Doctoral Thesis

Metabolic decision making by protein-metabolite interactions in Escherichia coli

Author(s):
Kochanowski, Karl Andreas

Publication Date:
2015

Permanent Link:
https://doi.org/10.3929/ethz-a-010604550

Rights / License:
In Copyright - Non-Commercial Use Permitted
Metabolic decision making by protein-metabolite interactions in *Escherichia coli*

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

*Karl Andreas Kochanowski*

Dipl. Biol., RWTH Aachen, Germany
born on 24.10.1983, citizen of Germany

Accepted on the recommendation of:

*Prof. Dr. Uwe Sauer, examiner*
*Prof. Dr. Matthias Heinemann, co-examiner*
*Prof. Dr. Martin Ackermann, co-examiner*

2015
# Contents

**Abstract** ................................................................................................................................. 4  
**Zusammenfassung** .......................................................................................................................... 6  
**Chapter 1 - General Introduction** ........................................................................................... 9  
(Microbial) life is full of tough metabolic decisions ................................................................. 10  
Metabolic decisions are made by regulatory circuits through protein-metabolite interactions...... 10  
Regulation of substrate uptake ................................................................................................. 11  
Catabolite repression ................................................................................................................. 11  
Regulation of central metabolism ............................................................................................ 12  
Regulation of energy metabolism ............................................................................................. 14  
Regulation of amino acid uptake and metabolism ................................................................. 16  
Regulation of protein biosynthesis ......................................................................................... 17  
Approaches to study metabolic regulatory circuits .................................................................. 18  
Estimating metabolic fluxes ....................................................................................................... 18  
Quantifying metabolite concentrations .................................................................................... 19  
Quantifying protein expression and activity ........................................................................... 20  
Identifying regulatory interactions between proteins, DNA and metabolites........................... 21  
Focus of this thesis ..................................................................................................................... 23  
**Tables** .................................................................................................................................. 26  
**Figures** .................................................................................................................................. 27  
**References** ............................................................................................................................. 31  
**Chapter 2 – the role of transcription in regulating microbial metabolic fluxes** ..................... 41  
Abstract ....................................................................................................................................... 42  
Introduction ................................................................................................................................. 43  
Effect of genetic perturbations on metabolic fluxes .................................................................. 43  
Enzyme overabundance as buffer .............................................................................................. 44  
Inferring metabolic state from transcriptomics ......................................................................... 46  
Suboptimality of gene expression .............................................................................................. 46  
Conclusion .................................................................................................................................... 48  
**Figures** .................................................................................................................................. 49  
**References** ............................................................................................................................. 50  
**Chapter 3 – Dissecting specific and global transcriptional regulation of bacterial gene expression.** 54  
Abstract ....................................................................................................................................... 55  
Introduction ................................................................................................................................... 56  
Results ......................................................................................................................................... 58
Mapping posttranslational regulatory events .......................................................... 163
Distinguishing regulatory signals from regulatory chatter ........................................ 163
Coarse-grained posttranslational regulation? ............................................................. 165
Metabolism as a source of phenotypic variability ...................................................... 165
Conclusion .................................................................................................................. 166
Figures ......................................................................................................................... 167
References ..................................................................................................................... 169

Chapter 7 – Systematic identification of protein-metabolite interactions in complex metabolite mixtures by ligand-detected NMR .......................................................... 172
Abstract ....................................................................................................................... 173
Introduction .................................................................................................................. 174
Results ......................................................................................................................... 176
Discussion .................................................................................................................... 179
Acknowledgements ...................................................................................................... 183
Material and Methods .................................................................................................. 184
Tables ........................................................................................................................... 187
Figures ........................................................................................................................ 189
Supplementary figures ............................................................................................... 193
Supplementary Tables ................................................................................................. 197
References ..................................................................................................................... 198

Chapter 8 - Concluding remarks ............................................................................... 201
Findings ........................................................................................................................ 202
Outlook ......................................................................................................................... 205
References ..................................................................................................................... 208

Acknowledgements .................................................................................................... 211
Curriculum Vitae ......................................................................................................... 212
Publications .................................................................................................................. 213
Abstract

Metabolism lies at the core of microbial life and fuels all cellular activities with building blocks, reducing power, and energy. To regulate metabolic activity, microbes utilize a complex network of regulatory layers, such as transcriptional regulation, covalent posttranslational modifications, and allosteric. Importantly, metabolism is not merely the endpoint of regulation, but can also provide feedback in form of metabolites that modulate the activity of regulatory proteins. However, how this complex interplay of regulatory layers and metabolic feedback ultimately gives rise to a coordinated metabolic response is currently only poorly understood.

The aim of this thesis is to elucidate the role of protein-metabolite interactions in making metabolic decisions in the model bacterium *Escherichia coli*. Specifically, we focus on protein-metabolite interactions in two regulatory layers, namely transcriptional (chapters 2 to 5) and posttranslational (chapters 5 to 7) enzyme regulation.

In chapter 2, we review our current understanding of the role of transcription in regulating microbial fluxes, focusing on central metabolism. We highlight recent works that investigated the impact of altering enzyme abundance, for example through perturbations of the transcriptional regulatory network, on central metabolic fluxes, and discuss the trade-off between efficient metabolic operation in a given environment and resilience against inevitable internal and external fluctuations.

In chapter 3, we develop a combined experimental-computational approach to dissect the contribution of specific transcription factors and global transcriptional regulation to *E. coli*’s gene expression. Using the L-arginine biosynthesis pathway as an example, we show that global transcriptional regulation sets each promoter’s maximal capacity in a largely growth-dependent manner. Specific regulation then decides how much of this capacity is being used based on the current demand for the pathway’s end product.

In chapter 4, we expand our analysis to quantify the impact of specific and global transcriptional regulation on central metabolic promoters. We find that global transcriptional regulation dominates the steady state response of many metabolic promoters, explaining about 70% of the total variance across various conditions. We further relate each promoter’s remaining specific transcriptional regulation with the cell’s metabolome response across the same conditions to identify metabolites which serve as potential regulatory signals. We show that few metabolites - cyclic AMP, fructose-1,6-bisphosphate (FBP) and fructose-1-phosphate (F1P) - explain most of this specific regulation through their interaction with the transcription factors Crcp and Cra.

In chapter 5, we aim to unravel the regulatory mechanisms that coordinate *E. coli*’s metabolic response to nutrient limitation. By integrating *E. coli*’s steady state response to genetically implemented
limitation of external glucose supply and internal glutamate production in the framework of regulation analysis, we obtain a quantitative picture of the relationship between metabolic fluxes, metabolites, and proteins, at single reaction resolution. We find that this metabolic response is largely established by two mechanisms, namely an approximate transcriptional program, which rarely controls fluxes alone, as well as passive regulation of enzyme activity through changes in enzyme saturation. Surprisingly, this approximate program is implemented by a single transcription factor, Crp, that directly activates catabolic enzymes, and indirectly represses anabolic enzymes by sequestering the cellular resources available for their expression.

In chapter 6, we review our current understanding of how microbial metabolism is shaped by posttranslational regulation, focusing on covalent posttranslational modifications and allosteric interactions. Recent technological advances have now made the systematic genome-wide mapping of covalent posttranslational modifications possible. In contrast, a lack of analogous methods for the investigation of allosteric interactions remains a major obstacle in understanding their regulatory role. For both regulatory layers, the current key challenge is to identify those posttranslational regulatory events that are actually relevant in a given condition.

In chapter 7, we develop an experimental in vitro approach to systematically identify protein-metabolite interactions. In a proof-of-concept study, we use ligand-detected Nuclear Magnetic Resonance (NMR) to identify metabolite binders of four well-characterized proteins (AroG, Eno, PfkA, BSA) among mixtures comprising up to 33 metabolites. We retrieve most previously reported interactions, and identify several novel ones, such as promiscuous binding of nucleotide phosphates and citrate to all tested proteins, and binding of L-tryptophan and L-tyrosine to AroG. We functionally validate these novel interactions for AroG using in vitro enzymatic assays and find that both L-tryptophan and L-tyrosine inhibit AroG activity.

Finally, in chapter 8 we summarize the key findings in this thesis and conclude that metabolic decisions in E. coli emerge from a regulatory division-of-labor between transcriptional and posttranslational regulation. An approximate transcriptional program partitions the cellular resources based on information provided by few intracellular regulatory metabolites, such as FBP, cyclic AMP, and ketoads, and uses only a small fraction of the cell’s repertoire of transcription factors in a given condition. To ultimately establish the metabolic response, additional posttranscriptional regulation is necessary. We envisage that the approaches and concepts developed in this thesis may allow to systematically identify and characterize novel regulatory metabolites in E. coli as well as other organisms, and may further help to guide efforts in the metabolic engineering of microbes.
Zusammenfassung

Stoffwechsel spielt eine zentrale Rolle in Mikroorganismen und versorgt sie mit biosynthetischen Bausteinen, Reduktionsäquivalenten und Energie. Um ihren Stoffwechsel zu regulieren, können Zellen auf ein komplexes Netzwerk mit verschiedenen Regulationsebenen, wie zum Beispiel Transkription, posttranslationale Modifikationen und Allosterie, zurückgreifen. Dabei fungiert der mikrobielle Stoffwechsel nicht nur als regulatorischer Endpunkt, sondern kann auch selbst Informationen an das die anderen zellulären Netzwerke senden, beispielsweise in Form von Metaboliten, die die Aktivität von regulatorischen Proteinen verändern. Allerdings ist weitgehend unklar, wie Zellen basierend auf dieser Vielzahl an molekularen Interaktionen koordinierte metabolische Entscheidungen treffen.

Das Ziel dieser Arbeit ist es, die Rolle von Protein-Metabolit Interaktionen in solchen metabolischen Entscheidungen am Beispiel des Modellbakteriums *Escherichia coli* aufzuschlüsseln. Insbesondere fokussieren wir uns auf zwei Regulationsebenen: Transkription (Kapitel 2 bis 5) und posttranslationale Regulation (Kapitel 5 bis 7).


In Kapitel 3 entwickeln wir eine neue Methode die es erlaubt, die regulatorischen Anteile spezifischer Transkriptionsfaktoren, sowie globaler transkriptioneller Regulation, an der Expression von Genen in *E. coli* getrennt voneinander zu quantifizieren. Wir nehmen den Biosyntheseweg der Aminosäure Arginin als Beispiel und zeigen, dass globale transkriptionelle Regulation die maximale Kapazität dieser Promotoren basierend auf der Wachstumsrate einstellt. Wieviel von dieser Kapazität die Zelle letztlich benutzt, wird durch spezifische transkriptionelle Regulation bestimmt und hängt vom aktuellen Bedarf am Stoffwechselprodukt ab.

In Kapitel 4 erweitern wir unseren Blickwinkel auf den gesamten Zentralstoffwechsel, um den Anteil spezifischer und globaler transkriptioneller Regulation zu bestimmen. Unsere Analysen zeigen, dass globale transkriptionelle Regulation einen dominierenden Einfluss auf viele Promotoren hat und etwa 70% aller gemessenen Veränderungen in einer Vielzahl an Bedingungen erklären kann. Zudem identifizieren wir Metabolite die als potentielle regulatorische Signale wirken können, indem wir sie systematisch gegen die spezifische Regulationskomponente jedes Promotors testen. Wir zeigen dass
drei Metabolite – zyklisches AMP, Fruktose-1,6-Bisphosphat (FBP) und Fruktose-1-Phosphat (F1P) ausreichen, um den Grossteil dieser spezifischen Regulation zu erklären.

In Kapitel 5 untersuchen wir die Mechanismen, die die metabolische Antwort von E. coli auf Nährstofflimitierung koordinieren. Indem wir die stationäre Antwort auf genetisch implementierte Limitierung der Glukoseaufnahme und Glutamatproduktion mittels integrierter Regulationsanalyse untersuchen, erhalten wir ein quantitatives Abbild des Wechselspiels zwischen Stoffwechselflüssen, Proteinen und Metaboliten für jede einzelne Reaktion. Wir zeigen dass E. coli metabolic response weitgehend durch eine Kombination aus ungenauen transkriptionellen Regulationsprogramm, sowie passiver Regulation der Enzysättigung beeinflusst wird. Interessanterweise wird dieses ungenaue Regulationsprogramm durch einen einzelnen Transkriptionsfaktor, Crp, umgesetzt, der direkt katabolische Enzyme induziert, und zudem auf indirekte Weise auch die Konzentration anabolischer Enzyme reduziert, in dem er die verfügbaren zellulären Ressourcen für ihre Expression beschränkt.

In Kapitel 6 fassen wir unser Verständnis des Effekts posttranslationaler Regulation auf den mikrobiellen Stoffwechsel zusammen, mit Fokus auf kovalente posttranslationale Modifikationen und Allosterie. Während technologische Fortschritte es uns jetzt ermöglichen, posttranslationale Modifikationen genaue zu kartieren, herrscht immer noch Mangel an entsprechenden Methoden für die Kartierung allosterischer Interaktionen. In beiden Fällen liegt die Herausforderung im Moment darin, herauszufinden, welche der beobachteten Modifikationen/Interaktionen tatsächlich relevant sind in einer bestimmten Umgebung.

In Kapitel 7 entwickeln wir einen experimentellen in vitro Ansatz, um Protein-Metabolit Interaktionen systematisch aufzuspüren. In dieser Konzeptstudie benutzen wir Kernspinresonanzspektroskopie, um für vier bereits gut charakterisierte Proteine (AroG, Eno, PfkA, BSA) aus einer Mischung bis zu 33 Metaboliten diejenigen zu identifizieren, die an die Proteine binden. Wir können die Mehrzahl bereits bekannter Interaktionen wiederfinden, und finden zudem mehrere neue, beispielsweise die unspezifische Bindung von Nukleotidphosphaten und Citrat, sowie die spezifische Bindung von Tryptophan und Tyrosin an AroG. Wir nutzen Enzymassays, um diese neuen Interaktionen für AroG funktionell zu validieren, und zeigen dass sowohl Tryptophan als auch Tyrosin AroG inhibieren.

Im abschliessenden Kapitel 8 fassen wir die Ergebnisse dieser Arbeit zusammen und schlussfolgern dass E. coli metabolic Entscheidungen weitgehend das Ergebnis einer regulatorischen Arbeitsteilung zwischen Transkription und posttranslationaler Regulation sind. Ein ungenaues Transkriptionsprogramm teilt die zellulären Ressourcen basierend auf wenigen internen Metabolitsignalen, wie zum FBP, zyklisches AMP, und Ketosäuren, ein, und nutzt nur einen Bruchteil
Zusammenfassung

Part of this chapter has been published in:


*: equal contribution
(Microbial) life is full of tough metabolic decisions

Microbial life occupies virtually all niches on earth, from the upper troposphere to deep subseafloor sediments, ecosystems deep within earth, and desolate man-made niches such as high-level nuclear waste sites and the ruined reactors of the Chernobyl nuclear power plant. Mirroring the enormous variation in occupied niches and life styles, cells draw from a large set of metabolic reactions: around 6000 unique reactions constitute the full known metabolic repertoire across all domains of life, and fuel all cellular activities with building blocks, reducing power and energy.

Despite their variation in occupied niches, life styles, and metabolic capacity, the fundamental metabolic tasks are highly similar across divergent species (Figure 1A). For example, all organisms must scavenge nutrients and then coordinate central metabolism, monomer synthesis, and macromolecule polymerization for biomass synthesis and proliferation. Furthermore, information about the metabolic state has to be transmitted to other cellular processes, to coordinate the availability of nutrients and energy with cellular functions. Ultimately, these tasks can be broken down into metabolic decisions which cells need to make to establish the flux magnitude of linear pathways and to partition the incoming or outgoing fluxes at branch points (Figure 1B).

Metabolic decisions are made by regulatory circuits through protein-metabolite interactions

Given the large and densely connected network of metabolites, enzymatic reactions, and regulatory interactions, it is a challenge to understand the intertwined metabolic and regulatory network in its entirety. Consequently, understanding why specific metabolic decisions are being made, and how they are implemented at a molecular level, is difficult even for well characterized organisms. However, it is possible to define individual regulatory circuits with specific information inputs and regulatory outputs. These circuits can then be studied with a combination of biological and theoretical approaches to unravel the molecular components and mechanisms that control them, as well as the resulting metabolic decisions (Figure 2). In this section, we review a number of regulatory circuits that regulate metabolic decision making in substrate uptake, operation of central carbon and energy metabolism, metabolism of amino acids, and ultimately protein biosynthesis, mostly focusing on the Gram negative model bacterium Escherichia coli. A key commonality of these circuits is their reliance on interactions between regulatory metabolites and proteins. Some of these regulatory metabolites, such as fructose-1,6-bisphosphate (FBP), α-ketoglutarate, or adenosine-triphosphate (ATP), occupy key positions in the metabolic network and inform cells about their global metabolic status (table 1), whereas other regulatory metabolites, such as arginine, report on the activity of a specific pathway, but all of these metabolites exert their function through their interaction with enzymes, transcription
factors, and other regulatory proteins. Thus, protein-metabolite interactions lie at the core of metabolic decision making in microorganisms.

**Regulation of substrate uptake**

Heterotrophic microbes such as *E. coli* thrive on a large number of carbon and energy sources\(^{14,15}\). Since permanent expression of all transporters would consume valuable cellular resources and occupy limited membrane space, cells selectively express transport systems based on extracellular and intracellular signals. Extracellular signals are primarily detected by two-component systems in which a membrane-bound sensor activates a cytoplasmic regulator in response to an external stimulus\(^{16}\). About 30 such two-component systems are encoded in the *E. coli* genome, including sensors of phosphate, nitrate\(^{17,18}\), and at least one class of externally sensed carbon sources (phosphorylated sugars\(^{19}\)). For the vast majority of carbon sources, however, *E. coli* and many other heterotrophs rely on an intracellular sensor, typically a transcription factor, that both senses the signal and provides a regulatory output\(^{20}\). In this one-component internal sensing scheme transporters and enzymes for utilization of various carbon sources are expressed at basal levels, leading to an increase in pathway intermediates upon nutrient availability. This leads to upregulation of transporter and enzyme expression through a transcription factor that senses these intermediates. The classic example of this scheme is the repressor LacI of the lac operon, which is released from the lac promoter upon binding of intracellular allolactose\(^{21}\) ([Figure 3](#)). Other examples include uptake of glucosamine, trehalose, fucose and maltose in *E. coli*\(^{20}\), and phylogenetic evidence suggests that such one-component sensors are in fact the dominant nutrient sensing mechanism in prokaryotes\(^{22}\). The regulatory logic of the internal sensing scheme results in a positive feedback loop, which enables small changes in substrate abundance to trigger a large transcriptional response\(^{13,23}\). Thus, the common regulatory principle of these circuits that enable demand-based uptake of alternative nutrients is well understood: accumulation of a pathway intermediate signals nutrient availability and this information then is transferred via a transcription factor to increase the magnitude of uptake flux.

**Catabolite repression**

The positive feedback circuits described above enable carbon source-specific regulation but do not allow prioritization amongst multiple substrates. The existence of such prioritization is evident from ‘diauxic growth’ which results from sequential substrate consumption\(^{24}\). Several regulatory circuits, collectively known as carbon catabolite repression, are known to achieve this prioritization\(^{25}\) by sensing the presence of preferred carbon sources and reducing uptake of alternative carbon sources. In *E. coli*, one of the most common preferred sources is glucose, which is transported into the cell via the phosphotransferase uptake system (PTS) ([Figure 3B](#)). When glucose is taken up by the PTS, one of
the PTS components (EIIA) is dephosphorylated, and directly inhibits transporters for several non-preferred carbon sources. This mechanism is referred to as 'inducer exclusion'\textsuperscript{25,26}.

While inducer exclusion is, at least in some cases, sufficient for achieving carbon source prioritization, \textit{E. coli} encodes an additional system to downregulate expression of genes responsible for transport and catabolism of non-preferred substrates. This system centers around the transcription factor Crp, which positively regulates the expression of a number of carbon uptake systems along with a suite of other genes involved in carbon catabolism\textsuperscript{27,28}. Crp is activated by the intracellular messenger cyclic AMP (cAMP), and cAMP synthesis by the enzyme adenylate cyclase is in turn activated by phosphorylated EIIA\textsuperscript{26,29}. As described above, phosphorylated EIIA is the dominant form only in the absence of glucose, so by this mechanism, external glucose prevents cAMP synthesis and thus Crp activation of alternative carbon uptake genes.

However, carbon sources that allow high growth rates but are not transported through the PTS also cause catabolite repression through Crp, suggesting that Crp is sensitive not just to the presence of a set of particular preferred sugars. A partial explanation of this glucose-independent repression was the finding that EIIA phosphorylation depends not only on the availability of glucose but also on the ratio of the central metabolites phosphoenolpyruvate (PEP) and pyruvate\textsuperscript{30}. However, several results remain unexplained, which led the authors of a recent review to postulate an unknown “factor X” as a regulator of catabolite repression\textsuperscript{25}. The demonstration that Crp activity is not only induced by carbon limitation, but also repressed by nitrogen or sulfur limitation, suggested that the information transferred to Crp is not general carbon availability, but rather the balance between carbon catabolism and anabolic capacity\textsuperscript{31}. Previous theoretical work had suggested\textsuperscript{32} that α-ketoglutarate and other α-ketoacids such as pyruvate and oxaloacetate, which are the direct carbon precursors for the transamination reaction in amino acid biosynthesis, could be effective regulators of carbon catabolic flux. Indeed, \textit{in vitro} experiments with permeabilized cells showed that α-ketoglutarate and related α-ketoacids inhibited the cAMP producing enzyme adenylate cyclase, closing the regulatory circuit between carbon availability and Crp activity in an elegant negative feedback loop (Figure 3C). In this regulatory circuit the information on the balance between catabolism and anabolic capacity is transferred to the activity of the transcription factor Crp through the concentration of α-ketoacids, resulting in a general shut down of catabolic gene expression when the ratio of carbon to nitrogen availability is high and induction when this ratio is low\textsuperscript{31}.

\textbf{Regulation of central metabolism}

In contrast to the independently operating and mostly well-understood regulatory circuits of uptake pathways, the densely connected network of regulatory interactions within central carbon metabolism has hampered the investigation of its regulatory circuits. Knowing the transcriptional regulatory
network is largely insufficient for understanding central metabolic operations because metabolic control relies heavily on allosteric regulation through metabolite binding and post-translational protein modifications, which is unsurprising given that metabolism might need to change rapidly—*E. coli* can adapt to environmental changes that reverse central fluxes in a matter of seconds. Nevertheless, an emerging theme is that cells rely on internal signals that are largely independent of the exact source of carbon, and the critical cues seem to come from a limited number of central metabolites (Table 2).

A particularly well-characterized example of signaling that uses both allosteric and transcriptional regulation is the substrate-dependent switch of glycolytic operation (Figure 3D). The transcription factor Cra is largely responsible for activation of gluconeogenic enzymes and repression of glycolytic enzymes, and is inactivated during growth on glycolytic carbon sources by binding to the glycolysis intermediate FBP. It is not immediately obvious why FBP should be more abundant during glycolytic growth. The key insight into the role of FBP as a sensor of glycolytic flux came from the realization that FBP is an allosteric activator of the downstream enzymes pyruvate kinase and PEP carboxylase, and thus FBP accumulates until enzyme activity in the lower part of glycolysis matches the upper glycolytic flux. Interestingly, the information transfer from glycolytic flux to the activity of Cra emerges from the topology of the circuit’s transcriptional and allosteric regulatory interactions and appears to be largely independent of the exact kinetic parameters. One situation in which this glycolytic flux sensor was found to have a massive impact on cellular physiology are environmental nutrient shifts. During a rapid shift from a glycolytic (such as glucose) to a gluconeogenic (such as acetate) carbon source, only a fraction of the otherwise isogenic population resumes growth on the gluconeogenic carbon source, while the rest of the cell population enters a non-growing dormant state which also shows increased resistance to antibiotic treatment. The size of the switching fraction in this responsive-diversification mechanism depends on the magnitude of the glycolytic flux prior to the shift. In conditions supporting high glycolytic fluxes, such as growth in glucose excess cultures, only a small fraction manages the switch, whereas in conditions supporting low glycolytic fluxes, such as glucose-limited chemostats, the fraction of switching cells is dramatically increased, demonstrating that a central metabolic regulatory circuit can affect cellular decision-making well beyond the regulation of metabolism as such.

The allosteric activation of pyruvate kinase by FBP is conserved even in higher eukaryotes including humans, suggesting the potential utility of the flux-sensing circuit described above. Curiously, in *B. subtilis* this feedforward activation appears to be absent. Nevertheless, *B. subtilis* also uses FBP to modulate the activity of several key transcription factors, notably CcpA and CggR, to control the expression of enzymes in central metabolism and as in *E. coli*, FBP concentration in *B. subtilis* correlates with glycolytic flux. The mechanism that is responsible for this correlation in the absence
of the pyruvate kinase-FBP interaction is unclear, since *B. subtilis* also does not encode PEP carboxylase, the other FBP-sensitive enzyme in *E. coli*. Apparently, the relationship between FBP and glycolytic flux is also maintained in yeast, in which FBP also activates pyruvate kinase. Although a role in transcriptional regulation has not yet been elucidated for FBP in yeast, other regulatory roles for this metabolite have been proposed such as inhibition of respiratory energy generation. The ultrasensitive FBP regulation of pyruvate kinase and PEP carboxylase, both of which act on the common substrate PEP, results in an inverse relationship between FBP and PEP concentrations. Accordingly, upon glucose depletion, when glycolytic flux reaches zero, PEP accumulates. This accumulation seems to ensure that there is sufficient PEP to serve as a substrate for phosphorylation of newly taken up glucose through the PTS system when glucose becomes available again. In *S. cerevisiae*, where glucose phosphorylation does not rely on PEP as a phosphate donor, PEP accumulation might simply be a way to store ATP equivalents. In *S. cerevisiae* PEP also accumulates upon oxidative stress and fosters the production of the redox protectant NADPH in the pentose phosphate pathway by inhibiting the glycolytic enzyme triosephosphate isomerase.

