed behavior.

In conclusion, our analysis demonstrates that a model without additional regulatory elements falls short in capturing the short-term dynamics observed after a carbon source switch. We defer a more comprehensive discussion of these results and potential avenues for further exploration to the dedicated "Discussion" section.

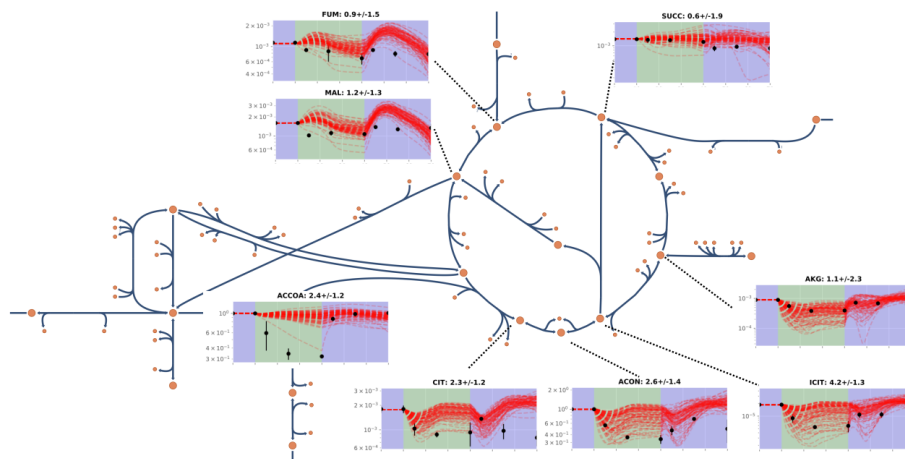


Figure 3.3: Best one percent of predictions without allosteric regulation. On top of the simulated time series depicted for each of the metabolite the geometric mean and standard deviation of the log residuals is provided.

3.3.2.5 Ensemble Modeling

In order to assess the relevance of putative allosteric regulators, we evaluated the impact of incorporating these interactions on the model’s ability to fit the data. To investigate the potential of allosteric regulation, we expanded the model by simulating additional regulatory interactions involving all 57 metabolites for which time series data were available. We considered both single and pairwise combinations of metabolite-enzyme interactions, resulting in a total of 2,850 models with a single allosteric interaction and 3,898,800 models with pairwise interactions.

Among these models, we conducted simulations for the single interaction models 1,000 times each with newly sampled initial conditions. For the pairwise interaction models, we performed simulations once. In the final step of our modeling procedure, we aggregated the outcomes to calculate a score for each individual allosteric interaction that was tested.

More precisely, the information content of each of the models was compared using the Akaike information criterion [2], which penalizes for additional parameters, needed to describe additional interactions, in the models. We then ranked the interactions based on the following two metrics: 1. the frequency at which an individual interaction occurred in a model that outperformed a model without regulation, and 2. the score of the information content of the achieved by the models containing this interaction (Figure 3.4).

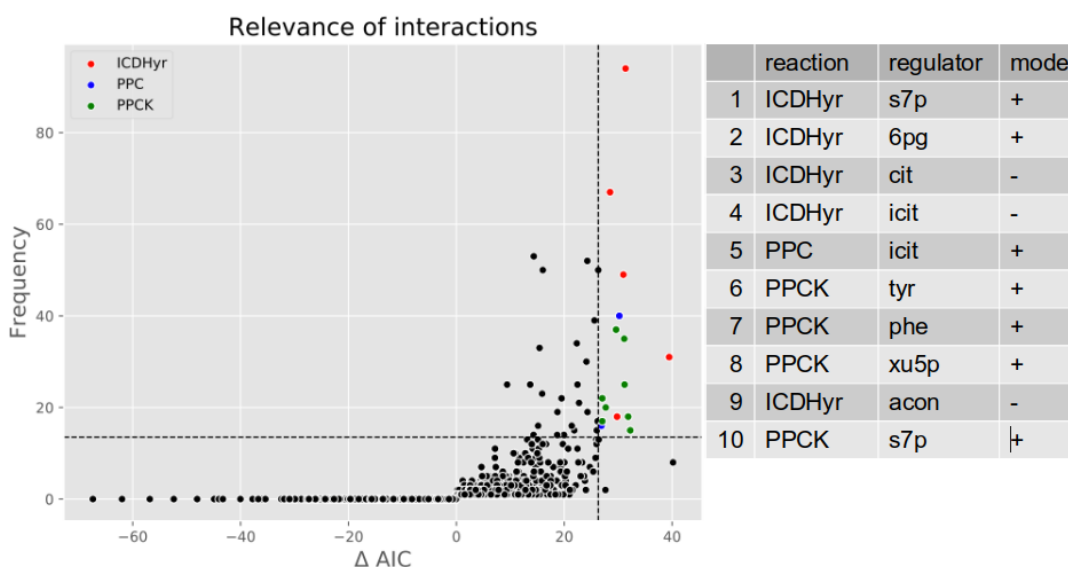


Figure 3.4: Top ranking allosteric interaction predictions. The results of all single and pairwise interaction models was combined in order to derive a ranking for each of the individual interactions tested. Abbreviations: isocitrate dehydrogenase (ICDHr), phosphoenolpyruvate carboxylase (PPC), phosphoenolpyruvate carboxykinase (PPCK), sedoheptulose 7-phosphate (s7p), 6-phospho-D-gluconate (6pg), citrate (cit), isocitrate (icit), L-tyrosine (tyr), L-phenylalanine (phe), D-xylulose 5-phosphate (xu5p), cis-aconitate (acon).

Figure 3.4 presents the results of the individual allosteric interaction tests conducted in our study. Among the highest-ranking interactions, notable entries include isocitrate dehydrogenase (ICDHr), phosphoenolpyruvate carboxykinase (PPCK), and phosphoenolpyruvate carboxylase (PPC). An intriguing observation is the qualitative similarity in the predicted behaviors of allosteric regulators for various reactions (Figure A.19). For instance, the dynamics of metabolites such as citrate, aconitate, and isocitrate are remarkably consistent, a consequence of their interconversion via an isomerase. Similarly, phenylalanine and tyrosine display closely resembling dynamics, owing to the latter’s formation from the former through a single hydroxylation reaction.

While the congruence in predictions for related metabolites underscores the robustness of our approach, it also accentuates a notable limitation. The method, by design, lacks the capacity to definitively differentiate between candidates exhibiting comparable dynamics. In instances where our predictions suggest that citrate, aconitate, and isocitrate act as inhibitors of ICDHr, or where tyrosine and phenylalanine activate PPCK, we are presented with reliable predictions that also illuminate the challenge of disentangling such closely related candidates. Although the possibility

exists that one of these predictions represents a true positive while others are false positives due to the similarity in dynamics, it's important to consider that their structural resemblance at the molecular level introduces the prospect of multiple regulatory roles.

Taking a closer look at a high-ranking prediction involving sedoheptulose 7-phosphate (s7p) inhibiting isocitrate dehydrogenase (ICDHr), we observe substantial improvements, especially in the fitting of citrate, aconitate, and isocitrate, as compared to the base model (Figure 3.5). This enhancement is particularly understandable considering the substantial errors that these metabolites exhibited in the base model's predictions. Remarkably, the positioning of ICDHr within the metabolic network as a pivotal control point aligns well with the logical rationale for this regulatory interaction.

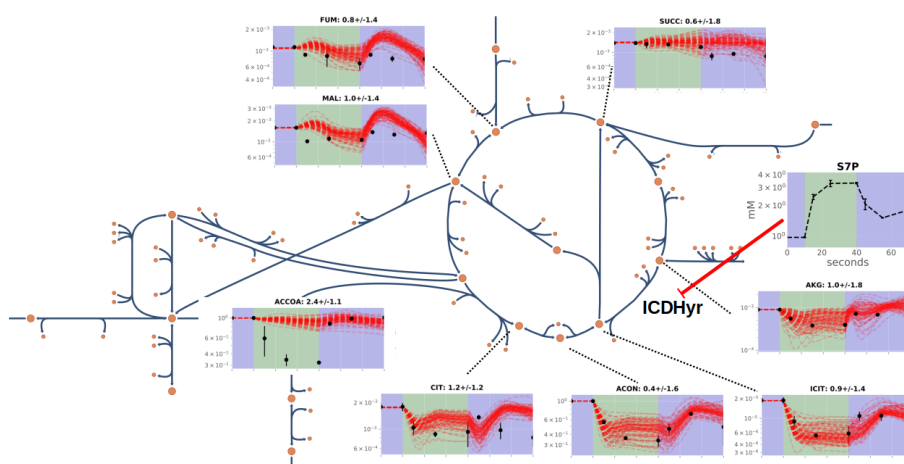


Figure 3.5: Best one percent of predictions with sedoheptulose 7-phosphate (s7p) inhibiting isocitrate dehydrogenase (ICDHr). On top of the simulated time series depicted for each of the metabolite the geometric mean and standard deviation of the log residuals is provided.

3.3.2.6 Predictions vs. Reported Interactions

The predictions extracted from our study’s analysis showcase specific allosteric interactions within key enzymes of the metabolic network. In comparison to the previously reported interactions (Supplementary Table A.9), our findings reveal some noteworthy disparities. For instance, our evaluation of Isocitrate Dehydrogenase (ICDHr) predicted interactions with sedoheptulose 7-phosphate (s7p), 6-phospho-D-gluconate (6pg), citrate (cit), isocitrate (icit), and cis-aconitate (acon). Notably, among the previously reported regulators, phosphoenolpyruvate (PEP) was the sole interaction present in both the reported list and the candidates we evaluated; however, it wasn’t predicted by our analysis. In the case of Phosphoenolpyruvate Carboxylase (PPC), our study anticipates an interaction with isocitrate (icit), diverging from the six previously reported interactions that were included in our study, encompassing citrate (cit), fructose 1,6-bisphosphate (fdp), fumarate (fum), guanosine triphosphate (GTP), and succinate (succ). We note, however, the high correlation between isocitrate (icit) and citrate (cit) due to their conversion by an isomerase. Notwithstanding, our analysis predicts an inverse mode of action compared to the reported interactions. Phosphoenolpyruvate Carboxykinase (PPCK) presents a distinct scenario, with our predictions highlighting tyrosine (tyr), phenylalanine (phe), D-xylulose 5-phosphate (xu5p), and sedoheptulose 7-phosphate (s7p) as potential regulators. Similar to our prior findings, none of these align with the previously reported regulators, which encompass compounds such as 3-phosphoglycerate (3PG), acetyl-coenzyme A (ACCOA), adenosine triphosphate (ATP), dihydroxyacetone phosphate (DHAP), fructose 6-phosphate (F6P), fructose 1,6-bisphosphate (FDP), reduced nicotinamide adenine dinucleotide (NADH), and phosphoenolpyruvate (PEP).

These discrepancies and divergences between our predictions and the previously reported interactions raise several plausible factors to consider. Firstly, false positives might arise due to strong correlations in metabolite time series data, leading to predictive overlaps that don’t necessarily denote direct regulatory relationships. Secondly, conditions reported in literature might have been obtained under *in vitro* conditions that aren’t physiologically relevant, potentially skewing the observed interactions. Additionally, reported interactions might indeed be valid but under different experimental conditions not replicated in our study. Moreover, inherent issues with the inference procedure, including model limitations or inherent biases, might contribute to discrepancies between predictions and known interactions.

In our evaluation of 2850 metabolite-reaction pairs as potential regulatory interactions, we found supporting evidence in the literature for 43 of them. With respect to the top 10 interactions we predicted, we found:

- **Isocitrate Dehydrogenase (ICDHr):** Our model covers two known interactions, while all five of our predictions are novel.
- **Phosphoenolpyruvate Carboxylase (PPC):** Among the six interactions covered by our model, our predicted interaction is novel.
- **Phosphoenolpyruvate Carboxykinase (PPCK):** eight known interactions are covered by our model, while all four predicted interactions are novel.

While these reasons warrant careful consideration, there remains the intriguing possibility of entirely novel regulatory roles or previously unidentified interactions within the metabolic network. Verifying these conjectures experimentally would be imperative to ascertain the true nature of these predictions.

3.4 Discussion

In this discussion, we delve into the implications and interpretations of our modeling results, focusing on the putative allosteric regulators and their potential impact on the dynamic behavior of reactions within or closely associated with the tricarboxylic acid (TCA) cycle. As in previous works [75, 17], our modeling framework primarily serves as a hypothesis generation tool, aiding in the identification of candidate regulatory interactions that could contribute to observed metabolic dynamics.

It’s crucial to highlight the pronounced divergence between the observed and simulated dynamics of Acetyl-Coenzyme A (AcCoA). An essential aspect to consider is that our available data for this metabolite was solely in the form of relative measurements. Despite this constraint, our model simulations were only capable of reproducing up to 30% of the substantial 70% decrease observed in the initial steady-state concentration. This discrepancy raises significant questions, the answers to which could provide deeper insights into our parameterization strategy. While one contributing factor might indeed be inaccuracies in parameter values, it’s important to acknowledge that other complexities could also contribute. AcCoA is not merely an average metabolite; rather, it occupies a central regulatory role and participates in various metabolic pathways, much like other cofactors. Unaccounted interactions and regulatory dynamics involving AcCoA could further contribute to the observed inconsistency between the data and the model simulations.

Our modeling results highlight the necessity of regulatory mechanisms to explain the metabolic response dynamics observed in the initial segment of the TCA cycle, specifically from oxaloacetate (OAA), the entry point of acetyl-CoA, to the bifurcation between isocitrate dehydrogenase (ICDHr) and isocitrate lyase (ICL). Among the candidate enzymes, our analysis singles out isocitrate dehydrogenase (ICDHr), phosphoenolpyruvate carboxylase (PPC), and phosphoenolpyruvate carboxykinase (PPCK) as the most likely targets of regulation.

Of particular note are the dynamics associated with the predicted allosteric interactions involving ICDHr. The predicted allosteric activators, sedoheptulose 7-phosphate (s7p), and 6-phospho-D-gluconate (6pg), exhibit rapid increases in cellular concentration upon the switch to glucose. Conversely, predicted allosteric inhibitors such as citrate, aconitase, and isocitrate display swift declines in intracellular levels following the perturbation. These findings align logically, suggesting that an augmented flux through ICDHr, as indicated by both activation and inhibition scenarios, could account for the observed dynamics.

It’s worth noting that our modeling efforts revealed that the observed dynamics of isocitrate cannot be adequately described by Michaelis-Menten kinetics alone. Although substrate inhibition was considered, the established knowledge about ICDHr and the lack of previous observations of substrate inhibition in this context [83, 21, 22, 99] render it unlikely as the primary cause.

Another explanation for the inability to capture these dynamics is phosphoregulation of ICDHr. This enzyme was among the first bacterial enzymes shown to be regulated by phosphorylation [117, 52], and this phosphoregulation has been implicated in regulating the glyoxylate shunt [52, 51, 18]. However, our method does not account for phosphoregulation, and distinguishing between allosteric and post-translational mechanisms remains a challenge, as both processes operate on similar timescales [110]. Moreover, detecting changes in the phosphorylation state of the enzyme requires additional techniques, such as phosphoproteomics, and is further complicated by the short sampling time interval required. Future research employing phosphomutants could offer insights, provided they are carefully designed to mitigate potential confounding factors. Such confounding factors do not only pertain to the possibility of affected enzyme kinetics, as the cells may also attain a different steady-state than the wild type. If such a mutation in central metabolism alters cellular physiology severely, the comparison of mutant and wildtype becomes impossible.

Among the top-ranking predictions, PPC and PPCK also emerge as key enzymes deserving attention. These enzymes govern the flux from phosphoenolpyruvate to oxaloacetate, with opposite catalytic directions. Allosteric activation of PPC by isocitrate could lead to decreased flux through PPC due to the perturbation-induced drop in isocitrate concentration. Similarly, the dynamics of predicted allosteric activators of PPCK, tyrosine, and phenylalanine, exhibit comparable spikes followed by rapid recovery. Their dynamics coincide with those of xu5p and s7p, which also show a prompt concentration increase and sustained levels. This concordance suggests a potential requirement for increased flux through PPCK, a conclusion that aligns logically with the findings for PPC.

It also may be noted that neither of these mechanisms by itself is sufficient to explain the

observed dynamics. It is therefore well possible that more complex regulation is needed, or that a flux-carrying reaction is missing. For example, secretion of malate could help explain our observations, however, *E. coli* BW25113 is not known to naturally secrete this, nor is it observed during steady-state growth on medium containing pyruvate or glucose.

Notably, none of the top-ranking interactions we identified have been previously reported in literature. To substantiate these predictions, further experimental assays are essential. Enzyme assays examining the effects of the predicted allosteric regulators *in vitro* could provide supportive evidence. However, given the complex cellular context, and the possibility to obtain false positives by using non-physiological test conditions *in vitro*, such a result cannot be considered definitive by itself for the existence of the allosteric interaction, let alone its relevance in a given condition. Using allosteric mutants as an avenue for potential validation of functional relevance may offer a way forward. However, it's important to acknowledge that this approach is susceptible to similar challenges as those encountered in phosphomutant experiments. The potential confounding effects of allosteric- and phosphomutants underscore the need for careful experimental design to mitigate unintended physiological alterations. The limitations of any single approach, including computational modeling and *in vitro* assays, suggest that a multifaceted approach involving orthogonal methods is necessary for a comprehensive understanding of regulatory interactions in the intricate milieu of *E. coli*.

3.4.1 Differences in methodological approaches

In this section, we delve into the discrepancies between our current methodological approach and those employed in previous studies. We discuss the implications of these differences on our predictions and propose suggestions for future research.

3.4.1.1 Experimental procedure

In previously work by Link *et al.* (2013) a perturbation starting from growth on pyruvate, and another from growth on fructose, to glucose, was performed [75]. In Christodoulou *et al.* (2018) a perturbation with hydrogen peroxide was performed [17]. In this work we focus on carbon source perturbations, starting from growth on pyruvate, succinate or acetate, and then switching to glucose.

Although carbon source switches may be considered less disruptive than exposure to hydrogen peroxide, the perturbation procedure itself introduces an intrinsic effect. During the experiments conducted in each of these studies, cells were placed on a filter in order to perfuse them with a secondary medium in order to perturb them. This filter is placed on top of a Buchner funnel that is connected to a vacuum pump to help facilitate the exchange of medium, which is necessary to prevent the cells from washing off. This approach, however, has several downsides associated, pertaining to the perturbation event and manual operating procedure.

Perturbation event: With respect to the perturbation media, the first thing to note is that this is fresh as opposed to preconditioned. The medium where the cells are sampled from is the medium these cells have been cultivated in, therefore, outside of a difference in carbon source, any other changes introduced in its content by the cultivation process constitute perturbation factors. This encompasses a wide range of factors, such as other nutrients taken up, as well accumulation of secreted metabolites, which may affect the extracellular pH, as well as signalling molecules. It is also worth noting that cells are expected to experience increased exposure to air. To mitigate such confounding effects, cells are perfused with fresh medium containing the original carbon source the cells were growing on for the first 10 seconds, before switching to medium containing glucose. However, whether this is sufficient to establish a new pseudo steady-state is unlikely. Experiments conducted by Ryback *et al.* with the yeast *S. cerevisiae* have shown that the filtration procedure itself has a considerable effect [101], providing credence to this claim.

It's important to note that even controls taken after perfusion with fresh medium, mimicking the perturbation conditions without a carbon source switch, exhibited substantial changes, raising questions about the extent to which the perturbation itself contributes to the observed effects. This control experiment was also included in the work by Link *et al.* and Christodoulou *et al.*, however it is not clear how this data was used, if at all. We note that in the case of several metabolites the change in the control is as significant or larger than that observed in the samples with the

perturbation of interest [75, 17], and we have observed the same in our experiments (see Figures A.20, A.21, A.22).

This does not mean that the signal of the effect of the intended perturbation was lost in the noise introduced by experimental handling. In future work, however, it is advisable to perform the experiment with and without the perturbation of interest, such that the effect of the experimental handling can be disentangled from those of the perturbation of interest.

Manual operating procedure: The human element in experimental procedures introduces an inherent degree of variability and potential error. Manual handling is an integral part of the current experimental procedure, encompassing the sampling of cells, applying the perturbation once cells are on the filter, and the transfer of the filter into extraction medium. The experimental design is further shaped by the necessity to initiate the experiment at an OD of 0.5 to ensure sufficient cell mass for quantification, while avoiding the onset of an OD of 1.0 to retain cells in mid-exponential growth. This dual requirement creates a time-sensitive window for conducting the experiment. The duration of this window varies based on the carbon source and growth rate, necessitating meticulous timing and execution.

Sampling cells manually poses certain challenges. Throughout the course of the experiment, cell density undergoes changes. While corrections are applied based on OD measurements taken before and after the experiment, with interpolation for specific sampling times to account for variations in cell density, this correction does not address the impact of cell density on the flow rate of perfusion medium through the filter.

In the context of manual perfusion, we utilized syringes to administer medium to the cells. Achieving and maintaining a consistent pressure for a uniform flow rate from the syringe is inherently challenging. While efforts were made to automate this process using a peristaltic pump, ensuring a steady flow rate that avoids both cell drying and medium overflow is complex. This challenge might be attributed to the increasing cell density during the experiment. It's noteworthy that employing a larger filter could aid in mitigating overflow concerns, yet it could simultaneously exacerbate the risk of cell drying, given the larger perfusion area required.

In addition to addressing the intricacies of manual operations, achieving high precision in both manual perfusion and timing simultaneously presents challenges, particularly considering the necessity to prevent cell drying and medium overflow during the procedure. This dual requirement places a significant cognitive burden on the experimentalist, who must not only monitor the cells undergoing perfusion but also keep track of the precise timing using a stopwatch. This multitasking endeavor adds another layer of complexity to the experimental process, as the experimentalist must carefully synchronize these tasks to ensure accurate timing of sampling intervals.

A distinct concern arises at the conclusion of the perturbation procedure. Failure to terminate perfusion at precisely the right moment will result in variable volumes of perfusion medium remaining with the cells on the filter. Moreover, the presence of excess medium on the filter can significantly affect downstream mass spectrometry measurements. Any additional time required to drain excess perfusion medium from the filter will inevitably lead to inaccuracies in the sampling interval.

However, our experimental setup incorporated a video camera to record the entire procedure, offering valuable insights and aids in overcoming some of these challenges. The recorded footage allowed for subsequent analysis, enabling the identification and correction of errors in the timing of sampling intervals, as well as other potential experimental issues. This approach enhances the experimental accuracy and reliability, as it offers the opportunity to review the procedure in detail, ensure proper execution, and identify and address any discrepancies that might have occurred during the experiment. Furthermore, the use of audio commentary in the video recording allowed for real-time note-taking, eliminating the need for manual recording and reducing the risk of errors during the process.

In future experimental designs, meticulous attention to the synchronization of manual procedures and accurate timing of sampling intervals should be a priority. Innovative methods to optimize these aspects could include automated perfusion systems and refined filter designs that minimize medium retention while ensuring consistent cell perfusion rates. By addressing these challenges, researchers can minimize the confounding effects of imprecise timing and manual handling, leading to more reliable and reproducible experimental outcomes.

Data quality and analytical variability: An integral aspect of our investigation pertained to the quality of the data acquired, and its comparison to previous studies. Strikingly, despite conducting the experiment with meticulous care and a range of replicates, our data quality fell notably short of that reported by Link *et al.* (2013), raising pertinent questions. The signal-to-noise ratio, a key metric reflecting data robustness, exhibited significant reduction in our results. Expert support from individuals well-versed in the experimental and analytical techniques did not amend this outcome.

Our study incorporated technical replicates to discern the extent of variability attributed solely to experimental procedures. These technical replicates involved secondary sample injections from the same well on a 96-well plate, with variations introduced only by liquid chromatography, mass spectrometry, and subsequent analysis. In a analysis of the top 15 metabolites, constituting 25% of the total dataset, with the highest signal-to-noise ratios, employing ^{13}C normalization only when it reduced sample variance, yielded an average coefficient of variation (CVs) of 0.15 ± 0.07 in technical replicates and 0.39 ± 0.28 in biological replicates.

Comparing these metrics with the previous study by Link *et al.*, we find an average CV of 0.12 ± 0.10 across all metabolite time series for biological replicates. The absence of technical replicates in the previous studies occludes direct comparison of technical variability. However, it is notable that our technical replicates' variance exceeded the total variance seen in biological replicates of Link *et al.*. We furthermore noticed that the variance in biological replicates in our study, after subtraction of technical variance, also exceeded that of Link *et al.* by more than two-fold. Looking into other previous studies using the experimental procedure by Christodoulou *et al.*, and Buffing *et al.*, we find average CVs of 0.45 ± 0.35 and 0.38 ± 0.26 , respectively [14, 17]. It is essential to acknowledge that multiple factors, ranging from variations in experimental setups to differences in analytical techniques, could contribute to such disparities.

We sought insights from experts in the field and carefully compared our results to those published by Link *et al.* While we are cautious in making direct claims about the data generated by others, the substantial differences observed raise questions about the reproducibility and robustness of the original findings. These observations have significant implications for the feasibility of model inference and predictions based on the obtained data. The discrepancies in data quality necessitated a reevaluation of the complexity of the models we could employ, ultimately challenging our ability to draw meaningful conclusions from the noisy data. In fact, the data quality was deemed unsuitable for model inference, rendering our original research goals unattainable. This unexpected outcome compelled us to critically examine and reevaluate the legitimacy of the data we generated.

In the face of the inherent challenges presented by the inconsistencies in data quality and the lack of reproducibility, we were compelled to consider alternative approaches. Intriguingly, we turned to the same dataset published by Link *et al.*, acknowledging the contentious nature of its validity, in our pursuit to perform model fitting. Our reluctance to rely on such data underscores our profound concerns regarding the reliability and reproducibility of our experimental procedure. It is not our intention to cast aspersions on the work of our colleagues; rather, we emphasize the importance of open and transparent discussions surrounding data quality and research methodology within the scientific community.

3.4.1.2 Modeling procedure

There are also some notable differences with previous approaches, which will be presented and discussed to guide future trailblazers. These are the three main issues addressed in this work:

1. reproducibility of models and their predictions
2. reusability of models and pipelines
3. errors in the construction process

Model construction

Model format: In previous work models were manually constructed and hard-coded into MATLAB. The model in Link *et al.* (2013) was constructed using Simulink, a MATLAB-based proprietary graphical programming environment, and does not exist in SBML format [48]. The

model in Christodoulou *et al.* (2018) was provided in SBML format. Although it can be simulated, it fails to pass the SBML validator checks. The SBML validator consists of various tests, and consists of checks for the following:

- consistency of measurement units associated with quantities
- correctness and consistency of identifiers used for model entities
- syntax of mathematical expressions [53]
- validity of systems biology ontology annotations [55]
- static analysis of whether the model is overdetermined
- various other good modeling practices and general SBML consistency checks

In contrast, models in this work are constructed programmatically with the use of libSBML [11], adheres to the specifications set out for level 3 version 2 [49, 50] and passes SBML validator tests.

To facilitate this automated model construction process, we have developed a comprehensive toolkit available in the repository at: https://gitlab.ethz.ch/karrenbelt/sbml_tools. This toolkit offers a suite of tools designed to work with SBML models and streamline various aspects of the modeling workflow. Some of the tools included are:

- An annotator for model annotation, ensuring consistent and accurate annotation of model entities.
- A COBRA wrapper to simplify constraint-based modeling tasks, enhancing the ease of working with these models.
- A data integrator allowing for various forms of interpolation and assignment of experimental measurement data to model variables.
- Equilibrator integration for estimating thermodynamic data for the model after annotation.
- A parameter balancer to aid in balancing model parameters.
- A RoadRunner wrapper for fast and efficient model simulations.
- A module for symbolic manipulation of mathematical equations within the model, enabling advanced mathematical operations.
- A kinetic rate law factory for automatic generation of kinetic rate equations based on reaction types.
- A parameter factory for generating model parameters with proper units and annotations.
- A unit and unit definition factory to handle proper unit assignments for model components.
- Automatic data retrieval of BRENDA data based on model annotation, facilitating integration of experimental enzyme kinetics data.
- And more tools tailored to enhance the modeling process.

We will illustrate the importance hereof more concretely in the following paragraphs.

system demarcation: In previous work, stoichiometric models were constructed in an *ad-hoc* manner. Modelers made decisions on lumping reactions and metabolites, often ignoring co-factors. Consequently, these systems lacked mass or charge balance, which is essential for accurately representing cellular metabolism and ensuring the model’s physical feasibility. In this work the full metabolic subsystem as we understand it today was used, ensuring that all reactions in the system are, and as a consequence the system as a whole is, both mass and charge balanced.

model annotation: Metabolites and reactions in the previous work were identified with alphanumeric abbreviations of a non-standardized format. The interpretation of non-standardized names is ambiguous and can lead to confusion (e.g. fdp and fbp being used by different people to describe the same metabolite). Moreover, it prevents the automatic retrieval and linking of information from external databases. In this work we have adopted the well-known identifier standard of the BiGG database [105], and annotated entities in the model using a multitude of external well-established database identifiers as well (see section 3.2.2.1 for details). Annotation of these entities with these identifiers was performed programmatically, validated to be compliant with the Minimal Information Requested In the Annotation of biochemical Models (MIRIAM) standard [68], and used for automated data retrieval from databases. Furthermore, elements of the model have also been annotated using the Systems Biology Ontology [55]. This further clarifies the nature of elements in the model and the intention of the modeler when it comes to entities such as species, parameters, reactions and rate laws, which in turn facilitates the interpretation and reusability.

metabolic rate laws: In previous work reactions were assumed to be either reversible or irreversible, and rate laws were chosen in accordance to obtain a kinetic model. In the case of thermodynamically unfavorable reactions irreversible Michaelis-Menten kinetics were chosen. For reactions deemed reversible, enzyme kinetics were assumed to be irrelevant and a simplified mass-action kinetics rate law was chosen. Whereas irreversible Michaelis-Menten kinetics may be considered suitable to describe the initial enzyme kinetics in an *in silico* setting where there initially is no product, it is not a realistic equation for the describing *in vivo* kinetics of a biochemical system. That is because in a metabolic pathway there must always be product available for the next enzymatic step, thus in order to model a pathway some degree of product inhibition or reversibility must be explicitly included [104]. In contrast, this work assumes that enzyme kinetics can be described by convenience kinetics, a generalized enzymatic rate law that is fully reversible.

In earlier work, rate laws were manually crafted and hardcoded into MATLAB scripts. This manual process could be error-prone. Moreover, in both Link *et al.* (2013) and Christodoulou *et al.* (2018) the apparent assumption is made that co-factors are irrelevant to describe reaction kinetics. Interestingly, when a reaction model is used to accommodate the use of an *in vitro* enzyme assay a far more elaborate rate law is used, which includes a co-factor and its associated dissociation constant (see supplementary material on method details [17]). In this work rate laws are constructed using a programmatic framework, which, among others, checks certain basic assumptions (e.g. that the stoichiometry of the reaction matches the assumptions of the selected rate law), and ensures that there are no errors in the rate laws.

Model initialization

steady-state flux analysis: In all three approaches steady-state fluxes were used to initialize the model around a steady-state. In previous work these steady-state fluxes were derived through flux balance analysis (FBA) [88], using the maximum uptake rate of the carbon source and sampling flux ratios for pathway branches and futile cycles. In this work the available data from ^{13}C metabolic flux analysis (MFA) by Gerosa *et al.* (2015) was used and a minimization problem was formulated to find a flux solution as close as possible to the reported steady-state flux [37, 109]. Using flux estimates from MFA offers a key advantage: flux ratios no longer require sampling, as labeling data is utilized for their estimation.

thermodynamics: Equilibrium constants were not used in previous work; reactions were either deemed irreversible or otherwise assumed to be in equilibrium. In this work we used Equilibrator [34] to obtain estimates for the Gibbs free energy of the metabolic reactions of our systems. In order to obtain proper estimates for these Gibbs free energies of reactions, it is important that reactions are mass and charge balanced, which is something we ensured when we demarcated our subsystem of interest. Since Gibbs free energies are estimated for the entire network, the resulting covariance matrix becomes available. Multivariate sampling of this matrix allows us to account for dependencies between the equilibrium constants, allowing us to obtain a thermodynamically consistent set of parameters.

metabolite concentrations: The metabolite concentrations of the first time point of the times series are used to initialize the model. In previous work metabolites for which no data was available were either lumped or sampled independently from a uniform distribution between 0 and 2 mM [75, 17]. We utilize the estimated equilibrium constants to define the sampling space of the metabolome more precisely, allowing us to constrain it to a smaller, thermodynamically feasible domain. In other words, the initial metabolite concentrations sampled from this space yield a thermodynamically consistent set of initial conditions.

enzyme concentrations: Enzyme concentrations were estimated from absolute protein quantification data [106] and cellular volume data [121]. Protein data was mapped to the model using gene identifiers (Supplementary Table A.3), and the gene-reaction-rules (Supplementary Table A.4) present in the genome-scale metabolic model of *E. coli* [88] to estimate steady-state enzyme concentrations for 22 out of the 23 reactions. An uninformative prior is constructed using all available data from the observed population of enzymes, which is then used to establish a comprehensive prior distribution over the enzymes.

kinetic parameters: In previous work, parameter data from BRENDA [107] was used to compute the mean value of Michaelis-Menten constants (K_m) for irreversible reactions. To account for uncertainty, K_m values were randomly sampled from an interval of 0.1 to 10 times the reported literature value. This range is arbitrarily chosen and samples are drawn from a uniform linear distribution. However, kinetic parameters are log-normally distributed [107, 76], hence the sampling was heavily biased towards smaller values. In this work the BRENDA database on enzyme kinetic parameters was parsed. This parser allows to customize building of the kinetic parameter prior by including data from species at different taxonomy levels by specifying a weighting factor that weighs data from more distantly related species less, allowing this information to be used in case no information at the species level is available. The same holds for mutant data. With the current pipeline settings only *E. coli*-specific non-mutant data are used, however, an uninformative prior over all available data for a particular type of parameter was also constructed and used as an uninformative prior in case no data is available for a specific parameter in the model. In previous work, only a single substrate was assumed to be relevant to describe irreversible reaction kinetics, and hence these equations involve a single Michaelis-Menten constant. In this work, we consider all substrates and products as relevant players. The former strategy yields a model with 8 Michaelis-Menten constants, for each of which a single literature reference was found. In this work, our model encompasses 92 Michaelis-Menten constants, with species-specific data available for 50 of them.

In previous work, irreversible kinetics were described using Michaelis-Menten kinetics with a V_{max} term, while reversible reaction kinetics were represented by a k^+ and k^- mass-action constant. The equation was solved for V_{max} given the previously determined steady-state flux for irreversible reactions. In Link *et al.*, (2013) the k^+ and k^- were fixed at an arbitrary 100 s^{-1} and $100 \text{ s}^{-1} \text{ mol}^{-1}$, consequently leading to a new steady-state. In Christodoulou *et al.* (2018) a back flux ratio between 0 and 1 was sampled, then forward and backward flux were balanced to maintain a consistent net flux, from which subsequently the k^+ and k^- were determined, preserving the original steady-state. In this work, we decompose V_{max} into a catalytic constant term and an enzyme term. Catalytic forward rate constants were obtained for 8 out of 23 reactions. Data were available for 2 out of the 23 backward catalytic rate constants. The data on backward catalytic rate constants is currently not used; instead we opted to rewrite the kinetic rate laws in their thermodynamic form, effectively replacing the need for backward catalytic rate constants with equilibrium constants (see section 3.2.3.4). Since enzyme concentration is assumed constant through the duration of the experiment, these act merely as a scaling factor. Together with the previously determined steady-state flux solution, we solve the systems for each reaction by drawing a sample of the enzyme and forward catalytic rate constant that matches this steady-state flux.

Ensemble modeling

Simulation throughput: Previous approaches employed MATLAB’s built-in solvers for model simulations. Christodoulou *et al.* (2018) recognized the potential of parallelization, which spread the computational workload across multiple workers, thus reducing the overall runtime. In our work, we further optimized individual simulations by leveraging a just-in-time (JIT) compiler

[8, 70, 112]. This compilation strategy significantly improved simulation speeds. The average simulation time of the base model: 7.0 s, 2.3 s and 0.025 s, in Link *et al.* (2013), Christodoulou *et al.* (2018) and this work, respectively. This substantial increase in simulation throughput was crucial to enable the numerical optimization of the models, without which the exploration of both allosteric topologies and their associated parameter space would have been computationally unfeasible. For instance, in the study by Link *et al.*, simulating the ensemble of models required approximately three days, while Christodoulou *et al.* needed around one day. Given that the optimization process requires several hundred times more simulations per parameter set than a single simulation utilizing randomly sampled initial conditions, undertaking such optimizations without the acceleration provided by JIT compilation would have been computationally unfeasible.

Optimization strategies: The utility of optimization strategies in scientific modeling has been extensively validated within the field. A substantial body of research has consistently demonstrated that the employment of optimization techniques contributes to the identification of lower minima, resulting in enhanced model performance. This assertion is supported by empirical evidence spanning various domains of science and engineering, where optimization has been instrumental in achieving higher accuracy and reliability [113, 42].

One key advantage of optimization is its ability to refine parameter space exploration. Through iterative refinement, optimization methods systematically guide the search towards regions of parameter space that yield improved fits to the available data. This iterative nature, often combined with a multi-start approach, ensures a broader exploration of the solution landscape and reduces the likelihood of convergence to local minima. Consequently, compared to approaches reliant solely on randomly sampled initial conditions, optimization-driven methodologies provide heightened robustness and a more comprehensive assessment of the parameter space. The scientific consensus established by these studies reinforces the efficacy of optimization as a means to enhance the fidelity of model predictions and the robustness of outcomes [84, 98].

In previous studies, a random sampling methodology was employed to identify the optimal parameter set. In these earlier approaches, the steady-state flux ratios, Michaelis-Menten constants, and metabolite concentrations without available data were uniformly sampled within a linear space. Specifically, the Michaelis-Menten constants were sampled over a range spanning 10-fold both above and below their literature-referenced values, while missing metabolite concentrations were drawn from a range between 0.1 and 2.0 mM. Notably, the model lacking allosteric regulation was subjected to a large number of simulations: 10,000 times in Link *et al.* (2013) and 2,000 times in Christodoulou *et al.* (2018), ultimately leading to the selection of the best performing model.

In contrast, our approach takes a different route. Equilibrium constants are sampled from a multivariate distribution, followed by the application of a thermodynamic sampling approach to metabolite concentrations, leveraging available data. Fluxes are acquired using the flux projection procedure [109, 85]. Enzyme and catalytic constant initial values are sampled based on available data and steady-state flux solutions. To optimize kinetic parameters, we employ an evolutionary algorithm, setting bounds either according to available data or within the limits defined by the uninformative prior. The use of a just-in-time (JIT) compiler significantly accelerates this process, reducing the simulation time by two orders of magnitude. As each iteration of the optimization requires a simulation run, the JIT compilation's speed boost becomes indispensable in handling the several hundred iterations needed for optimization. The procedure is halted once a solution converges to a (local) minimum. Importantly, this comprehensive procedure is replicated 1,000 times for the base model.

Allosteric search: An exhaustive search was undertaken to evaluate the allosteric interactions that enhance the base model's capability to depict observed dynamics. This exploration involved augmenting the models with either single or pairwise allosteric interactions. The introduction of an additional term, encompassing an extra parameter to represent the interaction, effectively incorporates the supplementary interaction into the model.

Diverging from previous methodologies, where only species included within the model were considered as potential regulators, our approach directly incorporates measured data as putative regulatory input signals. This strategy involves interpolating experimentally determined metabolite values and integrating them into the model. This extension enables the assessment of regulators not explicitly represented in the system of differential equations. Additionally, unlike prior investigations that confined candidate allosteric regulation to irreversible reactions, our study embraces all

reactions as prospective targets. Previously the assumption was made that reversible reactions are less likely to be regulated, however, no discernible correlation exists between the thermodynamic driving force of a reaction and the number of reported allosteric regulators influencing enzymes governing such reactions [100].

In the study by Link *et al.* (2013), 1,000 samples were drawn for the single interaction model, while 500 were drawn for pairwise interaction models. In contrast, Christodoulou *et al.* (2018) drew 2,000 parameters for all model topologies. Our simulations encompassed 1,000 iterations for models with a single interaction, and once for pairwise interaction models (as detailed in Table 3.3).

Table 3.3: Overview of numbers relating to the models and allosteric search space of previous and current studies. The glycolytic gluconeogenic switch in glycolysis was studied in Link *et al.* (2013), the effect of reactive oxygen species on the penthose phosphate pathway (PPP) in Christodoulou *et al.* (2018), and the dynamics resulting from a pyruvate to glucose switch on tricarboxylic acid (TCA) cycle metabolism in this study.

	Link <i>et al.</i> (2013)	Christodoulou <i>et al.</i> (2018)	this work
Base Model			
metabolites	7	12	30
reactions	14	16	23
Allosteric interactions			
single interaction	126	162	2,850
pairwise interaction	3,600	12,000	3,898,800
Simulations			
single interaction	126,000	324,000	2,850,000
pairwise interaction	6,200,000	24,000,000	3,898,800

time series analysis: In prior studies, metabolite data and simulations underwent normalization to their initial steady-state value, leading to relative error calculations using the residual sum of squares. In this work we deviated from this normalization approach for two key reasons. Firstly, normalization to the initial steady-state concentration of metabolites results in the loss of information, as it no longer penalizes absolute differences. Secondly, such normalization overly sensitizes the scoring to the initial observation. To mitigate these issues, we opted to calculate the residual sum of squares of the log residuals. Similar to previous methodologies, our model comparisons rely on two main metrics:

1. The Akaike information criterion (AIC) [2], which introduces penalties for models with additional parameters, and
2. The frequency with which specific allosteric interactions contribute to enhancing the models' explanatory power over the observed data.

When it comes to deriving a final ranking, our approach remains largely similar, with one minor distinction: we employ an extended version of the AIC that corrects for finite sample sizes, ensuring more robust model comparison outcomes. It is essential to note that the utilization of the AIC rests upon the assumption that the provided score constitutes a maximum likelihood estimate. This assumption, however, is violated in previous approaches where models are simulated via random initialization rather than optimization procedures. Drawing a conclusion from this, it becomes evident that employing optimization procedures rather than random initialization is essential not only for improving model fits but also for ensuring the validity of the AIC-based comparisons.

3.4.2 Suggestions for future work

With regard to the experimental work, we operated under the assumption that we would be able to obtain data of a similar quality as had been reported by Link *et al.* (2013). We suggest that a future experimentalist ensures that they can measure intracellular metabolite concentration reliably and consistently, such that they establish such expectation based on their own work, and not the work of others.

3.4.2.1 Automation

The current fast filtration experiment in which the perturbation is performed requires a lot of manual handling. As such, it is highly involved and challenging to perform as it requires precise coordination of cell dispensation, alternation between perfusion media while ensuring cells do not dry out or wash off the filter, all the while measuring time intervals of 5 seconds precisely. This error on the time-axis has thus far been ignored. To mitigate these challenges and enhance experimental accuracy, we suggest exploring the implementation of advanced automation technologies. The adoption of automated cell dispensation systems, coupled with robotic media switching mechanisms, can minimize human-related inconsistencies and errors. Furthermore, integrating high-speed imaging systems can provide real-time monitoring and data acquisition, alleviating the need for manual sampling. This integration of technology has the potential to significantly enhance experimental precision and data quality, ultimately leading to more reliable insights into cellular dynamics.

Automation of the experiment, particularly the automation of cell sampling and perfusion for perturbation, holds the potential to significantly minimize technical variability introduced by human involvement. Beyond enhancing data consistency, automation also paves the way for scalability in experimental design. An intriguing prospect is the application of a continuous cultivation setup, such as a chemostat, where the culture undergoes constant monitoring. Continuous media outflow from the culture provides a reservoir of cells that can be subjected to perturbations and subsequent measurement. By maintaining the culture at mid-exponential growth, the pseudo steady-state assumption can be upheld, offering advantageous conditions for model initialization.

The utility of a chemostat setup extends further due to its inherent capacity to finely control growth rates through dilution rate modulation. This precise growth control adds an extra layer of advantage, enabling the initiation of experiments from diverse steady-state conditions. The wealth of information garnered from distinct steady-states can prove invaluable for systems identification. We have partially explored the potential of such steady-states in the second chapter of this thesis and delved into dynamic response post-perturbation in the third chapter. Intertwining these approaches emerges as a logical progression for subsequent experimental work, promising an even deeper understanding of metabolic dynamics.

3.4.2.2 Observable states

With respect to metabolite quantification, it is important to carefully consider the states of interest that require observation. While our mass spectrometry method comprehensively covers most metabolites within central carbon metabolism, specific compounds like co-factors and glyoxylate present challenges in quantification, demanding additional efforts for accurate measurement. Given that central carbon metabolism plays a crucial role in energy generation and cofactor regeneration, it is reasonable to expect that their dynamics are of significant interest to those that wish to understand behavior of this system.

The work by Link *et al.* underscores substantial changes in the relative abundance of co-factors triggered by carbon source shifts. However, the quantification of co-factors presents difficulties due to their instability and propensity to react with other molecules, leading to degradation. To accurately measure these species, fresh standards must be prepared on the day of the experiment.

From the perspective of metabolic modeling, cofactors pose an equally formidable challenge. They are involved in so many reactions that it is either practically impossible or becomes trivially easy to model, depending on the amount of free parameters one is willing to introduce. In this study, we therefore chose to assign the dynamics of these species instead using the collected observational times series data. This highlights an interesting consideration in the design of our approach: we decide which dynamics we hold to be true, and what others we wish to estimate based on the data of the former. However, it's essential to acknowledge that we haven't yet quantified the extent of the impact of this choice on our model predictions, which stands as an important avenue for future investigation.

Moreover, the selection of relevant states to observe hinges on the specific behavior under study, the initial conditions, and the chosen perturbation. Although selecting an appropriate method for metabolite quantification necessitates informed judgment, identifying the states that carry the most informative content is inherently uncertain. For instance, during steady-state growth on pyruvate or glucose, the glyoxylate shunt does not exhibit significant flux. However, this absence of flux does not eliminate the potential for the pathway to play a role during flux adaptation. Notably, the

dataset from Link *et al.* that we utilized lacks glyoxylate measurements. A subsequent study could include glyoxylate quantification to assess the utilization of the glyoxylate shunt. Investigating glyoxylate’s dynamics would be particularly relevant in scenarios involving carbon sources known to activate the glyoxylate shunt, such as during growth on acetate or galactose. Moreover, a specialized method tailored for quantifying α -keto acids, such as glyoxylate, in central carbon metabolism, has been developed [132], further supporting its investigation.

3.4.2.3 Long-term adaptation

When contemplating long-term adaptation, such as transitioning between distinct steady-state growth conditions, a different approach to perturbation can be explored. One strategy involves perturbing the entire culture by manipulating the dilution rate of the chemostat, and subsequently monitoring the culture throughout the transition until a new steady-state is reached. This approach offers a straightforward means to study adaptations over an extended timescale.

This type of perturbation may be considered gentler than others, potentially not revealing all the attainable states of the system. Nevertheless, it provides a more detailed understanding of the underlying regulation during growth on a specific carbon source. Importantly, this perturbation impacts only a single exogenous variable, in contrast to the carbon source switch in our fast filtration experiment that affects at least two exogenous variables due to changes in both the original carbon source and the perfusion media composition.

A noteworthy aspect pertains to the media composition during cultivation. In our experiments, media preconditioning was not applied, leading to variations in the media composition. Consequently, transitioning to a different carbon source brought about alterations in metabolic by-products, proteins, and signaling molecules. To mitigate this effect, an alternative approach could be considered—modifying the inflow into the chemostat with an additional nutrient, for instance. However, this alteration would affect only a solitary exogenous variable within the system.

Incorporating enzyme levels as additional states in the model would escalate its complexity, particularly when considering dynamic changes beyond the initial minute. Nonetheless, continuous measurement alongside supplementary techniques like transcriptomics and proteomics could offset this complexity to some extent. Embracing increased complexity becomes paramount when aiming to delve into the intricacies of long-term adaptation.

Ultimately, delving into long-term adaptation necessitates navigating a balance between system complexity and the depth of insight gained, while simultaneously leveraging advancements in measurement techniques to enhance the accuracy and fidelity of experimental observations.

3.4.2.4 Input signals

In the context of system identification, it’s noteworthy that a precise knowledge of the perturbation mechanism isn’t an absolute requirement for comprehending resultant dynamics. What holds greater significance is the capacity to observe key states and the ability to perturb them, either directly or indirectly. This approach can yield the essential information necessary for understanding complex systems.

While the inherent nature of an input signal is crucial, it’s equally vital to take into account the frequency and amplitude of the signal. For instance, employing a single step- or pulse-like input, such as the glucose administration in our experiment, might not be exhaustive enough to unveil the complete spectrum of attainable system states and dynamic behaviors.

A compelling alternative is the use of a random binary input signal, a technique commonly embraced in systems identification methodologies [58]. This input signal possesses a unique attribute: it encompasses all the frequencies constituting the Bode plot at uniform amplitude levels, rendering it ideal for estimating frequency responses. Consequently, tracking cellular responses when subjected to diverse input signals—such as randomized doses of carbon sources administered at irregular intervals—can yield a wealth of valuable information.

Besides covering a large spectrum of frequencies, such observations may be considered more interesting from a biological perspective. This is because experimental conditions can be simulated that more closely resemble those *E. coli* may face in the wild, where cells are exposed to both gradual and sudden changes in their environment.

Within the realm of modeling procedures, the direct integration of measurement data as input holds notable advantages. As evident in our work, this approach facilitates the exploration of

extracellular perturbations on a metabolic subsystem without necessitating the inclusion of intricate descriptions linking the perturbing agent to the subsystem. The challenges associated with describing such connections can often be cumbersome and occasionally fraught with difficulties.