Whereas the regulatory interactions of FBP shed a great deal of light on the regulation of glycolysis, the network of hundreds of other interactions in central metabolism has so far eluded comprehensive understanding. Several molecules, such as PEP, pyruvate, glyoxylate and oxaloacetate, are highly connected, activating or inhibiting dozens of reactions and the activity of several transcription factors. However, precisely because of this large number of interactions, many of which might be of little relevance, little intuitive understanding of their role has emerged. For instance pyruvate, an allosteric effector of several glycolytic enzymes and two transcription factors (PdhR and IclR), not only activates its own consumption through the induction of pyruvate dehydrogenase, but also regulates genes involved in cell division, peptidoglycan synthesis, and other distant metabolic reactions, and it is unclear which information pyruvate concentration transfers to those processes. Meanwhile, quantitative models of metabolism are hampered by poor characterization of the biochemical parameters. The best predictions of central metabolic fluxes under environmental or genetic perturbations are therefore still based on heuristics or optimality principles as opposed to biochemical kinetics.

**Regulation of energy metabolism**

As the thermodynamic driving force of all cellular processes, biochemical energy, primarily in the form of ATP, is central to life. Two major processes convert the energy in carbon substrates to ATP: substrate-level phosphorylation (for example ATP produced in glycolysis) and oxidative phosphorylation (respiration) (Figure 3E-G). In the latter process, electrons obtained from carbon redox reactions are transferred to membrane transport chains that eventually reduce oxygen or, in
more specialized cases, other oxidized compounds such as nitrate, nitrite, or sulfate. This process of respiration yields more ATP molecules per unit carbon source than substrate-level phosphorylation, but also requires more proteins. As a result, in response to energy starvation, cells can respond in two ways: increase the total amount of carbon catabolism, or direct more flux to oxidative phosphorylation.

An increase in carbon catabolism is mostly mediated by the aforementioned Crp-cAMP circuit, which senses, among several signals, drops in α-ketoacid concentration, which indicate a lack of carbon relative to nitrogen or other nutrients (a situation that typically leads to energy limitation). However, glycolysis is also sensitive to ATP levels, which are a more direct sensor of energy limitation. Introducing an artificial “ATP sink” by overexpressing an ATP-dissipating ATPase lowers ATP concentrations and increases glycolytic flux, and ATP and its congeners ADP and AMP can modulate the activity of a large number of enzymes, but regulation of the reaction between fructose-6-phosphate and FBP is of particular importance. ATP inhibits phosphorylation and AMP dephosphorylation, leading to increased glycolysis under conditions of low energy charge (Figure 3E), a mechanism conserved also in humans. However, across a large number of conditions, nucleotide phosphate concentrations and energy charge remain rather constant in E. coli despite wide variation in glycolytic flux, suggesting that other mechanisms also have key roles in the regulation of glycolytic flux.

Compared to glycolysis, respiration requires a large number of different proteins, from tricarboxylic acid (TCA) cycle enzymes to electron transfer chain components. As such it is unsurprising that transcriptional regulation has a key role in the control of respiration. Two key transcription factors in E. coli, FNR and ArcA, coordinate transcription of TCA cycle enzymes with the availability of electron acceptors like oxygen (Figure 3F-G). FNR directly senses intracellular oxygen, repressing genes involved in aerobic respiration and inducing those for anaerobic metabolism in absence of oxygen. ArcA, as part of the ArcAB two component system, responds to the redox state of membrane-associated redox carriers in the respiratory chain, namely quinones, which accumulate when respiration becomes limited, for example by low oxygen availability. However, oxygen is not the exclusive regulatory input: transcriptional regulation of TCA enzymes by Crp strongly suggests that carbon source availability plays a crucial role. Allosteric regulation, such as α-ketoglutarate inhibition of citrate synthase, the first step in the TCA cycle, may also transfer carbon availability information to TCA cycle activity. Thus, E. coli appears to use information on carbon, oxygen, and energy availability to regulate the TCA cycle and respiration. However, quantitative studies are lacking, and much remains to be learned about how these signals are integrated.

The emerging picture is that E. coli’s decision to invest in the energy-efficient pathway of oxidative phosphorylation and TCA cycle is largely transcriptionally regulated. Recent evidence from B. subtilis...
and *S. cerevisiae*\(^{83,84}\) highlighting the strong relationship between TCA cycle gene expression and corresponding metabolic flux reveals a similar picture. A likely explanation for this transcriptional control is the large cost of expressing the many proteins required for respiration. With increasing growth rates cells must devote a larger fraction of their proteome to the production of ribosomes\(^{85}\), hence optimized resource allocation would favor stining expression of respiratory enzymes in rapidly growing cells as long as sufficient energy can be provided\(^{86,87}\).

**Regulation of amino acid uptake and metabolism**

Above we discussed how cells regulate the conversion of nutrients into metabolic intermediates, which next have to be converted into monomers such as amino acids, nucleotides, and lipids for macromolecule synthesis. Exemplarily we discuss here the coordination of individual amino acid biosynthesis and degradation pathways (Figure 4A). This coordination is mostly achieved by end-product inhibition, a ubiquitous regulatory mechanism that balances production of a specific amino acid with its demand while minimally affecting the rest of metabolism\(^{88,89}\). All 20 amino acids in *E. coli* have either been shown to inhibit the first committed step in their synthesis through allosteric regulation\(^{60}\) or, in the case of single-reaction pathways, can be assumed to do this through product inhibition. This mechanism ensures a rapid increase of synthesis in response to a rise in demand, or repression of synthesis in response to an excess supply. Since the inhibition typically affects only the branch of the pathway specific to the particular amino acid, parallel branches can mostly be tuned independently. To a large extent, similar principles also govern transcriptional regulation of amino acid biosynthesis, with at least 10 amino acids negatively regulating transcription of their own biosynthesis pathways, either through transcription factors\(^{90}\) or transcriptional attenuators\(^{91,92}\). However, in several cases, transcriptional regulation is mediated by a factor that binds a pathway intermediate, which builds up when allosteric regulation of the first step is relieved\(^{93}\). Nevertheless in each case the regulation follows the straightforward logic of increasing flux when amino acid usage exceeds supply, and hence the transferred information is the balance between supply and demand.

Such pathway-specific transcriptional regulation can be coopted to regulate degradation as well as synthesis. For instance, the transcription factor ArgR, which binds arginine to represses arginine biosynthesis enzymes when arginine is abundant\(^{84}\), can also activate arginine degradation enzymes\(^{95}\). Such regulation is also subject to NtrC regulation, thus integrating a global sensor of nitrogen demand with a local sensor of arginine availability.

Nevertheless, not all parts of the amino acid metabolism regulatory network are cleanly separated into specific branches. Crosstalk exists at both the allosteric\(^{96}\) and transcriptional\(^{90}\) level, whereby amino acids affect not only their own synthesis but also that of other amino acids. Moreover, some amino acids act as global regulators, affecting the activity of transcription factors that target hundreds of
genes. One such transcription factor in *E. coli* is Lrp, which binds leucine but regulates hundreds of genes involved not only in amino acid biosynthesis, but also in preparation for stationary phase\(^{97,98}\). Curiously, in *B. subtilis*, the global regulator CodY regulates many genes involved in the transition between growth and starvation, and is also sensitive to levels of branched chain amino acids like leucine\(^{99}\). The yeast TOR complex, which similarly regulates a large number of genes related to growth and starvation, also seems to be particularly sensitive to leucine levels\(^{100}\). No concrete theory has been proposed for why leucine should be such a commonly used signal of general starvation.

**Regulation of protein biosynthesis**

One of the endpoints of metabolism is the assembly of amino acids into proteins (Figure 4B). In a fast growing cell, protein synthesis and ribosome production account for the majority of nutrient and energy consumption. However, while fast growth can require up to 75% of cellular transcription devoted to production of ribosomes\(^{101}\), such a program would be highly deleterious when nutrients are limited\(^{85,102}\) and thus *E. coli* devotes resources to ribosome biogenesis only when resources for protein synthesis are abundant. Making this decision requires integration of several metabolic signals, and *E. coli* uses the availability of both energy sources (ATP and GTP) and amino acids, the major substrates of protein synthesis, to determine the rate of ribosome biogenesis. The concentrations of ATP and GTP directly affect transcription at ribosomal RNA (rRNA) promoters through availability of initiating nucleotides\(^{103}\). Moreover, in starvation conditions a drop in ATP also affects ribosome biogenesis indirectly by inhibiting the degradation of the stress sigma factor RpoS, thereby redirecting the transcriptional machinery from ribosomal genes to stress response genes\(^{104}\). The signal of amino acid availability is channeled through the small molecule (p)ppGpp, the synthesis of which is activated by the presence of uncharged (lacking amino acid) transfer RNA (tRNA) molecules through allosteric regulation of the enzyme RelA\(^{105}\). (p)ppGpp can then repress rRNA transcription both directly\(^{106}\) and through binding of the transcription factor DksA\(^{107}\). Although few quantitative studies of the precise input function exist, it is likely that the combination of these two inputs allows rRNA production only in conditions where both ATP and amino acids are available. Similar control is exerted over ribosomal protein expression\(^{108}\). (p)ppGpp also induces synthesis of a number of amino acid biosynthesis enzymes, serving as a global regulator in addition to pathway-specific signals.

The reason that superfluous ribosome biosynthesis would be deleterious is that cells can only modulate their total protein allocation, taking resources from one protein’s expression and devoting them to another\(^{11,85}\). Given that ribosomal proteins and metabolic enzymes are the major protein fractions at high growth rates\(^{109}\), decreased ribosome synthesis would allow increased enzyme synthesis. Conversely, this constraint of total protein allocation is a plausible explanation for why rapidly growing cells rely primarily on glycolytic energy generation and not on the more protein-
intensive respiration\textsuperscript{31,86}. Complementary to the common perception of protein synthesis regulating metabolism, \textit{E. coli} thus coordinates proteome partitioning between ribosome and metabolic enzyme synthesis via metabolic signal-dependent transcription factors, for example through the global carbon to nitrogen availability reporter \(\alpha\)-ketoglutarate and CRP\textsuperscript{31}.

\textbf{Approaches to study metabolic regulatory circuits}

Key to identifying the regulatory circuits operating across metabolism and its regulatory networks are methods for quantification of the different cellular components, such as metabolites and proteins. Clearly, knowing how much there is of every component in a cell is not enough to identify and understand regulatory circuits. A second key requirement is to know the cell’s regulatory ‘wiring diagram’, that is the regulatory interactions connecting these cellular components. Knowing a regulatory circuit’s wiring diagram can already provide some understanding of its potential \textit{in vivo} functioning, for example by identifying the prototypical network motifs – e.g. negative feedback or feedforward loops - which are part of the respective circuit\textsuperscript{110}. An example of such comparatively simple regulatory circuits is the prevalent product inhibition of amino acid biosynthesis pathways\textsuperscript{88,89}. However, as the examples in the previous sections illustrate, more complex regulatory circuits typically cannot be identified and characterized just by studying the cell’s wiring diagram. In these cases, a third key requirement is to quantify the relationship between a cellular phenotype, such as metabolic flux or gene expression, and the signals which serve as regulatory inputs, such as regulatory metabolites (figure 2). Knowing the quantitative relationship between observed phenotype and underlying regulatory signals allows to identify the key components and interactions of a regulatory circuit, and to generate testable hypotheses on its functioning. In this section, we review the most common approaches for the quantification of metabolic fluxes, metabolite concentrations, protein expression and activity, as well as approaches to identify regulatory interactions between proteins, DNA and metabolites.

\textit{Estimating metabolic fluxes}

There are two common approaches to estimate metabolic fluxes. The first approach utilizes computational models to predict metabolic fluxes based on the stoichiometry of the metabolic network and optimality criteria\textsuperscript{111}. Here, the chosen optimality criterion, such as maximized growth or ATP production, has a strong impact on the predicted flux distribution\textsuperscript{63}. Flux predictions can be further refined with physiological measurements on uptake and secretion rates\textsuperscript{112}, additional global constraints such as available cytosolic or membrane space\textsuperscript{113,114}, and information on enzyme expression\textsuperscript{115} or kinetics\textsuperscript{116}. The second approach does not invoke any particular optimality criterion, but rather relies on monitoring the propagation of isotopic label (such as \(^{13}\text{C}\)) from fed labeled substrates\textsuperscript{117–121}. The most common such approach, termed \(^{13}\text{C}\)-metabolic flux analysis (\(^{13}\text{C}\)-MFA), is
Chapter 1 - General Introduction

based on the fact that the molecular conversions within different metabolic pathways often yield a distinct labeling pattern of a particular metabolite intermediate depending on the used pathway, which in turn can be detected by mass spectrometry or NMR\textsuperscript{122,123}. For example, \textsuperscript{13}C-MFA can be used to determine the flux ratios at key branch points in metabolism using empirically derived equations relating \textsuperscript{13}C-labeling patterns to local flux ratios\textsuperscript{124}. Moreover, when supplemented with data on physiological measurements, measured \textsuperscript{13}C-labeling patterns can be used to not only infer flux ratios, but also absolute metabolic fluxes\textsuperscript{125–127}. However, the underlying assumptions and constraints of the approaches described above make the inference of metabolic fluxes by \textsuperscript{13}C-MFA often challenging for more complex systems. For example, these approaches typically require both metabolic (that is, all metabolic fluxes are balanced) and isotopic (that is, stable \textsuperscript{13}C enrichment over time) steady state and can therefore not be used in cases where cells don’t reach metabolic steady state (i.e. dynamic experiments), or where labeling propagation is very slow (i.e. slow growing organisms). Here, alternative approaches, such as non-stationary \textsuperscript{13}C flux ratio analysis\textsuperscript{120} or kinetic flux profiling\textsuperscript{121}, can be used to at least infer local flux information at branch points or pathways of interest.

\textit{Quantifying metabolite concentrations}

The rapid turn-over rates of many metabolic intermediates, together with their vast chemical diversity and concentration range across metabolites, makes the quantification of intracellular metabolite concentrations extraordinarily challenging. To tackle these challenges, two principal schools of thought have emerged. The first school of thought focuses on quantifying the absolute concentration (in moles per cell volume) of metabolites as closely as possible to their ‘true’ \textit{in vivo} concentration\textsuperscript{128}. Despite advances in our ability to measure metabolites by mass spectrometry\textsuperscript{129–131}, numerous hurdles have to be overcome to accurately quantify absolute metabolite concentrations. These hurdles include fast and reproducible quenching of intracellular metabolism while avoiding leakage of intracellular metabolites into the surrounding, quantitative extraction of chemically diverse metabolites, processing of metabolite extracts to be amendable to quantification e.g. in mass spectrometers while avoiding the degradation of metabolites, accurate quantification of cell volume for normalization\textsuperscript{132}, to name but a few\textsuperscript{133}. Overcoming these hurdles requires extensive experimental protocols, which dramatically reduce the throughput and scope of metabolomics experiments, and it has been argued that systematic errors are likely to haunt absolute quantitative metabolomics measurements in the foreseeable future\textsuperscript{128}. Nevertheless, there is evidence suggesting that in reality the picture may not be quite that dim after all: for example, thermodynamic analysis typically shows very good agreement of flux directions as inferred from absolute metabolite concentrations with known flux directions (as determined e.g. by flux analysis)\textsuperscript{134,135}, even in short term dynamic experiments\textsuperscript{36}. 

19
Contrary to this ‘moles-matter’ school of thought, the second one uses comparative approaches to identify relative differences in metabolite concentrations between strains/cell types\textsuperscript{136}, conditions\textsuperscript{137,138}, or time points\textsuperscript{57}. In analogy to transcriptomics analyses, which typically report relative changes in mRNA abundance, comparative metabolomics aims at obtaining a comprehensive picture of the metabolic changes induced by a particular internal or external perturbation at the expense of decreased measurement accuracy for each individual metabolite\textsuperscript{139}. A common application of comparative metabolomics is the identification of metabolites which are associated to a phenotype of interest, such as a disease\textsuperscript{140,141}. Beyond the identification of such metabolite biomarkers, comparative metabolomics can also be used to ask fundamental questions about metabolic regulation. For example, relative quantification of central metabolites revealed a highly similar dynamic response of two divergent microbes, namely \textit{E. coli} and \textit{S. cerevisiae}, to carbon and nitrogen starvation, suggesting a similar underlying regulatory scheme\textsuperscript{57}.

A key requirement of metabolomics is the ability to cover a large fraction of a cell’s metabolome. Recent technological advances towards this end, in particular in the field of mass spectrometry\textsuperscript{142}, have made the rapid measurement of metabolite concentrations at full metabolome scale possible, which allows to unearth novel links between metabolism and the phenotype of interest. A seminal example is the discovery of the ‘onco-metabolite’ 2-hydroxyglutarate - which is produced by isocitrate dehydrogenase mutants often found in human brain cancers - by untargeted mass spectrometry coupled with liquid chromatography\textsuperscript{143}. Further technological advances increased the sample throughput dramatically by omitting the chromatographic separation of metabolites altogether. One such example is flow-injection time-of-flight mass spectrometry\textsuperscript{136}, which was used to unravel the unexpected role of ubiquinones as osmoprotectants in \textit{E. coli}\textsuperscript{137}. Thus, the rapid untargeted measurement of relative metabolite concentrations in hundreds of samples at full metabolome scale is now becoming feasible. Consequently, focus has now shifted towards the development of approaches to systematically extract biological knowledge from such large-scale data sets\textsuperscript{144}.

Despite recent advances\textsuperscript{145}, virtually all of the available methods for large-scale quantification of metabolite concentrations are restricted to bulk population measurements. An alternative to bulk measurement methods is provided by Förster resonance energy transfer (FRET) sensors. These genetically encoded fusion proteins translate metabolite binding into a fluorescent readout of their conformational change in single cells\textsuperscript{146}, which can be used to quantify the concentration of individual metabolites at single-cell, or even sub-cellular compartment\textsuperscript{147}, resolution.

\textbf{Quantifying protein expression and activity}

A plethora of methods enable direct quantification of protein expression both at mRNA- and protein-level, including microarrays\textsuperscript{148,149}, RNA-seq\textsuperscript{150}, ribosome footprinting\textsuperscript{151}, and proteomics\textsuperscript{152–154}. Protein
expression can also be quantified indirectly using (typically fluorescent) expression reporters\textsuperscript{155–157}. However, although these advances have made it possible to routinely quantify protein expression at genome scale, measuring the expression of proteins is often not sufficient to infer their \textit{in vivo} activity: posttranslational modifications such as phosphorylation\textsuperscript{24,158} and acetylation\textsuperscript{35} are prevalent in microbes and can drastically alter protein activity. Proteins are also subject to nonenzymatic covalent modification through intrinsically reactive metabolites\textsuperscript{159,160}, as well as to non-covalent allosteric regulation. For example, about a third of the 200 transcription factors are known to bind to metabolites\textsuperscript{161}, and many more are expected to bind metabolites\textsuperscript{162}, raising questions about methods which solely rely of co-expression of transcription factors and target genes to reconstruct transcriptional regulatory networks\textsuperscript{163}. Consequently, efforts to infer \textit{in vivo} protein activity rely on the integration of several types of information, often aided by computational approaches. One such method is termed network component analysis and combines gene expression data and the topology of the transcriptional regulatory network to infer the activity of transcription factors\textsuperscript{164}. Other methods use kinetic modeling together with metabolomics\textsuperscript{36,165}, or the correlation between fluxes and protein abundance\textsuperscript{34}, to deduce enzyme activity. However, critical for the success of these methods is accurate information on the regulatory interactions between proteins, DNA and metabolites.

\textit{Identifying regulatory interactions between proteins, DNA and metabolites}

So far, much of the effort in identifying regulatory interactions has been focused on mapping protein-protein and protein-DNA interactions. Protein-protein interactions can be identified using a variety of biochemical and genetic methods, and examples include tandem affinity purification, FRET, various yeast-two hybrid systems, and phage display\textsuperscript{166}. Likewise, methods such as Chip-seq\textsuperscript{167} and SELEX\textsuperscript{168} have been used for the genome-wide identification of target genes of various transcription factors, cumulating in highly overlapping and dense transcriptional regulatory networks\textsuperscript{90,169–174}. In contrast, the development of methods to identify protein-metabolite interactions has so far lagged behind\textsuperscript{10,175}. Nevertheless, recent years have seen a surge of experimental methods to systematically identify such interactions\textsuperscript{175,176} \textit{in vivo} as well as \textit{in vitro}, which can be broadly categorized into three categories.

The first category comprises of methods which detect changes in protein activity in presence of potential allosteric metabolite effectors. Classical examples for such methods are biochemical \textit{in vitro} enzyme assays and related assays for non-catalytical proteins such as transcription factors. Even though such assays have been instrumental in the identification of many allosteric interactions in metabolism and are still being widely employed (e.g. \textsuperscript{177–180}), they typically need to be performed separately for each protein and each potential allosteric effector, making the systematic identification of allosteric interactions highly laborious. There are efforts to overcome this limitation using competitive activity-based protein profiling (ABPP), which allows to quantify the activity of whole
Chapter 1 - General Introduction

enzyme super-families with the same molecular probe, but so far these efforts have been limited to the screening of drug libraries. Thus, it is currently not clear how well ABPP translates to investigations of endogenous protein-metabolite interactions or non-catalytical proteins.

The second category of methods to identify protein-metabolite interactions relies on the fact that the binding of an allosteric effector often induces structural alterations of the protein, such as changes in protein conformation. Another type of structural alteration is exploited by a method termed drug affinity responsive target stability (DARTS). This method is based on the premise that binding of an allosteric effector can enhance a proteins’ resistance to proteolytic degradation, and was recently used to identify the inhibition of ATP synthase by alpha-ketoglutarate in C. elegans. These and related methods provide an exciting new avenue for the systematic identification of protein-metabolite interactions, but they are inherently limited to those interactions which indeed cause major structural alterations of the protein, which is not always the case.

Probably the biggest progress in the systematic identification of protein-metabolite interactions has been achieved with methods which directly detect binding between metabolites and proteins. These methods, which are largely restricted to in vitro conditions, come in various flavors. One subset of methods relies on the immobilization of one of the binding partners. Examples include metabolite or protein microarrays which are probed with fluorescently labeled proteins or metabolites, respectively. In another variant proteins are immobilized on affinity columns and used to retain binding metabolites from a metabolite extract, which are subsequently identified by mass spectrometry. An obvious drawback of these methods is that the chemical modification necessary for immobilization can impair the binding between protein and metabolite, in particular in the case of immobilized small metabolites. Consequently, assays using immobilized metabolites have been largely restricted to lipids and libraries of large synthetic compounds. One alternative method, termed MIDAS, does not require any chemical protein or metabolite modifications, but rather relies on detecting the depletion of metabolites through stoichiometric binding to a protein. Since the protein needs to be present in excess, this method is restricted to proteins that can easily be purified and remain stable at high concentrations. A common limitation of all these in vitro assays is that they do not allow to assess which of the identified allosteric interactions that are actually relevant in vivo. Efforts by Snyder et al. to circumvent this limitation have led to the development of a method which utilizes co-purification of in vivo protein metabolite complexes and subsequent identification of co-eluted metabolites by mass spectrometry. However, due to of protein purification conditions that typically involve washing steps in polar environments this method is much less suited for polar metabolites, although it was recently used to identify at least one such allosteric interaction, namely between human PKM2 and the polar nucleotide biosynthesis intermediate SAICAR.
Complementary to the aforementioned approaches, which probe protein-metabolite interactions purely experimentally, such interactions can also be identified computationally. For example, if a protein’s structure is known, molecular docking can be used to predict binding metabolites\textsuperscript{202}. Moreover, computational inference methods can be used to infer protein-metabolite interactions directly from experimental data\textsuperscript{36,203}. For example, we developed an approach which combines quantitative metabolomics with ensemble kinetic modeling to identify allosteric interactions that control glycolysis during rapid nutrient shifts\textsuperscript{36}. The key advantage of such approaches over the aforementioned purely experimental methods is that they allow to identify those interactions that are actually relevant \textit{in vivo} without prior knowledge on regulatory interactions, but they are currently restricted to rather small metabolic networks. Alternatively, if an organism’s regulatory interactions have already been reasonably well characterized, computational inference can help to identify those interactions that were relevant in a given condition. For example, Yugi et al. reconstructed the signal flow during from insulin signaling to the regulation of glycolytic enzymes in mammalian cells by integrating time-course phospho-proteomics and metabolomics data with the reported phospho-signaling network\textsuperscript{204}.

In summary, a flurry of activity in recent years has led to the development of many experimental methods for the systematic identification protein-protein as well as protein-DNA interactions, giving us a fairly comprehensive picture of the regulatory interactions between proteins and DNA. In contrast, the development of methods to identify protein-metabolite interactions is only slowly catching up, and many of these methods are restricted to specific types of interactions, such as protein-lipid interactions, or to interactions with synthetic drug libraries. Nevertheless, complementary computational approaches to infer protein-metabolite interactions offer some hope, and may moreover be used to identify those interactions that are actually relevant \textit{in vivo}.

**Focus of this thesis**

As the aforementioned examples of regulatory circuits illustrate, we now have a quite good understanding of how cells make some metabolic decisions, such as the prioritization between available carbon sources. However, for many other metabolic decisions, such as the decision how to distribute metabolic fluxes at the various branch points in central carbon metabolism\textsuperscript{13}, we currently only have a poor understanding of the underlying regulatory circuits, and the signals these circuits receive. Elucidation of these regulatory circuits has so far been hampered by three major challenges. Firstly, these circuits are highly intertwined, with often multiple regulatory inputs from specific as well as global signals. For example, recent works have demonstrated that bacterial gene expression is not only subject to regulation by a dense network of specific transcription factors\textsuperscript{90,169–174}, but also by the global physiology of the cell\textsuperscript{85,205–209}, for example through growth-rate dependent changes in free RNA
polymerase availability or ribosome abundance\textsuperscript{85,101,210}. However, we currently lack approaches to dissect the contribution of such specific and global regulatory inputs. Secondly, the available comprehensive, but static, regulatory interaction networks do not allow to identify which of the many reported interactions are relevant in a given environment. Thirdly, despite the fact that interactions between proteins and regulatory metabolites are critical for the functioning of many regulatory circuits, we lack approaches to systematically detect such interactions. In this thesis, we attempt to tackle these challenges to elucidate the role of protein-metabolite interactions in making metabolic decisions in the model bacterium \emph{E. coli}. Specifically, we focus on protein-metabolite interactions in two regulatory layers, namely transcriptional and posttranslational enzyme regulation.