A pertinent illustration arises from the study by Christodoulou *et al.* (2018), which investigated the dynamics of the pentose phosphate pathway in response to reactive oxygen species (ROS). In this model, NADPH is produced by two reactions that are described by irreversible Michaelis-Menten kinetics, and scavenged by ROS. ROS is also present in the lumped reactions describing lower glycolysis, indirectly describing scavenging of NADH. Since NADPH is not part of any kinetic rate law description, the response of a model without regulation is characterized by a steep drop in NADPH, an almost insignificant effect on PEP, while the other metabolite levels remain constant (see Supplementary Figure S1 [17]). Only once a regulatory interaction is added to the model that links NADPH to the kinetics of other enzymes is the model capable of simulating dynamics in the rest of the network. Given that the model without regulation is treated as a null hypothesis to address the question of whether a model without regulation can sufficiently describe the observed dynamics, the answer was trivial.

This raises the question if the model without regulation is a proper null hypothesis to assess whether additional regulation is required. While there is no definitive way to answer what constitutes a better null hypothesis, it would have been interesting to assess to what extent the remaining states could be predicted using different combinations of metabolites as input signals. More specifically, different non-empty subsets of the species could have been assigned to follow the trajectory of its observed dynamics, and the question of whether the remaining dynamics can be sufficiently recapitulated could have been assessed in this context. This couples back to our earlier point of being able to decide what constitutes an input to the model, and which remaining states need to be estimated. Furthermore, the ability to assess the regulatory potential of metabolites outside of those described in the kinetic equations extends the range of putative regulators that can be assessed.

3.4.2.5 Evaluation metrics

The prediction of potential activators or inhibitors can be considered as a classification approach. This raises the question in what way we may best assess the quality of the predictions derived from our modeling approach. In statistical classification an error matrix is often used. True positives, in this regard, are the interactions that were predicted and either previously reported or validated in follow-up *in vitro* experiments, such as enzyme essays. Oppositely, false positives would be predicted allosteric interactions where an effect on enzyme kinetics cannot be observed experimentally. False positives can, among others, result from the correlation between metabolites, missing post-translation regulation, or the absence of an unknown but necessary cofactor in an enzyme essay. True negatives are interesting to consider, however, negatives are commonly not reported - even when reported they do not end up in the BRENDA enzyme database - and newly predicted negatives are generally not followed up on. Lastly there are false negatives, which would be interactions predicted not to be relevant in the model, but for which evidence in the literature exists.

There is an important distinction to be made here between interactions that exist and those that are relevant in a given condition. An allosteric interaction may simply not play an important regulatory role in one condition, yet be of regulatory importance in another. A reported interaction may also have resulted from the use of test tube conditions that do not resemble the physiological conditions, such as the use of concentrations that exceed those in the cell. A bias towards publishing positive results further complicates the situation by increasing the number of false positives and decreasing the number of negatives reported in studies. Ultimately, deciding what metric is meaningful depends on the specific use case and goals of the model.

Molecules in a cellular environment are in constant motion and their collision frequency is high since the intracellular environment is crowded. Yet, most of these physical interactions are (in)elastic collisions that do not lead to a chemical reaction or regulatory binding event. If we want to understand the behavior of central carbon metabolism, we want to find models containing the minimal complexity required not only to explain observations, but also to provide us with falsifiable predictions. In light of these considerations, true positives are likely the most relevant to consider as a metric to assess the quality or usefulness of our ensemble modeling approach. For example, the ratio of positive results among the top ranking predictions can be compared to the number of positives one would expect to find in a random selection of enzyme-metabolite pairs.

The best approximation of a random selection that is currently available are large-scale physical interaction studies [96, 25]. Using the overlap in the detected physical interactions as the ground truth, a sample size calculation can be performed to determine how many predictions need to be experimentally tested in order to address this question at a desired confidence level [28].

Bibliography

- [1] Guy Aidelberg et al. “Hierarchy of non-glucose sugars in *Escherichia coli*”. In: *BMC systems biology* 8.1 (2014), pp. 1–12.
- [2] Hirotugu Akaike. “A new look at the statistical model identification”. In: *IEEE transactions on automatic control* 19.6 (1974), pp. 716–723.
- [3] Rafael Alcántara et al. “Rhea—a manually curated resource of biochemical reactions”. In: *Nucleic acids research* 40.D1 (2012), pp. D754–D760.
- [4] Amos Bairoch. “The ENZYME database in 2000”. In: *Nucleic acids research* 28.1 (2000), pp. 304–305.
- [5] Ariel Balter et al. “The Glazier-Graner-Hogeweg model: extensions, future directions, and opportunities for further study”. In: *Single-Cell-Based Models in Biology and Medicine*. Springer, 2007, pp. 151–167.
- [6] Lucia Banci et al. “Affinity gradients drive copper to cellular destinations”. In: *Nature* 465.7298 (2010), pp. 645–648.
- [7] BP Belousov. “An oscillating reaction and its mechanism”. In: *Sborn. referat. radiat. med.* 145 (1959).
- [8] Giancarlo Benettin et al. “Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; a method for computing all of them. Part 1: Theory”. In: *Meccanica* 15.1 (1980), pp. 9–20.
- [9] Bryson D Bennett et al. “Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach”. In: *Nature protocols* 3.8 (2008), pp. 1299–1311.
- [10] Paul Blainey, Martin Krzywinski, and Naomi Altman. “Points of significance: replication”. In: *Nature methods* 11.9 (2014), p. 879.
- [11] Benjamin J Bornstein et al. “LibSBML: an API library for SBML”. In: *Bioinformatics* 24.6 (2008), pp. 880–881.
- [12] George Edward Briggs and John Burdon Sanderson Haldane. “A note on the kinetics of enzyme action”. In: *Biochemical journal* 19.2 (1925), p. 338.
- [13] Joerg Martin Buescher et al. “Ultrahigh performance liquid chromatography- tandem mass spectrometry method for fast and robust quantification of anionic and aromatic metabolites”. In: *Analytical chemistry* 82.11 (2010), pp. 4403–4412.
- [14] Marieke Francisca Buffing. “In Vivo and in Vitro Identification of Allosteric Enzyme Regulation by Metabolites”. PhD thesis. ETH Zurich, 2018.
- [15] Ron Caspi et al. “The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases”. In: *Nucleic acids research* 36.suppl_1 (2007), pp. D623–D631.
- [16] Christophe Chassagnole et al. “Dynamic modeling of the central carbon metabolism of *Escherichia coli*”. In: *Biotechnology and bioengineering* 79.1 (2002), pp. 53–73.
- [17] Dimitris Christodoulou et al. “Reserve flux capacity in the pentose phosphate pathway enables *Escherichia coli*’s rapid response to oxidative stress”. In: *Cell systems* 6.5 (2018), pp. 569–578.
- [18] Alain J Cozzone and Mansi El-Mansi. “Control of isocitrate dehydrogenase catalytic activity by protein phosphorylation in *Escherichia coli*”. In: *Microbial Physiology* 9.3-4 (2005), pp. 132–146.

-
- [19] David Croft et al. “Reactome: a database of reactions, pathways and biological processes”. In: *Nucleic acids research* 39.suppl_1 (2010), pp. D691–D697.
- [20] Dan Davidi et al. “Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k cat measurements”. In: *Proceedings of the National Academy of Sciences* 113.12 (2016), pp. 3401–3406.
- [21] Antony M Dean and Daniel E Koshland Jr. “Kinetic mechanism of Escherichia coli isocitrate dehydrogenase”. In: *Biochemistry* 32.36 (1993), pp. 9302–9309.
- [22] Antony M Dean, Andrew K Shiau, and Daniel E Koshland Jr. “Determinants of performance in the isocitrate dehydrogenase of Escherichia coli”. In: *Protein science* 5.2 (1996), pp. 341–347.
- [23] Kirill Degtyarenko et al. “ChEBI: a database and ontology for chemical entities of biological interest”. In: *Nucleic acids research* 36.suppl_1 (2007), pp. D344–D350.
- [24] Frank DeVilbiss and Doraiswami Ramkrishna. “Addressing the need for a model selection framework in systems biology using information theory”. In: *Proceedings of the IEEE* 105.2 (2016), pp. 330–339.
- [25] Maren Diether et al. “Systematic mapping of protein-metabolite interactions in central metabolism of Escherichia coli”. In: *Molecular systems biology* 15.8 (2019), e9008.
- [26] Ivan Dikic. “Proteasomal and autophagic degradation systems”. In: *Annual review of biochemistry* 86 (2017), pp. 193–224.
- [27] J-OC Dunn, MG Mythen, and MP Grocott. “Physiology of oxygen transport”. In: *Bja Education* 16.10 (2016), pp. 341–348.
- [28] William D Dupont and Walton D Plummer Jr. “Power and sample size calculations: a review and computer program”. In: *Controlled clinical trials* 11.2 (1990), pp. 116–128.
- [29] KJ Edwards and DA Bazylinski. “Intracellular minerals and metal deposits in prokaryotes”. In: *Geobiology* 6.3 (2008), pp. 309–317.
- [30] Adrian H Elcock. “Models of macromolecular crowding effects and the need for quantitative comparisons with experiment”. In: *Current opinion in structural biology* 20.2 (2010), pp. 196–206.
- [31] Vitaly Epshtein et al. “An allosteric mechanism of Rho-dependent transcription termination”. In: *Nature* 463.7278 (2010), pp. 245–249.
- [32] Karen van Eunen et al. “Testing biochemistry revisited: how in vivo metabolism can be understood from in vitro enzyme kinetics”. In: *PLoS computational biology* 8.4 (2012), e1002483.
- [33] Eoin Fahy et al. “Update of the LIPID MAPS comprehensive classification system for lipids1”. In: *Journal of lipid research* 50 (2009), S9–S14.
- [34] Avi Flamholz et al. “eQuilibrator—the biochemical thermodynamics calculator”. In: *Nucleic acids research* 40.D1 (2012), pp. D770–D775.
- [35] Yoav Freund and Robert E Schapire. “A decision-theoretic generalization of on-line learning and an application to boosting”. In: *Journal of computer and system sciences* 55.1 (1997), pp. 119–139.
- [36] Mathias Ganter et al. “MetaNetX. org: a website and repository for accessing, analysing and manipulating metabolic networks”. In: *Bioinformatics* 29.6 (2013), pp. 815–816.
- [37] Luca Gerosa et al. “Pseudo-transition analysis identifies the key regulators of dynamic metabolic adaptations from steady-state data”. In: *Cell systems* 1.4 (2015), pp. 270–282.
- [38] Jürgen Gmehling et al. *Chemical thermodynamics for process simulation*. John Wiley & Sons, 2019.
- [39] Jeremy Gunawardena. “Models in systems biology: the parameter problem and the meanings of robustness”. In: *Elements of computational systems biology* 1 (2010), pp. 21–47.
- [40] Apoorv Gupta et al. “Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit”. In: *Nature biotechnology* 35.3 (2017), pp. 273–279.

-
- [41] Sean R Hackett et al. “Systems-level analysis of mechanisms regulating yeast metabolic flux”. In: *Science* 354.6311 (2016).
- [42] Nikolaus Hansen and Andreas Ostermeier. “Completely derandomized self-adaptation in evolution strategies”. In: *Evolutionary computation* 9.2 (2001), pp. 159–195.
- [43] Hulda S Haraldsdóttir et al. “CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models”. In: *Bioinformatics* 33.11 (2017), pp. 1741–1743.
- [44] Christopher J Hartline et al. “Dynamic control in metabolic engineering: Theories, tools, and applications”. In: *Metabolic engineering* 63 (2021), pp. 126–140.
- [45] John Z Hearon. “Nonlinear diffusion in metabolic systems”. In: *The bulletin of mathematical biophysics* 15.1 (1953), pp. 15–21.
- [46] Laurent Heirendt et al. “Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0”. In: *Nature protocols* 14.3 (2019), pp. 639–702.
- [47] Matthias W Hentze and Andreas E Kulozik. “A perfect message: RNA surveillance and nonsense-mediated decay”. In: *Cell* 96.3 (1999), pp. 307–310.
- [48] Michael Hucka et al. “The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models”. In: *Bioinformatics* 19.4 (2003), pp. 524–531.
- [49] Michael Hucka et al. “The Systems Biology Markup Language (SBML): language specification for level 3 version 2 core”. In: *Journal of integrative bioinformatics* 15.1 (2018).
- [50] Michael Hucka et al. “The systems biology markup language (SBML): language specification for level 3 version 2 core release 2”. In: *Journal of integrative bioinformatics* 16.2 (2019).
- [51] James H Hurley et al. “Catalytic mechanism of NADP+-dependent isocitrate dehydrogenase: implications from the structures of magnesium-isocitrate and NADP+ complexes”. In: *Biochemistry* 30.35 (1991), pp. 8671–8678.
- [52] James H Hurley et al. “Regulation of an enzyme by phosphorylation at the active site”. In: *Science* 249.4972 (1990), pp. 1012–1016.
- [53] Patrick Ion et al. *Mathematical Markup Language (MathML) 1.0 Specification*. World Wide Web Consortium (W3C) Cambridge, MA, USA, 1998.
- [54] Nusrat Jahan et al. “Development of an accurate kinetic model for the central carbon metabolism of *Escherichia coli*”. In: *Microbial cell factories* 15.1 (2016), pp. 1–19.
- [55] Nick Juty and Nicolas le Novère. “Systems biology ontology”. In: *Encyclopedia of Systems Biology* (2013), pp. 2063–2063.
- [56] Tuty Asmawaty Abdul Kadir et al. “Modeling and simulation of the main metabolism in *Escherichia coli* and its several single-gene knockout mutants with experimental verification”. In: *Microbial cell factories* 9.1 (2010), pp. 1–21.
- [57] Minoru Kanehisa et al. “The KEGG database”. In: *Novartis found symp.* Vol. 247. 2002, pp. 91–103.
- [58] Karel J Keesman and Karel J Keesman. *System identification: an introduction*. Vol. 2. Springer, 2011.
- [59] Ali Khodayari et al. “A kinetic model of *Escherichia coli* core metabolism satisfying multiple sets of mutant flux data”. In: *Metabolic engineering* 25 (2014), pp. 50–62.
- [60] Daniela Kiekebusch and Martin Thanbichler. “Spatiotemporal organization of microbial cells by protein concentration gradients”. In: *Trends in microbiology* 22.2 (2014), pp. 65–73.
- [61] Zachary A King et al. “BiGG Models: A platform for integrating, standardizing and sharing genome-scale models”. In: *Nucleic acids research* 44.D1 (2016), pp. D515–D522.
- [62] Zachary A King et al. “Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways”. In: *PLoS computational biology* 11.8 (2015), e1004321.
- [63] Karl Kochanowski et al. “Few regulatory metabolites coordinate expression of central metabolic genes in *Escherichia coli*”. In: *Molecular Systems Biology* 13.1 (2017), p. 903.

-
- [64] Karl Kochanowski et al. “Global coordination of metabolic pathways in *Escherichia coli* by active and passive regulation”. In: *Molecular systems biology* 17.4 (2021), e10064.
- [65] HL Kornberg. “Anaplerotic sequences and their role in metabolism.” In: *Essays Biochem.* 2 (1966), pp. 1–31.
- [66] HL Kornberg and Hans Adolf Krebs. “Synthesis of cell constituents from C 2-units by a modified tricarboxylic acid cycle”. In: *Nature* 179.4568 (1957), pp. 988–991.
- [67] R Krishna and JA Wesselingh. “The Maxwell-Stefan approach to mass transfer”. In: *Chemical engineering science* 52.6 (1997), pp. 861–911.
- [68] Camille Laibe and Nicolas Le Novère. “MIRIAM Resources: tools to generate and resolve robust cross-references in Systems Biology”. In: *BMC Systems Biology* 1.1 (2007), pp. 1–9.
- [69] Jay I Lakkis and Mathew R Weir. “Obesity and kidney disease”. In: *Progress in cardiovascular diseases* 61.2 (2018), pp. 157–167.
- [70] Siu Kwan Lam, Antoine Pitrou, and Stanley Seibert. “Numba: A llvm-based python jit compiler”. In: *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*. 2015, pp. 1–6.
- [71] Eric O. Lebigot. *Uncertainties: a Python package for calculations with uncertainties*. <http://pythonhosted.org/uncertainties/>. Adding the version number is optional.
- [72] Wolfram Liebermeister and Edda Klipp. “Bringing metabolic networks to life: convenience rate law and thermodynamic constraints”. In: *Theoretical Biology and Medical Modelling* 3.1 (2006), pp. 1–13.
- [73] Wolfram Liebermeister, Jannis Uhlendorf, and Edda Klipp. “Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation”. In: *Bioinformatics* 26.12 (2010), pp. 1528–1534.
- [74] An-Ping Lin, Mark T McCammon, and Lee McAlister-Henn. “Kinetic and physiological effects of alterations in homologous isocitrate-binding sites of yeast NAD⁺-specific isocitrate dehydrogenase”. In: *Biochemistry* 40.47 (2001), pp. 14291–14301.
- [75] Hannes Link, Karl Kochanowski, and Uwe Sauer. “Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo”. In: *Nature biotechnology* 31.4 (2013), pp. 357–361.
- [76] Timo Lubitz et al. “Parameter balancing in kinetic models of cell metabolism”. In: *The Journal of Physical Chemistry B* 114.49 (2010), pp. 16298–16303.
- [77] Paul M Mathias et al. “Data quality and assessment, validation methods and error propagation through the simulation software: Report from the Round-Table Discussion at the 10th World Congress of Chemical Engineering in Barcelona (October 1-5, 2017)”. In: *Chemical engineering research & design: transactions of the Institution of Chemical Engineers* 137 (2018).
- [78] Andrew G McDonald and Keith F Tipton. “Parameter Reliability and Understanding Enzyme Function”. In: *Molecules* 27.1 (2022), p. 263.
- [79] Aaron Meurer et al. “SymPy: symbolic computing in Python”. In: *PeerJ Computer Science* 3 (2017), e103.
- [80] Pierre Millard, Kieran Smallbone, and Pedro Mendes. “Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in *Escherichia coli*”. In: *PLoS computational biology* 13.2 (2017), e1005396.
- [81] Ron Milo et al. “BioNumbers—the database of key numbers in molecular and cell biology”. In: *Nucleic acids research* 38.suppl_1 (2010), pp. D750–D753.
- [82] Jens Nielsen and Jay D Keasling. “Engineering cellular metabolism”. In: *Cell* 164.6 (2016), pp. 1185–1197.
- [83] Hugh G Nimmo. “Kinetic mechanism of *Escherichia coli* isocitrate dehydrogenase and its inhibition by glyoxylate and oxaloacetate”. In: *Biochemical Journal* 234.2 (1986), pp. 317–323.
- [84] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 1999.

-
- [85] Elad Noor et al. “Pathway thermodynamics highlights kinetic obstacles in central metabolism”. In: *PLoS computational biology* 10.2 (2014), e1003483.
- [86] Brian Olson et al. “Basin Hopping as a General and Versatile Optimization Framework for the Characterization of Biological Macromolecules.” In: *Advances in Artificial Intelligence (16877470)* (2012).
- [87] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. “What is flux balance analysis?” In: *Nature biotechnology* 28.3 (2010), pp. 245–248.
- [88] Jeffrey D Orth et al. “A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011”. In: *Molecular systems biology* 7.1 (2011), p. 535.
- [89] Natalia Ostrowska, Michael Feig, and Joanna Trylska. “Modeling crowded environment in molecular simulations”. In: *Frontiers in molecular biosciences* 6 (2019), p. 86.
- [90] Ross Overbeek et al. “The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)”. In: *Nucleic acids research* 42.D1 (2014), pp. D206–D214.
- [91] Marta Pajares et al. “Redox control of protein degradation”. In: *Redox biology* 6 (2015), pp. 409–420.
- [92] Ellen A Panisko and Lee McAlister-Henn. “Subunit interactions of yeast NAD⁺-specific isocitrate dehydrogenase”. In: *Journal of Biological Chemistry* 276.2 (2001), pp. 1204–1210.
- [93] Natalya N Pavlova and Craig B Thompson. “The emerging hallmarks of cancer metabolism”. In: *Cell metabolism* 23.1 (2016), pp. 27–47.
- [94] Kirill Peskov, Ekaterina Mogilevskaya, and Oleg Demin. “Kinetic modelling of central carbon metabolism in Escherichia coli”. In: *The FEBS journal* 279.18 (2012), pp. 3374–3385.
- [95] Prashant S Phale, Harshit Malhotra, and Bhavik A Shah. “Degradation strategies and associated regulatory mechanisms/features for aromatic compound metabolism in bacteria”. In: *Advances in Applied Microbiology* 112 (2020), pp. 1–65.
- [96] Ilaria Piazza et al. “A map of protein-metabolite interactions reveals principles of chemical communication”. In: *Cell* 172.1-2 (2018), pp. 358–372.
- [97] Cecile M Pickart. “Mechanisms underlying ubiquitination”. In: *Annual review of biochemistry* 70.1 (2001), pp. 503–533.
- [98] William H Press. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.
- [99] Michael C Reed, Anna Lieb, and H Frederik Nijhout. “The biological significance of substrate inhibition: a mechanism with diverse functions”. In: *Bioessays* 32.5 (2010), pp. 422–429.
- [100] Ed Reznik et al. “Genome-scale architecture of small molecule regulatory networks and the fundamental trade-off between regulation and enzymatic activity”. In: *Cell reports* 20.11 (2017), pp. 2666–2677.
- [101] Brendan Michael Ryback. “Systematic identification of regulatory interactions between signaling and metabolic networks in *Saccharomyces cerevisiae*”. PhD thesis. ETH Zurich, 2020.
- [102] Fatemeh Saberi et al. “Natural antisense RNAs as mRNA regulatory elements in bacteria: a review on function and applications”. In: *Cellular & molecular biology letters* 21.1 (2016), pp. 1–17.
- [103] Andrea Saltelli et al. “Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices”. In: *Environmental modelling & software* 114 (2019), pp. 29–39.
- [104] Herbert M Sauro. *Enzyme kinetics for systems biology*. Future Skill Software, 2011.
- [105] Jan Schellenberger et al. “BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions”. In: *BMC bioinformatics* 11.1 (2010), pp. 1–10.
- [106] Alexander Schmidt et al. “The quantitative and condition-dependent Escherichia coli proteome”. In: *Nature biotechnology* 34.1 (2016), pp. 104–110.
- [107] Ida Schomburg, Antje Chang, and Dietmar Schomburg. “BRENDA, enzyme data and metabolic information”. In: *Nucleic acids research* 30.1 (2002), pp. 47–49.

-
- [108] Heide N Schulz and Bo Barker Jørgensen. “Big bacteria”. In: *Annual Reviews in Microbiology* 55.1 (2001), pp. 105–137.
- [109] Daniel Segre, Dennis Vitkup, and George M Church. “Analysis of optimality in natural and perturbed metabolic networks”. In: *Proceedings of the National Academy of Sciences* 99.23 (2002), pp. 15112–15117.
- [110] Maya Shamir et al. “SnapShot: timescales in cell biology”. In: *Cell* 164.6 (2016), pp. 1302–1302.
- [111] Siowling Soh et al. “Reaction-diffusion systems in intracellular molecular transport and control”. In: *Angewandte Chemie International Edition* 49.25 (2010), pp. 4170–4198.
- [112] Endre T Somogyi et al. “libRoadRunner: a high performance SBML simulation and analysis library”. In: *Bioinformatics* 31.20 (2015), pp. 3315–3321.
- [113] Mandavilli Srinivas and Lalit M Patnaik. “Adaptive probabilities of crossover and mutation in genetic algorithms”. In: *IEEE Transactions on Systems, Man, and Cybernetics* 24.4 (1994), pp. 656–667.
- [114] Katie L Stewart, Andrew M Stewart, and Thomas A Bobik. “Prokaryotic organelles: bacterial microcompartments in *E. coli* and *Salmonella*”. In: *EcoSal Plus* 9.1 (2020).
- [115] Rainer Storn and Kenneth Price. “Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces”. In: *Journal of global optimization* 11.4 (1997), pp. 341–359.
- [116] Shiho Tanaka, Michael R Sawaya, and Todd O Yeates. “Structure and mechanisms of a protein-based organelle in *Escherichia coli*”. In: *Science* 327.5961 (2010), pp. 81–84.
- [117] Peter E Thorsness and Daniel E Koshland. “Inactivation of isocitrate dehydrogenase by phosphorylation is mediated by the negative charge of the phosphate.” In: *Journal of Biological Chemistry* 262.22 (1987), pp. 10422–10425.
- [118] Daisuke Tsuchiya, Nobutaka Shimizu, and Masaru Tomita. “Cooperativity of two active sites in bacterial homodimeric aconitases”. In: *Biochemical and biophysical research communications* 379.2 (2009), pp. 485–488.
- [119] Charles L Turnbough Jr. “Regulation of bacterial gene expression by transcription attenuation”. In: *Microbiology and Molecular Biology Reviews* 83.3 (2019), e00019–19.
- [120] Wouter RL Van Der Star et al. “An intracellular pH gradient in the anammox bacterium *Kuenenia stuttgartiensis* as evaluated by ³¹P NMR”. In: *Applied microbiology and biotechnology* 86.1 (2010), pp. 311–317.
- [121] Benjamin Volkmer and Matthias Heinemann. “Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling”. In: *PloS one* 6.7 (2011), e23126.
- [122] Larry Wasserman. *All of statistics: a concise course in statistical inference*. Vol. 26. Springer, 2004.
- [123] J Kay-PDN Weitzman. “Krebs citric acid cycle: half a century and still turning”. In: *Biochem. Soc. Symp.* Vol. 54. 1987, pp. 1–198.
- [124] JA Wesselingh, P Vonk, and G Kraaijeveld. “Exploring the Maxwell-Stefan description of ion exchange”. In: *The Chemical Engineering Journal and The Biochemical Engineering Journal* 57.2 (1995), pp. 75–89.
- [125] Jessica C Wilks and Joan L Slonczewski. “pH of the cytoplasm and periplasm of *Escherichia coli*: rapid measurement by green fluorescent protein fluorimetry”. In: *Journal of bacteriology* 189.15 (2007), pp. 5601–5607.
- [126] Wade C Winkler and Ronald R Breaker. “Regulation of bacterial gene expression by riboswitches”. In: *Annu. Rev. Microbiol.* 59 (2005), pp. 487–517.
- [127] David S Wishart et al. “HMDB: the human metabolome database”. In: *Nucleic acids research* 35.suppl_1 (2007), pp. D521–D526.
- [128] Xu Yan et al. “Metabolic dysregulation contributes to the progression of Alzheimer’s disease”. In: *Frontiers in Neuroscience* 14 (2020), p. 1107.

-
- [129] Qin Yang, Archana Vijayakumar, and Barbara B Kahn. “Metabolites as regulators of insulin sensitivity and metabolism”. In: *Nature reviews Molecular cell biology* 19.10 (2018), pp. 654–672.
- [130] Katja Zieske and Petra Schulle. “Reconstitution of self-organizing protein gradients as spatial cues in cell-free systems”. In: *Elife* 3 (2014), e03949.
- [131] Dan Zilberstein et al. “Escherichia coli intracellular pH, membrane potential, and cell growth”. In: *Journal of bacteriology* 158.1 (1984), pp. 246–252.
- [132] Michael Zimmermann, Uwe Sauer, and Nicola Zamboni. “Quantification and mass isotopomer profiling of α -keto acids in central carbon metabolism”. In: *Analytical chemistry* 86.6 (2014), pp. 3232–3237.

Chapter 4

Conclusions and outlook

M.A.P. Karrenbelt wrote the chapter.
Prof. Dr. Uwe Sauer supervised and contributed to the writing.

4.1 Conclusions

In this thesis, we embarked on a journey to unravel the intricate regulatory mechanisms governing central carbon metabolism in *Escherichia coli*. Guided by a systems biology perspective, we addressed key challenges in understanding the dynamic behavior of this essential cellular process. Our contributions can be summarized as follows:

- We began by revisiting the foundations of systems biology and critically analyzing the prevailing definitions in the field. Through a meticulous exploration of its history and philosophy, we attempt to establish an operational framework that provides a clear and coherent conceptualization of systems biology. This foundational step allowed us to navigate subsequent investigations within a well-defined systems biology perspective.
- In our first study we used a single-reaction modeling approach, which we used to predict metabolite-protein interactions from steady-state data. Analogous to the way in which experimental measurement of metabolite variation *in vitro* is used for the inference of regulators and enzyme kinetic parameters, physiological changes in fluxes bring about a change in metabolite concentrations that can be used to infer reaction kinetics from cellular data. Using measurement data on metabolic fluxes, enzyme and metabolite concentrations across multiple series of steady-state growth conditions, we determined how well a Michaelis-Menten equation was able to capture these observations, and to what extent the inclusion of allosteric regulatory interactions increases this ability. A significant benefit is that the physiological importance of an allosteric regulator is implicit from its role in determining metabolic flux. Following up on the top ranking allosteric interactions predicted, we found experimental evidence for the existence of 11 new interactions.

Contextualizing these findings involves considering the broader landscape of knowledge. Comparing the 11 new interactions to previous research, where Reznik *et al.* extracted 1699 metabolite-enzyme interactions involving 321 metabolites and 364 enzymes for *E. coli* from databases, the addition of 11 interactions might appear initially modest in scale [4]. The EcoCyc database, as it stands, documents around 100 regulatory interactions related to the 35 major isoenzymes of central metabolism [2].

However, the significance of these 11 new interactions becomes more apparent when considered in light of Diether *et al.*'s NMR-based physical interaction study [1]. This study aimed to enhance our understanding of complex protein-metabolite interactions within central carbon metabolism. By systematically analyzing ligand-detected NMR profiles of 29 purified enzymes from *E. coli* central metabolism and 55 selected metabolites, they explored a total of 1595 potential interactions. Notably, their investigation uncovered a total of 98 interactions, with 76 of them being newly predicted. The identification of a physical interaction is a *sine qua non* for recognizing an allosteric regulator of an enzyme. Consequently, it is only logical to anticipate a higher count of physical interactions compared to allosteric regulators. In this light, the hit rate of 11 out of 22 stands as remarkably elevated when contrasted with the 92 out of 1595 interactions unveiled by Diether *et al.* [1].

In the exploration of more distantly related pairs, it's noteworthy to consider the work by Piazza *et al.*, which employs limited proteolysis-coupled mass spectrometry (LiP-MS) to examine protein structural changes and perturbations in a whole proteome [3]. It provides a comprehensive view of the bacterial metabolite-protein interactome, shedding light on both known and novel binding events that may govern various cellular processes. The work's peptide-level resolution approach also offers valuable insights into the promiscuity of enzyme active sites and the potential functional relevance of binding sites beyond active sites. In our work we also find more distally related pairs, such as the inhibition of AroG and PfkA by indolepyruvate, inhibition of serB by hypoxanthine, and the inhibition of purA by D-pantothenate and N-Acetyl-L-glutamate.

In evaluating why 11 new interactions are noteworthy despite their seemingly limited number, it's important to recognize that they represent an incremental expansion of our knowledge within a well-studied domain. The fact that these interactions were successfully predicted and subsequently supported by experimental evidence highlights the precision and relevance of the modeling approach. Moreover, given that the interactions were discovered within central carbon metabolism, a core aspect of cellular function, their impact on cellular processes could

be substantial. By integrating these data-driven predictions into a comprehensive framework, we bridge the gap between physical interactions and their potential functional roles. This approach places these findings in a different category as those of functional studies that rely solely on *in vitro* validation, as it offers both experimental evidence and the foundation for plausible *in vivo* functional impact.

It's noteworthy that among these 11 interactions, alignment in terms of their *modus operandi* (i.e., inhibition or activation) was observed in only 9 instances. To bolster the robustness of these findings, additional replication or, ideally, orthogonal evidence is imperative. At the forefront, the existence of these interactions propels the notion that allosteric regulation may wield a more extensive influence on CCM than previously envisioned. The potential contribution of distantly connected enzyme-metabolite pairs, as yet largely unexplored, holds promise for further research, unveiling uncharted territories within this regulatory network.

Nonetheless, we acknowledge that while these interactions are compelling, their physiological relevance remains speculative when considered in isolation. The exclusive reliance on single-reaction models and static network analyses risks presenting an incomplete perspective. Importantly, the predictions stemming from these newly found interactions can serve as valuable guiding principles for constructing and refining these larger, more comprehensive dynamic models. By integrating these predictions into the models, we can strategically focus our explorations, enabling simulations that illuminate how these interactions influence the behavior of the entire system over time.

- In the second phase of our study, we focused on constructing a comprehensive model of the TCA cycle using coupled differential equations. Our approach aimed to ensure reliability, reusability, and reproducibility in model generation through a systematic pipeline. By simulating metabolic dynamics in response to a carbon source switch, we investigated the role of allosteric interactions in shaping these responses. However, while our effort in establishing a standardized model construction process is commendable, we fell short of reaching substantial biological insights within the given timeframe.

Navigating the intricate landscape of model construction prompts a pivotal question: how do we harmonize models that yield insights, albeit challenging to reproduce, with meticulously structured models that fall short in delivering substantial biological insights? From my standpoint, the essence of science inherently resides in reproducibility. A model's validity hinges on its ability to be independently verified, which is a fundamental element of the scientific method. Complexity in models is indeed a double-edged sword – while it can enhance fidelity, it can also hinder applicability. This underlines the true challenge: finding the equilibrium between complexity and functionality. Yet, the pursuit of this balance isn't a binary choice; it's a dynamic process. It involves optimizing the trade-off between complexity and functionality at different research stages. During the initial phases, emphasizing functionality and feasibility is paramount. Simplified models, while grounded in a larger number of assumptions and less extensive data reliance, can pave the way for the development of a more comprehensive functional system. This approach facilitates the rapid creation of a starting point, which can then be systematically and iteratively enhanced over time. A simpler model, albeit heavily relying on assumptions, remains valuable as long as it's grounded in rigorous reproducibility. This approach enables us to progressively delve into the intricate details, scrutinize assumptions, and dissect the impact of various factors. By embracing a dynamic and iterative approach, we can navigate the complexities of scientific inquiry while nurturing a foundation rooted in rigorous reproducibility.

This leads us to a fundamental question: even in a scenario where our data were ideal, would the standardized model construction be the sole determinant of success? The reality is more nuanced. The journey of scientific discovery is riddled with complexities that extend beyond model construction. The impeccable quality of data, while important, is not the sole factor determining success. Next hurdles emerge, including the incorporation of biological context and validation against a wider array of experimental observations. This brings to light the necessity of a broader perspective and engagement with ongoing research, methodologies, and insights. Our work, which predominantly centered around model construction, is merely the initial step in a much larger process. Demonstrating the full potential of our standardized pipeline requires a multifaceted approach. While the pipeline itself streamlines construction, a true demonstration would involve showcasing how it accelerates the discovery process.

This entails demonstrating the pipeline's efficacy in revealing novel biological insights that can guide experimental studies and facilitate a deeper understanding of complex cellular systems.

- Contributions to model construction:
 - Model construction framework: A significant advancement achieved in this work was the development of an automated model construction framework. This approach harnessed the capabilities of various libraries, enabling us to build models with a high degree of reliability and consistency. By automating the model construction process, we eliminated the risk of manual errors and ensured adherence to the SBML specifications, resulting in standardized models that meet the highest quality standards.
 - Model annotation: The adoption of a standardized identifier system, exemplified by the BiGG database, coupled with the integration of external database identifiers, marked a pivotal enhancement in model annotation. This approach provides accurate and consistent annotation of model entities, a critical step that streamlined automated data retrieval, which improves the model's reusability and interpretation of the model.
 - Rate law construction: another improvement of this work was the implementation of a programmatic framework for constructing rate laws. This approach replaces error-prone manual procedures with automated processes, ensuring the precision and reliability of rate law formulation. Although not explored in this work, the framework's flexibility allows the exploration of a diverse range of rate laws, extending its utility across various biochemical systems.
 - Initialization strategies: our research redefined model initialization by leveraging available data on species, kinetic parameters and equilibrium constants. This shift yielded multiple benefits, including the establishment of thermodynamically consistent initial conditions and the utilization of more accurate and informed data-driven sampling strategies.

To conclude, this thesis earnestly explores the intricate regulatory mechanisms governing central carbon metabolism in *E. coli*. Guided by a systems biology perspective, we address key challenges in understanding the dynamic behavior of this essential cellular process. However, it's important to note that the vast complexities of biological systems are far from fully understood, and our work offers a modest contribution to the ongoing discourse. The journey embarked upon in this thesis is not a final destination but rather a stepping stone in the continuous quest to uncover the mysteries of life at the molecular level. Recognizing the expansive terrain that lies ahead, we strive to contribute to a foundational understanding that may inspire future exploration in the realms of systems biology and metabolic research.

In this context, I wish to extend my heartfelt appreciation to the "Computational Modeling in Biology" Network (COMBINE, <https://co.mbine.org/>). Their commendable efforts in coordinating the development of community standards, formats, and software tools in systems biology and related fields (<http://co.mbine.org/standards>) deserve significant credit and recognition. I believe that their work is instrumental in advancing our collective understanding and capabilities in the realm of computational biology.

In summary, this thesis reflects our modest attempt to contribute to the ongoing exploration of central carbon metabolism within the framework of systems biology.

4.2 Outlook

While our research primarily concentrated on unveiling allosteric interactions within the TCA cycle, the broader landscape of central carbon metabolism presents numerous opportunities for further exploration. Future studies could expand upon our findings to investigate the regulatory intricacies in other metabolic pathways, bridging different regulatory layers and considering longer time scales. Additionally, incorporating additional experimental data and advanced modeling techniques could enhance the accuracy of predictive models and provide a more comprehensive understanding of metabolic responses. Furthermore, as this research sheds light on the intricate regulatory mechanisms of central carbon metabolism in *E. coli*, several promising avenues for future investigations emerge:

- **Incorporating post-translational modifications:** To enhance our understanding of allosteric interactions and their modulation, investigating the impact of post-translational modifications (PTMs) is an exciting avenue. PTMs, such as phosphorylation and acetylation, play a crucial role in fine-tuning protein function and interaction dynamics. Incorporating PTMs into the modeling framework could uncover novel regulatory mechanisms and their influence on metabolic dynamics.
- **Integration of transcriptional regulation:** While this study predominantly focuses on the post-translational regulation of metabolic pathways, the integration of transcriptional regulation remains a critical aspect. Future research could explore how transcriptional and post-translational control mechanisms cooperate to orchestrate the metabolic response to environmental changes and cellular demands. This could involve the development of integrated models that encompass both layers of regulation, providing a more holistic view of cellular behavior.
- **Dynamic responses to multiple perturbations:** The robustness and adaptability of metabolic networks are evident in their responses to various perturbations. Future studies could explore how central carbon metabolism dynamically responds to multiple environmental changes, such as shifts in nutrient availability and temperature variations. Investigating these scenarios could reveal how the network's regulatory mechanisms prioritize different metabolic objectives under diverse conditions, providing insights into cellular versatility.
- **Advanced model validation techniques:** Develop and employ advanced validation techniques for models, such as uncertainty quantification and sensitivity analysis. These techniques help you identify critical model parameters that should be better characterized experimentally and provide insights into the variability in model predictions under different conditions.
- **Machine learning approaches:** Integrate machine learning approaches to analyze complex metabolic datasets and identify novel regulatory patterns. Machine learning algorithms can uncover hidden relationships and assist in predicting allosteric interactions based on large-scale data analysis. By analyzing complex metabolic datasets using machine learning, one can generate hypotheses about regulatory patterns, interactions, and dependencies that were not previously considered. They can guide the formulation of new mechanistic models or the modification of existing ones, allowing you to test and refine your assumptions based on data-derived patterns.
- **Synthetic biology applications:** Design synthetic circuits or genetic modifications that exploit allosteric interactions to engineer desired metabolic outcomes, such as increased production of specific metabolites or improved pathway efficiency. For example, one may design synthetic genetic circuits that enable the controlled expression of allosteric regulators in response to specific cues, such as specific external signals.

These directions for future research have the potential to deepen our comprehension of cellular regulation and contribute to broader advancements in the field of systems biology. In moving forward, it is essential to build upon the recommendations provided in earlier chapters without reiterating them, ensuring a seamless integration of ideas as we continue to uncover the complexities of cellular regulation.

4.3 Personal reflections

Throughout this research journey, I've embarked on a quest to unravel the intricate mechanisms of central carbon metabolism in *E. coli* and navigated the labyrinthine complexities of systems biology. How much this expedition deepened my understanding of cellular regulation pales in comparison with the invaluable insights it has provided into the challenges and intricacies that characterize the landscape of scientific research and academia.

Indeed, my foray into the world of systems biology has been a double-edged sword—while I haven't gleaned transcendental knowledge about metabolic regulation, my path has been illuminated by a different facet of discovery. It is the academic system itself, with its aspirations, pitfalls, and quirks, that I've come to understand with an ever-deepening clarity.

In engaging with the academic community, reviewing research proposals, and reflecting upon the scientific process, I've witnessed both the brilliance of inquiry and the shadows that frequently accompany it. I've come face-to-face with the underbelly of academia—the tangled webs of incentive structures, the reproducibility crisis, and the interplay between genuine discovery and the strategic pursuit of recognition.

4.3.1 A fractured foundation

Amidst the grandeur of scientific pursuit and the quest for knowledge lies a disquieting truth—an erosion of trust in the reproducibility of scientific findings. The reproducibility crisis, a widespread phenomenon plaguing various fields, exposes the vulnerability of the scientific method to biases, errors, and undue influences.

In the heart of scientific inquiry, the ability to replicate and verify results is a cornerstone of credibility. Yet, unsettlingly, a growing body of evidence suggests that many research findings are difficult to reproduce, casting shadows of doubt over the reliability of published studies. Whether due to undisclosed methodological choices, data manipulation, publication bias, or the failure to publish raw data and the code used to analyze it altogether, the reproducibility crisis challenges the very bedrock upon which science is built.

This crisis is not a mere academic abstraction; it resonates deeply with the scientific community, stakeholders, and society at large. It prompts introspection about the systemic factors that contribute to this predicament and compels us to address the root causes. Is it the pressure to publish prolifically, the allure of striking but unsubstantiated results, or the lack of incentives for reproducibility that fuels this phenomenon?

The reproducibility crisis underscores the need for transparency, rigor, and humility in scientific practice. It exhorts us to embrace open science, where data, methodologies, and analyses are shared openly, fostering a culture of collaboration and accountability. Moreover, it compels us to reconsider the incentive structures that shape the conduct of research, steering us away from the perilous path of prioritizing quantity over quality.

4.3.2 A dichotomy of motivations

Within the realm of academia, the pursuit of knowledge and the propagation of truth have stood as guiding principles. However, the incentive structures that underlie academic research often bear the weight of a more complex reality. The duality of motivations within these structures can shape research trajectories, yielding both inspiring advancements and concerning pitfalls.

The pressure to publish prolifically and secure research funding can inadvertently steer researchers towards a precarious path. Quantity can take precedence over quality, fostering a culture where novelty and attention-grabbing results are prioritized over rigorous investigation. This inclination towards eye-catching findings can lead to a distortion of priorities, encouraging the pursuit of provocative results.

In this landscape, a potential misalignment arises between the pursuit of truth and the quest for recognition. The allure of high-impact publications, prestigious awards, and coveted tenure positions can eclipse the fundamental responsibility to ensure the validity and reproducibility of research findings and undermine the scientific process itself.

This phenomenon echoes the sentiment of "trust, but verify"¹, a guiding principle in the scientific world. It highlights the importance of skepticism and the critical examination of findings,

¹From the Russian proverb: Первое правило: доверяй, но проверяй, which translates to "Rule number one: trust but verify", made popular by Ronald Reagan, who adopted it during the Cold War era.

regardless of the source. However, in the modern research environment, the adage may face challenges in practice, as the pressures of the academic system can lead to a somewhat paradoxical situation—where verification is overshadowed by the need to move forward rapidly.

While the incentive structures that govern academia may at times appear flawed, they also hold the potential for rejuvenation. The recent emergence of the Web3 movement with its maxim of "don't trust, verify" offers a glimpse of hope and a potential blueprint for reshaping the incentive structures of academia. As decentralized network technologies gain prominence, the emphasis on transparency, decentralization, and collaboration provides a fresh perspective on how systems can be designed to prioritize truth over expedience.

Bibliography

- [1] Maren Diether et al. “Systematic mapping of protein-metabolite interactions in central metabolism of *Escherichia coli*”. In: *Molecular systems biology* 15.8 (2019), e9008.
- [2] Ingrid M Keseler et al. “The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12”. In: *Nucleic acids research* 45.D1 (2017), pp. D543–D550.
- [3] Ilaria Piazza et al. “A map of protein-metabolite interactions reveals principles of chemical communication”. In: *Cell* 172.1-2 (2018), pp. 358–372.
- [4] Ed Reznik et al. “Genome-scale architecture of small molecule regulatory networks and the fundamental trade-off between regulation and enzymatic activity”. In: *Cell reports* 20.11 (2017), pp. 2666–2677.

Appendix A

Appendix

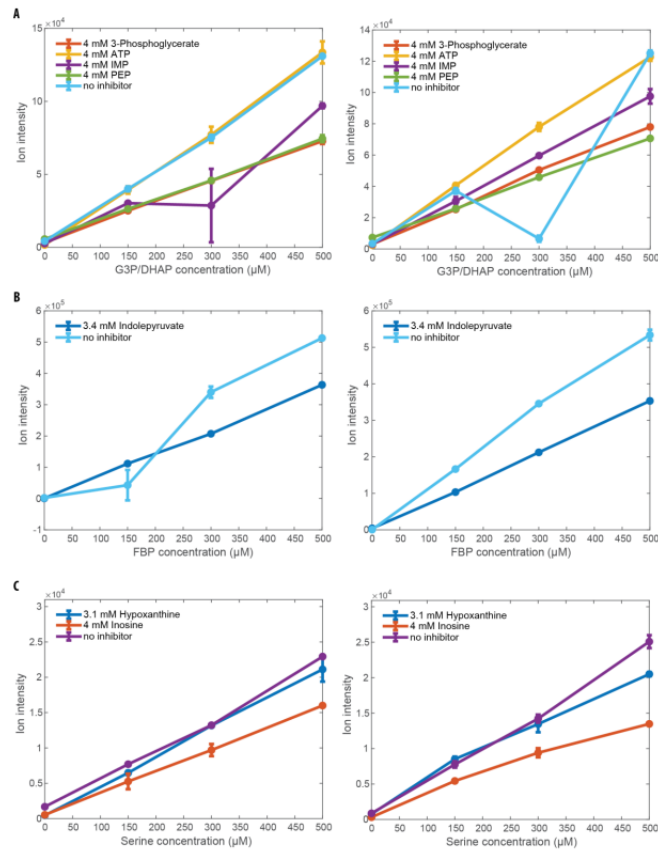


Figure A.1: Calibration curves to correct the effect of the regulation presence in the FIA-TOF. A. Duplicates of calibration curves of the FbaA products glyceraldehyde-3-phosphate and dihydroxyacetone-phosphate (1:1). **B.** Duplicates of calibration curves of the PfkA product fructose-1,6-bisphosphate. **C.** Duplicates of calibration curves of the SerB product serine. The ratio of the slope between curves with and without the regulator present is used as the correction factor.