In the first part of this thesis, comprising chapters 2 to 5, we focus on the transcriptional regulatory program governing \emph{E. coli}'s metabolism. We start off by reviewing our current understanding of the role of transcriptional regulation in regulating microbial metabolic fluxes (\textit{chapter 2}). Next, we develop a combined computational-experimental approach to dissect global and specific transcriptional regulation, using the arginine biosynthesis pathway as a test case (\textit{chapter 3}). In \textit{chapter 4}, we expand our analysis to unravel the transcriptional regulatory program of \emph{E. coli}'s central carbon metabolism in response to various nutrient conditions. In particular, we systematically identify metabolites that serve as regulatory signals in \emph{E. coli}'s central carbon metabolism, and relate these to the transcription factors ultimately exerting the regulatory function. To respond to environmental changes, cells need not only to regulate their central carbon metabolism, but also coordinate the regulation of all other metabolic sectors. In \textit{chapter 5}, we elucidate the transcriptional regulatory program achieving this coordination in \emph{E. coli} in response to two orthogonal nutrient limitations. Based on the genome-scale quantification of proteins, metabolites, and metabolic fluxes, we further explore the role of passive compensatory mechanisms, such as changes in substrate saturation, in establishing \emph{E. coli}'s metabolic response to nutrient limitation.

As the examples in this introduction illustrate, transcriptional enzyme regulation alone is often not sufficient to make metabolic decisions. In the second part of this thesis, comprising chapters 6 and 7, we therefore focus on a second regulatory layer, namely posttranslational enzyme regulation. In \textit{chapter 6}, we outline our current understanding of the role of posttranslational regulation, i.e. posttranslational protein modifications (PTM's) and allosteric protein-metabolite interactions, in regulating microbial metabolism. A recent surge in experimental efforts has now made the genome-wide mapping of PTM's possible, and consequently focus has shifted to identifying those PTM's that are relevant in a given condition. In contrast, the lack of similar methods to systematically identify allosteric protein-metabolite interactions remains a major obstacle in understanding the role of such interactions in regulating metabolism. To address this challenge, we explore a NMR-based
experimental approach to systematically identify protein-metabolite interactions in vitro (chapter 7). While NMR-based approaches have long been successfully used in the identification of drug interactions\textsuperscript{211}, they have rarely been used to identify interactions between endogenous metabolites and proteins\textsuperscript{212}. In a proof-of-concept study, we use ligand-detected-NMR to detect which metabolites out of a diverse metabolite mix interact with a selection of purified \textit{Escherichia coli} enzymes.
## Chapter 1 - General Introduction

### Tables

**Table 1.** Global regulatory metabolites in *E. coli* and the information they transfer.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Information transfer</th>
<th>Key regulatory interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>fructose-1,6-bisphosphate (FBP)</td>
<td>Glycolytic flux(^{40,42,213})</td>
<td><strong>Enzyme:</strong> Pyruvate kinase</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Enzyme:</strong> PEP carboxylase</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Transcription factor:</strong> Cra</td>
</tr>
</tbody>
</table>
| cAMP | \(\alpha\)-ketoacid concentration\(^{31}\)  
Phosphorylation state of PTS system\(^{29,30}\) | **Transcription factor:** Crp  |
| L-glutamine | Nitrogen availability\(^{165,214}\) | **Signalling:** GlnBK  
**Transcription factor:** NtrC (via GlnBK)  |
| \(\alpha\)-ketoglutarate | Ratio of carbon to nitrogen availability for amino acid biosynthesis\(^{31}\) | **Enzyme:** EI (PTS component)  
**Enzyme:** Adenylate cyclase (cAMP forming)  
**Signalling:** GlnBK  |
| other \(\alpha\)-ketoacids (e.g. pyruvate, oxaloacetate) | Ratio of carbon to nitrogen availability for amino acid biosynthesis\(^{31}\) | **Enzyme:** several enzymes in TCA cycle (oxaloacetate)  
**Enzyme:** Adenylate cyclase (cAMP forming)  
**Transcription factor:** PdhR (pyruvate)  |
| L-leucine | Balance of L-leucine production, uptake and protein biosynthesis\(^{97}\)  
General nutrient abundance\(^{97,98}\) | **Enzyme:** several steps in branched-chain amino acid biosynthesis  
**Transcription factor:** Lrp  |
| ppGpp | Amino acid starvation\(^{105}\) | **Transcription:** RNA polymerase\(^{105}\)  
**Transcription factor:** DksA  |
| ATP | Energy starvation\(^{104}\) | **Enzyme:** numerous enzymes in metabolism  
**Transcription factor:** RpoS via ClpXP  
**Transcription:** RNA polymerase (transcription of ribosomal promoters)  |
| Quinones | Balance of respiratory capacity and oxygen supply\(^{80}\) | **Signalling:** ArcAB  |
Figures

Figure 1. A) Coarse-grained view of different sectors that compose large parts of metabolism in many bacteria. Microorganisms need to carry out a range of metabolic tasks to ensure a supply of metabolic fluxes through the sectors and thus sustain cell maintenance and growth. All organisms must regulate the uptake of nutrients and coordinate carbon, energy and nitrogen metabolism to balance monomer synthesis and macromolecule polymerization. B) Each metabolic task can ultimately be broken down into decisions of establishing the flux magnitude through a linear pathway and the partitioning of incoming or outgoing fluxes at a branch point. These metabolic decisions are made by regulatory circuits. Figure adapted from [215].
**Figure 2.** The regulatory network that controls cellular metabolism consists of interactions between metabolites, enzymes and regulators. From this highly complex network, more tractable individual circuits can be delineated as modules that carry out concise functions of information input and regulatory effects. A critical step in understanding the function of a given circuit within the larger network is the analysis of its regulatory logic, that is, the relationship between the input and output of the circuit. This process typically involves iterations of hypothesis formulation and the generation of experimental evidence to identify the relevant active interactions that determine the wiring of the circuit. For the example shown in the figure, the regulatory logic that is implemented is the balancing of supply (given by flux $F_2$) and demand (given by flux $F_3$) by an integral feedback circuit. An imbalance between supply and demand perturbs the level of the metabolite $M_2$, which in turn begins to mitigate the imbalance via several regulatory links that affect the enzymes $E_2$ and $E_3$. Figure adapted from \textsuperscript{215}.
Figure 3. A) Schema of carbon and energy metabolism. Inset panels display regulatory modules controlling parts of the network. B) Preferential use of glucose as a carbon source. C) Coordination of local (lactose) and global (carbon supply) signals for carbon uptake by LacI and Crp, respectively. D) Fructose-1,6-bisphosphate (FBP)–Cra circuit which regulates the switch between glycolysis and gluconeogenesis. E) Control of carbon catabolism via ATP demand. F) Regulation of respiration by the availability of electron acceptors. G) Oxygen-sensing switch between aerobic and anaerobic respiration.

Abbreviations: αKG, α-ketoglutarate; acetyl-P, acetyl-phosphate; cAMP, cyclic AMP; EMP, Embden–Meyerhof pathway; G6P, glucose-6-phosphate; GLX, glyoxylate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PTM, post-translational modification. Figure adapted from 215.
Figure 4. Regulatory circuits that control amino acid metabolism and protein biosynthesis in *Escherichia coli*. Insets show regulatory modules controlling parts of the network. A) Amino acids typically control their production by both allosteric regulation of the first committed step and by transcriptional regulation of biosynthetic enzymes — arginine and cysteine are shown here as two examples. Although the regulatory topology differs between the two cases, both networks regulate the transcription of pathway enzymes according to end-product demand. B) Uncharged transfer RNA (tRNA) builds up when either amino acids or energy are unavailable and coordinates many processes, in particular ribosome biogenesis, through the small molecule regulator guanosine tetraphosphate (ppGpp). Abbreviations: PTM, post-translational modification; rP, ribosomal proteins. Figure adapted from [215].
References


Chapter 1 - General Introduction


Chapter 2 – the role of transcription in regulating microbial metabolic fluxes

This chapter has been published in:

Abstract

The most common way for microbes to control their metabolism is by controlling enzyme level through transcriptional regulation. Yet recent studies have shown that in many cases, perturbations to the transcriptional regulatory network do not result in altered metabolic phenotypes on the level of the flux distribution. We suggest that this may be a consequence of cells protecting their metabolism against stochastic fluctuations in expression as well as enabling a fast response for those fluxes that may need to be changed quickly. Furthermore, it is impossible for a regulatory program to guarantee optimal expression levels in all conditions. Several studies have found examples of demonstrably suboptimal regulation of gene expression, and improvements to the regulatory network have been investigated in laboratory evolution experiments.
Chapter 2 – the role of transcription in regulating microbial metabolic fluxes

Introduction

Transcriptional regulation is one of the most prevalent regulatory mechanisms across all kingdoms of life and affects virtually all cellular processes. Perhaps the most classic paradigm of transcriptional regulation is the induction of metabolic enzymes in response to changing environments\(^1\). Since the seminal work of Jacob and Monod on the lac operon, countless more examples have been observed, and the unraveling of the entire regulatory network has been facilitated by new experimental methods. These include direct identification of physical interactions among proteins\(^2,3\) or between proteins and DNA\(^4,5\), as well as a variety of methods to infer regulatory links from high-throughput transcriptomic or proteomic data\(^6,7\).

Such work has revealed thousands of regulatory interactions affecting metabolic enzyme expression in even the simplest model microbes. However, knowing that transcription of a particular enzyme is affected under a particular condition is only suggestive of changes in the metabolic state, since many other layers, such as posttranscriptional modifications, allosteric regulation, or thermodynamics play a regulatory role\(^8\). This metabolic state is defined by fluxes, the *in vivo* rates of conversion of substrate to product. For many branches of metabolism, these fluxes can be inferred, either by feeding cells with isotopically labeled substrates and measuring the propagation of the label\(^9\), or by presuming a metabolic objective function and finding the fluxes that maximize it purely *in silico*\(^10\). Numerous studies have revealed the remarkable plasticity of the cellular metabolic state\(^11,12\) and in this review we ask how this state is affected by the aforementioned plethora of transcriptional regulatory interactions.

Effect of genetic perturbations on metabolic fluxes

The simplest question to ask is what happens when the transcriptional regulation of metabolism is perturbed. A common result of such perturbations is that cells cannot upregulate enzymes necessary under a particular condition, and the low basal expression is unable to sustain sufficient flux for optimal growth. For instance, a recent example showed that deletion of the transcription factor hexR in *Shewanella oneidensis* leaves the bacterium unable to induce the expression of the key gluconeogenic enzyme ppsA and unable to grow on pyruvate or lactate\(^13\). Similar examples exist for utilization and biosynthesis of various other nutrients\(^14,15\). An example of the reverse situation, where cells are unable to repress key enzymes, was given by deletion of the repressor ccpN in *Bacillus subtilis*\(^16\). The expression of its gluconeogenic targets gapB and pckA even under glycolytic conditions caused a large futile cycle that impaired growth. Such results indicate that transcriptional regulation is often used by cells to enable large changes in fluxes in response to drastic changes in the environment.

It may therefore seem logical that most changes in enzyme abundance from the wild-type levels should affect the flux distribution. However, several studies give strong evidence against this. Perhaps the
most canonical one dates back over a decade and attempted to increase the flux through glycolysis in *S.cerevisiae* by overexpressing the corresponding enzymes\(^\text{17}\). Surprisingly, despite concurrent overexpression of as many as seven enzymes, no increase in flux was observed. Further recent studies showed that yeast can compensate for such changes in enzyme abundance with changes in metabolite levels to keep flux constant\(^\text{18}\), which also agrees with observations in *E.coli* that almost no enzymes in central carbon metabolism are fully saturated, based on metabolite levels and substrate affinity (K\(_m\)) measurements\(^\text{19}\).

Further perturbations to the transcriptional regulatory network have been explored on a more global scale by two works – one in *Saccharomyces cerevisiae*\(^\text{20}\), and one in *Escherichia coli*\(^\text{21}\), that directly measured fluxes through the branch points of central carbon metabolism in hundreds of strains with deletions of individual transcription factors. Since the transcription factors under study had dozens of targets among metabolic enzymes, significant changes in fluxes were expected. The striking results showed that while many TF deletions indeed affected fluxes, the vast majority of these effects were strongly correlated with substrate uptake rates (and consequently also with growth rates), meaning that most deletions likely affected cell growth as opposed to specific metabolic fluxes. Few TF deletions, including those of factors with known targets in the corresponding metabolic pathways, actually changed the flux distribution, that is, the split ratios at key branch points of central carbon metabolism. Put another way, TF deletions that induce moderate changes in enzyme levels do not seem to affect the distribution of fluxes, making metabolism resilient against perturbations in the transcriptional network.

While it’s possible that the lack of flux distribution changes upon transcription factor deletion was due partially to the redundancy of the transcriptional regulation of metabolic enzymes, the relationship between fluxes and transcriptional regulation was also tested without making any genetic perturbations. This was done in a number of studies involving “regulatory analysis” from the Westerhoff group which analyzed the relative change in pathway flux as compared to the relative change in either individual enzyme activity or abundance across different environments\(^\text{22,23}\). The basic rationale is that if fluxes are indeed controlled transcriptionally, or more precisely, by regulation of enzyme abundance, then the relative change in abundance should be exactly equal to the relative change in flux. However, these studies consistently found a mismatch between the two ratios, implying a significant role for both posttranscriptional modifications\(^\text{22}\) and metabolic control\(^\text{24}\), that is, control of flux either through substrate availability or allosteric interactions.

**Enzyme overabundance as buffer**

Taken together, the evidence in the previous section strongly suggests that metabolic fluxes are not very sensitive to moderate perturbations in enzyme abundance, implying that the enzymes are
expressed at higher levels than necessary. Why would such a program exist when it is well known that spurious expression of unnecessary proteins has a measurable negative effect on fitness?\textsuperscript{25,26} (though see Eames and Kortemme\textsuperscript{27} for a new perspective on the \textit{lac} operon commonly studied in this context) One clear hypothesis is that cells are protecting themselves against inevitable variation in protein levels stemming from the noisiness of gene expression\textsuperscript{28}. While unnecessary protein production may have some cost, the cost from insufficient metabolic flux, especially through central pathways that affect a large number of metabolic processes, is likely to be much higher, making a certain buffer of protein expression desirable. This is in good agreement with the observation that while single cells in a population exhibit some variation in their individual growth rates, this variation has not been found to correlate with the expression of key metabolic genes\textsuperscript{29,30}. A second hypothesis has to do with the timing of metabolic flux changes. While transcriptional regulation will change enzyme abundance on the scale of minutes, it may be necessary or desirable to change some fluxes much faster than that. For instance, the response to oxidative stress in yeast involves the upregulation of the flux through the pentose phosphate pathway branch\textsuperscript{31}. In a series of elegant experiments, Ralser et al. showed that these fluxes in fact change on the timescale of seconds upon oxidative stress through allosteric regulatory mechanisms, much faster than the transcriptional changes that follow later\textsuperscript{32}. Yet another hypothesis for the apparent enzyme buffer is simply that it is very difficult to design a regulatory strategy that always produces the optimal level of enzyme. We will explore this aspect further in the “suboptimality of gene expression” section.

Nevertheless, some exceptions to the general theme of overabundant enzyme expression do exist. In some cases, overexpression of enzymes does lead to significant changes in flux through the corresponding pathway, as was observed for overexpression of phosphoglucomutase for increased galactose uptake in \textit{S.cerevisiae}\textsuperscript{33} or overexpression of fumarate hydratase for decreased fumarate secretion in \textit{Yarrowia lipolytica}\textsuperscript{34}. Furthermore, for one node in central metabolism, the pyruvate/acetyl-CoA node, perturbations to the transcriptional regulatory network in\textsuperscript{21} and \textsuperscript{20} did lead to small but significant changes in the flux ratios. This node involves the key decision between respirotive metabolism through the TCA cycle and overflow metabolism through ethanol, acetate or lactate secretion. In fact, recent work has suggested possible incentives for this branch to be regulated at the transcriptional level. For instance, two recent models that attempt to explain the prevalence of overflow metabolism rely on the high cost of respiratory enzyme expression, either through the metabolic cost\textsuperscript{35} or through the limitation of available membrane space\textsuperscript{36}. Beyond the high cost of respiratory enzyme expression, there is also indication that the cellular tolerance to perturbations in this seemingly important ratio is quite high. Firstly, even very closely related species often have large differences in this flux ratio while maintaining otherwise similar metabolism\textsuperscript{37}. Secondly, in \textit{E.coli}, the knockout of the key acetate production pathway had only a mild growth defect on glucose media, and
in laboratory evolution quickly recovered wild-type growth rates\textsuperscript{38}. If it is in fact the case that fluctuations in this branch point are easily tolerated and TCA cycle enzyme production is particularly costly, it may indeed be desirable for cells to favor reduced enzyme cost, as achieved by transcriptional regulation, over fast response time, as achieved by posttranscriptional regulation.

**Inferring metabolic state from transcriptomics**

The picture we have outlined is that at steady-state, enzymes tend to be expressed at overabundant levels, meaning that small to moderate changes in enzyme abundance tend not to affect fluxes significantly, but drastic rearrangements of the flux distribution do often involve large transcriptional changes (Figure 1). Such changes can be on a local level, such as the induction of a biosynthesis pathway in response to depletion of an amino acid in the medium\textsuperscript{39}, or on a more global level involving many pathways, such as the switch between aerobic and anaerobic metabolism\textsuperscript{40} or the transition to stationary phase\textsuperscript{41}. Naturally, this notion on the role of transcriptional regulation in regulating fluxes invokes the reverse question of whether the changes induced by such shifts accurately reflect the cell’s new metabolic requirements: do cells indeed only change the abundance of those enzymes which are necessary to enable drastic rearrangements of the flux distribution? Intriguingly, many of the now ubiquitous genome-wide transcriptomics\textsuperscript{42,43} and high-throughput proteomics\textsuperscript{44,45} analyses have led to the striking findings that even seemingly simple environmental perturbations lead to a broad array of gene expression changes. A particularly exhaustive study that monitored the response of *B.subtilis* to the transition between the carbon sources glucose, malate, and combinations thereof found over 2000 genes were found to change in at least one of the shifts, suggesting massive rearrangements of the transcriptional program\textsuperscript{46}. Yet constitutive expression of just a few genes was found to be sufficient for adaptation to the new environment. While such constitutive expression may well trigger other changes, it nevertheless seems likely that a large fraction of the transcriptional response is not directly related to the stimulus or necessary for adaptation.

**Suboptimality of gene expression**

Such supposedly unnecessary changes in gene expression may point to hidden complexity in the system, but a more parsimonious explanation may also be available. While modulation of metabolic gene expression is invaluable to adapting to new environments, an equally evident but more often ignored point is that cells cannot optimize their gene expression programs for all possible environments. Through combinations of available substrates, environmental stresses, and inter-species competition, microbes can encounter an infinite number of conditions both in nature and in the laboratory. Given the limited number of available sensors and regulators, it perhaps should not come as a surprise that gene expression programs are probably suboptimal in certain lab environments. This was observed upon construction of the yeast genome-wide knockout library\textsuperscript{47}
when the fitness of a gene knockout in a particular environment did not always anticorrelate with its expression in the wild type in the same environment. Recently, profiling of a transposon mutant library in \textit{S. oneidensis} revealed the same result – disruption of many genes was deleterious despite the fact that their expression was upregulated in the wild type in the same condition, and vice versa\textsuperscript{48}. Such analysis was then extended to three different bacteria and 15 conditions, finding hundreds of genes that are regulated suboptimally \textsuperscript{49}. One attractive hypothesis for this behavior would be anticipation of environmental changes, as proposed previously in yeast\textsuperscript{50}, but no evidence for such anticipatory programs was found in these conditions. Furthermore, suboptimality of expression programs was more directly demonstrated when rational rewiring of the regulatory network led to improved fitness under stress conditions\textsuperscript{51}.

One way that microbes cope with the problem of insufficient sensors is by using a single indicator that can take on a range of values corresponding to a variety of conditions. A particularly effective example of this could be a sensor for the magnitude of a central metabolic flux. Kotte et al.\textsuperscript{52} proposed that cells could sense the nutrients present in the environment not by dedicated nutrient receptors, but rather by directly measuring central metabolic fluxes via the interaction of metabolites with transcription factors. One metabolite that has been demonstrated to likely play exactly such a role is the glycolytic intermediate fructose-bis-phosphate as a sensor of the glycolytic flux in \textit{E. coli}\textsuperscript{53}. An important consequence of employing such flux sensors would be that cells become somewhat blind for the exact nature of the nutrients, as long as they lead to a similar flux distribution. Thus, such a mechanism would lead to a “just-in-case” model of transcriptional regulation, where many genes need to be expressed even though they are not necessary, since the signal upstream of the sensed flux would be unknown.

The fact that the wild-type program is suboptimal for many laboratory environments means that one can expect to see improvements in fitness in laboratory evolution experiments. Indeed, this is often found, even under supposedly “preferred” conditions like glucose media for \textit{E. coli} and \textit{S. cerevisiae}\textsuperscript{54,55}. With the impressive recent technological improvements in efficiency of strain sequencing and phenotypic characterization, such experiments have become more common, and in many cases have found a large number of gene expression changes that are consistent with the upregulation of metabolic pathways required for optimal growth\textsuperscript{56,57} (reviewed in \textsuperscript{58}). However, one other common result of such studies is large variability among evolved populations, both in fitness and in gene expression levels\textsuperscript{59,60}. This points to a situation in which the fitness surface is rough, that is, there are many local minima that can capture the evolutionary trajectory and reduce the chances of finding a global minimum. While some exceptions exist\textsuperscript{61,62}, it seems safe to assume that even in the natural
environment, the rough fitness landscape means that some optimal gene regulation solutions may never be found by evolution\textsuperscript{63,64}.

How does one evaluate the optimality of metabolic gene expression? One of the more common ways is to use genome-scale metabolic models\textsuperscript{10} to predict metabolic pathway usage that produces optimal biomass yields, and then assume that genes coding for those enzymes need to be upregulated, whereas genes in less efficient pathways should be downregulated\textsuperscript{65,66}. Recently developed methods that penalize longer pathways (which presumably need more enzymes) are likely a good proxy for the metabolic cost of gene expression and are valuable for this type of analysis\textsuperscript{57}. However, to predict gene regulation, more detailed models of costs and benefits of gene expression have been necessary to determine optimal regulatory strategies\textsuperscript{35,39,67}. A similar approach was also used by Poelwijk et.al.\textsuperscript{68} in a laboratory evolution experiment where artificial penalties were imposed on gene expression in particular phases of a cyclical environment. This resulted in the evolution of a reversed regulatory logic, as predicted by the model.

Conclusion

In summary, the picture that we have presented is that most metabolic enzymes in growing wild-type cells are present at high enough levels to not limit flux through the corresponding reaction. This protects the cell from random fluctuations in enzyme level, and allows it to adjust flux quickly when forced to by an external stress. In contrast, environmental changes that require significant flux redistribution or activation of new pathways typically do rely on transcriptional regulation to produce large changes in enzyme abundance. However, even these changes are often accompanied by other, apparently unnecessary changes, making the transcriptional response seem suboptimal. This could reflect the differences between lab and natural environments as well as the natural ruggedness of the fitness landscape. Clearly, the hypothesis on the role of transcriptional regulation in the control of metabolic fluxes put forward here needs further testing. Advances in high-throughput measurements of metabolic fluxes, proteins, and transcripts will allow for a more rigorous analysis of the relationship between metabolic fluxes and enzyme abundance across different conditions and organisms. In particular, dramatic improvements in our ability to further investigate the relationship between transcriptional regulation and metabolic state could come from technological advances in single-cell measurements. By correlating gene expression with metabolic fluxes at the single cell level, one could directly test how fluctuations in enzyme abundance affect metabolism. Recent advances in next generation sequencing combined with microfluidics have made quantification of the entire transcriptome in single cells realistic\textsuperscript{69,70}. While no one has yet proposed a method to measure fluxes in single cells, single-cell metabolomics has seen significant progress in recent years\textsuperscript{71,72}, and could also be an effective way to assess the metabolic state\textsuperscript{73}. 

48
Chapter 2 – the role of transcription in regulating microbial metabolic fluxes

Figures

Figure 1: Role of transcriptional regulation in regulating metabolic fluxes.

At low levels of gene expression (1,2) fluxes are enzyme-limited, i.e. any change in enzyme abundance will proportionally affect fluxes. Much evidence suggests that most wild-type enzymes are expressed at overabundant levels (3), protecting fluxes from fluctuations in gene expression and providing a mechanism for fast flux changes by enzyme modification.
Chapter 2 – the role of transcription in regulating microbial metabolic fluxes

References


Chapter 3 – Dissecting specific and global transcriptional regulation of bacterial gene expression

This chapter has been published in:


*: equal contribution
Abstract

Gene expression is regulated by specific transcriptional circuits but also by the global expression machinery as a function of growth. Simultaneous specific and global regulation thus constitutes an additional - but often neglected - layer of complexity in gene expression. Here, we develop an experimental-computational approach to dissect specific and global regulation in the bacterium *Escherichia coli*. By using fluorescent promoter reporters, we show that global regulation is growth rate-dependent not only during steady state but also during dynamic changes in growth rate and can be quantified through two promoter-specific parameters. By applying our approach to arginine biosynthesis, we obtain a quantitative understanding of both specific and global regulation that allows accurate prediction of the temporal response to simultaneous perturbations in arginine availability and growth rate. We thereby uncover two principles of joint regulation: i) specific regulation by repression dominates the transcriptional response during metabolic steady states, largely repressing the biosynthesis genes even when biosynthesis is required, and ii) global regulation sets the maximum promoter activity that is exploited during the transition between steady states.
Chapter 3 – Dissecting specific and global transcriptional regulation of bacterial gene expression

Introduction

Specific transcription factors regulate fundamental biological functions including metabolism, development and differentiation\(^1\)\(^-\)\(^4\). Departing from discovery and intuitive reasoning, the current focus is on understanding the quantitative and dynamic responses of transcriptional circuits to perturbations, for example to uncover DNA repair dynamics\(^5\) or to design de novo synthetic circuits\(^6\). A hallmark of such studies is the use of mathematical models to place experimental measurements on firm theoretical footing\(^7\). Typically, the dialogue between theory and experiment considers only interactions strictly within the transcriptional circuit, such as gene-transcription factor or ligand-transcription factor interactions. Transcriptional circuits are thus implicitly considered to function independently of their host organism status, an assumption that is both simplistic and incomplete. In fact, early\(^8\)\(^-\)\(^10\), and recent works\(^11\),\(^12\) demonstrated that the overall process of gene expression in bacteria is tightly coupled to the physiological growth status of the cell. Specifically, growth-dependent parameters related to cellular physiology and the global expression machinery, such as transcription rate and gene copy number increase due to multiple replication forks, strongly link gene expression to the growth rate\(^13\). Because environmental perturbations typically trigger both a change in growth rate and a specific transcriptional response, simultaneous regulation of gene expression by transcriptional circuits and the global expression machinery constitutes an unavoidable – but often neglected – layer of complexity in gene expression. To fully understand the principles by which bacterial gene expression is a function of the specific (‘by the transcriptional circuit’) and the global (‘by the expression machinery’) regulation, the contribution of expression machinery must therefore be quantified and included into analysis of gene expression regulation.

The expression machinery is a complex molecular network that performs the necessary steps of gene expression from transcription to translation, a process that involves hundreds of components and interactions\(^14,\)\(^15\). As a consequence, quantification of expression machinery activity cannot be obtained by simply measuring component abundances, such as RNA polymerases (RNAP) or ribosomes, but must be achieved by model-based interpretation of expression data. The natural starting point to understand expression machinery regulation is studying constitutive (‘not specifically regulated’) gene expression. Foundational results from Escherichia coli and Salmonella demonstrated expression of constitutive genes to be a function of the specific growth rate as the only parameter of gene expression, established at the level of transcription that increases with growth rate\(^9,\)\(^10,\)\(^12,\)\(^16\). Consequently, proposed models of constitutive gene expression focus at the promoter level and postulate mechanisms of growth rate-dependent increase in transcription, for example by an increase in the pool of RNAP\(\sigma^{70}\) holoenzyme that is available to initiate transcription\(^7,\)\(^18\). So far, two major limitations hampered the inclusion of global regulation into the analysis of transcriptional circuits. First,
growth rate dependency of gene expression has so far only been shown for steady state growth, hence it remains unclear whether this relationship is applicable to the general case of dynamic changes in growth rate. Second, determining the impact of expression machinery regulation on individual genes requires promoter-specific parameters that are difficult to estimate. Extending our quantitative understanding of global expression machinery regulation to dynamic changes in growth rate and obtaining the relevant governing parameters is necessary to unravel the interplay between transcriptional circuits and the global expression machinery in regulating dynamic cellular processes.

Here, we develop an approach to quantify global expression machinery regulation in the bacterium *E. coli*. We employ fluorescent reporters to quantify promoter activity\(^\text{19}\) as a measure of gene expression for a set of constitutive and specific regulated promoters. Using a Michaelis-Menten type rate law to capture the promoter-specific growth rate dependency, we can quantitatively describe and predict constitutive promoter activity not only during steady state but also during dynamic changes in growth rate. We then unravel the joint regulation by a transcriptional circuit and the global expression machinery for the specific case of arginine biosynthesis\(^\text{20,21}\). Accurate model-based predictions of the complex temporal responses to simultaneous perturbations in arginine availability and growth rate, indicate that we achieved a quantitative understanding of both specific and global transcriptional regulation processes. We conclude that our approach allows including the unavoidable and ubiquitous global regulation in the analysis and simulation of bacterial gene expression.
Chapter 3 – Dissecting specific and global transcriptional regulation of bacterial gene expression

Results

A model for specific and global regulation of bacterial gene expression. Here, we present a model of bacterial gene expression to interpret promoter activity data obtained from plasmid-borne fluorescence promoter reporters in the bacterium *E. coli*. Promoter activity (pa) is routinely determined as the production rate of a stable green fluorescent protein (GFP) expressed as a promoter fusion from a reporter plasmid, normalized by the optical density of the cell population (OD) and thus constitutes the aggregated output of all steps in the gene expression cascade. To relate this phenomenological measurement to specific and global regulatory mechanisms in the gene expression cascade, we used a standard model of bacterial gene expression in which GFP concentration is determined by transcription rate ($\alpha_m$), translation rate ($\alpha_p$), gene copy number ($g$), cell volume ($v$), dilution by growth rate ($\mu$), mRNA ($\beta_m$) and protein ($\beta_p$) degradation.

All above parameters could influence the measured promoter activity and thereby, in principle, must be accounted for in the mechanistic interpretation of experimental results. During exponential growth, however, mRNA degradation $\beta_m$ and translation rate $\alpha_p$ were previously shown to be constant and GFP degradation $\beta_p$ to be negligible compared with GFP dilution due to cell growth ($\beta_p<<\mu$). Although gene copy number ($g$) and cell volume ($v$) are functions of the growth rate, analysis of their reported dependency on the growth rate, complemented with plasmid copy numbers measured by us and others, revealed a constant plasmid concentration for the plasmids used in this study (see SI Text 2, Fig. S4). Thus, the only growth-dependent parameter of the expression machinery that affects promoter activity is the transcription rate $\alpha_m$.

Superimposed to global transcriptional regulation that effects essentially all genes is specific regulation of genes through transcription factors that can regulate the recruitment of RNAP polymerase as well as other steps in transcription. Promoter activity under joint global and specific transcriptional regulation can thus be expressed as function of the growth rate ($\mu$) and of the activity of the specific transcription factors (TF), with a proportionality term to transcription rate that is given by the constant plasmid concentration ($g/v$), translation rate ($\alpha_p$) and mRNA degradation ($\beta_m$):

$$\text{[1]} \quad \text{pa} = \frac{d\text{GFP}}{dt \cdot \text{OD}} = \alpha_m(\mu, \text{TF}) \cdot \frac{g\alpha_p}{v\beta_m}$$

For practical purposes, the above relationship implies that to dissect specific and global regulation an explicit formulation of the transcription rate $\alpha_m$ is necessary. As a first step toward dissecting the two regulatory sources, we develop a transcription rate function $\alpha_m$ for promoters that are solely under global regulation.
Chapter 3 – Dissecting specific and global transcriptional regulation of bacterial gene expression

In exponential growth, global regulation has been shown to increase transcription rate as a function of the growth rate by a trend well represented by a Michaelis-Menten rate law\textsuperscript{17,18}. Mechanistic interpretations of this observed relationship suggested a growth rate-dependent increase in the availability of the RNAP α\textsuperscript{70} holoenzyme that is free to initiate transcription\textsuperscript{17,18}, from now on referred to as free RNAP. For a constitutive promoter, we thus expect promoter activity to be described by a Michaelis-Menten type rate law as a function of the growth rate (μ) and two promoter-specific parameters $V_{\text{max}}$ and $K_m$:

$$[2] \quad p_a = V_{\text{max}} \cdot \frac{\mu/K_m}{1 + \mu/K_m} \propto a_m(\mu,-)$$

In the above relationship, $V_{\text{max}}$ quantifies the maximal promoter activity sustained by the promoter and $K_m$ the growth rate at which promoter activity is half-maximal. Similar to standard representations of specific transcriptional regulation\textsuperscript{28,29}, the global regulation is therefore described by the set of promoter-specific parameters $V_{\text{max}}$ and $K_m$, as well as by a variable signal, the expression machinery activity, captured here in the form of growth rate (μ). To dissect the contribution of global regulation to promoter activity, Eq. 2 should be validate against systematic experimental data and the corresponding $V_{\text{max}}$ and $K_m$ parameters should be quantified. In the next section, we test Eq. 2 and estimate $V_{\text{max}}$ and $K_m$ parameters from quantitative promoter activity and growth rate data for 12 novel constitutive promoters.

**Quantification of Global Expression Machinery Regulation.** To quantify the global expression machinery regulation using the developed model of constitutive promoter activity in Eq. 2, we constructed 12 constitutive promoter-GFP reporter strains that are not regulated by specific transcription factors anymore. Specifically, we choose three promoters in central carbon metabolism (pykF, kbl, and epd) and eight promoters in arginine biosynthesis for which well-characterized specific transcription factor and RNAP binding sites were known\textsuperscript{30–33}. Three constitutive promoter reporters were constructed for the pykF, kbl, and epd promoters by replacing their specific transcription factor binding sites that did not overlap with the RNAP binding site with a non-functional sequence. A fourth constitutive promoter was obtained by replacing the RNAP binding site of the epd promoter with the RNAP binding site of the icd promoter to generate the constitutive hybrid epd-icd promoter. Additionally, we made all eight promoters of the arginine biosynthesis pathway constitutive by transforming their promoter reporter plasmids\textsuperscript{39} into a strain with a deletion of their only regulator, the repressor ArgR (termed ΔArgR)\textsuperscript{34}. Expression from these 12 constitutive promoter-GFP constructs is thus exclusively subject to global regulation. We determined constitutive promoter activity and growth rate in the 12 strains during exponential growth under 18 nutritional conditions with growth
Chapter 3 – Dissecting specific and global transcriptional regulation of bacterial gene expression

rates between 0.2 to 1.5 h⁻¹; i.e. minimal medium with acetate, pyruvate, galactose, fructose, glucose, gluconate, glycerol, mannose, or succinate, either with or without supplemented amino acids.

The activity of these 12 constitutive promoters varied up to four-fold across conditions and showed a positive correlation with the growth rate, a trend that was well recapitulated by the least square fit to Eq. 2 with a mean percentile error of 16% (Fig. 1, Fig. S5). The estimated promoter-specific constitutive parameters varied between 373 and 8217 GFP-OD⁻¹ h⁻¹ for the $V_{\text{max}}$ values and between 0.20 and 1.90 h⁻¹ for the $K_m$ values (see Table 1). The two parameters $V_{\text{max}}$ and $K_m$ were independent of each other (Pearson correlation -0.2, p-value 0.45), and their estimation was robust to point elimination (up to 3 points removed, Fig. S6) and measurement noise (up to 20% additional noise, Fig. S7).

We evaluated the parameter space of $V_{\text{max}}$ and $K_m$ pairs that fit data within 5%, 10% and 20% of the optimal sum square error and found that promoters have distinct parameter spaces and thus are arguably promoter-specific even under parameter estimation uncertainty (Fig. S8). The different $K_m$ parameters suggest a promoter-specific component, rather than solely an unspecific one, in the non-linear relationship between promoter activity and growth rate. Following previously proposed mechanistic interpretations, this finding suggests a saturation trend of promoters by the free RNApo⁷⁻⁰ concentration at higher growth rates. To confirm that this Michaelis-Menten type relationship is specific to constitutive promoters, we constructed or retrieved the existing transcriptionally regulated GFP-reporters of the constitutive promoters (see Table S1), and measured their activity during steady state exponential growth under the 18 previously employed conditions (Fig. 1). Visual inspection and Pearson correlation on the linearized Lineweaver-Burk reciprocal revealed that the Michaelis-Menten type relationship between growth rate and promoter activity was statistically significant (p-value<0.01) for all constitutive and only for two, presumably weakly or growth-dependent specifically regulated promoters (Fig. S9).

Consistent with existing data, the growth rate-dependent expression machinery regulation recapitulates the activity of constitutive promoters in steady state. This relationship, however, has not yet been tested during dynamic changes in growth rate, where an array of regulatory mechanisms, such as the alarmone ppGpp or alternative sigma factors, could possibly affect the activity of the expression machinery. According to our model of global regulation, activation of alternative regulatory mechanisms would be observable as a disruption of the growth rate dependency of constitutive promoter activity. To identify the possible occurrence of such regulatory mechanisms, we monitored promoter activity and growth rate of the 12 constitutive promoters throughout the entire batch growth cycle under the 18 previously described conditions, from inoculation to stationary phase. In addition, we imposed a shift in growth rate by a diauxic nutritional downshift from glucose to succinate. We simulated constitutive promoter activities using calculated instantaneous growth rates.
and the earlier determined promoter-specific $V_{\text{max}}$ and $K_m$ parameters within Eq. 2 (Fig. 2). Throughout the whole time course of the investigated conditions, measured and simulated promoter activities exhibited a surprising agreement not only in their trend of higher activity at higher growth rate (Pearson correlation = 0.87 and p-value < 10^{-5}, see Table S6), but also in their quantitative levels ($R^2$=0.75, see Table S7). This consistency indicates that our constitutive promoters are predominantly regulated by the global expression machinery in a strictly growth rate-dependent relationship.

Thus, we confirmed and extended previous results based on fewer promoters and conditions that global regulation of promoter activity in steady state is relevant and can be quantitatively described by a Michaelis-Menten type rate law as a function of growth rate and two promoter-specific parameters $V_{\text{max}}$ and $K_m$. Our observation of different $K_m$ values is consistent with the notion that the increase in promoter activity is the result of a promoter-specific response to growth rate increases.

Further, we demonstrated that this quantitative relationship holds more generally also during dynamic changes in growth rate. For practical purposes, our findings indicate that growth rate, an easily accessible experimental parameter, can be employed as an approximation to quantitatively predict the global regulation of constitutive promoters given the parameters $V_{\text{max}}$ and $K_m$.

**Quantifying Specific Transcriptional Regulation of the Arginine Repressor Circuit.** Equipped with our approach to quantify global regulation, we next aimed at understanding the interplay of specific and global transcriptional regulation in controlling biological function. As a test case, we focused on the single input repressor circuit that controls arginine biosynthesis through a feedback loop from the pathway product to the repressor ArgR in *E. coli* (Fig. 3a)\(^{20,21}\). We developed a promoter activity equation using the thermodynamic framework\(^{25,37}\) that includes regulation by the ArgR repressor along with the constitutive characterization by $V_{\text{max}}$ and $K_m$ parameters (see Table 1). To link the global regulation with the repression mechanisms known to operate on the arginine promoters\(^{20}\), we interpreted growth rate as a proxy to the free RNAP concentration ($P\sigma^\sigma\propto\mu$),\(^{13,18}\). Under this assumption and given that ArgR and RNAP\(^{70}\) binding is mutually exclusive due to steric hindrance by overlapping ArgR and RNAP binding sites\(^{20}\), transcriptional regulation by *arg* promoters can be written as a function of the condition- and time-dependent arginine repressor activity ArgR\(^*$ and its promoter-specific binding dissociation constant, $K_r$:

$$[3] \quad p_{\text{argin}} = V_{\text{max}}^x \cdot \frac{\mu/K_m^x}{1+\mu/K_m^x+\text{ArgR}^*} \propto \alpha_m(\mu, \text{ArgR}^*)$$

where $V_{\text{max}}^x$ and $K_m^x$ are the constitutive parameters and $K_r^x$ the repressor binding affinity for the arginine promoter argx with $x\in\{A, \text{CBH}, D, E, F, G, I, R\}$. Under the above mechanistic interpretation, the parameter $K_m$ becomes a measure of the affinity between free RNAP and the promoter. As becomes apparent from Eq. 3, the expression machinery activity given by the growth rate sets the
maximum promoter activity that is achievable at a given point, in the following referred to as promoter capacity. The repressor activity $\text{ArgR}^*$ modulates promoter activity between promoter capacity and full repression.

We inferred the promoter-specific repressor affinities $K_r$ and the condition-dependent repressor activity $\text{ArgR}^*$ from activities of the eight regulated promoters under the 18 steady state growth conditions by least square fitting minimizing the overall percentage error (see SI Text 5). Specifically, we used the previously estimated $V_{\text{max}}$ and $K_m$ parameters (see Table 1) and the measured growth rate in Eq. 3. Since $\text{ArgR}^*$ activity and the $K_r$ parameters can be scaled, we set the repressor affinity for the argA promoter to unity. As expected, substantially lower $\text{ArgR}^*$ activity was found in conditions with biosynthetic production of arginine compared to those with externally supplemented arginine (Fig. S11c). Repressor affinity varied at most three-fold among the seven enzymatic promoters, but was ten-fold weaker for its own argR promoter (see Table 1, Fig. S11a). Consistent with an optimally efficient response to arginine depletion, such lower affinity for the argR promoter ensures basal repressor expression even under full repression. A striking quantitative aspect was much lower promoter activity than the promoter capacity (Fig. 1), showing that these promoters are strongly repressed, even when cells actually need to synthesize arginine. Calculations assuming mutual exclusion of ArgR and RNA polymerase, as described above, showed the ArgR repressor to be bound to enzymatic promoters more than 80% of the time during arginine biosynthesis (Fig. S11b).

Why have the arginine promoters evolved promoter capacities that are much larger than the promoter activity required in either the repressed or the unpressed steady state regime? To test whether this apparent excess of promoter capacity is exploited during dynamic metabolic adaptations, we determined activities of the eight regulated arginine promoters upon arginine depletion during exponential growth of *E. coli* batch cultures growing at 0.2 h$^{-1}$ (galactose), 0.65 h$^{-1}$ (glucose), and 1.5 h$^{-1}$ (glucose supplemented with amino acids except arginine) (Fig. 4a). We thereby triggered a dynamic response of pathway-specific transcriptional regulation while maintaining the expression machinery regulation at three different activity levels. During the transition from external supplementation to intracellular biosynthesis of arginine, the activity of the arginine promoters transiently peaks far above the steady state levels (Fig. 4a, Fig. S13 and Fig. S14). The only exception was the argR promoter with a higher baseline level and a lower peak, presumably due to its lower repressor affinity. Calculating promoter capacity of the eight arginine promoters from growth rate and the constitutive parameters $V_{\text{max}}$ and $K_m$ using Eq. 2 revealed that the regulated promoter activity approaches the promoter capacity during the transient adaptation (Fig. 4a, Fig. S13 and Fig. S14). Notably, the promoter activity bursts at fast growth could not be achieved by the promoter capacity set by the expression machinery regulation at slow growth. Thus, the promoter capacity is only transiently exploited during adaptation.
and the growth rate-dependent expression machinery regulation ensures increasing maximum capacity for rapid enzyme synthesis at higher growth rates.

These dynamic patterns of promoter activity bursts achieve a smooth increase and constant steady state concentration of the biosynthesis enzymes at all growth rates (Fig. 4b), indicating that increasing promoter capacity counter balances the faster dilution of enzymes in more rapidly dividing cells. Consistent with the postulated just-in-time regulatory program \(^7\), we observed an ordering of enzyme concentration that follows the pathway order with \( \text{argA}> \text{argCBH}>\text{argD}>\text{argE}>\text{argF}>\text{argG} \) and the isoenzyme \( \text{argI} \) as the only exception (Fig. 4b). However, we did not find evidence for the proposed ordering of response times \(^7\); \( i.e. \) all enzymes reached half-maximum concentration roughly at the same time for a given growth rate (Fig. 4b). Overall, our results highlight two distinct principles of arginine pathway regulation. First, the specific repressor circuit determines the fraction of promoter capacity that is exploited. Full derepression is confined to short transient phases whereas in steady state the biosynthesis promoters are largely repressed, even when required. Second, the global regulation by the expression machinery sets a growth-dependent promoter capacity that allows higher bursting of promoter activity during fast growth, to rapidly produce enzymes and counter balance the increased growth dilution.

**Predicting the Response to Simultaneous Dynamic Specific and Global Perturbations.** With a quantification of the specific and global regulation interplay at the promoter level in hand, we finally attempted to obtain a comprehensive, pathway-level understanding of regulation in the arginine pathway (Fig. 3a) by investigating the response to simultaneous and orthogonal dynamic perturbations in the specific and global regulation. For this purpose, we developed an ordinary differential equation (ODE) model that includes arginine production, consumption and feedback to the repressor ArgR (Fig. 3b). Since we focus on gene expression regulation rather than detailed pathway biochemistry, the ArgA-catalyzed first of the nine linear reactions converting glutamate to arginine was modelled as rate-limiting, assuming the other reactions to operate instantaneously. The regulatory and metabolic pathway structure was converted into a set of ODEs that describe concentrations of enzymes and repressor as dictated by their promoter activity, arginine concentration and repressor activity. The model has 30 parameters, 24 of which were previously inferred as the 8 promoter-specific \( V_{\text{max}} \) and \( K_m \) parameters and ArgR dissociation constants \( K_r \) (see Table 1). To infer the six missing parameters, we fitted the ODE model to reproduce the ArgR activity underlying the three arginine shifts by least-square criteria (Fig. 4c). ArgR activity was obtained from the measured promoter activity using Eq. 3 and, as expected, showed an initially high activity and a transient minimum that increases to reach a second slightly lower steady-state level. As the combined output of arginine production, consumption and repressor activation, ArgR activity represents an ideal readout to infer the relevant metabolic-feedback
parameters in the model. We fitted the ODE model to the dynamic ArgR activity from the onset of promoter up-regulation, thus discarding the initial fully repressed state that is not informative, and identified a parameter set that optimally reproduces the dynamic ArgR activity (parameter values \( k_{\text{cat}}=0.32, K_{\text{deg}}=0.64, K_{\text{arg}}=1.3, r_b=68, n=2.6 \) and \( k_p=148 \), initial conditions ArgR=800, ArgA=0, arg=71).

To challenge our parameterized ODE model, we investigated the pathway response to a novel scenario on which the model has not been trained: simultaneous and orthogonal dynamic perturbations in both specific and global regulation. In a set of experiments *E. coli* cultures experienced arginine depletion at different phases of growth during a diauxic shift from the preferred carbon source glucose to succinate, whose consumption is repressed in the presence of glucose. The obtained promoter activities of arginine promoters for the entire seven hours of growth are consistent with the above principles (Fig. 5a); *i.e.* a transient burst in arginine promoter activity that increases with the growth rate, and a growth rate-dependent promoter capacity set by the expression machinery that is essential for reaching the observed burst levels. To test whether the quantitative knowledge acquired by our approach can capture such a complex transcriptional response, we predicted *in silico* the regulatory response of arginine promoters to both simultaneous perturbations using the ODE model without any refitting of model parameters. The inputs given were the measured growth rate during diauxic growth and the 12 onset times at which arginine runs out, as identified from the experimental data (see SI text 7 for simulation details). The simulated promoter activities accurately reproduced the measured data (Fig. 5a-b, Fig. S15). Since our mathematical model of arginine pathway regulation is precise in recapitulating the dynamic responses (Fig. 5b), we conclude that our approach is able to achieve a satisfactory quantitative understanding of joint specific and global regulation.
Discussion

Cellular functions arise from the coordinated interplay between molecular networks. While reconstruction of topologies is thriving, the challenge shifts to identify the principles of network coordination 38,39. Here, we unravelled two principles of coordination between a repressor circuit with a single input and the global expression machinery for the conjoint control of arginine biosynthesis. First, substantially reduced ArgR repressor activity occurs only transiently during a switch from external uptake to arginine biosynthesis. Together with a strong affinity of the repressor for enzymatic promoters, this leads to strongly repressed arginine promoters, even when cells need to produce arginine, intervened by only short bursts of near maximum expression at the onset of the transition. Second, the global expression machinery sets a growth rate-dependent maximum expression level that we interpreted as promoter capacity, such that transient promoter activity can peak higher when de-repressed at fast growth. Thus, specific regulation by repression is dominant during metabolic steady states, and global regulation becomes relevant during transitions between metabolic steady states. Together, these two coordination principles between the metabolic-repressor circuit and the global expression machinery enable rapid pathway induction at any growth regime. The surprising 27% reduction of growth rate in the ArgR knockout strain during biosynthetic conditions (Fig. S1) might thus be explained by an abnormal upregulation of enzyme levels, since we showed that de-repressed biosynthesis genes are quantitatively tuned to foster enzyme bursts and not to maintain steady state levels.

The uncovered coordination principles are likely to be general for the large class of biosynthetic pathways that are controlled by single input repressor circuits, because metabolic pathways are unavoidably challenged with the growth dilution of enzymes during transient upregulation 40. For example, growth-rate dependent scaling of promoter capacity can also be observed in other biosynthetic pathways, e.g. for L-methionine (appendix figure 1) and purine biosynthesis (appendix figure 2). Quantitative functioning of the abundant single input repressor circuit motif 41 was previously assessed through parameterizing interactions 5,40 or quantifying regulator abundances 42. Here, we added a layer of complexity that was previously neglected by explicitly including the global expression machinery regulation. Our results demonstrate that single input repressor circuit performance is not simply dictated by its own input-output relation 42, but also by the activity of the global expression machinery that sets the promoter capacity. Thus, understanding the quantitative role of network motifs in gene regulation 41 might require to characterize their relationship with global regulation by the expression machinery. As an example for the much larger class of biosynthetic genes, we found specific regulation to dominate arginine biosynthesis through a single-input repressor. By focusing on one enzyme-encoding and two transcription factor genes, a recent paper postulates predominance of

65
global regulation in central metabolism with only a minor, fine-tuning contribution from the pleiotropic transcription factors Crp and Fis\textsuperscript{43}. The here proposed approach can now be used to evaluate how common such specific and global transcriptional paradigms of regulation are across cellular functions and transcriptional circuit configurations.

Akin to activity inference of metabolic\textsuperscript{44} and transcription\textsuperscript{18,45} networks, the activity of the global expression machinery was accessible only by model-based interpretation of a measurable proxy, in our case through GFP-promoter fusions of synthetic constitutive promoters\textsuperscript{39}. For data interpretation we used a model based on foundational work showing i) that steady state expression of constitutive genes is strictly growth rate-dependent\textsuperscript{9,10,12,16} and ii) that a Michaelis-Menten type rate law describes steady state constitutive promoter activity dependency on growth rate\textsuperscript{17,18}. We extended this work by demonstrating that the governing $V_{\text{max}}$ and $K_m$ parameters are promoter-specific and that the phenomenological growth rate dependency largely applies also to dynamic changes in growth. Specifically, we found that $V_{\text{max}}$ and $K_m$ estimates obtained from steady state data can be used to predict constitutive promoter activity from experimentally accessible growth rates. Akin to enzymes, promoter regions could thus be routinely characterized by their $V_{\text{max}}$ and $K_m$ values and made available in databases of standardized biological parts\textsuperscript{46}. Such promoter parameters could potentially be determined \textit{in vitro}\textsuperscript{47} or inferred computationally from \textit{in vivo} expression and sequence information\textsuperscript{29,48–50}.

To include global regulation within transcriptional circuits, we leveraged a previously proposed mechanistic interpretation in terms of free RNAP\textsuperscript{70} availability and promoter saturation\textsuperscript{17,18} which is consistent with the physiological and gene expression parameters of \textit{E. coli} reported in literature\textsuperscript{13}. This view is also consistent with the reported growth dependency of parameters in the gene expression cascade\textsuperscript{13}, which highlight transcription rate as the main driver of growth dependency in gene expression. An implication is that the translation rate is growth independent, even though the ribosome content is known to increase with faster growth\textsuperscript{51}. Several authors proposed a constant free ribosome concentration as a likely explanation\textsuperscript{13,22,23}. However, it can be argued that the translation rate has not been characterized with the systematic rigor of the other parameters\textsuperscript{22,23} and a growth rate dependent translation rate might eventually be included within formulations of constitutive gene expression. Our approach and the relative conclusions on the mainly repressed biosynthesis promoters with only transient expression bursts are independent of this putative influence of the translation rate, as is the phenomenological shown predictive power for constitutive and regulated promoter activity. Solely the mechanistic interpretation of increased constitutive gene expression in terms of increased promoter capacity might have to be amended by a growth dependent translation rate. This
amendment would require to additionally dissect the individual contributions of the possibly tightly and complexly coupled processes of transcription and translation in global regulation.