A.1 Supplement Chapter 2

A.2 Supplement Chapter 3

Table A.1: Reaction data coverage

reaction	formula	enzymes	substrates	products
A5PISO	ru5p_D_c <=> ara5p_c	kdsD	ru5p_D	
ACACT1r	2.0 accoa_c <=> aacoa_c + coa_c	fadA	accoa	coa
ACACT2r	accoa_c + btcoa_c <=> 3ohcoa_c + coa_c	fadA	accoa, btcoa	coa
ACGK	acglu_c + atp_c -> acg5p_c + adp_c	argB	acglu, atp	adp
ACGS	accoa_c + glu_L_c -> acglu_c + coa_c + h_c	argA	accoa, glu_L	acglu, coa
ACONTa	cit_c <=> acon_C_c + h2o_c	acnA, acnB	cit	acon_C
ACONTb	acon_C_c + h2o_c <=> icit_c	acnA, acnB	acon_C	
ADSL1r	dcamp_c <=> amp_c + fum_c	purB	dcamp	amp, fum
ADSS	asp_L_c + gtp_c + imp_c -> dcamp_c + gdp_c + 2.0 h_c + pi_c	purA	asp_L, gtp, imp	dcamp, gdp
AGMHE	adpheap_DD_c -> adpheap_LD_c	hldD	adpheap_DD	
AIRC3	5aize_c <=> 5caiz_c	purE	5aize	
AKGDH	akg_c + coa_c + nad_c -> co2_c + nadh_c + succoa_c	lpdA, sucB, sucA	akg, coa, nad	nadh, succoa
ARGSS	asp_L_c + atp_c + citr_L_c -> amp_c + argsuc_c + h_c + ppi_c	argG	asp_L, atp, citr_L	amp, ppi
ASP1DC	asp_L_c + h_c -> ala_B_c + co2_c	panD	asp_L	
ASPECT	asp_L_c + cbp_c -> chasp_c + h_c + pi_c	pyrB	asp_L, cbp	
ASPK	asp_L_c + atp_c <=> 4pasp_c + adp_c	thrA, metL, lysC	asp_L, atp	4pasp, adp
ASPTA	akg_c + asp_L_c <=> glu_L_c + oaa_c	aspC	akg, asp_L	glu_L
ATPPPT	atp_c + prpp_c -> ppi_c + prbatp_c	hisG	atp, prpp	ppi
CPMPS	gtp_c + h2o_c -> cpmp_c + ppi_c	moaC	gtp	ppi
CTPS2	atp_c + glu_L_c + h2o_c + utp_c -> adp_c + ctp_c + glu_L_c + 2.0 h_c + pi_c	pyrG	atp, glu_L, utp	adp, ctp, glu_L
CYTK1	atp_c + cmp_c <=> adp_c + cdp_c	cmk	atp, cmp	adp, cdp
DDPA	e4p_c + h2o_c + pep_c -> 2dda7p_c + pi_c	aroF, aroG, aroH	e4p, pep	2dda7p
DHBS	23dlb_c + atp_c + h_c -> 23dhba_c + ppi_c	entE	23dlb, atp	ppi
DHORTS	dhor_S_c + h2o_c <=> chasp_c + h_c	pyrC	dhor_S	
DHQS	2dda7p_c -> 3dhq_c + pi_c	aroB	2dda7p	3dhq
DHQT1	3dhq_c -> 3dhsk_c + h2o_c	aroD	3dhq	3dhsk
DUTPDP	dutp_c + h2o_c -> dump_c + h_c + ppi_c	dut	dutp	ppi
E4PD	e4p_c + h2o_c + nad_c <=> 4per_c + 2.0 h_c + nadh_c	gapA	e4p, nad	nadh
FBA	fdp_c <=> dhap_c + g3p_c	fbaB, fbaA	fdp	dhap
FUM	fum_c + h2o_c <=> mal_L_c	fumC, fumA	fum	mal_L
G3PD2	glyc3p_c + nadp_c <=> dhap_c + h_c + nadph_c	gpsA	glyc3p, nadp	dhap, nadph
G6PDH2r	g6p_c + nadp_c <=> 6pgl_c + h_c + nadph_c	zwf	g6p, nadp	6pgl, nadph
GF6PTA	f6p_c + gln_L_c -> gam6p_c + glu_L_c	glmS	f6p, gln_L	gam6p, glu_L
GK1	atp_c + gmp_c <=> adp_c + gdp_c	gmK	atp, gmp	adp, gdp
GLU5K	atp_c + glu_L_c -> adp_c + glu5p_c	proB	atp, glu_L	adp
GLUDy	glu_L_c + h2o_c + nadp_c <=> akg_c + h_c + nadph_c + nh4_c	gdhA	glu_L, nadp	akg, nadph
GLUPRT	glu_L_c + h2o_c + prpp_c -> glu_L_c + ppi_c + pram_c	purF	glu_L, prpp	glu_L, ppi, pram
GLYK	atp_c + glyc_c -> adp_c + glyc3p_c + h_c	glpK	atp, glyc	adp, glyc3p
GND	6pgc_c + nadp_c -> co2_c + nadph_c + ru5p_D_c	gnd	6pgc, nadp	nadh, ru5p_D
GTPCI	gtp_c + h2o_c -> ahdt_c + for_c + h_c	folE	gtp	
HISTP	h2o_c + hisp_c -> histd_c + pi_c	hisB	hisp	
IMPC	h2o_c + imp_c <=> fprica_c	purH	imp	
IMPD	h2o_c + imp_c + nad_c -> h_c + nadh_c + xmp_c	guaB	imp, nad	nadh
IPPM1b	2ippm_c + h2o_c <=> 3c3hmp_c	leuC, leuD	2ippm	
IPPS	3mob_c + accoa_c + h2o_c -> 3c3hmp_c + coa_c + h_c	leuA	3mob, accoa	coa
KDOPP	h2o_c + kdo8p_c -> kdo_c + pi_c	kdsC	kdo8p	
LEUTA1	4mop_c + glu_L_c <=> akg_c + leu_L_c	ilvE, tyrB	4mop, glu_L	akg, leu_L
MDH	mal_L_c + nad_c <=> h_c + nadh_c + oaa_c	mdh	mal_L, nad	nadh
METAT	atp_c + h2o_c + met_L_c -> amet_c + pi_c + ppi_c	metK	atp, met_L	ppi
NDPK2	atp_c + udp_c <=> adp_c + utp_c	ndk, adk	atp, udp	adp, utp
NDPK3	atp_c + cdp_c <=> adp_c + ctp_c	ndk, adk	atp, cdp	adp, ctp
PFK	atp_c + f6p_c -> adp_c + fdp_c + h_c	pfkA, pfkB	atp, f6p	adp, fdp
PGCD	3pg_c + nad_c -> 3php_c + h_c + nadh_c	serA	3pg, nad	3php, nadh
PGI	g6p_c <=> f6p_c	pgi	g6p	f6p
PGK	3pg_c + atp_c <=> 13dpg_c + adp_c	pgk	3pg, atp	adp
PGL	6pgl_c + h2o_c -> 6pgc_c + h_c	pgl	6pgl	6pgc
PHETA1	akg_c + phe_L_c <=> glu_L_c + phpyr_c	ilvE, tyrB, aspC	akg, phe_L	glu_L, phpyr
PMPK	4ampm_c + atp_c -> 2mahmp_c + adp_c	thiD	4ampm, atp	adp
PPBNGS	2.0 5aop_c -> 2.0 h2o_c + h_c + ppbng_c	hemB	5aop	
PRAGSr	atp_c + gly_c + pram_c <=> adp_c + gar_c + h_c + pi_c	purD	atp, gly, pram	adp
PRASCSi	5aize_c + asp_L_c + atp_c -> 25aics_c + adp_c + h_c + pi_c	purC	5aize, asp_L, atp	adp
PRFGS	atp_c + fgam_c + gln_L_c + h2o_c -> adp_c + fpram_c + glu_L_c + h_c + pi_c	purL	atp, fgam, gln_L	adp, glu_L
PSCVT	pep_c + skm5p_c <=> 3psme_c + pi_c	aroA	pep, skm5p	
PSERT	3php_c + glu_L_c -> akg_c + pser_L_c	serC	3php, glu_L	akg, pser_L
PSP_L	h2o_c + pser_L_c -> pi_c + ser_L_c	serB	pser_L	ser_L
PYK	adp_c + h_c + pep_c -> atp_c + pyr_c	pykF, pykA	adp, pep	atp
RHCCE	rhcys_c -> dhptd_c + hcys_L_c	luxS	rhcys	
RPE	ru5p_D_c <=> xu5p_D_c	rpe	ru5p_D	xu5p_D
RPI	r5p_c <=> ru5p_D_c	rpiA	r5p	ru5p_D
S7P1	s7p_c -> gmhep7p_c	gmhA	s7p	
SADT2	atp_c + gtp_c + h2o_c + so4_c -> aps_c + gdp_c + pi_c + ppi_c	cysN, cysD	atp, gtp, so4	gdp, ppi
SERAT	accoa_c + ser_L_c <=> acser_c + coa_c	cysE	accoa, ser_L	coa
SHKK	atp_c + skm_c -> adp_c + h_c + skm5p_c	aroK	atp, skm	adp, skm5p
THD2pp	2.0 h_p + nadh_c + nadp_c -> 2.0 h_c + nad_c + nadph_c	pntA, pntB	nadh, nadp	nad, nadph
THDPS	h2o_c + succoa_c + thdp_c -> coa_c + sl2a6o_c	dapD	succoa, thdp	coa, sl2a6o
THRD_L	thr_L_c -> 2obut_c + nh4_c	ilvA	thr_L	2obut
THRS	h2o_c + phom_c -> pi_c + thr_L_c	thrC	phom	thr_L
TKT1	r5p_c + xu5p_D_c <=> g3p_c + s7p_c	tktB, tktA	r5p, xu5p_D	s7p
TKT2	e4p_c + xu5p_D_c <=> f6p_c + g3p_c	tktB, tktA	e4p, xu5p_D	f6p
TPI	dhap_c <=> g3p_c	tpiA	dhap	
TRPS3	3ig3p_c -> g3p_c + indole_c	trpB, trpA	3ig3p	indole
TYRTA	akg_c + tyr_L_c <=> 34hpp_c + glu_L_c	tyrB, aspC	akg, tyr_L	glu_L
UAGCVT	pep_c + uacgam_c -> pi_c + uaccg_c	murA	pep, uacgam	uaccg
UMPK	atp_c + ump_c <=> adp_c + udp_c	cmk, pyrH	atp, ump	adp, udp

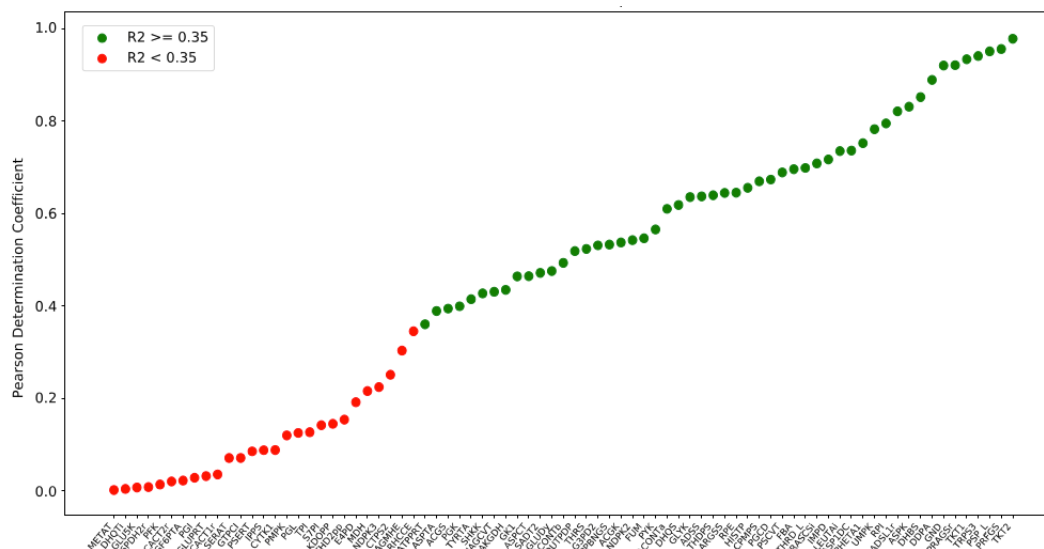


Figure A.2: A model without allosteric regulation can explain a lot of the fluxes in 62% of the reactions. The Pearson correlation coefficient indicates correspondence of observed and predicted fluxes in nine different conditions. Reactions for which $R^2 \geq 0.35$ are shown in green, those below in red.

Table A.2: *In vitro* enzyme assays

BIGG ID	Reaction	Major isoenzyme	Purified	Active	Enzyme assay
DDPA	3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase	AroG			
PFK	Phosphofructokinase	PfkA			
FBA	Fructose-bisphosphate aldolase	FbaA			
FUM	Fumarase	FumA			
ATPPRT	ATP phosphoribosyltransferase	HisG			
GLUPRT	Glutamine phosphoribosyldiphosphate amidotransferase	PurF			
TKT2	Transketolase	TktA			
PSP_L	Phosphoserine phosphatase (L-serine)	SerB			
THRD_L	L-threonine deaminase	IlvA			
GF6PTA	Glutamine-fructose-6-phosphate transaminase	GlmS			
ADSS	Adenylosuccinate synthase	PurA			
THRS	Threonine synthase	ThrC			

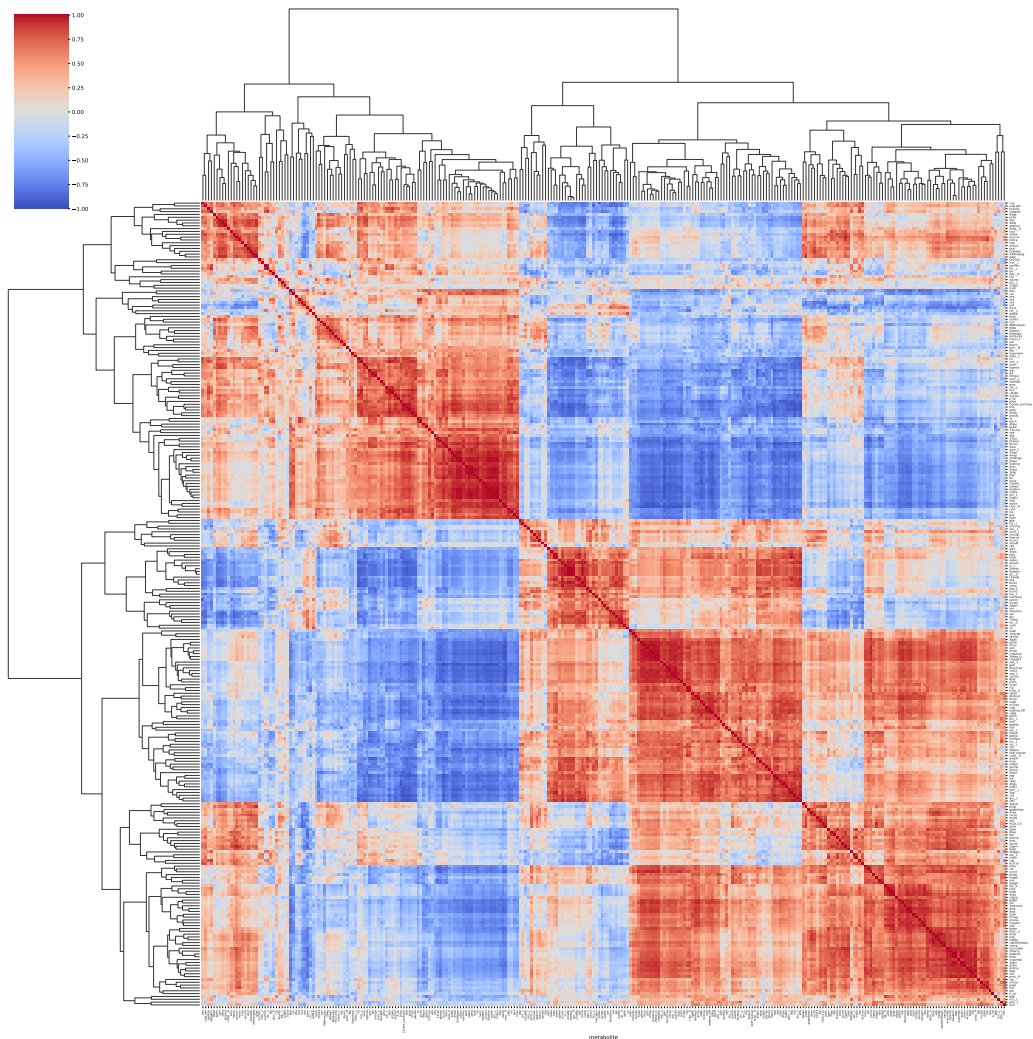


Figure A.3: correlations among the metabolites across nine conditions. In our reaction fitting procedure high correlations among metabolites that are tested as putative regulators will result in similar prediction scores, and hence reflect chances of false positive results of predictions are made solely using this data to differentiate candidate regulators.

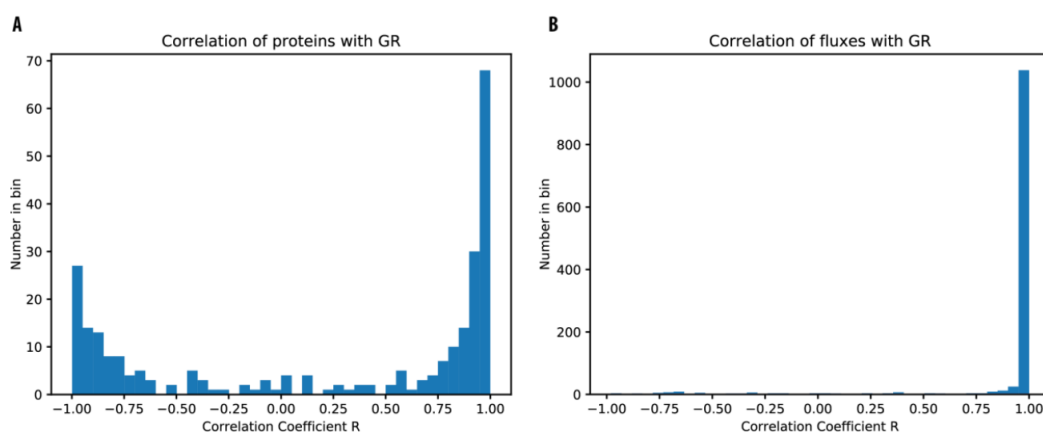


Figure A.4: Most fluxes and protein concentrations tend to depend linearly on the growth rate. **A.** Correlation between protein concentrations of the 84 analyzes reactions and growth rate. **B.** Correlation of fluxes with growth rate. These linear dependencies are used to inter/extrapolate proteomic and fluxomic values and compare data at the same growth rate.

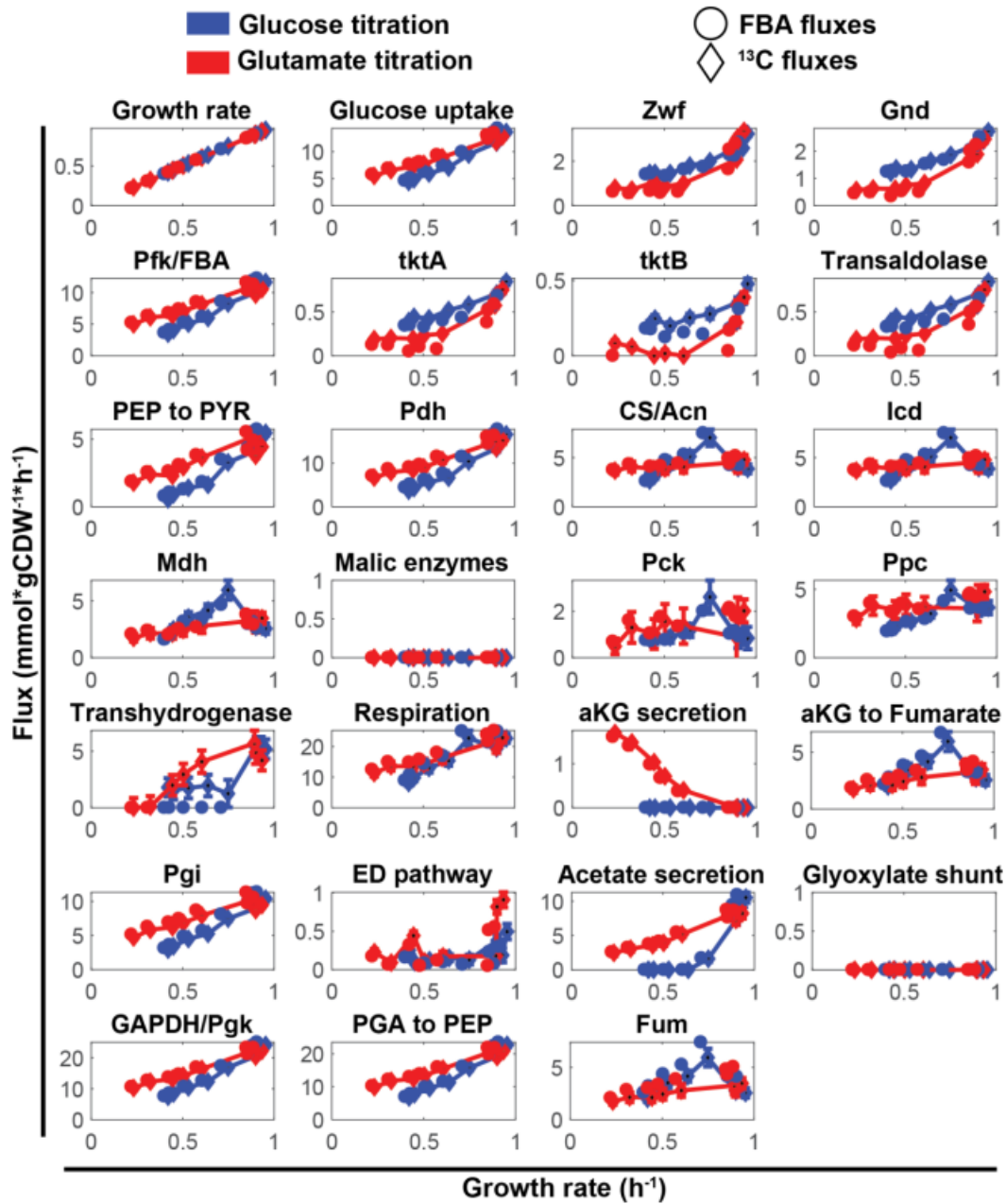


Figure A.5: Correlation between ¹³C flux analysis and CCM fluxes determine by FBA. Each figure plots the two flux sources against the growth rate across the eight glucose-limitation (blue) and eight glutama-limitation (red) conditions. The correspondence of FBA-derived (circles) and ¹³C flux analysis (diamonds) is high for the growth rate and the 26 measured CCM fluxes.

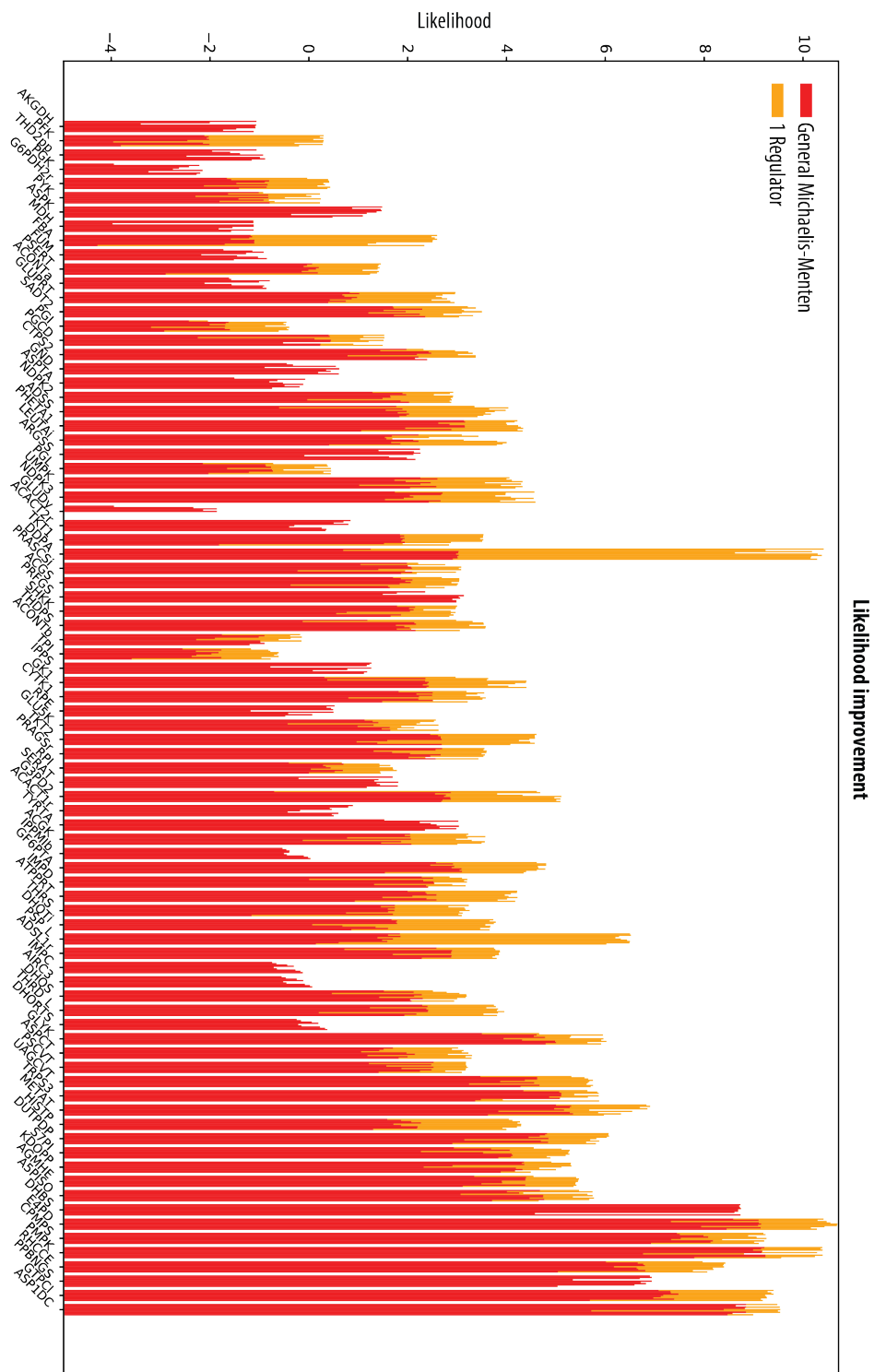


Figure A.6: Likelihood improvement upon addition of single allosteric regulators across the 84 reactions. The likelihood of the unregulated reaction (red) is shown, as well as the best likelihood upon inclusion of single regulators (orange). For each reaction, nine adjacent smaller bars correspond to the nine conditions. Only regulators that improve the akaike information criterion (AIC) with respect to the unregulated model are considered. Reaction abbreviations correspond to BiGG [6] identifiers.