Overall, we developed an experimentally validated approach to include global expression machinery regulation in the analysis and simulation of bacterial gene expression. This is particularly relevant for promoter activity data that is increasingly employed to quantify transcriptional regulation, but less so for the growth-independent mRNA abundance as a measure of gene expression. Unless global regulation is properly considered, inferring specific transcriptional regulation from promoter activities obtained at different growth rates will be misleading. While some authors have used constitutive reporters to normalize promoter activity data, we developed a superior approach based on first principles to explicitly include the promoter-specific growth dependency into models, under the constraint that parameters in the gene expression system can be expressed as a function of the growth rate. Our approach not only allows to pin-point the regulatory role of global expression machinery instead of concealing it under a heuristic normalization, but additionally allows precise simulation of a pathway under joint specific and global regulation. We envisage that the here described principles will help including the ubiquitous and unavoidable role of global regulation in studies aiming for holistic understanding of interacting cellular networks.

Acknowledgements
We gratefully acknowledge Iftach Nachman for advising on the modelling, Tobias Bollenbach for kindly providing strains and Bart R. B. Haverkorn van Rijsewijk and Victor Chubukov for helpful discussions.
Material and Methods

**Strains, plasmids and media.** The *E. coli* K-12 strain BW25113 was used throughout. The *argR* deletion mutant was obtained from a knockout library\(^1\) and deprived of its kanamycin resistance as previously described\(^2\). Green fluorescent protein (GFP)-based promoter reporter plasmids were obtained from a library\(^2\) or constructed by PCR following the procedures of the original study\(^2\) (see Table S1). Growth experiments were performed in M9 minimal medium to which carbon sources were added from sterilized stock solutions (adjusted to pH 7) to a final concentration 5 g/L (see SI Text 1). For amino acid addition, a stock solution containing all 20 amino acids (adjusted to pH 7) was added to yield final concentrations as described elsewhere\(^3\), unless stated otherwise (see Table S2). All chemicals were purchased from Sigma-Aldrich unless stated otherwise.

**Cultivation.** M9 medium batch cultures in 96 deep-well format plates (Kuehner AG, Birsfeld, Switzerland) were inoculated 1:50 from LB precultures and incubated overnight at 37°C under shaking. Subsequently, 96 well flat transparent plates (Nunc, Roskilde, Denmark) containing M9 medium (fill volume 200 µL) were inoculated 1:200 with overnight cultures and sealed with parafilm to reduce evaporation. On-line measurements of optical density at 600 nm (OD\(_{600}\)) and fluorescence (excitation wavelength: 500 nm, emission wavelength: 530 nm) were performed at 37°C with shaking using a plate reader (TECAN infinite M200, Tecan Group Ltd., Männedorf, Switzerland) at 10 min intervals.

**Data processing and modelling.** Fluorescence (GFP) and OD\(_{600}\) measurements were analyzed using custom MATLAB software. Both signals were processed in a well-specific manner to obtain promoter activity (dGFP/(dt-OD)), growth rate (dln(OD)/dt) and expression profile (GFP/OD) time-course data. We note that GFP values represent raw reading values from the measuring device and are thus in arbitrary units, not in protein counts. OD and GFP signals were normalized to blanks by subtracting values prior to inoculation of cells and smoothed using a moving average window with size 3. GFP and ln(OD) derivatives were obtained by two-point finite difference numerical approximation. Promoter activity and expression profile signals were corrected for fluorescence background by subtracting the corresponding signal of the promoter-less plasmid reporter strain p139\(^2\). Promoter activity and growth rate values for steady state growth were calculated as the average value in the time range visually identified as exponential phase. In arginine depletion experiments, signals from different promoters were aligned by the growth rate to correct for differences in inoculation OD. Estimation of parameters was performed using the *fmincon* MATLAB function.
Figures

**Figure 1. Global expression machinery regulation of promoter activity during exponential growth.**
Promoter activity of constitutive (blue dots) and native, specific regulated (green dots) promoters as a function of the steady state growth rate under 18 nutritional conditions. Red lines show optimal least-square fitting of a Michaelis-Menten rate law. Grey shaded areas illustrate the margins of fits within 20% of the optimal sum squared error. Constitutive promoter activity was measured in wild type, except promoters in the arginine pathway which are measured in the ΔArgR background strain. Promoter activity of the native specific regulated version of each promoter is shown (green dots). In the case of the *epd-icd* promoter, promoter activity was also measured in a ΔArgR strain (white dots) to evaluate ΔArgR knockout effects on promoter activity and growth rate.
**Figure 2.** Global expression machinery regulation of promoter activity during dynamic changes in growth rate. Measured (blue) and simulated (red) constitutive promoter activities for three representative promoters and nine growth conditions in rows and columns, respectively. Simulations are based on time-course measured growth rate and the constitutive, promoter-specific $V_{\text{max}}$ and $K_m$ parameters inferred in steady state. The Pearson correlation (c) and the coefficient of determination (r) between measured and simulated promoter activities are given in each box. The *epd-icd* and *kbl* promoters were measured in a wild type background and the *argI* promoter is in the ArgR knockout background. Except for the diauxic shift, growth conditions are ordered by increasing maximum growth rate.
Chapter 3 – Dissecting specific and global transcriptional regulation of bacterial gene expression

Figure 3. Model of specific and global regulation of the arginine biosynthesis pathway in *E. coli*. a) Schematic representation of specific transcription (red lines) and global expression machinery (green lines) regulation of gene expression in the arginine biosynthesis pathway (blue lines). The pathway synthesizes arginine through linear reactions catalyzed by enzymes transcribed from seven $\sigma^{70}$ promoters. The transcriptional circuit is coupled to arginine metabolism through activation of the ArgR repressor by arginine (blue line). Global expression machinery regulation regulates constitutive promoter activity as a function of growth rate. Parameters that govern the interaction strength between molecular components as defined in our model are shown. b) Ordinary differential equation model for the arginine biosynthesis pathway including metabolic biosynthesis and feedback, transcriptional regulation and expression machinery regulation.
Figure 4. Specific and global regulation of the arginine pathway upon dynamic arginine depletion during exponential growth. The three growth phases during the switch from external uptake to biosynthesis of arginine are separated by vertical dotted lines: growth on externally supplemented (E), transition from uptake to biosynthesis (T) and steady biosynthesis (B) of arginine. a) Activities of the eight arginine promoters in response to depletion of externally supplemented arginine during exponential batch growth. The red continuous line represents the measured regulated promoter activity and the red dotted line represents the simulated growth rate dependent promoter capacity for argA. b) The fluorescence expression profile, a proxy to enzyme concentration, of the eight arginine biosynthesis enzymes as measured during the shifts. c) Activity of the arginine repressor ArgR (blue continuous line) as inferred from measured promoter activity of the eight arginine promoters and as fit by the ODE model (purple dashed line).
Figure 5. Measured and simulated argA promoter activity under simultaneous dynamic perturbations in specific and global regulation. In 12 independent experiments a different arginine concentration (see legend) was supplemented and cells depleted it at different time points during the course of growth in a diauxic shifts from glucose to succinate. a) Promoter activity of the argA promoter as measured in the 12 experiments. b) Promoter activity of the argA promoter as simulated by the ODE model starting simulation at each of 12 different upregulation onset times (see legend). The simulated constitutive promoter activity for argA promoter is shown (dotted black line).
Tables

Table 1. Promoter-specific constitutive parameters and ArgR repressor dissociation constants.

<table>
<thead>
<tr>
<th>Promoter</th>
<th>$V_{\text{max}}$ [GFP-OD$^{-1}$·h$^{-1}$]</th>
<th>$K_m$ [h$^{-1}$]</th>
<th>$K_r$ [-]</th>
</tr>
</thead>
<tbody>
<tr>
<td>argA</td>
<td>2660</td>
<td>0.42</td>
<td>1</td>
</tr>
<tr>
<td>argCBH</td>
<td>1886</td>
<td>0.45</td>
<td>1.52</td>
</tr>
<tr>
<td>argD</td>
<td>1296</td>
<td>1.15</td>
<td>5.25</td>
</tr>
<tr>
<td>argE</td>
<td>1414</td>
<td>0.72</td>
<td>1.75</td>
</tr>
<tr>
<td>argF</td>
<td>1558</td>
<td>1.90</td>
<td>1.43</td>
</tr>
<tr>
<td>argG</td>
<td>373</td>
<td>1.77</td>
<td>2.85</td>
</tr>
<tr>
<td>argI</td>
<td>8217</td>
<td>1.02</td>
<td>0.60</td>
</tr>
<tr>
<td>argR</td>
<td>1146</td>
<td>0.20</td>
<td>25.70</td>
</tr>
<tr>
<td>pykF</td>
<td>683</td>
<td>1.72</td>
<td>-</td>
</tr>
<tr>
<td>kbl</td>
<td>877</td>
<td>1.63</td>
<td>-</td>
</tr>
<tr>
<td>epd</td>
<td>1185</td>
<td>0.39</td>
<td>-</td>
</tr>
<tr>
<td>epd-icd</td>
<td>5104</td>
<td>0.53</td>
<td>-</td>
</tr>
</tbody>
</table>
Supplementary material

Experimental measurement datasets, the ODE model (in MATLAB script format) and supplementary information are available at the Molecular Systems Biology website (www.nature.com/msb).

Appendix

Appendix figure 1. Activation of L-methionine biosynthesis pathway promoters upon L-methionine depletion during exponential growth on four different carbon sources. Strains bearing L-methionine biosynthesis promoters were obtained from\textsuperscript{19} and inoculated in M9 minimal medium supplemented with 2g/L of the respective carbon source and 16 µM L-methionine. The concentration of L-methionine as chosen such that it is completely depleted during mid-exponential phase. Upper panel: OD curves. Lower panel: corresponding promoter activity curves. Promoter activities were determined as described in the main text.
Chapter 3 – Dissecting specific and global transcriptional regulation of bacterial gene expression

Appendix figure 2. Activation of purine biosynthesis pathway promoters upon hypoxanthine and guanine depletion during exponential growth on four different carbon sources. Strains bearing purine biosynthesis promoters were obtained from^19^ and inoculated in M9 minimal medium supplemented with 2g/L of the respective carbon source and 24 µM hypoxanthine or guanine, two intermediates of the purine salvage pathway, respectively. These concentrations were chosen such that it is completely depleted during mid-exponential phase. Upper panel: OD curves. Middle panel: corresponding promoter activity curves for hypoxanthine depletion. Lower panel: corresponding promoter activity curves for guanine depletion. Promoter activities were determined as described in the main text.
References

Chapter 3 – Dissecting specific and global transcriptional regulation of bacterial gene expression

Chapter 3 – Dissecting specific and global transcriptional regulation of bacterial gene expression

Chapter 4 – Few regulatory metabolites coordinate expression of central metabolic genes in *E. coli*

Karl Kochanowski*, Luca Gerosa*, Dimitris Christodoulou, Simon Brunner, Uwe Sauer

*: equal contribution

Contributions:

KK, LG and US designed the study. KK wrote this chapter, with contribution from US and LG. KK performed all experiments/computational analyses, with contribution from LG, DC and SB.
Chapter 4 – Few regulatory metabolites coordinate expression of central metabolic genes in E. coli

Abstract
Microorganisms use transcriptional regulation as a key mechanism to adapt to diverse environments. However, it is unclear how the interplay of a cell’s complex transcriptional network and a multitude of potential regulatory signals gives rise to a coordinated transcriptional regulatory program. In this work, we unravel the transcriptional regulatory program of central carbon metabolism in the model bacterium *Escherichia coli*. Using a library of fluorescent transcriptional reporters, we comprehensively quantify the activity of central metabolic promoters in a wide range of environmental conditions. We find that few expression patterns dominate central carbon metabolism, in particular growth-rate dependent global regulation for most promoters, together with highly condition-specific activation for only few promoters. Using an approximate mathematical description of promoter activity, we dissect the contribution of global and specific transcriptional regulation, and find global transcriptional regulation accounts for about 70% of the total variance in promoter activity across conditions. By relating each promoter’s remaining specific transcriptional regulation with the cell’s metabolome response across the same conditions, we identify metabolites which serve as potential regulatory signals. We find that few metabolites, namely cyclic AMP, FBP and F1P, explain most of this specific transcriptional regulation across promoters through their interaction with two major specific transcription factors, namely Crp and Cra. Overall, this work suggests that the transcriptional coordination of *Escherichia coli*’s central carbon metabolism is achieved by a surprisingly simple regulatory program, which utilizes only a fraction of the cell’s transcriptional regulatory repertoire across a wide range of conditions and relies on information from few intracellular regulatory metabolites.
Introduction

Growth sustaining environments of most microbes encompass a wide range of nutritional conditions\textsuperscript{1–3}. Coping with such diverse environments requires coordinated regulation of metabolism to provide the biomass precursors, redox factors, and energy necessary\textsuperscript{4}. Towards this end, microorganisms use transcriptional regulation as a key regulatory mechanism. An example is the induction of a uptake and utilization pathway upon availability of a nutrient\textsuperscript{4}. Beyond such relatively simple control of particular pathways, transcriptional regulation of metabolism is typically much more complex, where even subtle environmental changes alter expression of hundreds of metabolic genes\textsuperscript{5–9}. How are such global transcriptional responses achieved? Many of the studies addressing this question have focused on the role of transcription factors in mediating these responses, and efforts in mapping out each transcription factor’s targets have cumulated in highly overlapping and dense transcriptional regulatory networks comprising major (with hundreds of target genes) and minor (with few target genes) transcription factors\textsuperscript{10–16}.

To exert their regulatory function, transcription factors in turn need to receive signals. These signals may for example be transmitted through signaling pathways. In bacteria, these signaling pathways are typically two-component systems, which relay the detection of internal or external signals by a designated receptor protein to the phosphorylation of the respective transcription factor\textsuperscript{17}. Alternatively, transcription factors may also respond directly to levels of regulatory metabolites, whose binding directly alters their activity. Examples of such regulatory metabolites include amino acids, which frequently regulate the expression of their own biosynthesis pathway, intermediates of nutrient utilization pathways such as glycerol-phosphate, and central metabolites such as pyruvate or fructose-1,6-bisphosphate\textsuperscript{4}. This ‘one-component’ type of regulation is prevalent in bacteria\textsuperscript{18}, allowing them to match their transcriptional response to their current metabolic state. Thus, a cell’s transcriptional response is the result of the complex interplay of many transcription factors whose activity is carefully adjusted to the environment by their respective regulatory signals\textsuperscript{10,11,13}.

Understanding how a cell’s transcriptional response emerges from this complex interplay of regulatory networks and signals is a daunting task, and two key limitations have hampered our ability to study such regulatory networks in a quantitative manner. The first limitation stems from the realization that the activity of a promoter is not only determined by the network of transcription factors (termed ‘specific transcriptional regulation’), but also by the global physiology of the cell\textsuperscript{19–26} (termed ‘global transcriptional regulation’), for example through growth-rate dependent changes in free RNA polymerase availability or ribosome abundance\textsuperscript{22,27,28}. This global transcriptional regulation adds another layer of complexity to the cellular transcriptional response, and was found to affect a large fraction of genes in bacteria and eukaryotes\textsuperscript{25}. Thus, to understand a cell’s transcriptional response,
we need to robustly dissect the contribution of global and specific transcriptional regulation for large scale networks across diverse conditions, but previous efforts have been restricted to detailed dynamic analysis of few promoters\textsuperscript{23,24}, or large-scale analysis for few conditions\textsuperscript{21,25}. The second limitation is the lack of methods to systematically identify signals which regulate transcription factor activity. In particular transcription factor-metabolite interactions are – as all protein-metabolite interactions\textsuperscript{29} – notoriously difficult to detect due to their non-covalent nature\textsuperscript{30}. Consequently such interactions have largely been identified through extensive biochemical analyses or serendipity on a case-by-case basis, and even for well-studied organisms such as \textit{Escherichia coli} the regulatory signal for the majority of transcription factors has remained elusive\textsuperscript{31}. Efforts to infer such interactions directly from \textit{in vivo} experimental data have been restricted to few case studies, without accounting for the confounding influence of global transcriptional regulation\textsuperscript{32,33}.

In this work, we aim to overcome these limitations to unravel the transcriptional regulatory program of central carbon metabolism in the model bacterium \textit{Escherichia coli}, which is regulated by a well-characterized and extensive network of transcription factors\textsuperscript{15}. Due to its pivotal role in providing the cell with energy and biomass precursors, central carbon metabolism is likely to be affected by a wide range of environmental perturbations. Using a library of fluorescent transcriptional reporters, we comprehensively quantify the activity of central metabolic promoters in a wide range of environmental conditions, and dissect the contribution of global and specific transcriptional regulation at single-promoter level. Moreover, we systematically infer metabolites which serve as potential regulatory signals. We find that the transcriptional coordination of \textit{Escherichia coli}'s central carbon metabolism is achieved by a surprisingly simple regulatory program, which utilizes only a fraction of the cell’s transcriptional regulatory repertoire and few intracellular regulatory metabolites.
Results

Quantifying gene expression of *Escherichia coli* using transcriptional fluorescent reporters. To unravel the transcriptional regulatory program governing *E. coli*'s central metabolic genes, we first needed to systematically quantify their expression. Towards this end we used a library of fluorescent transcriptional reporter plasmids, and expanded it with 28 additional promoters to cover 95% of the promoters in central carbon metabolism. We further included five previously characterized synthetic constitutive promoters that are only affected by global transcriptional regulation, hence allow to study its effect in isolation (see supplementary table 2 for full list of promoters). To study the activity of these 95 promoters in a wide range of physiological states, we selected 26 diverse conditions - including carbon sources, amino acid supplementation, and sub-lethal ribosome inhibition by chloramphenicol – which span growth rates between 0.1 h⁻¹ and 1.5 h⁻¹ (see supplementary table 1 for full list of conditions). Despite promoter activities spanning several orders of magnitude, there were no adverse effects of GFP expression on growth rate (supplementary figure S1). Day-to-day comparison showed that promoter activity measurements during exponential growth were reproducible within approximately 10-20% variation, which is comparable to previous studies (supplementary figure S2). 31 of the tested promoters, predominantly those of minor central metabolic iso-enzymes, were inactive under all tested conditions and were therefore discarded for subsequent analyses.

Hierarchical clustering of promoter activity identified surprisingly few distinct patterns that dominate across conditions (figure 1). Few small clusters, mostly consisting of promoters of carbon utilization pathways, showed highly condition-specific activation in one or two of the tested conditions (e.g. clusters 11 and 12). Approximately 15% of the tested promoters, mostly consisting of TCA cycle promoters, were activated on carbon sources supporting growth rates below 0.8 h⁻¹, but were not activated by chloramphenicol treatment that led to similarly low growth rates (clusters 14 and 15). The activity of the majority of tested promoters, however, was consistently related to the growth rate supported by each condition (clusters 1 and 13). Notably, all tested synthetic constitutive promoters, which are only affected by global transcriptional regulation, fell into the last category, suggesting that the other promoters in the respective clusters are also strongly affected by global transcriptional regulation.

Dissecting global and specific transcriptional regulation in central carbon metabolism. To dissect the measured promoter activity of each promoter into its respective global and specific transcriptional regulation components, we next quantified the contribution of global transcriptional regulation. Previous approaches toward this end relied on normalization by synthetic constitutive promoters, prior parameterization of each promoter’s global regulation by quantifying its promoter activity in absence of its specific regulation, or on a phenomenological scaling factor that captures pairwise
differences in global regulation between conditions\textsuperscript{25}. To dissect global and specific transcriptional regulation across many conditions while retaining a mechanistic description of gene expression, we expanded a previously described\textsuperscript{23} approximate mathematical description of promoter activity (pa):

\[ p_{a_{i,j}} = \left[ \frac{E^*_j}{K_{E,i}} \right]^{\alpha_{E,i}} \cdot \prod_{i \in \{TF\}} \left( \frac{T^*_{i,j}}{K_{i,j} + 1} \right)^{\alpha_{i,j}} \]  

(Eq. 1)

Here, \( E^* \) denotes the activity of the expression machinery (in condition j) with its promoter-specific parameters \( K_{E,j} \) and \( \alpha_{E,j} \), and \( T^* \) denotes the activity of each specific transcription factor (in condition j) that regulates the respective promoter with its promoter-specific parameters \( K_i \) and \( \alpha_i \). Upon normalization and transformation into log space, this equation can be simplified to:

\[ \Delta \log p_{a_{i,j}} \approx \alpha_{E,i} \cdot \Delta \log \left( \left[ E^*_j \right] \right) + \sum_{i \in \{TF\}} \alpha_{i,j} \cdot \Delta \log \left( \left[ T^*_{i,j} \right] \right) \]  

(Eq. 2)

As a result, log normalized promoter activity can be approximated by a linear combination of global (formalized as expression machinery activity) and specific (formalized as transcription factor activity) transcriptional regulation. In this approximation, dominant signals, such as global transcriptional regulation, manifest as projections of the data that capture most of the data set’s variability and can be determined as singular vectors by singular value decomposition\textsuperscript{35}. Expectedly, the first singular vector showed strong growth rate dependence and accounted for 68% of the total variability in our data set (figure 2A). Conversely, comparison of measured promoter activities with their respective reconstruction purely based on the first singular vector showed very good agreement for many promoters, such as the synthetic constitutive promoters, showing that their activity is dominated by global transcriptional regulation (figure 2B). To test whether this growth-dependent singular vector allows to also predict global transcriptional regulation in complex dynamic environments, we measured dynamic promoter activities during a diauxic shift from glucose to succinate and compared them to predictions utilizing the relationship between measured growth rate and the first singular vector (figure 2C). We found that promoters with dominant global transcriptional regulation (as determined in figure 2B) were indeed well predicted.

Thus, our approximate mathematical description of promoter activity in combination with singular value decomposition enables dissection of global and specific transcriptional regulation for large numbers of promoters without prior promoter-specific parameterization\textsuperscript{24}, and without the requirement for constitutive promoters to be used for normalization\textsuperscript{23}. Our analysis revealed that the contribution of global transcriptional regulation is considerable, accounting for 68% of the total variability in promoter activities across conditions.

**Systematic identification of metabolites affecting specific transcriptional regulation.** Given that we can quantify the contribution of global regulation, we now know for each promoter the fraction of its
activity that is governed by specific, presumably transcription factor driven, regulation. To understand this specific transcriptional regulation component, we next focused on inferring the metabolites serving as most probable regulatory signals, as well as the transcription factors that ultimately exert these regulatory functions. In the first step, we established the link between promoters and their respective regulatory metabolites (we will focus on identifying the corresponding transcription factors in the next section).

Towards this end, we started from the observation that the activity of bacterial transcription factors across conditions is typically regulated post-translationally, in particular through direct binding of regulatory metabolites\(^\text{18}\), rather than by changes in transcription factor expression itself\(^\text{16}\). In this case, a promoter’s specific transcriptional regulation can be described by the sum of metabolites \(M\) regulating the respective transcription factors, weighted by two condition-independent and promoter-specific parameters \(\alpha_i\) and \(\beta_{i,k}\) (denoting each metabolite’s impact on the respective transcription factor):

\[\text{specific regulation} \approx \sum_{i \in \text{TF}, k \in \text{MT}} \alpha_i \cdot \beta_{i,k} \cdot \Delta \log([M_{k,i}]) \]  
(Eq. 3)

We hypothesized that a promoter’s regulatory signals could therefore be identified by relating its specific transcriptional regulation component to a panel of potential regulatory metabolites. Inferring such regulatory signals from a promoter’s quantified specific transcriptional regulation can be computationally challenging because each promoter can, in principle, be regulated by several transcription factors, and in turn each transcription factor can receive more than one regulatory signal. Since not all transcription factors/regulatory signals are likely to affect a promoter to the same extent in a given set of conditions, we first wanted to systematically identify promoters whose specific transcriptional regulation component can be explained by one dominant regulatory signal. Given our focus on the regulation of central metabolic promoters, these regulatory signals are likely to be central metabolites as well. To test our hypothesis, we therefore quantified the concentration of 47 metabolites in central carbon metabolism during exponential growth in 23 of the 26 conditions by targeted metabolomics\(^\text{37}\) (supplementary figure 3). For each of the quantified metabolites, we tested whether it can explain the specific transcriptional regulation component of any of the tested promoters based on the model described in Eq. 3 (approach outlined in figure 3A). We used the Pearson correlation coefficient between measured specific transcriptional regulation and its reconstruction (based on Eq. 3) to assess the explanatory power of each metabolite. Non-parametric methods to assess the agreement between measured and reconstructed data, such as rank correlation, yielded highly similar results (supplementary figure 4).
Typically, only one or few single metabolites could potentially explain each promoter’s specific transcriptional regulation component (see examples in figure 3B and full description in supplementary figure S5). In cases where more than one metabolite was identified as a potential regulatory signal, the identified metabolites were strongly cross-correlated across conditions. For example, the six metabolites identified for *pykF*, namely F1P, FBP, DHAP, Ru5P, R5P, Xu5P, show a median cross-correlation of 0.83. Conversely, most metabolites could not explain the specific transcriptional regulation component of any of the promoters, and few metabolites were inferred to regulate more than one promoter. For example, F1P and FBP were identified as potential regulatory signals for several glycolytic promoters, and cyclic AMP was identified as a potential regulatory signal for a third of the tested promoters, in particular in TCA cycle and carbon uptake systems (supplementary figure 5A). Notably, repeating the analysis without removing the contribution of global transcriptional regulation yielded very little overlap in terms of identified potential regulatory metabolites (supplementary figure 5B).

In some cases a promoter’s specific transcriptional regulation component could not be explained by any of the tested metabolites alone. One possible explanation for such a result would be that the respective promoter receives more than one regulatory signal. In the second step, we therefore sought to systematically identify promoters whose specific transcriptional regulation component can be explained by a pairwise combination of regulatory signals. Towards this end, we identified for each measured promoter the best single metabolite and tested whether addition of any of the remaining metabolites can better explain its specific transcriptional regulation component. We assessed the statistical significance of the improvement by boot-strapping (see example in figure 4A-D). Notably, only four promoters could be explained significantly better by two metabolites (figure 4E). For example, the promoter of *Crp* could be best explained by F1P and G6P (figure 4A-D), and the promoter of *GlpD* could be well explained by combinations of glycerol-3-phosphate with several metabolites in glycolysis, as well as cyclic AMP. Thus, these results suggest that central metabolic promoters are mostly regulated by single dominating regulatory signals in the tested conditions.

By combining the identified metabolite regulatory signals and the corresponding global transcriptional regulation, we could draw a quantitative picture of the transcriptional program governing *E.coli*’s central carbon metabolism at single-promoter resolution (figure 5A). We quantified the contribution of specific transcriptional regulation (of the single metabolite or metabolite pair best explaining the respective specific transcriptional regulation component) as the improvement in agreement between measured and predicted promoter activity over global transcriptional regulation alone. For the majority of promoters, the contribution of specific transcriptional regulation is small (figure 5A, black circles), suggesting that it merely modulates a dominant global regulatory input. Exceptions are uptake
systems and TCA cycle promoters, which tend to have a stronger contribution of specific transcription
regulation, and which are largely affected by cyclic AMP. Notably, only few promoters, such as the
promoters associated with gluconate utilization (idnD, idnK gntK, gntT), could not be explained by
either global or specific transcriptional regulation, presumably because we could not quantify the
underlying metabolic signal. Moreover, few metabolites, namely cyclic AMP, F1P and FBP, explain the
majority of observed specific transcriptional regulation. Taken together, these results demonstrate
that the transcriptional program of E.coli’s central carbon metabolism is driven by few regulatory
signals in response nutrient variations or perturbations of the expression machinery itself.

The unbiased mathematical approach described above allows us to systematically identify potential
metabolic regulatory signals which affect a promoter’s specific transcriptional regulation component,
and thus provides a quantitative map linking metabolic regulatory signals to promoters. Since this
approach does not require information about the underlying regulatory network, it is applicable to
any, even poorly characterized, organism.

Relating metabolic regulatory signals to transcription factors. Finally, we asked how the inferred links
between regulatory signals and promoters are established mechanistically, i.e. which transcription
factors ultimately mediate this link. Towards this end, we used the known transcriptional regulatory
network of E.coli’s central carbon metabolism as reported in RegulonDB, which comprises over 30
different transcription factors, and determined the overlap between each metabolite’s target
promoters (as shown in figure 5A) and the frequency with which these promoters were regulated by a
given transcription factor (figure 5B). Reassuringly, this analysis correctly predicted well known
interactions between cyclic AMP and the transcription factor Crp (activation), as well as between the
two metabolites FBP and F1P and the transcription factor Cra (inhibition). Besides these previously
known transcription factor-metabolite interactions, our analysis also predicted a novel interaction,
namely between Crp and PEP. Notably, a considerable fraction of the promoters predicted to be
regulated by cyclic AMP had not been reported to be targets of Crp, or showed a discrepancy in
interaction sign. For example, our analysis predicted activation of talA and pck promoters by Crp-
cAMP, in contrast to previous reports. We could confirm the activation of talA and pck by Crp-cAMP
through external supplementation of cyclic AMP, which led to an increase in promoter activity in the
wild-type, but not in a Crp knockout strain (supplementary figure S6A). Moreover, activation of these
promoters in carbon sources supporting slow growth required Crp (supplementary figure S6B). These
results demonstrate that the approach presented here not only systematically identifies in vivo
relevant regulatory signals of transcriptional regulatory networks, but also provides mechanistic
insights by relating these regulatory signals to the transcription factors ultimately exerting these
regulatory functions, provided the topology of the underlying transcriptional regulatory network is at least partially known.
Discussion

In this work, we aimed to unravel the transcriptional program governing central carbon metabolism for the model organism *Escherichia coli*. Starting from the measured activity of about 100 central metabolic promoters during exponential growth in various environmental conditions, we quantified the promoter- and condition-specific impact of global transcriptional regulation, which was the dominant regulatory input for the majority of promoters, in particular for glycolysis and pentose phosphate pathway promoters. Combining an approximate mathematical description of promoter activity with measurements of *E. coli*'s central carbon metabolome response, we further identified for each promoter those metabolites which could serve as potential regulatory signals for its specific transcriptional regulation component without the confounding influence of global transcriptional regulation, and linked these metabolites to the transcription factors which ultimately exert these regulatory functions. Thus, this work provides a quantitative picture of the transcriptional program, and in particular the interplay between global and specific transcriptional regulation, in *E.coli*'s central carbon metabolism at individual promoter level (figure 5A).

The emerging picture revealed a surprisingly simple transcriptional regulatory program of central carbon metabolism, in which global transcriptional regulation, together with few regulatory metabolites, was sufficient to explain the majority of changes in promoter activity across conditions. In particular, we found that global transcriptional regulation dominated the transcriptional response of most promoters, accounting for about 70% of the total variance across conditions. This finding is consistent with previous observations\(^{21,23,25}\) and may explain why transcriptional adaptation to environmental changes is typically accompanied by a large number of gene expression changes even for very closely related conditions\(^{5-9}\). Notably, the dominant contribution of global transcriptional regulation is not restricted to expression data obtained by fluorescent transcriptional reporters. For example, transcriptomics studies of *S. cerevisiae* in various nutrient limitation experiments have shown that the expression of a large fraction of genes strongly depends on the cell's growth rate regardless of the exact type of limitation\(^{40}\). While careful quantitative studies have provided a fairly good understanding of the molecular mechanisms underlying global transcriptional regulation in bacteria\(^{27,41}\), its molecular underpinnings in eukaryotes are less clear\(^{42}\), calling for further mechanistic studies to reveal the interplay of global and specific transcriptional regulation at the molecular level.

In addition to global transcriptional regulation, few metabolites emerged as dominant specific regulatory signals in *E.coli*'s central carbon metabolism. In particular, three metabolites, namely cyclic AMP, FBP, and F1P, were identified as dominant regulatory signals for the specific transcriptional regulation of glycolysis (FBP/F1P through the transcription factor Cra), as well as TCA cycle and the utilization of carbon sources (cyclic AMP through the transcription factor Crp). These findings are in
good agreement with previous studies, which have highlighted the importance of Cra for regulating the switch between glycolysis and gluconeogenesis\textsuperscript{43,44}, as well as for the sensing of glycolytic flux\textsuperscript{45,46}. Similarly, our work confirms previous studies which have demonstrated the importance of Crp for regulating TCA cycle\textsuperscript{47,48} and carbon utilization\textsuperscript{49,50}. Notably, F1P and FBP affected different sets of promoters more strongly through the same transcription factor Cra. This difference in effector specificity seems to be largely encoded in the respective Cra binding site, since addition of a Cra binding site from a F1P-regulated promoter rendered a synthetic promoter more specific for F1P, and vice versa (figure 5A). Although protein structural information about Cra is limited, these results suggest that binding of F1P and FBP may trigger distinct conformational changes in Cra.

Interestingly, we also found that the effect of Cra on glycolytic promoters was rather weak and only modulated the dominant global transcriptional regulation. What could be the physiological relevance of such a modulating regulatory signal? One attractive hypothesis emerges when considering the final output of gene expression, namely the protein concentration (figure S7). In case of promoters which are solely subject to global transcriptional regulation, variation of carbon source availability yields a negative relationship between growth rate and protein concentration. Consequently, adjusting a protein’s concentration through global regulation alone will lead to disproportionally high concentrations at slow growth if it is required at high concentration during fast growth (figure S7B, upper panel). Since glycolytic carbon sources tend to support faster growth, and result in higher FBP concentrations, regulation through Cra-FBP (i.e. repression by Cra which is alleviated by FBP) may counter this effect, causing more constant protein concentrations across different growth rates (figure S7B, middle panel). Conversely, the regulatory input of Crp-cAMP yields a previously described linear negative relationship between protein concentration and growth rate in carbon limitation\textsuperscript{51} (figure S7B, lower panel). Thus, few such regulatory signals may allow cells to coarsely allocate proteome resources based on the supported growth rate.

Besides these three major regulatory signals, our analysis also predicted several minor regulatory signals. Most notably, our analysis predicted a second, novel, regulatory signal, namely PEP, which was identified to interact with Crp. The fact that PEP and the main Crp-regulator cyclic AMP showed moderate cross-correlation across conditions (R=0.55) suggests that PEP may serve as a complementary, modulating, signal to regulate Crp activity in conditions where intracellular PEP concentrations were found to change dramatically, such as carbon starvation\textsuperscript{52}. Future biochemical efforts may allow to test this potential novel interaction experimentally.

The simple transcriptional program identified in this study suggests that \textit{E.coli} only uses a small fraction of its transcriptional regulatory network in a given environment, which is consistent with recent observations\textsuperscript{21,25}. How can this finding be reconciled with the reported dense transcriptional regulatory
network of *E. coli*’s central carbon metabolism, which comprises over 30 additional transcriptional factors\(^{15}\). Firstly, methods to physically map transcriptional regulatory networks, such as ChIP-chip\(^{14}\), ChIP-seq\(^{53}\), or SELEX\(^{38,54}\), typically provide binding information without directly assessing the identified transcription factor-gene interactions in different conditions. In contrast, our approach relies on detecting (metabolite-dependent) changes in transcription factor activity across conditions, instead of transcription factor binding as such. Consequently, if a transcription factor binding a promoter does not change its activity across conditions, we cannot capture its effect. Therefore, we envision that combining methods to physically map transcription factor-promoter interactions with the approach presented here can be used to find out not only which transcription factor-promoter interactions are present in an organism, but also which of these are active in a given set of conditions. Secondly, in this study we mostly focused on the transcriptional response to changes in nutrient availability (i.e. carbon supply), together with perturbations affecting the activity of the expression machinery itself. Clearly, different types of perturbations may activate other parts of the transcriptional regulatory network. We did test a number of additional perturbations, which are known to directly affect metabolism, such as oxidative stress, and found that these perturbations yielded largely comparable expression patterns, but obviously no selection of conditions is going to be exhaustive. Future studies may identify additional metabolic regulatory signals for central carbon metabolism which are relevant in other subsets of conditions.

In this proof-of-concept study, we focused on *E. coli*’s central carbon metabolism, which is already well characterized in terms of its transcriptional regulatory program as well as its metabolite regulatory signals\(^4\). However, the approach presented here can be easily extended to other cellular networks and organisms, as long as concurrent quantification of gene expression and metabolome is available. Given the recent advances in metabolomics and quantification of gene expression, such data is likely to be more readily available in the future (see appendix for an exemplary analysis of transcriptomics and metabolomics data for *S. cerevisiae*). This work may serve as a template for the data-driven systematic identification of novel metabolite regulatory signals and ultimately transcription factor-metabolite interactions.

**Acknowledgements**

We gratefully acknowledge Elad Noor and Terence Hwa for helpful discussions.
Material and Methods

Reagents and strains. Unless stated otherwise, all reagents were obtained from Sigma-Aldrich. Fluorescent transcriptional reporter plasmids were directly obtained from24,34 or constructed as described in the original study34, and subsequently transformed into the E.coli wild-type strain BW2511355. The Crp deletion strain was obtained from56 and cured from its antibiotic resistance as described previously57.

Cultivation. All experiments were performed using M9 minimal medium (see supplementary table 1 for full list of conditions). Cultivations for the quantification of promoter activity were performed as described previously24. Briefly, M9 medium batch cultures in 96 deep-well format plates (Kuehner AG, Birsfeld, Switzerland) were inoculated 1:50 from LB precultures and incubated overnight at 37°C under shaking. Subsequently, 96-well flat transparent plates (Nunc, Roskilde, Denmark) containing M9 medium (fill volume 200 ml) were inoculated 1:200 with overnight cultures and sealed with parafilm to reduce evaporation. On-line measurements of optical density at 600 nm (OD600) and fluorescence (excitation wavelength: 500 nm, emission wavelength: 530 nm) were performed at 37°C with shaking using a plate reader (TECAN infinite M200, Tecan Group Ltd, Männedorf, Switzerland) at 10 min intervals. Cultivations for the quantification of intracellular metabolite concentrations were performed as follows: M9 medium batch cultures in 96 deep-well format plates (Kuehner AG, Birsfeld, Switzerland) were inoculated 1:50 from LB precultures and incubated overnight at 37°C under shaking. Subsequently, 96 deep-well plate cultures were inoculated with overnight cultures to a starting OD600 of 0.03 (total fill volume per well: 1.2 ml) and incubated at 37°C under shaking. Culture OD600’s were monitored by OD600 sampling from parallel wells on the same deep-well plate and subsequent OD600 measurements using a plate reader (TECAN infinite M200, Tecan Group Ltd, Männedorf, Switzerland).

Quantification of intracellular metabolite concentrations. Metabolomics samples were taken during mid-exponential phase at OD’ s between 0.5 and 0.7 by fast filtration (sampling volume: 1 ml)58, and were immediately quenched in 4 ml quenching/extraction solution (40% methanol, 40% acetonitrile, 20% H2O) at -20°C59. To normalize for variations in sample processing, 100 mueL of a fully 13C-labeled yeast internal metabolome extract were added. Samples were incubated for 2h at -20°C, subsequently dried completely at 120 μbar (Christ RVC 2-33 CD centrifuge and Christ Alpha 2-4 CD freeze dryer) and stored at −80°C until measurements. Before measurements, samples were resuspended in 100 mueL water, centrifuged for 5 min (5000g, 4°C) to remove residual particles, and transferred to V-bottomed 96 well sample plates (Thermo Fisher Scientific). Measurement, data acquisition, and data analysis was performed as described previously37,44. Briefly, separation of compounds was achieved by ion-pairing ultrahigh performance liquid chromatography (UPLC) using a Waters Acquity UPLC with a Waters Acquity T3 end-capped reverse phase column (dimensions, 150 mm × 2.1 mm × 1.8 μm; Waters
Corporation) and coupled to compound detection using a tandem mass spectrometer (Thermo TSQ Quantum Ultra triple quadrupole; Thermo Fisher Scientific). Data acquisition and peak integration was performed with in-house software. To determine the absolute concentration of metabolites, a 1:3 dilution series of a standard solution (containing more than 80 metabolites of the central carbon metabolism) with 13C internal standard was prepared and measured in parallel.

**Data processing.** All data processing steps were performed with custom MATLAB software. Promoter activities and corresponding growth rates were determined as described previously. Briefly, raw GFP and OD600 time courses from each well were corrected for blank GFP and OD600 before cell addition and smoothed using a moving average window with size 3. From these time courses, promoter activity and growth rate were quantified as $d\text{GFP}/(dt \times \text{OD})$ and $d\ln(\text{OD})/dt$ by two point finite difference numerical approximation. Promoter activity was corrected for fluorescence background by subtracting the corresponding signal of the promoter-less plasmid reporter strain p139. Promoter activity and growth rate values for steady-state growth were calculated as the average value in the time range visually identified as exponential phase. Promoters whose activities were below the detection threshold (>2x above activity of promoter-less strain) in all tested conditions were discarded for further analysis. 1-dimensional hierarchical clustering of z-score normalized promoter activity data was performed using the Pearson correlation coefficient as the distance metric between promoters (cutoff: 0.225).

**Dissecting global and specific transcriptional regulation.** Promoter activity data was log-transformed and z-score normalized, and singular value decomposition was used to determine the first singular vector, which explains most of the data set’s variability. The specific transcriptional regulation component of each promoter was quantified by subtracting this singular vector.

**Identification of regulatory metabolites.** Each promoter’s specific transcriptional regulation component was related to each quantified metabolite by linear regression based on the equation $s=p*\text{metabolite}$, where $x$ is the specific transcriptional regulation component and $p$ denotes the promoter- and metabolite specific parameter to be determined in the regression. The goodness of fit was determined as the Pearson correlation coefficient between specific transcriptional regulation component and the corresponding reconstruction based on the fitted parameter $p$ and the metabolite concentrations using the MATLAB function `corr`. Conditions in which a metabolite was used as a carbon source, or secreted by the cells, were omitted. Identification of pairwise regulatory metabolites was performed by systematic linear regression of each promoter’s top correlating single metabolite as well as a second metabolite based on the equation $s=p1*\text{metabolite1} + p2*\text{metabolite2}$. The goodness of fit was again determined as the Pearson correlation coefficient between measured and reconstructed specific transcriptional regulation. The improvement in correlation coefficient by adding a second
metabolite was assessed by randomly permuting the second metabolite 5000 times using the MATLAB function `randperm` and calculating the fraction of correlation coefficients with randomly permuted metabolite data which are larger than the original correlation coefficient.
Chapter 4 – Few regulatory metabolites coordinate expression of central metabolic genes in E. coli

Figures

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Figures" /></td>
<td><img src="image" alt="Figures" /></td>
</tr>
</tbody>
</table>

**Figure 1. Steady state promoter activity of central carbon metabolic genes across 26 conditions shows few expression patterns.**

**A)** Steady state promoter activities in various carbon sources as well as different sub-lethal doses of chloramphenicol (full list of conditions given in supplementary table 1). Promoter activities were normalized by z-score normalization and sorted by 1-dimensional hierarchical clustering across promoters (clustering metric: Pearson correlation coefficient, cut-off 0.225). Carbon sources were sorted by increasing growth rate (from left to right), Chloramphenicol data were sorted by increasing chloramphenicol concentrations (from left to right). Last column: M9 glucose with 5 mM cyclic AMP (cAMP).

**B)** Selected clusters for each condition plotted against their respective steady state growth rate. Large filled circles: mean promoter activities of all promoters in respective cluster. Small non-filled circles: promoter activities of all individual promoters in respective cluster.
Figure 2. Singular value decomposition quantifies the contribution of global transcriptional regulation to promoter activity. A) Singular vector 1 (SV1, explaining 68% of the total variance) plotted against the respective growth rate. Black line: second-degree polynomial fit used for predictions in C), dashed lines: 95% confidence interval of fit. B) Pearson correlation between measured and reconstructed promoter activities across all promoters. Reconstruction was performed using only singular vector 1. Promoters were sorted according to their correlation coefficients. The majority of promoters are sufficiently explained by singular vector 1. C) Prediction of promoter activity during diauxic shift from glucose to succinate based on dynamic growth rate measurements for the 18 promoters with the strongest input from global transcriptional regulation as determined in B). M9 minimal medium cultures containing 0.5 g/L glucose and 5 g/L succinate were inoculated from M9 glucose pre-cultures. Black line: measured promoter activity in linear scale. Data show the mean of three biological replicates, and the grey shaded areas indicate the standard deviation across replicates for each time point. Red line: prediction of promoter activity based on measured dynamic growth rates. Dynamic SV1 values were calculated from growth rates using the polynomial fit in A), and then reverted back to the original linear scale. Red shaded areas show the corresponding 95% confidence intervals. Pearson correlation coefficients between measured and predicted promoter activity time courses of each promoter are shown in brackets.
Figure 3. Systematic identification of metabolites affecting specific transcriptional regulation. A) Outline of approach using pykF as an example: first, global regulation is removed from each promoter. Its remaining specific regulation is then related to each metabolite separately by linear regression to identify potential metabolic signals (goodness of fit assessed by Pearson correlation between measured data and the prediction based on its reconstruction). B) Summary of analysis for six promoters. Full circles: correlation coefficients when correlating each promoter’s specific regulation with each metabolite across all conditions. Empty circles: correlation coefficients when using normalized promoter activity instead (combined output of global and specific regulation). Red dashed lines mark correlation coefficient cut-offs of -0.75 and 0.75, respectively. Marked with red circles and bold names: reported regulatory metabolites (e.g. FBP and F1P, reported regulators of pykF through the transcription factor Cra). If metabolites pass the correlation coefficient cut-offs, their respective correlation coefficients without removal of global transcriptional regulation are shown in empty circles.
Figure 4. Identification of pairwise metabolic regulatory signals affecting the specific transcriptional regulation of promoters. A-D) Outline of approach with example promoter Crp. A) Predicted and measured specific regulation for best single metabolite (G6P), with correlation coefficient R in brackets. Blue circles: carbon sources. Red circles: chloramphenicol conditions. B) Correlation coefficients of all metabolite pairs of G6P with any of the remaining metabolites. To assess whether adding a second metabolite leads to a significantly improved prediction, each added metabolite was randomly permuted (5000 times), and the resulting correlation coefficient was calculated. The permutation p-value was calculated as the fraction of random permutations with a higher correlation coefficient than the original data, with a cut-off of <0.005. In the example, the metabolite pair with the highest correlation coefficient is G6P-F1P. C) Predicted versus measured specific regulation for best metabolite pair. Black line: best fit. p1 and p2 denote the respective fitting parameters for G6P and F1P, respectively. D) Distribution of correlation coefficients for random permutations for the best metabolite-pair (black line), together with the original correlation coefficient (red line). E) Summary of all pairwise tests. Promoters were sorted according to metabolic pathways, and the color code indicates the improvement (in percent) of the correlation coefficient compared to the best single metabolite. Only metabolite pairs with a correlation coefficient >0.75 and a permutation p-value < 0.005 were considered. For each promoter, the best metabolite pair is marked with thick black borders. Right panel: name of best single metabolite. Only those promoters, in which either the best single metabolite or the best metabolite pair had a correlation coefficient > 0.75, were considered.
Figure 5. Inferred global and specific transcriptional regulatory program of E.coli’s central carbon metabolism. A) Promoters were sorted according to metabolic pathways. Right panel: contribution of global (grey circles) and specific (black circles) transcriptional regulation across conditions for each promoter. Grey circle sizes denote the influence of global transcriptional regulation on the respective promoter as determined in figure 2B. Black circle sizes denote the influence of specific transcriptional regulation on the respective promoter. Left panel: metabolite regulatory signals as determined in figures 3 and 4. Activating and inhibiting interactions are shown in green and red, respectively. Previously reported interactions are shown with thick black edges. Only metabolites with at least one potential promoter target are shown. Number in brackets: number of potential promoter targets for respective metabolite. B) Inference of potential transcription factor-metabolite interactions based on the reported transcriptional regulatory network (as reported in RegulonDB15). For each metabolite with more than one target promoters, its target overlap which each of the transcription factors regulating central carbon metabolism was calculated (100% means that all of the target promoters of a metabolite are regulated by the respective transcription factor). For each metabolite, the transcription factor with the largest overlap in target is shown with a thick black edge. To account for
differences in the number of target promoters between transcription factors, the enrichment of transcription factor targets among each metabolite’s targets was assessed by hypergeometric testing. All four top transcription factor-metabolite interactions showed significant enrichment (p-value <0.05).
Supplementary figures

**Figure S1. GFP expression does not impair growth rate.** Promoter activities of all 95 tested promoter strains in all 26 tested conditions plotted against the growth rate of the respective strain (relative to the mean growth rate in the same condition).

**Figure S2. Day-to-day reproducibility of promoter activity measurements.** Promoter activity during exponential growth (condition: M9 glucose 2g/L) was determined in five independent experiments. Median day-to-day variation is below 15%.
Figure S3. Intracellular concentrations of central carbon metabolites during exponential growth.

Absolute metabolite concentrations of 47 central carbon metabolism metabolites in 23 conditions (log transformed and normalized by mean metabolite concentration across conditions) as quantified by targeted LC/LC mass spectrometry. Metabolites which were present in the medium were omitted from the quantification. For metabolites marked with (*), no absolute quantification was available. Their concentration was quantified relative to M9 glucose, and then log transformed.
Chapter 4 – Few regulatory metabolites coordinate expression of central metabolic genes in E. coli

Figure S4. Comparison of linear (Pearson) and rank (Spearman) correlation as metrics to identify promoter-metabolite interactions. The ability of each metabolite to explain each promoter’s specific transcriptional regulation component (see figure 3A) was assessed using Pearson and Spearman correlation coefficients, respectively. Both metrics yield highly comparable results.

Figure S5. Identification of promoter-metabolite interactions in absence or presence of global transcriptional regulation. A) Heatmap of Pearson correlation coefficients between measured specific transcriptional regulation (after removal of the respective global regulation component) and its reconstruction based on one metabolite (as described in figure 3A) for all promoter-metabolite pairs. Pairs with correlation coefficients above 0.75 or below -0.75 are shown with thick white edges. B)
Chapter 4 – Few regulatory metabolites coordinate expression of central metabolic genes in E. coli

Heatmap of Pearson correlation coefficients between log normalized promoter activity (consisting of global and specific transcriptional regulation) and its reconstruction based on one metabolite (as described in figure 3A) for all promoter-metabolite pairs. Pairs with correlation coefficients above 0.75 or below -0.75 are shown with thick white edges.