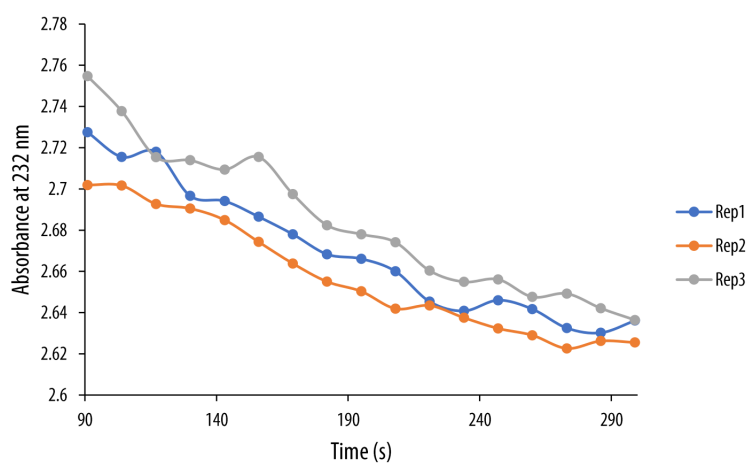


Figure A.7: AroG absorbance enzyme assay with 1mmol of 4-hydroxyphenylpyruvate and 250 μ mol erythrose-4-phosphate. The absorbance indicates consumption of 4-hydroxyphenylpyruvate suggesting that it acts as a competitive substrate. The lines correspond to three biological replicates with an average slope of $-4.765e4$ units s^{-1} , which is significantly different from the slope observed in the absence of enzyme $-6.276e5$ units s^{-1} (One-tailed t-test p-value < 0.001).

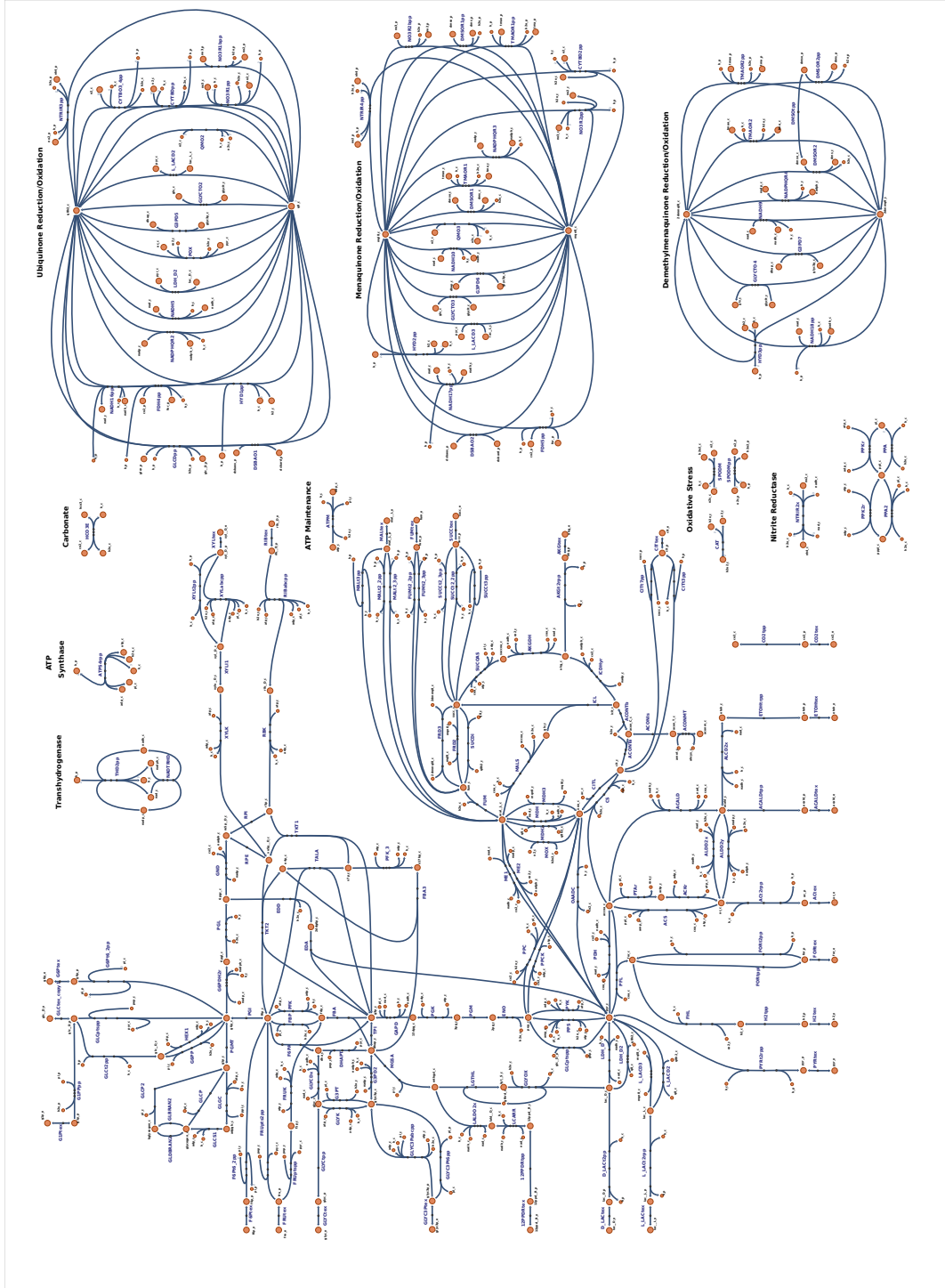


Figure A.8: Central metabolism of *Escherichia coli*. Biochemical pathways that constitute central metabolism from the genome scale metabolic model iJO1366 [12]. The graphic is created with the use of Escher [7].

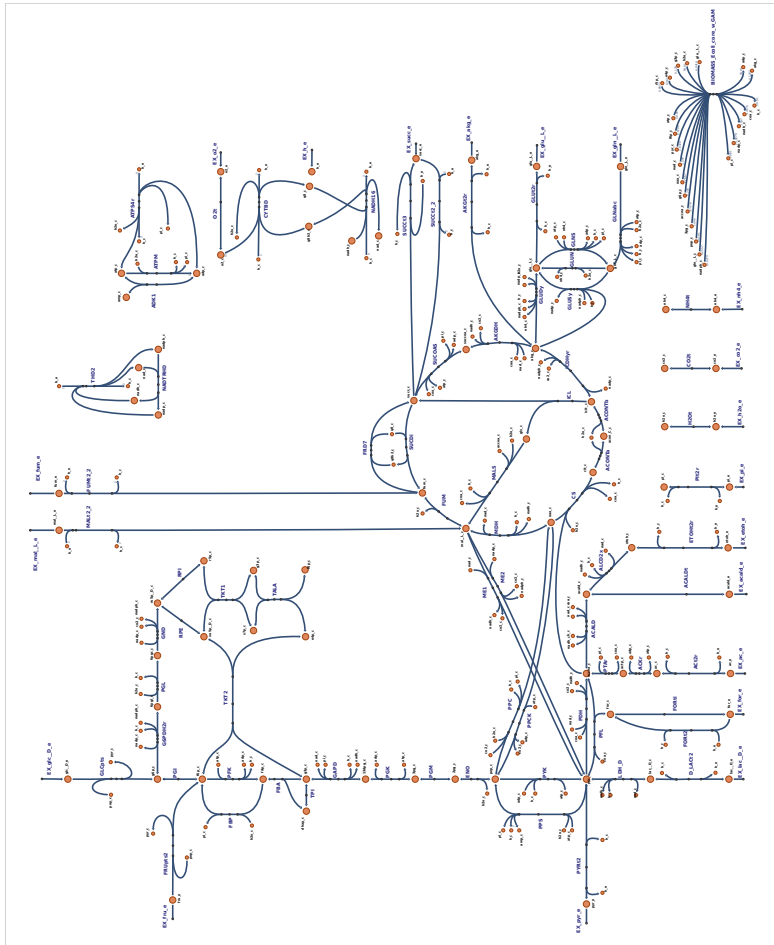


Figure A.9: Central metabolism of Escherichia coli. Biochemical pathways that constitute central metabolism from the core model [11]. The graphic is created with the use of Escher [7].

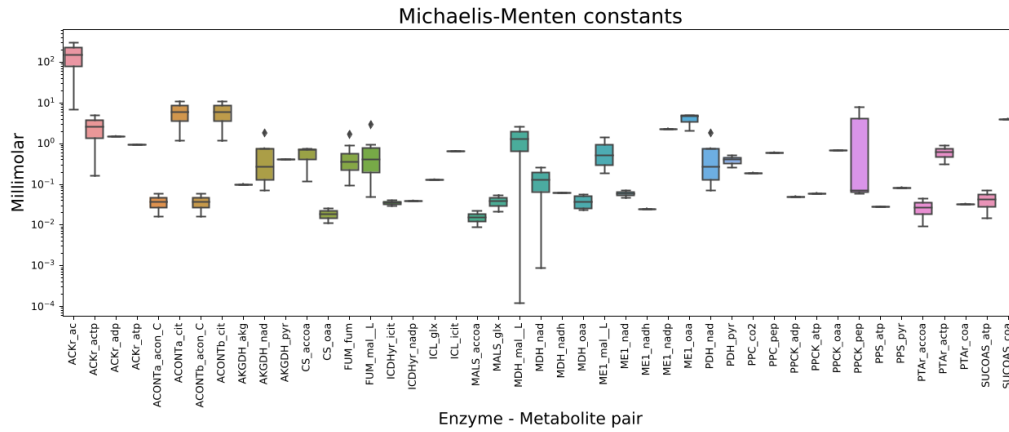


Figure A.13: Michaelis-Menten constants of the TCA cycle. Data are obtained from the BRENDA Enzyme Database [16]. The data covers 45 dissociation constants and for the majority of these we obtain multiple values from the published literature.

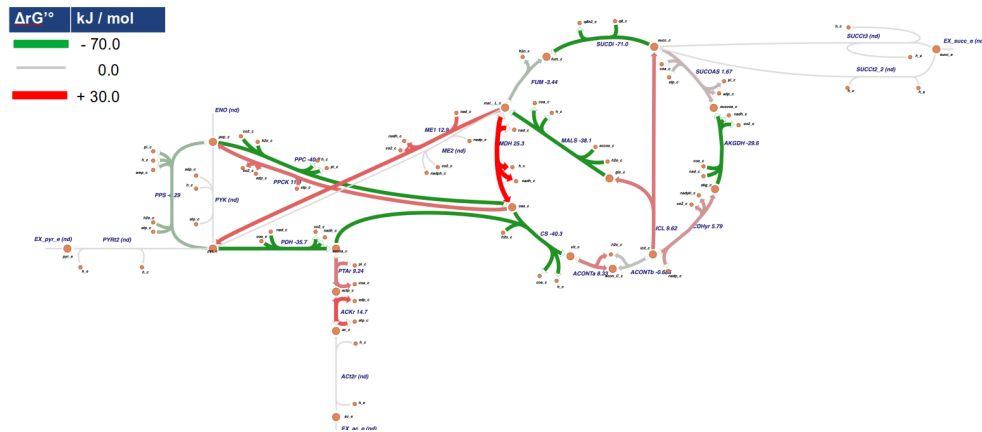


Figure A.14: Equilibrium constants of the TCA cycle. Estimates were obtained using the eEquibrator biochemical thermodynamics calculator [1]. Reactions with a negative Gibbs energy of formation are thermodynamically favorable and indicated in green. Those with a positive Gibbs free energy are thermodynamically unfavorable.

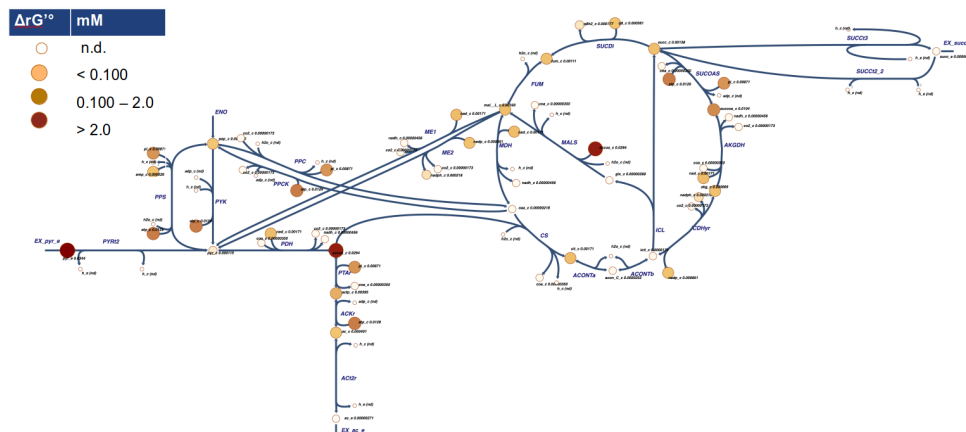


Figure A.15: Metabolite concentrations of the TCA cycle during steady-state growth on pyruvate. Data shown are from [9].

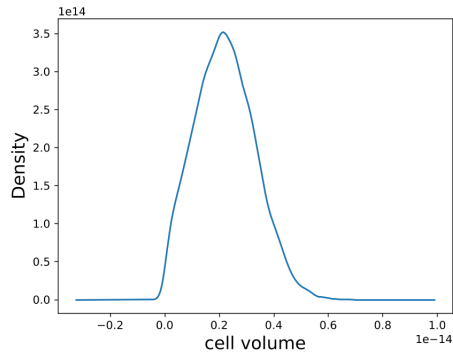


Figure A.16: Cell volume of *E. coli* grown on pyruvate. Data were obtained from [18] and the kernel density estimate was plotted to highlight the issue of computing mean and standard deviation in linear scale, and not supplying the raw data.

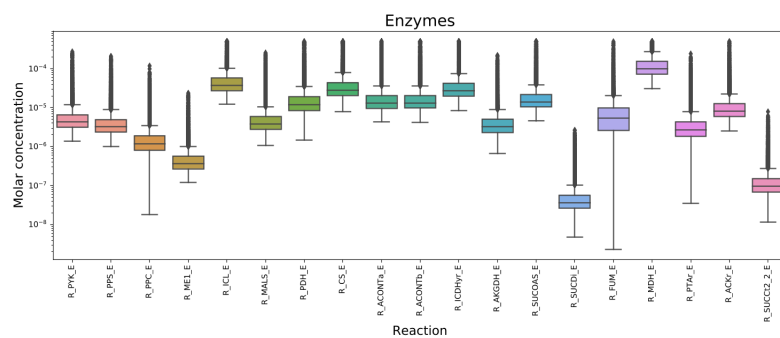


Figure A.17: Sampled enzyme concentrations. Absolute enzyme count data were obtained from [15] and cell volume data from [18]. The result of reporting only the mean and standard deviation on linear scale results in skewed sampling results, with unrealistically high enzyme concentrations.

Table A.3: Genes present in the *E. coli* TCA cycle model. These gene identifiers were used to map absolute protein quantification data from Schmidt *et. al.* [15] to the model.

b-number	gene	uniprot	asap	ecogene	ncbigene	ncbigi
b1676	pykF	P0AD61	ABE-0005600	EG10804	946179	gi:16129632
b1854	pykA	P21599	ABE-0006182	EG10803	946527	gi:16129807
b1702	ppsA	P23538	ABE-0005678	EG10759	946209	gi:16129658
b3956	ppc	P00864	ABE-0012950	EG10756	948457	gi:16131794
b3403	pck	P22259	ABE-0011106	EG10688	945667	gi:16131280
b1479	maeA	P26616	ABE-0004931	EG10948	946031	gi:90111281
b4015	aceA	P0A9G6	ABE-0013128	EG10022	948517	gi:16131841
b2976	glcB	P37330	ABE-0009767	EG20080	948857	gi:16130876
b4014	aceB	P08997	ABE-0013125	EG10023	948512	gi:16131840
b0116	lpd	P0A9P0	ABE-0000404	EG10543	944854	gi:16128109
b0114	aceE	P0AFG8	ABE-0000397	EG10024	944834	gi:16128107
b0115	aceF	P06959	ABE-0000400	EG10025	944794	gi:16128108
b0720	gltA	P0ABH7	ABE-0002451	EG10402	945323	gi:16128695
b1276	acnA	P25516	ABE-0004283	EG11325	946724	gi:16129237
b0118	acnB	P36683	ABE-0000411	EG12316	944864	gi:16128111
b1136	icd	P08200	ABE-0003823	EG10489	945702	gi:16129099
b0726	sucA	P0AFG3	ABE-0002478	EG10979	945303	gi:16128701
b0727	sucB	P0AFG6	ABE-0002480	EG10980	945307	gi:16128702
b0728	sucC	P0A836	ABE-0002483	EG10981	945312	gi:16128703
b0729	sucD	P0AGE9	ABE-0002485	EG10982	945314	gi:16128704
b0724	sdhB	P07014	ABE-0002468	EG10932	945300	gi:16128699
b0722	sdhD	P0AC44	ABE-0002464	EG10934	945322	gi:16128697
b0721	sdhC	P69054	ABE-0002460	EG10933	945316	gi:16128696
b0723	sdhA	P0AC41	ABE-0002466	EG10931	945402	gi:16128698
b1611	fumC	P05042	ABE-0005380	EG10358	946147	gi:16129569
b1612	fumA	P0AC33	ABE-0005392	EG10356	946826	gi:16129570
b4122	fumB	P14407	ABE-0013501	EG10357	948642	gi:16131948
b3236	mdh	P61889	ABE-0010613	EG10576	947854	gi:16131126
b2297	pta	P0A9M8	ABE-0007582	EG20173	946778	gi:16130232
b2458	eutD	P77218	ABE-0008097	EG14188	946940	gi:16130383
b1849	purT	P33221	ABE-0006162	EG11809	946368	gi:16129802
b3115	tdcD	P11868	ABE-0010245	EG11172	947635	gi:145698313
b2296	ackA	P0A6A3	ABE-0007579	EG10027	946775	gi:16130231
b3528	dctA	P0A830	ABE-0011527	EG20044	948039	gi:16131400
b1380	ldhA	P52643	ABE-0004619	EG13186	946315	gi:16129341
b2133	dld	P06149	ABE-0007048	EG10231	946653	gi:16130071
b1761	gdhA	P00370	ABE-0005865	EG10372	946802	gi:16129715

Table A.4: Gene reaction rules of the *E. coli* TCA cycle model. Gene-reaction-rules represent how genes, via the proteins they encode, are associated with metabolic reactions. These rules are represented using Boolean expressions: 'or' signifies that either (set of) gene can catalyze the reaction, while 'and' mandates the presence of both (sets of) genes.

reaction	gene rule
PYK	b1854 or b1676
PPS	b1702
PPC	b3956
PPCK	b3403
ME1	b1479
ICL	b4015
MALS	b4014 or b2976
PDH	b0115 and b0114 and b0116
CS	b0720
ACONTa	b0118 or b1276
ACONTb	b0118 or b1276
ICDHyr	b1136
AKGDH	b0726 and b0116 and b0727
SUCOAS	b0728 and b0729
SUCDi	b0723 and b0721 and b0722 and b0724
FUM	b4122 or b1612 or b1611
MDH	b3236
PYRt2	
PTAr	b2297 or b2458
ACKr	b2296 or b3115 or b1849
ACt2r	
SUCt2_2	b3528
LDH_D	b2133 or b1380
GLUDy	b1761
FUMt2_2	b3528

Table A.5: Reactions in the *E. coli* TCA cycle model. These reaction identifiers were used to retrieve data from the BRENDA database [16]. Reactions are BiGG identifiers [14]. Other identifier databases: metanetx [2], ec-code [19], kegg [5]. The following annotation are omitted here for clarity: SBO, biocyc, reactome, rhea.

reaction	metanetx	ec-code	kegg
PYK	MNXR103371	2.7.1.40	R00200
PPS	MNXR103140	2.7.9.2	R00199
PPC	MNXR103096	4.1.1.31	R00345
PPCK	MNXR103099	4.1.1.49	R00341
ME1	MNXR101446	1.1.1.39, 1.1.1.38	R00214
ICL	MNXR100789	4.1.3.1	R00479
MALS	MNXR101347	2.3.3.9	R00472
PDH	MNXR102425	2.3.1.12, 1.2.1.51, 1.8.1.4, 1.2.4.1	R00209
CS	MNXR96920	2.3.3.1, 2.3.3.16, 2.3.3.3	R00351
ACONT _a	MNXR95386	4.2.1.3	R01325
ACONT _b	MNXR95387	4.2.1.3	R01900
ICDH _{yr}	MNXR100781	1.1.1.42	R00267
AKGDH	MNXR95655	1.2.1.52, 1.2.4.2, 1.8.1.4, 2.3.1.61	R08549
SUCOAS	MNXR104635	6.2.1.5	R00405
SUCD _i	MNXR99641		
FUM	MNXR99705	4.2.1.2	R01082
MDH	MNXR101439	1.1.1.37, 1.1.1.299	R00342
PYR _{t2}	MNXR103385		
PTA _r	MNXR103319	2.3.1.8	R00230
ACK _r	MNXR95269	2.7.2.1, 2.7.2.15	R00315
AC _{t2r}	MNXR95429		
SUC _{Ct2_2}	MNXR104620		
LDH _{_D}	MNXR101037	1.1.1.28	R00704
GLUD _y	MNXR100086	1.4.1.13, 1.4.1.3, 1.4.1.4	R00248
FUM _{t2_2}	MNXR99711		

Table A.6: Reaction formulae of the *E. coli* TCA cycle model. This is the model used for flux balance analysis from which the kinetic model is derived. Reactions prefixed with 'EX_' are exchange reactions used for flux balance analysis. Metabolites are suffixed with a compartment identifier: '_c' for cytosol and '_e' for extracellular.

reaction	reaction
ACKr	ac_c + atp_c <=> actp_c + adp_c
ACONTa	cit_c <=> acon_C_c + h2o_c
ACONTb	acon_C_c + h2o_c <=> icit_c
ACt2r	ac_e + h_e <=> ac_c + h_c
AKGDH	akg_c + coa_c + nad_c -> co2_c + nadh_c + succoa_c
CS	accoa_c + h2o_c + oaa_c -> cit_c + coa_c + h_c
FUM	fum_c + h2o_c <=> mal_L_c
FUMt2_2	fum_e + 2.0 h_e <=> fum_c + 2.0 h_c
GLUDy	glu_L_c + h2o_c + nadp_c <=> akg_c + h_c + nadph_c + nh4_c
ICDHyr	icit_c + nadp_c <=> akg_c + co2_c + nadph_c
ICL	icit_c -> glx_c + succ_c
LDH_D	lac_D_c + nad_c <=> h_c + nadh_c + pyr_c
MALS	accoa_c + glx_c + h2o_c -> coa_c + h_c + mal_L_c
MDH	mal_L_c + nad_c <=> h_c + nadh_c + oaa_c
ME1	mal_L_c + nad_c -> co2_c + nadh_c + pyr_c
PDH	coa_c + nad_c + pyr_c -> accoa_c + co2_c + nadh_c
PPC	co2_c + h2o_c + pep_c -> h_c + oaa_c + pi_c
PPCK	atp_c + oaa_c -> adp_c + co2_c + pep_c
PPS	atp_c + h2o_c + pyr_c -> amp_c + 2.0 h_c + pep_c + pi_c
PTAr	accoa_c + pi_c <=> actp_c + coa_c
PYK	adp_c + h_c + pep_c -> atp_c + pyr_c
PYRt2	h_e + pyr_e <=> h_c + pyr_c
SUCct2_2	2.0 h_e + succ_e -> 2.0 h_c + succ_c
SUCDi	q8_c + succ_c -> fum_c + q8h2_c
SUCOAS	adp_c + pi_c + succoa_c <=> atp_c + coa_c + succ_c
EX_ac_e	ac_e <=>
EX_adp_c	adp_c <=>
EX_amp_c	amp_c <=>
EX_atp_c	atp_c <=>
EX_co2_c	co2_c <=>
EX_fum_e	fum_e <=>
EX_glu_L_c	glu_L_c <=>
EX_h2o_c	h2o_c <=>
EX_h_c	h_c <=>
EX_h_e	h_e <=>
EX_lac_D_c	lac_D_c <=>
EX_nad_c	nad_c <=>
EX_nadh_c	nadh_c <=>
EX_nadp_c	nadp_c <=>
EX_nadph_c	nadph_c <=>
EX_nh4_c	nh4_c <=>
EX_pep_c	pep_c <=>
EX_pi_c	pi_c <=>
EX_pyr_e	pyr_e <=>
EX_q8_c	q8_c <=>
EX_q8h2_c	q8h2_c <=>
EX_succ_e	succ_e <=>

Table A.7: Metabolites in the *E. coli* TCA cycle model. These metabolite identifiers were used to retrieve data from the BRENDA database [16]. Metabolites are BiGG identifiers [14]. Other identifier databases: seed [4], kegg [5] and metanetx [2]. The following annotation are omitted here for clarity: reactome, chebi, hmdb, biocyc, lipidmaps.

metabolite	seed	kegg	metanetx
adp	cpd00008	C00008	MNXM7
h	cpd00067	C00080	MNXM1
pep	cpd00061	C00074	MNXM73
atp	cpd00002	C00002	MNXM3
pyr	cpd00020	C00022	MNXM23
h2o	cpd00001, cpd15275, cpd27222	C00001, C01328, C18714	MNXM2
amp	cpd00018, cpd22272	C00020	MNXM14
pi	cpd00009, cpd27787	C00009, C13558	MNXM9
co2	cpd00011	C00011	MNXM13
oaa	cpd00032, cpd02469	C00036, C03981	MNXM46
malL	cpd00130	C00149	MNXM98
nad	cpd00003	C00003	MNXM8
nadh	cpd00004	C00004	MNXM10
icit	cpd00260	C00311	MNXM89661
glx	cpd00040	C00048	MNXM69
succ	cpd00036	C00042	MNXM25
accoa	cpd00022	C00024	MNXM21
coa	cpd00010, cpd22528	C00010	MNXM12
cit	cpd00137	C00158, C13660	MNXM131
aconC	cpd00331	C00417	MNXM813
nadp	cpd00006	C00006	MNXM5
akg	cpd00024	C00026	MNXM20
nadph	cpd00005	C00005	MNXM6
succoa	cpd00078	C00091	MNXM92
q8	cpd15560	C17569	MNXM232
fum	cpd00106	C00122	MNXM93
q8h2	cpd15561, cpd29608	C00390	MNXM191
actp	cpd00196	C00227	MNXM280
ac	cpd00029	C00033	MNXM26
lacD	cpd00221	C00256	MNXM285
gluL	cpd00023, cpd19002, cpd27177	C00025, C00302	MNXM89557
nh4	cpd00013, cpd19013	C00014, C01342	MNXM15

Table A.8: Metabolite formulae and charge of the *E. coli* TCA cycle model. Metabolites are BiGG identifiers [14]. Boundary condition indicates whether or not the metabolite was considered as such in the kinetic model. Rule assigned denotes whether a species were modeled using a linear interpolated rule and experimentally determined time series data.

metabolite	formula	charge	boundary condition	rule assigned
ac	C2H3O2	-1	False	False
accoa	C23H34N7O17P3S	-4	False	False
aconC	C6H3O6	-3	False	False
actp	C2H3O5P	-2	False	False
adp	C10H12N5O10P2	-3	True	True
akg	C5H4O5	-2	False	False
amp	C10H12N5O7P	-2	True	True
atp	C10H12N5O13P3	-4	True	True
cit	C6H5O7	-3	False	False
co2	CO2	0	True	False
coa	C21H32N7O16P3S	-4	False	False
fum	C4H2O4	-2	False	False
gluL	C5H8NO4	-1	True	True
glx	C2H1O3	-1	False	False
h2o	H2O	0	True	False
h	H	1	True	False
icit	C6H5O7	-3	False	False
lacD	C3H5O3	-1	True	True
malL	C4H4O5	-2	False	False
nad	C21H26N7O14P2	-1	True	True
nadh	C21H27N7O14P2	-2	True	True
nadp	C21H25N7O17P3	-3	True	True
nadph	C21H26N7O17P3	-4	True	True
nh4	H4N	1	True	False
oaa	C4H2O5	-2	False	False
pep	C3H2O6P	-3	True	True
pi	HO4P	-2	True	False
pyr	C3H3O3	-1	False	False
q8	C49H74O4	0	True	False
q8h2	C49H76O4	0	True	False
succ	C4H4O4	-2	False	False
succoa	C25H35N7O19P3S	-5	False	False

Table A.9: Small Molecule Regulatory Network Table of *E. coli*. This table presents data extracted from Reznik et al.’s supplement [13], featuring columns detailing biochemical reactions (Reaction), associated metabolites (Metabolite), Enzyme Commission numbers (E.C. number), Kyoto Encyclopedia of Genes and Genomes IDs (KEGG ID), mode of action (Mode), and the regulatory mechanisms (Mechanism). Only those reactions and metabolite that are part of our study are presented, using their respective BiGG identifiers.