Figure S6. Promoters talA and pck are activated by Crp. A) Promoter activity of talA, pck, and acs, during exponential growth in M9 glucose 2g/L with varying concentrations of externally supplemented cyclic AMP, which activates the transcription factor Crp. Red: wildtype strain, blue: Crp deletion strain. B) Promoter activity of talA, pck, and acs, during exponential growth on M9 minimal medium supplemented with 12 different single carbon sources (galactose, acetate, pyruvate, succinate, mannose, glycerol, fructose, gluconate, glucose, lactate, GlcNAC, G6P). Red: wildtype strain, blue: Crp deletion strain. Note that the Crp deletion strain shows impaired growth in all conditions.
Figure S7. Relationship between promoter activity and protein concentration. A) Promoter activities of promoters only affected by global transcriptional regulation (top panel), mostly by global transcriptional regulation with modulating specific input from fructose-1,6-bisphosphate (FBP, middle panel), and promoters with dominating specific transcriptional regulation through cyclic AMP (cAMP, lower panel). Grey circles: isolated global transcriptional regulation component of promoter activity. B) Corresponding GFP concentrations as calculated from promoter activities (GFP concentration = promoter activity divided by growth rate). Grey circles: isolated global transcriptional regulation component of GFP concentration.
### Supplementary Tables

**Table S1. Conditions used in this study.**

<table>
<thead>
<tr>
<th>Condition</th>
<th>mean growth rate $[\text{h}^{-1}]$</th>
<th>standard deviation growth rate $[\text{h}^{-1}]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M9 minimal medium with 5g/L glucose</td>
<td>0.64</td>
<td>0.015</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L pyruvate</td>
<td>0.40</td>
<td>0.008</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L gluconate</td>
<td>0.63</td>
<td>0.013</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L acetate</td>
<td>0.32</td>
<td>0.009</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L glycerol</td>
<td>0.48</td>
<td>0.006</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L fructose</td>
<td>0.53</td>
<td>0.009</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L succinate</td>
<td>0.53</td>
<td>0.009</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L galactose</td>
<td>0.18</td>
<td>0.006</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L glucose and 2g/L casamino acids</td>
<td>1.11</td>
<td>0.014</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L mannose</td>
<td>0.41</td>
<td>0.014</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L glucose-6-phosphate (G6P)</td>
<td>0.81</td>
<td>0.028</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L lactate</td>
<td>0.38</td>
<td>0.012</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L N-acetyl-glucosamin (glcNAc)</td>
<td>0.61</td>
<td>0.031</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L gluconate and all 20 amino acids added as described in$^{60}$</td>
<td>1.49</td>
<td>0.046</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L fructose and all 20 amino acids added as described in$^{60}$</td>
<td>1.33</td>
<td>0.043</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L sorbitol</td>
<td>0.49</td>
<td>0.014</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L mannitol</td>
<td>0.57</td>
<td>0.014</td>
</tr>
<tr>
<td>M9 minimal medium with 2g/L glucose without chloramphenicol</td>
<td>0.67</td>
<td>0.014</td>
</tr>
<tr>
<td>M9 minimal medium with 2g/L glucose with 0.75 $\mu$M chloramphenicol</td>
<td>0.62</td>
<td>0.011</td>
</tr>
<tr>
<td>M9 minimal medium with 2g/L glucose and 1 $\mu$M chloramphenicol</td>
<td>0.45</td>
<td>0.007</td>
</tr>
<tr>
<td>M9 minimal medium with 2g/L glucose and 2 $\mu$M chloramphenicol</td>
<td>0.33</td>
<td>0.010</td>
</tr>
<tr>
<td>M9 minimal medium with 2g/L glucose and 4 $\mu$M chloramphenicol</td>
<td>0.23</td>
<td>0.009</td>
</tr>
<tr>
<td>M9 minimal medium with 2g/L glucose and 6 $\mu$M chloramphenicol</td>
<td>0.17</td>
<td>0.006</td>
</tr>
<tr>
<td>M9 minimal medium with 2g/L glucose and 8 $\mu$M chloramphenicol</td>
<td>0.15</td>
<td>0.004</td>
</tr>
<tr>
<td>M9 minimal medium with 2g/L glucose and 10 $\mu$M chloramphenicol</td>
<td>0.09</td>
<td>0.004</td>
</tr>
<tr>
<td>M9 minimal medium with 5g/L glucose and 5 mM cyclic AMP</td>
<td>0.44</td>
<td>0.025</td>
</tr>
</tbody>
</table>
Table S2. Promoters used in this study.

<table>
<thead>
<tr>
<th>Promoter name</th>
<th>Promoter description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>p139</td>
<td>promoter-less plasmid</td>
<td></td>
</tr>
<tr>
<td>eda1</td>
<td>multifunctional 2-keto-3-deoxygluconate 6-phosphate aldolase and 2-keto-4-hydroxylutarate aldolase and oxaloacetate decarboxylase</td>
<td>this study</td>
</tr>
<tr>
<td>idnD</td>
<td>L-Idonate 5-dehydrogenase</td>
<td>this study</td>
</tr>
<tr>
<td>ackA</td>
<td>acetate kinase A (propionate kinase 2)</td>
<td>34</td>
</tr>
<tr>
<td>gltA</td>
<td>citrate synthase</td>
<td>34</td>
</tr>
<tr>
<td>mmo</td>
<td>malate dehydrogenase (NADP-dependent)</td>
<td>34</td>
</tr>
<tr>
<td>gnd</td>
<td>gluconate-6-phosphate dehydrogenase, decarboxylating (1st module)</td>
<td>34</td>
</tr>
<tr>
<td>tktA</td>
<td>transketolase 1 thiamin-binding, isozyme</td>
<td>34</td>
</tr>
<tr>
<td>dcuB</td>
<td>dicarboxylate transporter DcuB</td>
<td>this study</td>
</tr>
<tr>
<td>aceB</td>
<td>malate synthase A</td>
<td>34</td>
</tr>
<tr>
<td>mdh</td>
<td>malate dehydrogenase (NAD-dependent)</td>
<td>this study</td>
</tr>
<tr>
<td>rrnB</td>
<td>16S rRNA</td>
<td>34</td>
</tr>
<tr>
<td>pykA</td>
<td>pyruvate kinase II</td>
<td>this study</td>
</tr>
<tr>
<td>edd</td>
<td>6-phosphogluconate dehydratase</td>
<td>34</td>
</tr>
<tr>
<td>epd</td>
<td>erythrose 4-phosphate dehydrogenase</td>
<td>24</td>
</tr>
<tr>
<td>acnB</td>
<td>aconitate hydratase 2 (2nd module)</td>
<td>34</td>
</tr>
<tr>
<td>gntK</td>
<td>gluconate kinase 2 in GNT I system, thermoresistant</td>
<td>34</td>
</tr>
<tr>
<td>pck</td>
<td>phosphoenolpyruvate carboxykinase</td>
<td>34</td>
</tr>
<tr>
<td>pgi</td>
<td>Phospho glucose isomerase</td>
<td>34</td>
</tr>
<tr>
<td>tpiA</td>
<td>triosephosphate isomerase</td>
<td>34</td>
</tr>
<tr>
<td>ppc</td>
<td>PEP carboxlase</td>
<td>this study</td>
</tr>
<tr>
<td>acs</td>
<td>acetyl-CoA synthase</td>
<td>this study</td>
</tr>
<tr>
<td>ptsH</td>
<td>histidine protein (component of PTS)</td>
<td>this study</td>
</tr>
<tr>
<td>rrnD</td>
<td>16S rRNA</td>
<td>34</td>
</tr>
<tr>
<td>pfkB</td>
<td>6-phosphofructokinase II</td>
<td>34</td>
</tr>
<tr>
<td>ppsA</td>
<td>PEP synthetase</td>
<td>this study</td>
</tr>
<tr>
<td>epd+Crp-Cra</td>
<td>synthetic epd promoter with scrambled Cra binding site (=Crp-reporter)</td>
<td>this study</td>
</tr>
<tr>
<td>eutS</td>
<td>putative carboxysome structural protein, ethanol utilization</td>
<td>34</td>
</tr>
<tr>
<td>gntP</td>
<td>GntP family, gluconate transport protein, GNT III system (1st module)</td>
<td>34</td>
</tr>
<tr>
<td>pdhR</td>
<td>transcriptional repressor for pyruvate dehydrogenase complex (GntR family) (1st module)</td>
<td>34</td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>pgml</td>
<td>phosphoglycerate mutase III, cofactor independent (3rd module)</td>
<td></td>
</tr>
<tr>
<td>ybhA</td>
<td>putative phosphatase</td>
<td></td>
</tr>
<tr>
<td>fbp</td>
<td>fructose-1,6-bisphosphatase I</td>
<td></td>
</tr>
<tr>
<td>fumA</td>
<td>fumarase A</td>
<td></td>
</tr>
<tr>
<td>yggF</td>
<td>fructose-1,6-bisphosphatase II</td>
<td></td>
</tr>
<tr>
<td>aceE</td>
<td>pyruvate dehydrogenase (subunit of E1p component)</td>
<td></td>
</tr>
<tr>
<td>fbaB</td>
<td>fructose bisphosphate aldolase I</td>
<td></td>
</tr>
<tr>
<td>gapA</td>
<td>glyceraldehyde-3-phosphate dehydrogenase</td>
<td></td>
</tr>
<tr>
<td>zwf</td>
<td>glucose-6-phosphate dehydrogenase</td>
<td></td>
</tr>
<tr>
<td>fruB</td>
<td>Sugar Specific PTS family, fructose-specific enzyme IIA/FPr component (1st module)</td>
<td></td>
</tr>
<tr>
<td>gntR</td>
<td>transcriptional repressor for gluconate utilization (GalR/LacI family)</td>
<td></td>
</tr>
<tr>
<td>poxB</td>
<td>pyruvate dehydrogenase/oxidase FAD and thiamine PPi binding, cytoplasmic in absence of cofactors (1st module)</td>
<td></td>
</tr>
<tr>
<td>pyrG</td>
<td>CTP synthetase</td>
<td></td>
</tr>
<tr>
<td>ybhE</td>
<td>putative isomerase</td>
<td></td>
</tr>
<tr>
<td>kbl-const</td>
<td>kbl promoter without Lrp binding site (kbl-constitutive)</td>
<td></td>
</tr>
<tr>
<td>fumC</td>
<td>fumarase C</td>
<td></td>
</tr>
<tr>
<td>aroK</td>
<td>shikimate kinase I</td>
<td></td>
</tr>
<tr>
<td>fur</td>
<td>transcriptional repressor of iron transport (Fur family)</td>
<td></td>
</tr>
<tr>
<td>crp</td>
<td>catabolite activator protein (CAP), cyclic AMP receptor protein (CAMP-binding family)</td>
<td></td>
</tr>
<tr>
<td>pfkA</td>
<td>6-phosphofructokinase I</td>
<td></td>
</tr>
<tr>
<td>epdlcd+Cra(epd)</td>
<td>synthetic epd promoter, with -35 to -10 box from icd promoter and Cra binding site from epd (=Cra-reporter 1)</td>
<td></td>
</tr>
<tr>
<td>fumB</td>
<td>fumarase B (fumarate hydratase class I), anaerobic isozyme (1st module)</td>
<td></td>
</tr>
<tr>
<td>gntT</td>
<td>GntP family, high-affinity gluconate permease in GNT I system</td>
<td></td>
</tr>
<tr>
<td>ptsG</td>
<td>multimodular PtsG: PTS family enzyme IIC, glucose-specific (1st module)</td>
<td></td>
</tr>
<tr>
<td>rpe</td>
<td>D-ribulose-5-phosphate 3-epimerase</td>
<td></td>
</tr>
<tr>
<td>acnA</td>
<td>aconitate hydratase 1</td>
<td></td>
</tr>
<tr>
<td>zwf-const</td>
<td>native zwf promoter without transcription factor binding sites (=zwf-constitutive)</td>
<td></td>
</tr>
<tr>
<td>ptsA</td>
<td>General PTS family, enzyme I, phosphohistidine domain (1st module)</td>
<td></td>
</tr>
<tr>
<td>talA</td>
<td>transaldolase A (2nd module)</td>
<td></td>
</tr>
<tr>
<td>galR</td>
<td>transcriptional repressor for galactose utilization (GalR/LacI family)</td>
<td></td>
</tr>
<tr>
<td>idnR</td>
<td>IdnR transcriptional regulator</td>
<td></td>
</tr>
<tr>
<td>eno2</td>
<td>enolase</td>
<td></td>
</tr>
</tbody>
</table>
### Regulatory Metabolites and Their Functions

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Function and Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>epdIcd+Cra(pykF)</td>
<td>Synthetic epd promoter, with -35 to -10 box from icd promoter and Cra binding site from epd (Cra-reporter 2)</td>
</tr>
<tr>
<td>galE</td>
<td>UDP-galactose 4-epimerase (1st module)</td>
</tr>
<tr>
<td>icd</td>
<td>Isocitrate dehydrogenase in e14 prophage, specific for NADP+ (2nd module)</td>
</tr>
<tr>
<td>sdhC</td>
<td>Succinate dehydrogenase, cytochrome b556</td>
</tr>
<tr>
<td>rpiA</td>
<td>Ribosephosphate isomerase, constitutive</td>
</tr>
<tr>
<td>glpD</td>
<td>Sn-glycerol-3-phosphate dehydrogenase FAD/NAD(P)-binding (aerobic)</td>
</tr>
<tr>
<td>pykF-const</td>
<td>PykF promoter with scrambled Cra binding site (pykF-constitutive)</td>
</tr>
<tr>
<td>glpF</td>
<td>Glycerol channel</td>
</tr>
<tr>
<td>dam</td>
<td>DNA adenine methylase</td>
</tr>
<tr>
<td>hupA</td>
<td>DNA-binding protein HU-alpha (HU-2)</td>
</tr>
<tr>
<td>gntR2</td>
<td>Transcriptional repressor for gluconate utilization (GalR/LacI family)</td>
</tr>
<tr>
<td>pykF</td>
<td>Pyruvate kinase I (formerly F), fructose stimulated (2nd module)</td>
</tr>
<tr>
<td>rpoD</td>
<td>Sigma D (sigma 70) factor of RNA polymerase, major sigma factor during exponential growth (2nd module)</td>
</tr>
<tr>
<td>galU</td>
<td>Glucose-1-phosphate uridylyltransferase</td>
</tr>
<tr>
<td>idnK</td>
<td>Gluconokinase II</td>
</tr>
<tr>
<td>rpoS</td>
<td>Sigma S (sigma 38) factor of RNA polymerase, major sigma factor during stationary phase</td>
</tr>
<tr>
<td>seqA</td>
<td>Negative modulator of initiation of replication, inhibits open complex formation, mutation in gene alters cell membrane</td>
</tr>
<tr>
<td>sdhD</td>
<td>Succinate:quinone oxidoreductase</td>
</tr>
<tr>
<td>epd-const</td>
<td>Epd promoter without Crp binding site and with scrambled Cra binding site (=epd-constitutive)</td>
</tr>
<tr>
<td>glpK</td>
<td>Glycerol kinase</td>
</tr>
<tr>
<td>dctA</td>
<td>DAACS family, C4-dicarboxylic acids transport protein</td>
</tr>
<tr>
<td>ihfb</td>
<td>IHF transcriptional dual regulator</td>
</tr>
<tr>
<td>phoB</td>
<td>Response regulator in two-component regulatory system with PhoR (or CreC), regulates Pi uptake (OmpR family) (1st module)</td>
</tr>
<tr>
<td>pgk</td>
<td>Phosphoglycerate kinase</td>
</tr>
<tr>
<td>rpoH</td>
<td>Sigma H (sigma 32) factor of RNA polymerase; transcription of heat shock proteins induced by cytoplasmic stress</td>
</tr>
<tr>
<td>glpA</td>
<td>Sn-glycerol-3-phosphate dehydrogenase FAD/NAD(P)-binding (anaerobic), large subunit (1st module)</td>
</tr>
<tr>
<td>lpd</td>
<td>Dihydrolipoamide dehydrogenase, FAD/NAD(P)-binding component of 2-oxodehydrogenase and pyruvate complexes</td>
</tr>
<tr>
<td>rpoE</td>
<td>Sigma E (sigma 24) factor of RNA polymerase, response to periplasmic stress (TetR/ArcR family)</td>
</tr>
<tr>
<td>talB</td>
<td>Transaldolase B (2nd module)</td>
</tr>
<tr>
<td>sucA</td>
<td>2-oxoglutarate decarboxylase, thiamine requiring</td>
</tr>
</tbody>
</table>

This list highlights the regulatory metabolites involved in the expression of central metabolic genes in E. coli. The synthetic epdIcd+Cra(pykF) promoter is used to coordinate the expression of these genes, with specific regulatory elements such as the -35 to -10 box from the icd promoter and the Cra binding site from the epd promoter. The expression of genes like galE, icd, sdhC, rpiA, glpD, pykF, and rpoD is crucial for the metabolic pathways in E. coli. The study also includes the use of genes like glpK and dctA for specific metabolic functions, and regulatory proteins such as ihfb, phoB, and rpoH for the transcriptional regulation of these genes.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>epdIcd</td>
<td>synthetic epd promoter, with -35 to -10 box from icd promoter and without Cra and Crp binding sites</td>
<td>24</td>
</tr>
<tr>
<td>gpmA</td>
<td>2,3-bisphosphoglycerate-dependent phosphoglycerate mutase</td>
<td>this study</td>
</tr>
<tr>
<td>cyaA</td>
<td>adenylate cyclase</td>
<td>34</td>
</tr>
<tr>
<td>kdgr</td>
<td>putative transcriptional repressor (IclR family)</td>
<td>34</td>
</tr>
</tbody>
</table>
Chapter 4 – Few regulatory metabolites coordinate expression of central metabolic genes in E. coli

Appendix

Identification of gene-metabolite interactions in S. cerevisiae. In this section, we use the approach developed in this work to identify potential gene-metabolite interactions in S.cerevisiae based on published metabolomics and transcriptomics data sets. The metabolomics data set comprises 106 mostly central carbon metabolites which were quantified in chemostat experiments using three types of natural nutrient limitations (carbon, nitrogen, phosphorus), as well as two artificial limitations (leucine and uracil limitation in leucine-auxotroph and uracil-auxotroph strains, respectively), with five growth rates per limitation. The transcriptomics data set comprises genome-wide transcript data for the same five limitations at six growth rates per limitation, and was performed using the same experimental setup as the metabolomics experiments.

In the main text, we developed an approximate model which describes promoter activity as a function of global and specific transcriptional regulation. Since the relationship between promoter activity and transcript abundance is not well understood (in S. cerevisiae), we use the assumption that the same relationship between transcript abundance and global/specific transcriptional regulation exists, even if the mechanistic details might differ:

\[ [mRNA]_{i,j} = \frac{[E^*_i/K_{E,i}]^{\alpha_{E,i}} \cdot \prod_{E(TF)} \left( \frac{T^*_E/K_{l,i+1}}{1} \right) \alpha_{l,i}}{1} \quad (Eq. A1) \]

Here, \( E^* \) denotes the activity of the expression machinery (in condition \( j \)) with its promoter-specific parameters \( K_{E,i} \) and \( \alpha_{E,i} \), and \( T^* \) denotes the activity of each specific transcription factor (in condition \( j \)) that regulates the respective gene with its gene-specific parameters \( K_i \) and \( \alpha_i \). Using singular value decomposition, we identified the global regulation component across all limitations (figure A-1A). As was already described in the respective study, the first singular value showed strong correlation with the growth rate across all limitations. However, in contrast to the promoter activity data in the main text, this global regulation component was less pronounced (capturing 43% of the total variance in the data set, compared to 68% of the variance being explained by global transcriptional regulation in E.coli promoter activity data).

Identifying single gene-metabolite interactions. Next, we wanted to systematically identify metabolites which could serve as potential regulators of gene expression. Towards this end, we removed the aforementioned global transcriptional regulation component and tested for each gene, whether its remaining specific transcriptional regulation can be explained by any of the quantified metabolites (following the approach described in the main text). Since metabolome and transcript data were not generated at exactly the same growth rates (but in the same range of growth rates between 0.05 and 0.3 h\(^{-1}\)), we interpolated the transcript data to exactly match the growth rate of the
metabolome data. We found that the specific transcriptional regulation component of most genes (>60%) could not be explained by any metabolite, and only few genes had more than one potential metabolite regulators (figure A-1B). Conversely, about 30% of the quantified metabolites could not explain the specific transcriptional regulation component of any of genes, and few metabolites emerged as potential regulators of more than a hundred genes (figure A-1C). For example, the nucleotide biosynthesis intermediates 5-phospho-α-D-ribose 1-diphosphate (PRPP) and cytidine were identified as potential regulators for about 300 genes. Notably, in contrast to the analysis of E. coli promoter activity, the effect of removing global transcriptional regulation on the identified gene-metabolite interactions was less pronounced (figure A-1D).

**Linking metabolite regulators to metabolic pathways.** We next wanted to identify potential regulatory links between metabolites and metabolic pathways. Towards this end, we tested for each metabolite whether its target genes are enriched for genes to belonging to a specific metabolic pathway by pathway enrichment (figure A-2A). To categorize genes into specific metabolic pathways, we relied on a previously published genome-scale metabolic reconstruction comprising 904 genes\(^ {62}\). We found that the vast majority of metabolites were not associated to any particular metabolic pathway, with few notable exceptions. For example, our analysis identified a connection between asparagine and the expression of its own metabolic pathway. More strikingly, our analysis also revealed a strong activating link between cytidine and glycolysis/gluconeogenesis (see figure A2B for two examples and figure A-3 for full description). Cytidine was found to be most strongly affected by phosphorus limitation (figure A-1E), which causes cytidine accumulation presumably due to nucleotide phosphate salvage. Previous studies have shown that glycolysis and phosphate metabolism are tightly linked at the metabolic activity level\(^ {63}\), as well as by phosphate dependent activation of glucose PKA signaling\(^ {64}\). Our analysis suggests that this link extends to the transcriptional level as well, and the results suggest that this link may be mediated by cytidine, or related metabolites showing highly similar profiles that were not quantified in this study. A second, inhibiting, link was identified between oxidative phosphorylation and the metabolite histidinol, an intermediate of histidine biosynthesis. So far, there is very little evidence linking histidine biosynthesis and oxidative phosphorylation, and the reactions producing and consuming the metabolite are both located in the cytosol. Nevertheless, since histidinol is consumed by a NAD+ dependent oxidoreductase, it may serve as a potential indirect signal of cytosolic NAD+ availability.

**Additional remarks.** Here, we used the approach developed in this study to identify potential gene-metabolite interactions in S. cerevisiae based on published transcriptomics and metabolomics data in different nutrient limitations. Our analysis suggests that many of the findings regarding promoter-metabolite interactions in E. coli apply to S. cerevisiae as well. For example, in both organisms the
identified interactome is rather scarce: few metabolites dominate as inputs of specific transcriptional regulation (even if the identified dominant metabolites differ), and the expression of most genes cannot be explained by any metabolite. However, there are also some marked differences. For example, the influence of global transcriptional regulation on transcript abundances in *S. cerevisiae* seems to be less pronounced than on promoter activity in *E. coli*. Given that promoter activity measurements (which measure the expression of fluorescent reporter proteins) in *S. cerevisiae* have been found to be strongly affected by growth-dependent global regulation, these results suggest a considerable influence of global posttranscriptional regulation. Our analysis revealed a number of novel gene-metabolite links in *S. cerevisiae*, for example between glycolysis/gluconeogenesis and the nucleotide biosynthesis intermediate cytidine, providing a possible additional link between glycolysis, phosphate and nucleotide metabolism. However, a clear limitation of this analysis is the lack of known gene-metabolite interactions in *S. cerevisiae* which could serve as positive controls. Previous attempts to identify gene-metabolite interactions in *S. cerevisiae* have used the topology of the metabolic network as a reference by marking each metabolite’s consuming/producing enzyme as a potential gene-metabolite interaction\textsuperscript{32}. However, if the gene-metabolite interactome from *E. coli* is any indication, this approach may not be representative of the true gene-metabolite interactions in *S. cerevisiae*. 
Figure A-1. Identification of gene-metabolite interactions in *S. cerevisiae*. A) Global transcriptional regulation component as determined by singular value decomposition for five types of nutrient limitation. Depicted: singular vector 1 (SV1), which captures most of the variance in the data set (43%). B) Distribution of the number of identified potential metabolite regulators for each gene. C) Number of potential target genes for each tested metabolite. D) Density plot of all pairwise gene-metabolite Pearson correlation coefficients (which serve as a metric for each metabolite’s ability to explain a gene’s expression) before and after removal of the global transcriptional regulation component (blue: lowest density; orange: highest density). E) Profiles of the three metabolites with the largest number of potential gene targets. All identified gene-metabolite interactions were obtained after correction for multiple testing of the respective correlation p-values as described previously65.
Figure A-2. Identification of metabolites affecting transcriptional regulation of metabolic pathways.

A) Pathway enrichment analysis. For each metabolite, the enrichment of its potential target genes (as determined in figure 1) for specific metabolic pathways was calculated by hypergeometric testing as described previously.\textsuperscript{66} The p-value denotes the probability that (at least) the given number of genes in a metabolic pathway would be identified by chance. Metabolic pathway definitions were obtained from a \textit{S.cerevisiae} genome-scale metabolic model.\textsuperscript{62} B) Comparison of measured specific transcriptional regulation component and its reconstruction based on a single metabolite for two glycolytic genes. (+) denotes an activating interaction. FBA1: FBP aldolase 1; PGK1: 3-phosphoglycerate kinase. C) Comparison of measured specific transcriptional regulation component and its reconstruction based on a single metabolite for two genes belonging to oxidative phosphorylation (-) denotes an inhibiting interaction. COX5A: Subunit Va of cytochrome c oxidase; SDH3: cytochrome b subunit of succinate dehydrogenase.
Chapter 4 – Few regulatory metabolites coordinate expression of central metabolic genes in E. coli

Figure A-3. Gene-metabolite interactions in asparagine metabolism, glycolysis/gluconeogenesis, and oxidative phosphorylation. Heatmap of Pearson correlation coefficients between measured specific transcriptional regulation (after removal of the respective global regulation component) and its reconstruction based on one metabolite for all gene-metabolite pairs in the respective pathways. For clarity, only pairs with correlation coefficients above 0.75 or below -0.75 are shown. White edges denote the highest-scoring interaction for the respective promoter.
References


Chapter 5 – Global coordination of *Escherichia coli*’s metabolic response to nutrient limitation

Karl Kochanowski, Hiroyuki Okano, Vadim Patsalo, Uwe Sauer, Terence Hwa

Contributions:
KK, US and TH designed the study. KK performed all experiments and analyses, with contribution from TH, HO and VP. KK wrote the chapter, with contribution from US.
Abstract
To cope with diverse environments, microorganisms need to regulate their metabolic activity accordingly. While many recent studies have provided insights into how individual pathways are regulated, the mechanisms that ultimately give rise to a coordinated metabolic response are only poorly understood. In this study, we identify the mechanisms underlying this coordination in the model bacterium Escherichia coli. Towards this end, we use genetically implemented gradual limitations of external glucose supply and internal glutamate production, which allow to study E. coli’s metabolic response to nutrient limitation while maintaining constant environmental conditions. We quantify E. coli’s steady state metabolic response by flux analysis and untargeted metabolomics, and integrate these data with proteome quantification using regulation analysis to obtain a quantitative picture of the relationship between metabolic flux, metabolite, and enzyme concentrations at single-reaction resolution. Our analysis reveals two key findings. Firstly, to coordinate its metabolic response to nutrient limitation, E. coli largely relies on two mechanisms, namely an approximate transcriptional program that roughly allocates the proteome but rarely controls fluxes alone, as well as passive changes in enzyme saturation that adjust this approximate transcriptional response. Secondly, to implement this transcriptional program, E. coli uses a single transcription factor, Crp, which not only directly activates the expression of catabolic enzymes, but also indirectly reduces the expression of anabolic enzymes by sequestering the cellular resources available for their expression.
Chapter 5 – Global coordination of Escherichia coli’s metabolic response to nutrient limitation

Introduction

In nature, microbes typically encounter unpredictable changes in nutrient availability. To sustain growth in face of such environmental insults, microbes need to coordinate their metabolism for providing biomass precursors, energy, and redox factors. Towards this end, microbes can mobilize a vast arsenal of mechanisms, such as transcriptional regulation, covalent posttranslational modifications, and non-covalent binding of small molecules, and many recent case studies have elucidated the role of these mechanisms in regulating individual metabolic pathways. Since mounting an appropriate response typically not only requires regulation of a single metabolic pathway, coordination must extend to the entire metabolic network. Examples of such coordination are the bacterial stringent response to nutrient starvation that triggers the general activation of stress genes, the Crp-dependent activation of catabolic genes in carbon limitation, and the coordination of glucose and nitrogen uptake by 2-oxoglutarate. More recently, a proteomics study revealed a simple proteome allocation strategy of the bacterial response to orthogonal perturbations, in which proteins belonging to specific metabolome sectors, such as catabolism and anabolism, are expressed in a coordinated manner. However, while these, and other, examples shed light on some of the general regulatory mechanisms underlying the coordination of microbial metabolism, they do not explain how this coordination is ultimately established for individual reactions.