Reaction	Metabolite	E.C. number	KEGG ID	Mode	Mechanism
cs	accoa	2.3.3.16		+	Allosteric
cs	akg	2.3.3.16	C00026	-	Competitive
cs	atp	2.3.3.16		-	Competitive
cs	nad	2.3.3.16		-	Allosteric
cs	nadh	2.3.3.16	C00004	-	Allosteric
fum	cit	4.2.1.2	C00158	-	Competitive
icdhydr	pep	1.1.1.42	C00074	-	Allosteric
icdhydr	oaa	1.1.1.42	C00036	-	Competitive
icl	3pg	4.1.3.1	C00597	-	Competitive
icl	akg	4.1.3.1	C00026	-	
icl	pep	4.1.3.1	C00074	-	Uncompetitive
icl	succ	4.1.3.1	C00042	-	Uncompetitive
pdh	accoa	1.2.4.1	C00024	-	
ppc	accoa	4.1.1.31		+	Allosteric
ppc	cit	4.1.1.31		-	
ppc	fdp	4.1.1.31	C00354	+	Allosteric
ppc	fum	4.1.1.31	C00122	-	
ppc	gtp	4.1.1.31	C00044	+	Allosteric
ppc	succ	4.1.1.31	C00042	-	
ppck	3pg	4.1.1.49	C00597	-	
ppck	accoa	4.1.1.49	C00024	+	
ppck	atp	4.1.1.49	C00002	-	
ppck	dhap	4.1.1.49	C00111	-	
ppck	f6p	4.1.1.49	C00085	-	
ppck	fdp	4.1.1.49	C00354	-	
ppck	nadh	4.1.1.49	C00004	-	
ppck	pep	4.1.1.49		-	
pps	adp	2.7.9.2	C00008	-	
pps	akg	2.7.9.2	C00026	-	
pps	amp	2.7.9.2	C00020	-	
pps	atp	2.7.9.2	C00002	-	
pps	pep	2.7.9.2	C00074	-	
pyk	amp	2.7.1.40		+	Allosteric
pyk	atp	2.7.1.40		-	
pyk	fdp	2.7.1.40	C00354	+	Allosteric
pyk	g6p	2.7.1.40	C00092	+	
pyk	gtp	2.7.1.40		-	
pyk	pep	2.7.1.40	C00074	+	
sucoas	adp	6.2.1.5	C00008	-	
sucoas	adp	6.2.1.5	C00008	+	
sucoas	akg	6.2.1.5	C00026	-	
sucoas	atp	6.2.1.5	C00002	-	
sucoas	nadh	6.2.1.5	C00004	-	
sucoas	succ	6.2.1.5	C00042	-	

A.2.1 Rate Equations

The rate equations governing the dynamics of the metabolic network comprising the TCA cycle and anaplerotic reactions. Each rate equation represents the rate of a specific metabolic reaction, and it encapsulates the effects of enzyme concentrations, substrate concentrations, and reaction constants. In the following rate equations, various symbols and notations are used to represent enzymes, metabolites, constants, and mathematical relationships. A reaction-local nomenclature is used for enzymes and parameters to facilitate readability:

- **Enzyme symbols:** Enzymes are represented by E corresponding to the enzyme of the specific reaction.
- **Metabolite symbols:** Reactants and products are indicated by their BiGG identifiers. Only for extracellular metabolites has a subscript been used: ac_e , pyr_e , fum_e , $succ_e$, mal_e .
- **Catalytic rate constants:** forward catalytic constants are represented by k_{cat}^f .
- **Michaelis-Menten constants:** are denoted with their corresponding metabolite in the superscripts, for instance, K_m^{adp} .
- **Equilibrium constants:** are designated by K_{eq} and provide insight into the balance of products and reactants at equilibrium.

$$v_{PYK} = \frac{E \cdot k_{cat}^f \cdot adp \cdot pep \cdot \left(1.0 - \frac{atp \cdot pyr}{K_{eq} \cdot adp \cdot pep}\right)}{K_m^{adp} \cdot K_m^{pep} \cdot \left(1.0 + \frac{adp}{K_m^{adp}}\right) \cdot \left(1.0 + \frac{pep}{K_m^{pep}}\right) + \left(1.0 + \frac{atp}{K_m^{atp}}\right) \cdot \left(1.0 + \frac{pyr}{K_m^{pyr}}\right) - 1.0} \quad (A.1)$$

$$v_{PPS} = \frac{E \cdot k_{cat}^f \cdot atp \cdot pyr \cdot \left(1.0 - \frac{pi \cdot amp \cdot pep}{K_{eq} \cdot atp \cdot pyr}\right)}{K_m^{atp} \cdot K_m^{pyr} \cdot \left(1.0 + \frac{\pi}{K_m^{\pi}}\right) \cdot \left(1.0 + \frac{amp}{K_m^{amp}}\right) \cdot \left(1.0 + \frac{pep}{K_m^{pep}}\right) + \left(1.0 + \frac{atp}{K_m^{atp}}\right) \cdot \left(1.0 + \frac{pyr}{K_m^{pyr}}\right) - 1.0} \quad (A.2)$$

$$v_{PPC} = \frac{E \cdot k_{cat}^f \cdot co_2 \cdot pep \cdot \left(1.0 - \frac{pi \cdot oaa}{K_{eq} \cdot co_2 \cdot pep}\right)}{K_m^{co_2} \cdot K_m^{pep} \cdot \left(1.0 + \frac{\pi}{K_m^{\pi}}\right) \cdot \left(1.0 + \frac{oaa}{K_m^{oaa}}\right) + \left(1.0 + \frac{co_2}{K_m^{co_2}}\right) \cdot \left(1.0 + \frac{pep}{K_m^{pep}}\right) - 1.0} \quad (A.3)$$

$$v_{PPCK} = \frac{E \cdot k_{cat}^f \cdot atp \cdot oaa \cdot \left(1.0 - \frac{adp \cdot co_2 \cdot pep}{K_{eq} \cdot atp \cdot oaa}\right)}{K_m^{atp} \cdot K_m^{oaa} \cdot \left(1.0 + \frac{adp}{K_m^{adp}}\right) \cdot \left(1.0 + \frac{co_2}{K_m^{co_2}}\right) \cdot \left(1.0 + \frac{pep}{K_m^{pep}}\right) + \left(1.0 + \frac{atp}{K_m^{atp}}\right) \cdot \left(1.0 + \frac{oaa}{K_m^{oaa}}\right) - 1.0} \quad (A.4)$$

$$v_{ME1} = \frac{E \cdot k_{cat}^f \cdot mal-L \cdot nad \cdot \left(1.0 - \frac{co_2 \cdot nadh \cdot pyr}{K_{eq} \cdot mal-L \cdot nad}\right)}{K_{mal-L} \cdot K_m^{nad} \cdot \left(1.0 + \frac{co_2}{K_m^{co_2}}\right) \cdot \left(1.0 + \frac{nadh}{K_m^{nadh}}\right) \cdot \left(1.0 + \frac{pyr}{K_m^{pyr}}\right) + \left(1.0 + \frac{mal-L}{K_m^{mal-L}}\right) \cdot \left(1.0 + \frac{nad}{K_m^{nad}}\right) - 1.0} \quad (A.5)$$

$$v_{ICL} = \frac{E \cdot k_{cat}^f \cdot icit \cdot \left(1.0 - \frac{glx \cdot succ}{K_{eq} \cdot icit}\right)}{K_m^{icit} \cdot \left(1.0 + \frac{glx}{K_m^{glx}}\right) \cdot \left(1.0 + \frac{succ}{K_m^{succ}}\right) + K_m^{icit}} \quad (A.6)$$

$$v_{MALS} = \frac{E \cdot k_{cat}^f \cdot accoa \cdot glx \cdot \left(1.0 - \frac{coa \cdot mal-L}{K_{eq} \cdot accoa \cdot glx}\right)}{K_m^{accoa} \cdot K_m^{glx} \cdot \left(1.0 + \frac{accoa}{K_m^{accoa}}\right) \cdot \left(1.0 + \frac{glx}{K_m^{glx}}\right) + \left(1.0 + \frac{coa}{K_m^{coa}}\right) \cdot \left(1.0 + \frac{mal-L}{K_m^{mal-L}}\right) - 1.0} \quad (A.7)$$

$$v_{PDH} = \frac{E \cdot k_{cat}^f \cdot coa \cdot nad \cdot pyr \cdot \left(1.0 - \frac{accoa \cdot co_2 \cdot nadh}{K_{eq} \cdot coa \cdot nad \cdot pyr}\right)}{K_m^{coa} \cdot K_m^{nad} \cdot K_m^{pyr} \cdot \left(1.0 + \frac{accoa}{K_m^{accoa}}\right) \cdot \left(1.0 + \frac{co_2}{K_m^{co_2}}\right) \cdot \left(1.0 + \frac{nadh}{K_m^{nadh}}\right) + \left(1.0 + \frac{coa}{K_m^{coa}}\right) \cdot \left(1.0 + \frac{nad}{K_m^{nad}}\right) \cdot \left(1.0 + \frac{pyr}{K_m^{pyr}}\right) - 1.0} \quad (A.8)$$

$$v_{CS} = \frac{E \cdot k_{cat}^f \cdot accoa \cdot oaa \cdot \left(1.0 - \frac{cit \cdot coa}{K_{eq} \cdot accoa \cdot oaa}\right)}{K_m^{accoa} \cdot K_m^{oaa} \cdot \left(\left(1.0 + \frac{accoa}{K_m^{accoa}}\right) \cdot \left(1.0 + \frac{oaa}{K_m^{oaa}}\right) + \left(1.0 + \frac{cit}{K_m^{cit}}\right) \cdot \left(1.0 + \frac{coa}{K_m^{coa}}\right) - 1.0\right)} \quad (A.9)$$

$$v_{ACONTa} = \frac{E \cdot k_{cat}^f \cdot cit \cdot \left(1.0 - \frac{acon-C}{K_{eq} \cdot cit}\right)}{K_m^{cit} \cdot \left(1.0 + \frac{cit}{K_m^{cit}} + \frac{acon-C}{K_m^{acon-C}}\right)} \quad (A.10)$$

$$v_{ACONTb} = \frac{E \cdot k_{cat}^f \cdot acon-C \cdot \left(1.0 - \frac{icit}{K_{eq} \cdot acon-C}\right)}{K_m^{acon-C} \cdot \left(1.0 + \frac{icit}{K_m^{icit}} + \frac{acon-C}{K_m^{acon-C}}\right)} \quad (A.11)$$

$$v_{ICDHyr} = \frac{E \cdot k_{cat}^f \cdot igit \cdot nadp \cdot \left(1.0 - \frac{akg \cdot co_2 \cdot nadph}{K_{eq} \cdot igit \cdot nadp}\right)}{K_m^{igit} \cdot K_m^{nadp} \cdot \left(\left(1.0 + \frac{akg}{K_m^{akg}}\right) \cdot \left(1.0 + \frac{co_2}{K_m^{co_2}}\right) \cdot \left(1.0 + \frac{nadph}{K_m^{nadph}}\right) + \left(1.0 + \frac{icit}{K_m^{icit}}\right) \cdot \left(1.0 + \frac{nadp}{K_m^{nadp}}\right) - 1.0\right)} \quad (A.12)$$

$$v_{AKGDH} = \frac{E \cdot k_{cat}^f \cdot akg \cdot coa \cdot nad \cdot \left(1.0 - \frac{co_2 \cdot nadh \cdot succoa}{K_{eq} \cdot akg \cdot coa \cdot nad}\right)}{K_m^{akg} \cdot K_m^{coa} \cdot K_m^{nad} \cdot \left(\left(1.0 + \frac{akg}{K_m^{akg}}\right) \cdot \left(1.0 + \frac{coa}{K_m^{coa}}\right) \cdot \left(1.0 + \frac{nad}{K_m^{nad}}\right) + \left(1.0 + \frac{co_2}{K_m^{co_2}}\right) \cdot \left(1.0 + \frac{nadh}{K_m^{nadh}}\right) \cdot \left(1.0 + \frac{succoa}{K_m^{succoa}}\right) - 1.0\right)} \quad (A.13)$$

$$v_{SUCOAS} = \frac{E \cdot k_{cat}^f \cdot pi \cdot adp \cdot succoa \cdot \left(1.0 - \frac{atp \cdot coa \cdot succ}{pi \cdot K_{eq} \cdot adp \cdot succoa}\right)}{K_m^{adp} \cdot K_m^{pi} \cdot K_m^{succoa} \cdot \left(\left(1.0 + \frac{pi}{K_m^{pi}}\right) \cdot \left(1.0 + \frac{adp}{K_m^{adp}}\right) \cdot \left(1.0 + \frac{succoa}{K_m^{succoa}}\right) + \left(1.0 + \frac{atp}{K_m^{atp}}\right) \cdot \left(1.0 + \frac{coa}{K_m^{coa}}\right) \cdot \left(1.0 + \frac{succ}{K_m^{succ}}\right) - 1.0\right)} \quad (A.14)$$

$$v_{SUCDi} = \frac{E \cdot k_{cat}^f \cdot qs \cdot succ \cdot \left(1.0 - \frac{fum \cdot qsh2}{K_{eq} \cdot qs \cdot succ}\right)}{K_m^{qs} \cdot K_m^{succ} \cdot \left(\left(1.0 + \frac{fum}{K_m^{fum}}\right) \cdot \left(1.0 + \frac{qsh2}{K_m^{qsh2}}\right) + \left(1.0 + \frac{qs}{K_m^{qs}}\right) \cdot \left(1.0 + \frac{succ}{K_m^{succ}}\right) - 1.0\right)} \quad (A.15)$$

$$v_{FUM} = \frac{E \cdot k_{cat}^f \cdot fum \cdot \left(1.0 - \frac{mal-L}{K_{eq} \cdot fum}\right)}{K_m^{fum} \cdot \left(1.0 + \frac{mal-L}{K_m^{mal-L}} + \frac{fum}{K_m^{fum}}\right)} \quad (A.16)$$

$$v_{MDH} = \frac{E \cdot k_{cat}^f \cdot mal-L \cdot nad \cdot \left(1.0 - \frac{nadh \cdot oaa}{K_{eq} \cdot mal-L \cdot nad}\right)}{K_m^{mal-L} \cdot K_m^{nad} \cdot \left(\left(1.0 + \frac{mal-L}{K_m^{mal-L}}\right) \cdot \left(1.0 + \frac{nad}{K_m^{nad}}\right) + \left(1.0 + \frac{nadh}{K_m^{nadh}}\right) \cdot \left(1.0 + \frac{oaa}{K_m^{oaa}}\right) - 1.0\right)} \quad (A.17)$$

$$vPYRt2 = \frac{E \cdot k_{cat}^f \cdot pyr_e \cdot \left(1.0 - \frac{pyr}{K_{eq} \cdot pyr_e}\right)}{K_m^{pyr} \cdot \left(1.0 + \frac{pyr}{K_m^{pyr}} + \frac{pyr_e}{K_m^{pyr}}\right)} \quad (\text{A.18})$$

$$vPTAr = \frac{E \cdot k_{cat}^f \cdot accoa \cdot pi \cdot \left(1.0 - \frac{actp \cdot coa}{pi \cdot K_{eq} \cdot accoa}\right)}{K_m^{accoa} \cdot K_m^{pi} \cdot \left(\left(1.0 + \frac{\pi}{K_m^{pi}}\right) \cdot \left(1.0 + \frac{accoa}{K_m^{accoa}}\right) + \left(1.0 + \frac{actp}{K_m^{actp}}\right) \cdot \left(1.0 + \frac{coa}{K_m^{coa}}\right) - 1.0 \right)} \quad (\text{A.19})$$

$$vACKr = \frac{E \cdot k_{cat}^f \cdot ac \cdot atp \cdot \left(1.0 - \frac{actp \cdot adp}{K_{eq} \cdot ac \cdot atp}\right)}{K_m^{ac} \cdot K_m^{atp} \cdot \left(\left(1.0 + \frac{ac}{K_m^{ac}}\right) \cdot \left(1.0 + \frac{atp}{K_m^{atp}}\right) + \left(1.0 + \frac{actp}{K_m^{actp}}\right) \cdot \left(1.0 + \frac{adp}{K_m^{adp}}\right) - 1.0 \right)} \quad (\text{A.20})$$

$$vACTr = \frac{E \cdot k_{cat}^f \cdot ac_e \cdot \left(1.0 - \frac{ac}{K_{eq} \cdot ac_e}\right)}{K_m^{ac} \cdot \left(1.0 + \frac{ac}{K_m^{ac}} + \frac{ac_e}{K_m^{ac}}\right)} \quad (\text{A.21})$$

$$vSUCCt2-2 = \frac{E \cdot k_{cat}^f \cdot succ_e \cdot \left(1.0 - \frac{succ}{K_{eq} \cdot succ_e}\right)}{K_m^{succ} \cdot \left(1.0 + K_m^{succ} \cdot \frac{succ_e}{K_m^{succ}} + K_m^{succ}\right)} \quad (\text{A.22})$$

$$vLDH-D = \frac{E \cdot k_{cat}^f \cdot lac-D \cdot nad \cdot \left(1.0 - \frac{nadh \cdot pyr}{K_{eq} \cdot lac-D \cdot nad}\right)}{K_m^{lac-D} \cdot K_m^{nad} \cdot \left(\left(1.0 + \frac{lac-D}{K_m^{lac-D}}\right) \cdot \left(1.0 + \frac{nad}{K_m^{nad}}\right) + \left(1.0 + \frac{nadh}{K_m^{nadh}}\right) \cdot \left(1.0 + \frac{pyr}{K_m^{pyr}}\right) - 1.0 \right)} \quad (\text{A.23})$$

$$vGLUDy = \frac{E \cdot k_{cat}^f \cdot glu-L \cdot nadp \cdot \left(1.0 - \frac{akg \cdot nadph \cdot nh_4}{K_{eq} \cdot glu-L \cdot nadp}\right)}{K_m^{glu-L} \cdot K_m^{nadp} \cdot \left(\left(1.0 + \frac{akg}{K_m^{akg}}\right) \cdot \left(1.0 + \frac{nadph}{K_m^{nadph}}\right) \cdot \left(1.0 + \frac{nh_4}{K_m^{nh_4}}\right) + \left(1.0 + \frac{glu-L}{K_m^{glu-L}}\right) \cdot \left(1.0 + \frac{nadp}{K_m^{nadp}}\right) - 1.0 \right)} \quad (\text{A.24})$$

$$vFUMt2-2 = \frac{E \cdot k_{cat}^f \cdot fum_e \cdot \left(1.0 - \frac{fum}{K_{eq} \cdot fum_e}\right)}{K_m^{fum} \cdot \left(1.0 + \frac{fum}{K_m^{fum}} + \frac{fum_e}{K_m^{fum}}\right)} \quad (\text{A.25})$$

A.2.2 Differential Equations

The differential equations governing the dynamics of the metabolic network comprising the TCA cycle and anaplerotic reactions. Each equation encapsulates the rate of change of a specific metabolite concentration with respect to time. The notation employed in these differential equations builds upon the established conventions in biochemical modeling:

- **Metabolite symbols:** The symbols used for metabolites reflect the chemical compounds they represent, such as `pyr` for pyruvate, `oaa` for oxaloacetate, and `mal-L` for L-malate.
- **Reaction rates:** The terms in the differential equations correspond to the summed rates of various metabolic reactions that either produce or consume the respective metabolite.

$$\frac{d\text{pyr}}{dt} = v_{\text{PYK}} + v_{\text{ME1}} + v_{\text{PYRt2}} + v_{\text{LDH-D}} - v_{\text{PPS}} - v_{\text{PDH}} \quad (\text{A.26})$$

$$\frac{d\text{oaa}}{dt} = v_{\text{PPC}} + v_{\text{MDH}} - v_{\text{PPCK}} - v_{\text{CS}} \quad (\text{A.27})$$

$$\frac{d\text{mal-L}}{dt} = v_{\text{MALS}} + v_{\text{FUM}} - v_{\text{ME1}} - v_{\text{MDH}} \quad (\text{A.28})$$

$$\frac{d\text{cit}}{dt} = v_{\text{ACONTb}} - v_{\text{ICL}} - v_{\text{ICDHyr}} \quad (\text{A.29})$$

$$\frac{d\text{glx}}{dt} = v_{\text{ICL}} - v_{\text{MALS}} \quad (\text{A.30})$$

$$\frac{d\text{succ}}{dt} = v_{\text{ICL}} + v_{\text{SUCOAS}} + v_{\text{SUCCt2-2}} - v_{\text{SUCDi}} \quad (\text{A.31})$$

$$\frac{d\text{accoa}}{dt} = v_{\text{PDH}} - v_{\text{MALS}} - v_{\text{CS}} - v_{\text{PTAr}} \quad (\text{A.32})$$

$$\frac{d\text{coa}}{dt} = v_{\text{MALS}} + v_{\text{CS}} + v_{\text{SUCOAS}} + v_{\text{PTAr}} - v_{\text{PDH}} - v_{\text{AKGDH}} \quad (\text{A.33})$$

$$\frac{d\text{cit}}{dt} = v_{\text{CS}} - v_{\text{ACONTa}} \quad (\text{A.34})$$

$$\frac{d\text{acon-C}}{dt} = v_{\text{ACONTa}} - v_{\text{ACONTb}} \quad (\text{A.35})$$

$$\frac{d\text{akg}}{dt} = v_{\text{ICDHyr}} + v_{\text{GLUDy}} - v_{\text{AKGDH}} \quad (\text{A.36})$$

$$\frac{d\text{succoa}}{dt} = v_{\text{AKGDH}} - v_{\text{SUCOAS}} \quad (\text{A.37})$$

$$\frac{d\text{fum}}{dt} = v_{\text{SUCDi}} + v_{\text{FUMt2-2}} - v_{\text{FUM}} \quad (\text{A.38})$$

$$\frac{d\text{actp}}{dt} = v_{\text{PTAr}} + v_{\text{ACKr}} \quad (\text{A.39})$$

$$\frac{d\text{ac}}{dt} = v_{\text{Act2r}} - v_{\text{ACKr}} \quad (\text{A.40})$$

Table A.10: Generior prior on states and parameters. These values were taken from Lubitz *et al.* [10]. Symbols: k_{catf} : forward catalytic constant, k_{catr} : reverse catalytic constant, K_m : Michaelis-Menten constant, K_i : inhibition constant, K_a : activation constant, K_{eq} : equilibrium constant, u : enzyme concentration. c : metabolite concentration.

Symbol	Geometric Mean	Geometric Std. Dev.	Lower Bound	Upper Bound	Unit
k_{catf}	10.0	100.0	1.0×10^{-2}	1.0×10^7	1/s
k_{catr}	10.0	100.0	1.0×10^{-9}	1.0×10^7	1/s
K_m	1.0×10^{-4}	10.0	10×10^{-9}	1.0	M
K_i	1.0×10^{-4}	10.0	1.0×10^{-7}	1.0×10^{-1}	M
K_a	1.0×10^{-4}	10.0	1.0×10^{-7}	1.0×10^{-1}	M
K_{eq}	1.0	100.0	1.0×10^{-10}	1.0×10^8	dimensionless
u	1.0×10^{-6}	100.0	1.0×10^{-9}	5.0×10^{-4}	M
c	1.0×10^{-4}	10.0	1.0×10^{-9}	1.0	M

Table A.11: Enzyme counts on pyruvate for reactions in the *E. coli* TCA cycle model. Enzyme counts for each reaction were estimated using the absolute protein quantification data from Schmidt *et al.* [15] in conjunction with the gene-reaction-rules present in the model (See Supplementary Table A.4) [12]. These counts were aggregated on a per-reaction basis using these boolean rules: AND indicates both proteins are needed, hence we took the minimum, whereas OR indicates both can be used, and hence summation was performed, in order to approximate to total number of enzymes that could catalyze a given reaction.

reaction	mean	std	unit
ACKr	10637.0	793.78	counts
ACONTa	17180.0	749.37	counts
ACONTb	17180.0	749.37	counts
AKGDH	4196.0	625.62	counts
CS	37156.0	4265.51	counts
FUM	3903.0	NaN	counts
FUMt2_2	126.0	30.52	counts
GLUDy	3776.0	682.70	counts
ICDHyr	35445.0	1265.39	counts
ICL	48723.0	1476.31	counts
LDH_D	1123.0	107.59	counts
MALS	4989.0	578.44	counts
MDH	129097.0	9075.52	counts
ME1	480.0	14.208	counts
PDH	15791.0	3611.40	counts
PPC	1536.0	468.17	counts
PPS	4184.0	186.61	counts
PTAr	3512.0	1027.26	counts
PYK	5587.0	219.79	counts
SUCct2_2	126.0	30.51	counts
SUCDi	48.0	8.64	counts
SUCOAS	18432.0	263.58	counts

Table A.12: Metabolite concentrations on pyruvate for species in the *E. coli* TCA cycle model. Metabolite concentrations were estimated by combining and averaging the data from Gerosa *et al.* and Kochanowski *et al.* [3, 8].

metabolite	mean	std	unit
adp	9.876e-04	4.29e-05	M
akg	8.543e-04	2.364e-04	M
amp	7.719e-04	1.41e-05	M
atp	5.776e-03	1.741e-04	M
cit	1.117e-02	8.437e-04	M
fum	1.350e-03	2.062e-04	M
icit	5.620e-04	5.47e-05	M
mal-L	2.363e-03	8.93e-05	M
nad	6.751e-03	1.868e-04	M
nadh	2.018e-04	1.711e-04	M
nadp	1.84e-05	7e-07	M
nadph	1.33e-04	8e-06	M
pep	3.97e-04	3.17e-05	M
succ	1.857e-03	1.187e-03	M

Table A.13: Steady-state fluxes on pyruvate for reactions in the *E. coli* TCA cycle model. Fluxes data were obtained from Gerosa *et. al.* [3].

reaction	mean	std	unit
EX_ac_e	1.40e-03	6.40e-05	M/s
EX_succ_e	0.00e+00	0.00e+00	M/s
EX_pyr_e	3.20e-03	0.00e+00	M/s
PYK	-4.20e-04	3.00e-05	M/s
PDH	2.50e-03	1.30e-05	M/s
CS	9.80e-04	2.70e-06	M/s
ICDHyr	9.60e-04	1.20e-05	M/s
AKGDH	9.00e-04	1.20e-05	M/s
SUCDi	9.10e-04	2.70e-06	M/s
FUM	9.10e-04	2.70e-06	M/s
MDH	9.00e-04	3.00e-05	M/s
ME1	2.30e-05	3.00e-05	M/s
PPC	3.00e-04	1.50e-05	M/s
PPCK	1.40e-04	3.30e-05	M/s
ICL	1.20e-05	1.30e-05	M/s
FUMt2_2	0.00e+00	0.00e+00	M/s
LDH_D	-1.40e-04	8.30e-06	M/s

Table A.14: Standard Gibbs free energy for reactions in the *E. coli* TCA cycle model. Standard Gibbs free energy change (ΔG°) values and their corresponding standard deviations. These values were acquired using Equilibrator [1], where we assumed a constant temperature of 37°C, an intracellular pH of 7.4, and an ionic strength of 0.25M [21, 20]. We sampled the covariance matrix, which is not shown here for convenience. Units are kJ/mol.

	ΔG°	std
PYK	-23.448	0.186
PPS	-6.3	0.258
PPC	-40.211	10.057
PPCK	11.047	10.111
ME1	12.901	10.099
ICL	9.637	1.896
MALS	-38.073	4.501
PDH	-35.732	10.786
CS	-40.278	0.199
ACONTa	8.324	1.336
ACONTb	-0.697	1.338
ICDHyr	5.743	10.547
AKGDH	-29.617	15.675
SUCOAS	1.698	1.847
SUCDi	-71.138	42.165
FUM	-3.433	0.08
MDH	25.303	0.097
PYRt2	0.0	0.0
PTAr	9.247	0.382
ACKr	14.708	0.314
ACt2r	0.0	0.0
SUCt2_2	0.0	0.0
LDH_D	19.412	4.464
GLUDy	31.811	0.146
FUMt2_2	0.0	0.0

Table A.15: Forward catalytic rate constants for reactions of the *E. coli* TCA cycle model.
Data were obtained from the BRENDA database [16].

reaction	gmean	gstd	mean	std	unit
CS	8.10e+01	nan	8.10e+01	nan	1/s
FUM	4.10e+02	4.30e+00	7.80e+02	5.20e+02	1/s
GLUDy	3.70e+01	nan	3.70e+01	nan	1/s
ICDHyr	8.80e+01	nan	8.80e+01	nan	1/s
MALS	4.80e+01	nan	4.80e+01	nan	1/s
PDH	3.80e+01	nan	3.80e+01	nan	1/s
PPC	5.40e+02	nan	5.40e+02	nan	1/s
PTAr	6.00e+01	2.00e+00	7.50e+01	4.50e+01	1/s

	Technical replicates		Biological replicates	
	12C	12C / 13C	12	12C / 13C
Asparagine	0.12	0.13	0.24	0.15
Aspartate	0.13	0.14	0.26	0.12
Fumarate	0.15	0.24	0.39	0.43
Glutamate	0.03	0.04	0.2	0.09
Glutamine	0.11	0.1	0.19	0.1
GTTox	0.21	0.55	0.4	1.21
GTTred	0.47	0.24	0.7	0.59
Hexoses	0.35	0.41	0.62	0.8
Homoserine	0.16	0.41	0.28	1.15
Malate	0.19	0.27	0.49	0.76
Panthothenate	0.13	0.15	0.24	0.19
PEP	0.14	0.11	0.38	0.32
Phenylalanine	0.12	0.12	0.34	0.18
Succinate	0.1	0.09	0.27	0.12
Tyrosine	0.13	0.16	0.36	0.22

Figure A.18: Analysis of best peaks from the pyruvate-glucose-pyruvate shift experiment. Samples were measured as biological triplicates from independent experiments, each of which was injected twice from the same well on the same 96-well plate to assess technical variability.

Table A.16: Michaelis-Menten constants for reaction-metabolite pairs in the *E. coli* TCA cycle model. Data were obtained from the BRENDA database [16].

reaction	species	gmean	gstd	mean	std	unit
ACKr	ac	4.60e-02	6.50e+00	1.50e-01	1.50e-01	M
ACKr	actp	8.90e-04	5.60e+00	2.60e-03	2.40e-03	M
ACKr	adp	1.50e-03	nan	1.50e-03	nan	M
ACKr	atp	9.40e-04	nan	9.40e-04	nan	M
ACONTa	acon_C	3.00e-05	1.90e+00	3.70e-05	2.10e-05	M
ACONTa	cit	3.60e-03	3.10e+00	6.10e-03	4.90e-03	M
ACONTb	acon_C	3.00e-05	1.90e+00	3.70e-05	2.10e-05	M
ACONTb	cit	3.60e-03	3.10e+00	6.10e-03	4.90e-03	M
AKGDH	akg	1.00e-04	nan	1.00e-04	nan	M
AKGDH	nad	3.00e-04	3.40e+00	6.10e-04	7.10e-04	M
AKGDH	pyr	4.00e-04	nan	4.00e-04	nan	M
CS	accoa	4.00e-04	2.30e+00	5.20e-04	2.90e-04	M
CS	oaa	1.70e-05	1.50e+00	1.80e-05	7.50e-06	M
FUM	fum	3.70e-04	2.20e+00	5.10e-04	4.60e-04	M
FUM	mal__L	3.90e-04	3.10e+00	6.70e-04	7.20e-04	M
GLUDy	akg	1.10e-03	4.40e+01	1.00e-01	2.00e-01	M
GLUDy	glu__L	9.70e-03	1.70e+01	3.40e-01	5.80e-01	M
GLUDy	nadp	1.10e-04	9.20e+00	2.70e-03	6.40e-03	M
GLUDy	nadph	7.20e-05	1.60e+01	6.70e-03	1.90e-02	M
GLUDy	nh4	8.90e-03	8.50e+00	6.90e-02	1.20e-01	M
ICDHyr	icit	3.40e-05	1.20e+00	3.50e-05	5.80e-06	M
ICDHyr	nadp	3.90e-05	nan	3.90e-05	nan	M
ICL	glx	1.30e-04	nan	1.30e-04	nan	M
ICL	icit	6.50e-04	nan	6.50e-04	nan	M
MALS	accoa	1.40e-05	1.60e+00	1.50e-05	6.50e-06	M
MALS	glx	3.40e-05	1.60e+00	3.80e-05	1.70e-05	M
MDH	mal__L	1.80e-05	1.50e+02	1.30e-03	1.30e-03	M
MDH	nad	1.50e-05	1.70e+01	1.30e-04	1.30e-04	M
MDH	nadh	6.10e-05	nan	6.10e-05	nan	M
MDH	oaa	3.60e-05	1.50e+00	3.90e-05	1.40e-05	M
ME1	mal__L	5.20e-04	2.00e+00	6.40e-04	4.10e-04	M
ME1	nad	5.80e-05	1.20e+00	5.90e-05	8.30e-06	M
ME1	nadh	2.50e-05	nan	2.50e-05	nan	M
ME1	nadp	2.30e-03	nan	2.30e-03	nan	M
ME1	oaa	3.70e-03	1.50e+00	4.00e-03	1.30e-03	M
PDH	nad	3.00e-04	3.40e+00	6.10e-04	7.10e-04	M
PDH	pyr	3.80e-04	1.30e+00	3.90e-04	1.00e-04	M
PPC	co2	1.90e-04	nan	1.90e-04	nan	M
PPC	pep	6.00e-04	nan	6.00e-04	nan	M
PPCK	adp	5.00e-05	nan	5.00e-05	nan	M
PPCK	atp	6.00e-05	nan	6.00e-05	nan	M
PPCK	oaa	6.70e-04	nan	6.70e-04	nan	M
PPCK	pep	3.20e-04	9.70e+00	2.70e-03	3.70e-03	M
PPS	atp	2.80e-05	nan	2.80e-05	nan	M
PPS	pyr	8.30e-05	nan	8.30e-05	nan	M
PTAr	accoa	2.10e-05	2.20e+00	2.70e-05	1.80e-05	M
PTAr	actp	5.30e-04	1.70e+00	6.10e-04	2.90e-04	M
PTAr	coa	3.30e-05	nan	3.30e-05	nan	M
SUCOAS	atp	3.20e-05	2.20e+00	4.20e-05	2.80e-05	M
SUCOAS	coa	4.00e-03	nan	4.00e-03	nan	M

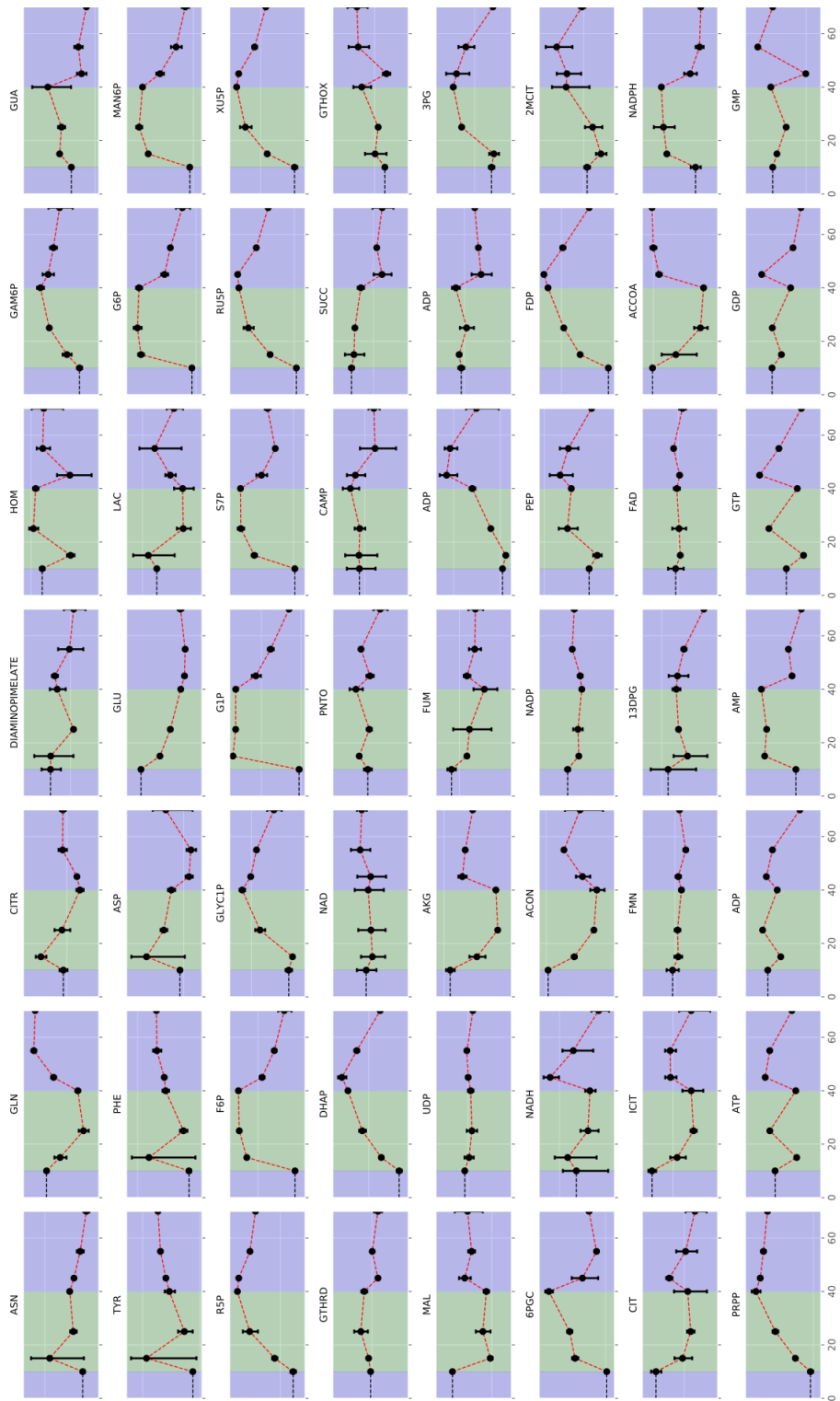


Figure A.19: Pyruvate-glucose-pyruvate switch. The black dots and error bars indicate the mean and standard deviations of the measurements, the red dashed line is a linear interpolation of the data. Data were obtained from [9].

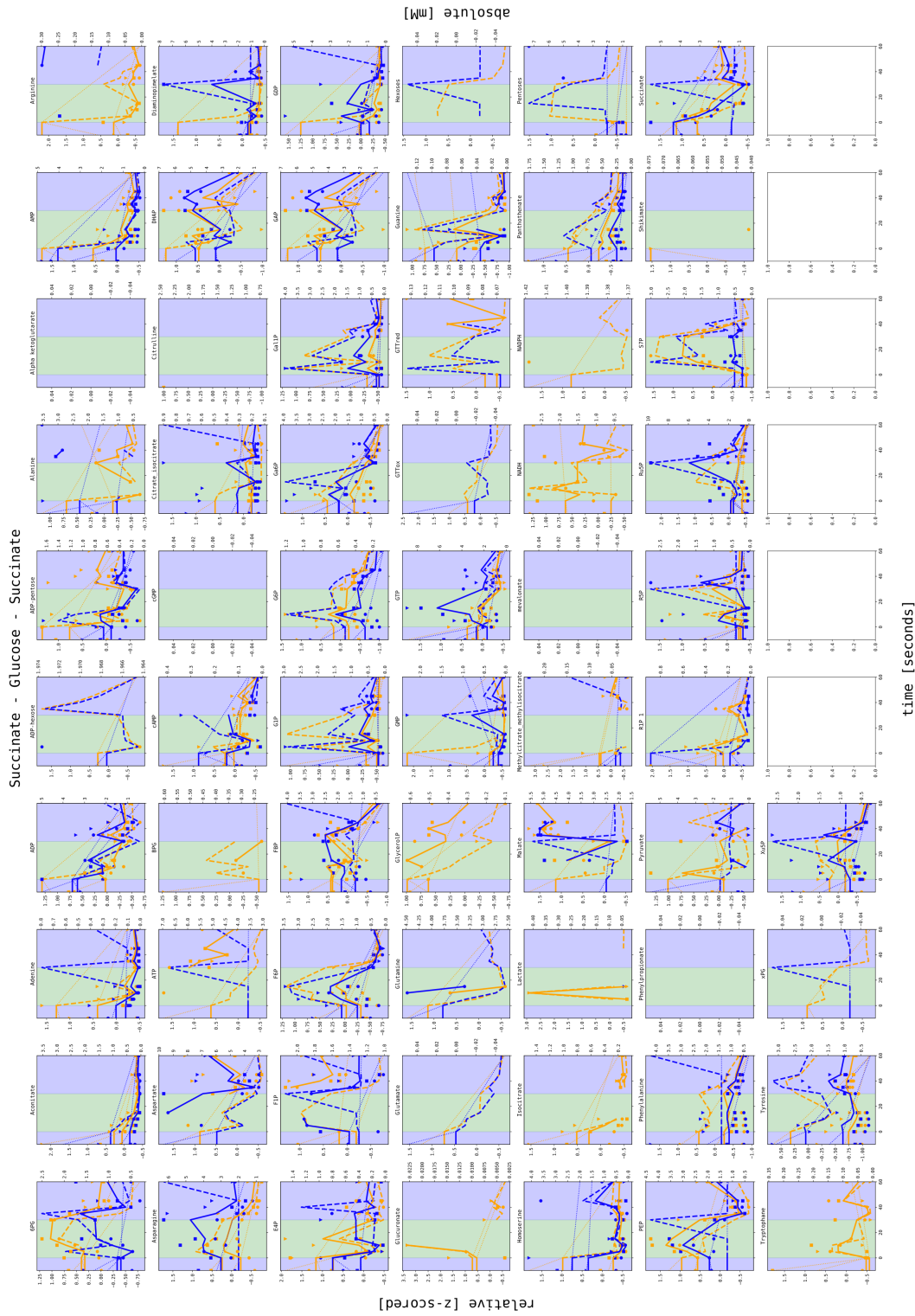


Figure A.22: Succinate-glucose-succinate switch. Among the 11 time points with 3 replicates each, there were 35 out of 65 metabolites for which more than half of time points could be integrated from the mass spectrometry data. For those metabolite present in the model, this covers aconitate, ADP, AMP, (Iso)citrate, PEP and succinate. Squares, circles and triangles indicate individual biological replicates, blue lines are non-normalized times series, yellow lines are ¹³C-normalized times series, dashed lines are relative quantification data, solid lines are absolute quantification data computed from the calibration curves, and dotted-lines are linear interpolation of the initial time point and the time point of the control sample where the cells were perfused by not perturbed.

Bibliography

- [1] Avi Flamholz et al. “eQuilibrator—the biochemical thermodynamics calculator”. In: *Nucleic acids research* 40.D1 (2012), pp. D770–D775.
- [2] Mathias Ganter et al. “MetaNetX. org: a website and repository for accessing, analysing and manipulating metabolic networks”. In: *Bioinformatics* 29.6 (2013), pp. 815–816.
- [3] Luca Gerosa et al. “Pseudo-transition analysis identifies the key regulators of dynamic metabolic adaptations from steady-state data”. In: *Cell systems* 1.4 (2015), pp. 270–282.
- [4] Christopher S Henry et al. “High-throughput generation, optimization and analysis of genome-scale metabolic models”. In: *Nature biotechnology* 28.9 (2010), pp. 977–982.
- [5] Minoru Kanehisa et al. “The KEGG database”. In: *Novartis found symp.* Vol. 247. 2002, pp. 91–103.
- [6] Zachary A King et al. “BiGG Models: A platform for integrating, standardizing and sharing genome-scale models”. In: *Nucleic acids research* 44.D1 (2016), pp. D515–D522.
- [7] Zachary A King et al. “Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways”. In: *PLoS computational biology* 11.8 (2015), e1004321.
- [8] Karl Kochanowski et al. “Global coordination of metabolic pathways in Escherichia coli by active and passive regulation”. In: *Molecular systems biology* 17.4 (2021), e10064.
- [9] Hannes Link, Karl Kochanowski, and Uwe Sauer. “Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo”. In: *Nature biotechnology* 31.4 (2013), pp. 357–361.
- [10] Timo Lubitz et al. “Parameter balancing in kinetic models of cell metabolism”. In: *The Journal of Physical Chemistry B* 114.49 (2010), pp. 16298–16303.
- [11] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. “What is flux balance analysis?” In: *Nature biotechnology* 28.3 (2010), pp. 245–248.
- [12] Jeffrey D Orth et al. “A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011”. In: *Molecular systems biology* 7.1 (2011), p. 535.
- [13] Ed Reznik et al. “Genome-scale architecture of small molecule regulatory networks and the fundamental trade-off between regulation and enzymatic activity”. In: *Cell reports* 20.11 (2017), pp. 2666–2677.
- [14] Jan Schellenberger et al. “BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions”. In: *BMC bioinformatics* 11.1 (2010), pp. 1–10.
- [15] Alexander Schmidt et al. “The quantitative and condition-dependent Escherichia coli proteome”. In: *Nature biotechnology* 34.1 (2016), pp. 104–110.
- [16] Ida Schomburg, Antje Chang, and Dietmar Schomburg. “BRENDA, enzyme data and metabolic information”. In: *Nucleic acids research* 30.1 (2002), pp. 47–49.
- [17] Daniel Segre, Dennis Vitkup, and George M Church. “Analysis of optimality in natural and perturbed metabolic networks”. In: *Proceedings of the National Academy of Sciences* 99.23 (2002), pp. 15112–15117.
- [18] Benjamin Volkmer and Matthias Heinemann. “Condition-dependent cell volume and concentration of Escherichia coli to facilitate data conversion for systems biology modeling”. In: *PloS one* 6.7 (2011), e23126.

-
- [19] Edwin C Webb et al. *Enzyme nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes*. Ed. 6. Academic Press, 1992.
- [20] Jessica C Wilks and Joan L Slonczewski. “pH of the cytoplasm and periplasm of *Escherichia coli*: rapid measurement by green fluorescent protein fluorimetry”. In: *Journal of bacteriology* 189.15 (2007), pp. 5601–5607.
- [21] Dan Zilberstein et al. “*Escherichia coli* intracellular pH, membrane potential, and cell growth”. In: *Journal of bacteriology* 158.1 (1984), pp. 246–252.