Here we attempt to identify the regulatory mechanisms of individual reactions that enable Escherichia coli to coordinate its metabolic response to nutrient limitation. Towards this end, the key challenge is to impose nutrient limitations in a gradual and controlled manner under identical environmental conditions that minimize confounding secondary effects. Previously, continuous cultures have been used for this purpose, which allow to control external nutrient supply, but they are technically challenging, cannot be performed at high-throughput, and are not directly amenable to standard metabolome sampling techniques. To circumvent this challenge, we implement nutrient limitations genetically, allowing the external gradual control of the degree of limitation while maintaining identical batch-like experimental conditions. In this study, we focus on E. coli’s global metabolic response to two orthogonal limitations, namely carbon limitation, as well as limitation of anabolic capacity while carbon is in excess. By integrating the quantification of metabolic fluxes, proteins, and metabolites, in the framework of regulation analysis, we focus on two questions: Firstly, how is E. coli’s metabolic response to these limitations established at single-reaction level? Secondly, how is this response coordinated globally?
Chapter 5 – Global coordination of Escherichia coli’s metabolic response to nutrient limitation

Results

Quantifying E.coli’s metabolic response to gradual nutrient limitation. We chose to study E. coli’s metabolic response to two orthogonal limitations (figure 1, left panel). To implement an artificial carbon limitation, we put one of the components of the phosphotransferase system (PTS), namely PtsG, under the control of a 3-methylbenzyl alcohol (3-MBA) inducible promoter, thereby coupling glucose uptake to the 3-MBA concentration. Conversely, to mimic an anabolic limitation, we put the enzyme glutamine synthase (GOGAT) under control of an IPTG-inducible promoter and deleted the second glutamate-producing enzyme glutamate dehydrogenase. Consequently, the internal production of glutamate, which serves as the substrate for most transamination reactions in biosynthetic pathways, is coupled to the IPTG concentration. Thus, tuning of the expression of PtsG or GOGAT allows to gradually limit carbon or glutamate - and thereby effectively nitrogen - supply.

To obtain a comprehensive picture of the steady state metabolic response to these orthogonal limitations, we quantified fluxes at eight different induction levels per limitation using 96-well format cultivation (figure 1, middle panel). Specifically, we quantified ten ratios of intracellular fluxes by 13C flux analysis (supplementary figure S1) and used them, together with quantified extracellular exchange rates (supplementary figure S2), as constraints to estimate absolute central carbon fluxes using the FiatFlux software. Using these estimates as constraints, we further inferred genome-scale fluxes by flux-balance analysis (FBA) (supplementary figure S3).

Moreover, we quantified the relative intracellular concentration of 430 metabolites covering most metabolic pathways by untargeted metabolomics, as well as the absolute concentration of 40 central metabolites by targeted metabolomics (supplementary figure S4). Both metabolomics platforms yielded highly comparable metabolome responses for overlapping metabolites (supplementary figure S4B). To obtain a picture of E.coli’s proteome response, we further utilized a previously published proteomics data set, which used equivalent limitations and comprised about 1000 proteins (supplementary figure S5).

Generally, E.coli’s metabolic response was highly limitation-specific (figure 2). For example, central carbon fluxes showed marked changes in metabolic strategy between the limitations. As the imposed glucose limitation became more severe, we observed a shift from mixed acetate fermentation and respiration to full respiration (figure 1A), which is consistent with many previous studies in glucose-limiting chemostats. Surprisingly, we did not see activation of the glyoxylate shunt, which is another hallmark of severe glucose limitation in E. coli. A possible explanation is that the lowest growth rate in the glucose limitation (0.4 h⁻¹) is still higher than the onset of glyoxylate shunt activation, which was found to be below 0.4 h⁻¹, albeit in a different E. coli strain. In contrast to this shift to full respiration in glucose limitation, glutamate-limited cells not only continued to secrete acetate even at
the lowest tested growth rate (0.2 h⁻¹), but also 2-oxoglutarate. Taken together, acetate and 2-oxoglutarate secretion in glutamate limitation accounted for up to 50% of the utilized glucose, resulting in a reduced biomass yield (supplementary figure S2B). Biosynthetic fluxes outside of central carbon metabolism were for the most part strictly growth rate dependent regardless of the imposed limitation.

The metabolome response also strongly depended on the imposed limitation, and Kmeans clustering revealed several groups of metabolites with distinct patterns (figure 2B). In some of these groups, only one limitation caused changes in metabolite concentrations, for example, the L-glutamine (cluster 3) and 2-oxoglutarate (cluster 5) groups accumulated only in glutamate limitation. Other metabolite groups, such as the one including fructose-1,6-bisphosphate (FBP, cluster 10), were affected by both limitations, albeit in some cases in opposite directions (e.g. clusters 2/4/9). Nevertheless, more than 50% of the quantified metabolites were not significantly affected by either limitation, including most amino acids (cluster 12). Surprisingly, L-glutamate itself proved to be quite resilient against either limitation. Particularly in the glutamate limitation, which caused an up to five-fold growth rate reduction, its concentration was only two-fold reduced and never dropped below 10 mM (supplementary figure S4C), which is above the reported $K_m$ values for most glutamate-consuming reactions⁹⁹.

Taken together, comprehensive quantification of fluxes and metabolites revealed complex metabolic responses to gradual glucose and glutamate limitation. In particular, these limitations triggered orthogonal changes in fermentation strategy and many metabolite concentrations. These findings recapitulate the behavior of E. coli's catabolic and anabolic proteome sectors, which respond in opposite directions to the imposed limitations (supplementary figure 5). In the next sections, we attempt to unravel the regulatory mechanisms underpinning E. coli's response to carbon and anabolic limitation at single-reaction level.

**Transcriptional regulation of metabolic fluxes.** The most obvious mechanism by which cells could mount a metabolic response, i.e. a change in metabolic fluxes, is by adjusting the expression of the catalyzing enzymes. In the first step, we asked whether the observed flux changes could indeed be explained by corresponding protein concentration changes. Towards this end, we relied on the previously established theoretical framework of regulation analysis, which allows to relate flux changes of a reaction of interest to the contribution of individual regulatory layers⁹,¹⁶,³². This relationship can be formalized as follows:

$$
\log(J/J_{ref}) = \log(P/P_{ref}) + X \quad \text{(Eq. 1)}
$$
where \( J \) denotes the flux (normalized to a reference condition), \( P \) denotes the normalized concentration of the catalyzing enzyme, and the term \( X \) incorporates all additional effects (such as changes in substrate concentration, allosteric effectors etc.). The slope between flux and protein changes, termed transcriptional regulation coefficient \( \rho_p \), quantifies the contribution of transcriptional regulation:

\[
\rho_p = \frac{\log(P/P_{ref})}{\log(J/J_{ref})}
\]  

(Eq. 2)

A regulation coefficient of one signifies that the observed flux changes can be fully explained by the observed changes in protein concentration. We determined transcriptional regulation coefficients separately for each limitation by linear regression, considering all reactions that carry flux in the tested conditions and for which data of at least one of the associated proteins is available (figure 3). Comparison of the two limitations showed marked differences in the distribution of transcriptional regulatory coefficients (figure 3A), in particular a shift towards negative coefficients in the glutamate limitation (signifying that protein and flux change in opposite directions). This shift was particularly pronounced in biosynthetic reactions (figure 3B, middle panel), but was also observed for central metabolic reactions (figure 3B, top panel). Moreover, in both limitations only a small fraction of reactions had transcriptional regulatory coefficients around one (16% and 6% for glucose and glutamate limitation, respectively), showing that only few of the observed flux changes could be explained by corresponding changes in protein concentrations. However, there were a few exceptions (figure 3C, bottom panel). One of these exceptions is the flux through the final reaction of methionine biosynthesis, which could be almost fully explained by the abundance of its catalyzing enzyme homocysteine transmethylase (MetE) in the glucose limitation. Given that MetE is one of the most expensive enzymes in \( E. coli \) (i.e. high copy number per cell with low specific activity) and was shown to operate close to its maximal capacity during exponential growth on glucose\(^{40}\), our results suggest that \( E. coli \) responds to carbon limitation by reducing the expression of MetE, thereby diminishing the cost of methionine biosynthesis.

Thus, transcriptional regulation analysis revealed that for most reactions the observed flux response to glucose and glutamate limitation could not be explained by matching changes in protein concentration. This finding is in line with previous observations showing that gene expression is typically a poor predictor of flux changes\(^{3,9,16,41}\), and suggests a rather approximate transcriptional regulatory program: in particular in the glutamate limitation we observed that fluxes and proteins tended to change in opposite directions for many biosynthetic reactions (i.e. decrease in biosynthetic flux that is accompanied by an increase in protein concentration). To understand how the observed
metabolic response could be established otherwise, we next turned to the effect of altered metabolite concentrations on flux.

**Passive flux regulation by changes in substrate concentration.** One way how altered protein concentrations could be adjusted is by changes in enzyme saturation through changes in the corresponding substrate concentrations. This ‘passive’ mechanism of flux regulation provides a means for cells to effectively buffer changes in enzyme concentration while keeping the flux constant. To assess the role of enzyme saturation in *E. coli*’s metabolic response, we expanded the regulation analysis outlined above to also account for changes in substrate concentration as described before. The nonlinear relationship between substrate concentration and reaction rate can be approximated by a power-law term, which upon log normalization relates flux, protein concentration, and substrate concentration(s) as follows:

\[
\log(J/J_{ref}) = \log(P/P_{ref}) + \sum_{i=1}^{n} \alpha_i \cdot \log(M_i/M_{i,Ref}) + X \tag{Eq. 3}
\]

where \(M_i\) denotes the normalized concentration of the \(i\)th reaction substrate with the reaction-specific parameter \(\alpha_i\) capturing the non-linearity of its impact on flux. Analogous to \(\rho_P\), the contribution of enzyme saturation is quantified by the saturation regulation coefficient \(\rho_S\):

\[
\rho_S = \frac{\sum_{i=1}^{n} \alpha_i \log(M_i/M_{i,Ref})}{\log(J/J_{ref})} \tag{Eq. 4}
\]

We quantified \(\rho_S\) for over 100 reactions (separately for each limitation), only considering those reactions for which all substrates were quantified. In both limitations, the vast majority of \(\rho_S\)’s were positive, meaning that flux and substrate concentrations changed in the same direction (figure 4A, middle column). Analysis of the combined effect of transcriptional regulation and enzyme saturation – defined as the sum of \(\rho_P\) and \(\rho_S\) – showed that these two mechanisms could explain about 50% of all flux changes in both limitations (figure 4A, right column). However, the relative contribution of transcriptional regulation and enzyme saturation differed between the two limitations: in glucose limitation, both mechanisms had a similar impact, whereas in the glutamate limitation the contribution of enzyme saturation was dominant. This difference in regulatory strategy was typically maintained within a metabolic pathway (see arginine biosynthesis pathway as an example, supplementary figure S6). Notably, in central carbon metabolism, the fraction of reactions that could be explained by transcriptional regulation and enzyme saturation alone was rather small (supplementary figure S7), in line with findings from previous studies focusing on the bacterial central metabolic response to different carbon sources.

Overall, our analysis revealed two mechanisms that coordinate *E. coli*’s metabolic response to nutrient limitation at single reaction level, namely an approximate transcriptional regulatory program, which
rarely sets protein concentrations to exactly match the respective flux, as well as passive regulation of enzyme saturation through changes in substrate concentration. Together, transcriptional regulation and enzyme saturation explain a large fraction of observed flux changes in glucose (46%) and glutamate (56%) limitation. Nevertheless, the results also illustrate that many reactions cannot be explained by transcriptional regulation and enzyme saturation alone. Specifically, closer inspection revealed that glutamate-dependent transamination reactions – in which the amination of a metabolite is coupled to the deamination of glutamate to 2-oxoglutarate - were only poorly explained (see IlvE as example in figure 4B). The rates of these reversible reactions depend not only on the concentration of their substrates, but also of their products. We next asked whether changes in product concentration could potentially account for the discrepancy between measured flux, substrate, and protein concentration in these reactions.

**Impact of thermodynamic driving force on E. coli’s metabolic response.** As a starting point, we relied on the flux-force relationship\(^\text{41}\), which captures the impact of the impact of the ratio between product and substrate concentrations, termed Q, on flux:

\[
\Delta G' = -RT \ln \left( \frac{J_f}{J_r} \right) \quad K_{eq}/Q = \frac{J_f}{J_r} \quad (\text{Eq. 5a and 5b})
\]

\(\Delta G'\) being the Gibbs free energy, R the gas constant, T the temperature, \(J_f\) the forward flux, \(J_r\) the reverse flux, and \(K_{eq}\) the equilibrium constant. As Q approaches \(K_{eq}\), a larger fraction of the enzyme is engaged in catalyzing exchange fluxes, until eventually the net flux becomes zero, thus coupling the net flux of a reaction to its thermodynamic driving force. Conversely, by quantifying Q (and therefore \(\Delta G'\)) for a given reaction, we can assess its impact on the net flux. However, the relative metabolite quantification by untargeted metabolomics used here does not allow to quantify \(\Delta G'\) directly. To address this issue, we exploited the fact that reactions operating near equilibrium – which are particularly susceptible to changes in driving force- can be approximated by Q-linear kinetics\(^\text{44}\), thereby directly related a reaction’s Q and net flux. If only relative metabolite data is available, this relationship is described as follows:

\[
Q^* \approx -\frac{J_f}{k_1} K_1 + K_2 \quad (\text{Eq. 6})
\]

Where \(Q^*\) denotes the ratio of relative product and substrate concentrations, and \(K_1\) and \(K_2\) denote two reaction-specific parameters. We hypothesized that reactions falling under this regime in either of the limitations could be identified based on their linear relationship between \(Q^*\) and normalized flux. For about 40 reactions, we had quantified all substrates and products. Only a small subset of these reactions (~20%) showed a strong negative linear relationship between \(Q^*\) and normalized flux, as determined by linear regression. Reassuringly, among these was the enolase reaction, which is known to be largely driven by changes in driving force\(^\text{43}\) (figure 5). Moreover, the three transamination
reactions for which we had quantified all reactants (phenylalanine, valine, (iso)leucine transaminases) showed a strong negative relationship between \( Q^* \) and normalized flux in the glutamate limitation (figure 5). These results suggest that in the glutamate limitation these reactions shift closer to thermodynamic equilibrium, ultimately causing a reduction in net flux, which could potentially explain the aforementioned discrepancy between flux, substrate, and protein concentration in transamination reactions. Notably, this shift is largely driven by a change in product concentration, given that the common transamination substrate glutamate only shows small concentration changes, whereas the common product 2-oxoglutarate massively accumulates in the glutamate limitation (figure 2B).

These findings indicate that the thermodynamic driving force serves as a second compensatory mechanism for the passive regulation of metabolic fluxes, as has been proposed previously\(^1\). Although the lack of metabolite data for many reactions does not allow us to assess the impact of driving force at a large scale, our results suggest that it only affects a small subset of reactions. In particular, transamination reactions are ‘thermodynamically choked’ in the glutamate limitation by the accumulation of 2-oxoglutarate and resulting reduction in thermodynamic driving force, offering a potential explanation for how a moderate decrease in glutamate concentration can cause a disproportionally large reduction in growth rate. In contrast, transamination reactions which use aspartate instead of glutamate a substrate (with fumarate instead of 2-oxoglutarate as a product), could already be explained well by transcription plus enzyme saturation (see ArgG in supplementary figure S6 as example).

**Regulatory program underlying *E. coli*’s coordinated transcriptional response.** The previous sections revealed that *E. coli*’s coordinated metabolic response to glucose and glutamate limitation is driven by an approximate transcriptional response that is further adjusted by enzyme saturation and thermodynamic driving force. One striking feature of this transcriptional program is the response of catabolic (C-sector) and anabolic (A-sector) proteins in the two limitations: carbon and glutamate limitations are accompanied by a general increase and decrease in abundance of C-sector proteins, respectively, a pattern that is mirrored for A-sector proteins (figure 6A). In this final section, we ask how this coordinated transcriptional response is established mechanistically.

Recently, the transcriptional regulation of *E. coli*’s C-sector proteins was shown to be largely driven by one transcriptional activator, namely Crp\(^{21}\): in carbon limitations, the accumulation of Crp’s small molecule activator cyclic AMP causes the general increase in the expression of C-sector proteins (figure 6A). Conversely, in anabolic limitations, the accumulation of keto-acids, such as 2-oxoglutarate, inhibits the production of cyclic AMP, thereby preventing the increase in C-sector proteins. In agreement with these previous findings, we found that cyclic AMP accumulates in the glucose and decreases in the glutamate limitation, respectively (figure 6B). In contrast, the mechanism underlying the response of
A-sector proteins is poorly understood. Given the mirrored behavior of C- and A-sectors, the most parsimonious mechanism would be Crp (which activates C-sector proteins) serving as a repressor of A-sector proteins. However, although Crp is known to be a dual transcriptional regulator, analysis of the reported transcriptional network showed that Crp is not a prevalent regulator (neither activator nor repressor) of anabolic genes (supplementary figure S8). Moreover, no other major transcription factor could be associated to the A-sector based on the reported transcriptional regulatory network (supplementary figure S8).

How else, then, could the A-sector response be achieved? We hypothesized that the proteome response of the A-sector may also be an indirect consequence of Crp regulation. Various lines of evidence suggest that genes compete with each other for the limited capacity of the cellular expression machinery. For example, expression of unneeded proteins was found to decrease the translation rate of other proteins, presumably due to competition for ribosomes\textsuperscript{45}. Since Crp exerts its activating effect by directly recruiting RNA polymerase\textsuperscript{46} and has more than 2000 copies per cell\textsuperscript{47}, we hypothesized that Crp activation may reduce the number of free RNA polymerases, estimated to have far less than 1000 copies per cell\textsuperscript{48,49}, available for the expression non-Crp targets. Thus, activation of Crp would favor the expression of C-sector proteins at the expense of all other proteins (in absence of additional dedicated regulation). Conversely, in absence of Crp activation, the other protein sectors would take over the resulting spare capacity. To test this hypothesis and mimic the divergent response of Crp in the two limitations, we constructed a strain lacking endogenous cyclic AMP production, in which Crp activity can be controlled externally through addition of cyclic AMP to the medium. Titration of external cyclic AMP led to a bell shaped growth response, in which both low and high cyclic AMP concentrations caused a reduction in growth rate (figure 6C). Using this strain, we quantified the activity of two synthetic constitutive promoters, which are not regulated by any specific transcription factor\textsuperscript{50}, as well as two exclusively Crp-activated promoters, at varying cyclic AMP concentrations using fluorescent reporter plasmids\textsuperscript{50}. As expected, addition of cyclic AMP caused a monotonous increase in the activity of Crp-regulated promoters (supplementary figure S10), yielding an expression pattern (figure 6D, lower panels) reminiscent of C-sector proteins in the two limitations (figure 6A, left panel). However, in agreement with our hypothesis, this pattern was reversed in the constitutive promoters, whose activity was lowest at the highest cyclic AMP concentration (figure 6D, upper panels, and supplementary figure S10). To further validate our hypothesis, we quantified this strain’s proteome response at varying cyclic AMP concentrations. Indeed, we found that gratuitous Crp activation caused a decrease in A-sector fraction, whereas diminished Crp activation led to an increase in A-sector fraction at similar growth rates (figure 6E). Notably, not all proteome sectors were affected: ribosomal proteins maintained their strict relationship with the growth rate, suggesting additional compensatory regulation. Overall, these findings indicate a novel transcriptional regulatory scheme in which Crp, the
main transcriptional regulator of catabolic proteins, indirectly regulates the expression of anabolic proteins through competition for expression machinery capacity, and thereby coordinates the expression of catabolism and anabolism.
Discussion

In this study, we aimed to identify the mechanisms underlying *E. coli*’s coordinated metabolic response to nutrient limitation. Using genetically implemented gradual limitations of external glucose supply and internal glutamate production, we quantified steady state metabolic fluxes and metabolite concentrations at genome-scale, and included previously published proteome quantification\(^{23}\) to obtain a comprehensive quantitative picture of *E. coli*’s metabolic response. To understand how this metabolic response was established at single-reaction level, we integrated these data using regulation analysis and quantified the contributions of transcriptional regulation and passive regulation by changes in metabolite concentrations for each reaction. Our analysis revealed that *E. coli*’s metabolic response is coordinated by an approximate transcriptional regulatory program, in which only a small fraction of reactions can be explained by the observed changes in protein concentration. This effect was particularly pronounced in the glutamate limitation, where for the majority of reactions proteins and metabolic fluxes changed in opposite directions. These results re-iterate observations from several previous studies, which showed that protein expression is generally a poor predictor of metabolic flux\(^{3,41}\), suggesting that the transcriptional regulation of many metabolic proteins is suboptimal.

While this suboptimal regulation of protein expression was found to be widespread in bacteria\(^{51}\), it appears to be at odds with numerous evolutionary studies showing that even small differences in fitness (i.e. growth rate) are selected against (e.g. \(^{52}\)). One, previously proposed\(^{51}\), explanation is that cells have not evolved to cope with the artificial lab environments they are subjected to, which is particularly true for the genetically implemented limitations used in this study. Nevertheless, the fact that *E. coli*’s metabolic response to glucose limitation recapitulates previous studies using complementary experimental setups\(^{26,28,36,37}\) suggests that we did capture a physiologically relevant metabolic response. A more parsimonious explanation is that this response emerges from an evolutionary trade-off: given the multitude of environments that cells could encounter, it is simply infeasible to have a dedicated optimized program for each of these\(^{51,53,54}\). Consequently, cells may rather rely on a ‘ballpark’ mode of regulation, which coarsely allocates the proteome according to few signals\(^1\). A good illustration of this principle is the proteome response to glutamate limitation: if cells cannot identify the exact reaction responsible for the reduction in anabolic capacity, the best alternative is across-the-board expression of anabolic proteins.

However, this approximate transcriptional regulatory program requires additional mechanisms to adjust the activity of each individual reaction and ultimately mount a coordinated metabolic response. Our analysis shows that this adjustment is largely achieved by changing an enzyme’s saturation level through adjustments in substrate concentration. This passive regulatory mechanism is an inherent property of enzyme kinetics and provides a biochemical buffer to make a reaction more robust against
fluctuations in substrate or enzyme concentrations, but it also makes each reaction more expensive, since enzymes do not work at full capacity. Thus, our results provide an example of an evolutionary solution to conflicting objectives, namely between optimized growth in a given condition and robustness against external and internal fluctuations: E. coli increases the robustness of its metabolic response at the cost of suboptimal performance in a given condition. Evolutionary experiments suggest that the selected solution may actually be quite plastic: prolonged cultivation in a constant environment is typically accompanied by the accumulation of mutations in a population that increase fitness (i.e. growth rate) in this condition. It is tempting to speculate that this adaptation may come at the price of reduced robustness, as the cell’s enzymes are likely to become more saturated, and future studies of the metabolome response of such evolved populations may shed more light on the underlying evolutionary trade-offs.

In total, transcriptional regulation and enzyme saturation accounted for about 50% of the observed flux changes in both limitations, which is higher than reported in comparable previous studies focusing on central metabolism. Nevertheless, also in our case the fraction of central metabolic fluxes that can be explained by these two mechanisms is lower compared to the rest of metabolism (supplementary figure S7), indicating that additional regulatory mechanisms, such as allosteric regulation, may generally play a larger role in central metabolism.

Surprisingly, our results suggest that the global transcriptional coordination of E. coli’s metabolic response is implemented by a single transcription factor, namely Crp. Specifically, we found that Crp not only directly activates the expression of catabolic proteins, but also indirectly represses the expression of anabolic proteins, presumably by reducing the number of free RNA polymerases available for the expression of other proteins. In the carbon and anabolic limitations tested here, the underlying regulatory signal is likely provided by keto-acids such as 2-oxoglutarate through a recently discovered regulatory circuit. In this circuit, imbalances between a cell’s catabolic and anabolic metabolic capacities cause the accumulation or depletion of keto-acids, which in turn inhibit adenylate cyclase and thereby adjust Crp activity. Although to our knowledge this is the first example of such indirect transcriptional regulation by a transcription factor, it may provide a possible explanation for the frequently observed condition-dependent differences in the activity of the global expression machinery. Notably, this regulatory mechanism can be overridden by additional designated regulation. An evident example is the regulation of the ribosomal proteome fraction, which maintains its strict dependency on growth rate in both tested limitations (figure 6A and E). Moreover, the observation that addition of sub-lethal dosages of chloramphenicol, which cause a massive increase in the ribosomal proteome fraction, is accompanied by a reduction in all other proteome sectors, suggest the existence of additional indirect transcriptional regulation mechanisms.
Acknowledgements

We thank Elad Noor for helpful discussions.
Material and Methods

Reagents and strains. Unless stated otherwise, all reagents were obtained from Sigma-Aldrich. All used strains were derived from NCM372259. 3-MBA-inducible ptsG (NQ1243 + NQ1390) and IPTG-inducible GOGAT (NQ39323) titration strains, as well as NQ1399, which lacks both adenylate cyclase (CyaA) and cyclic AMP phosphodiesterase (CpdA), were provided by the Hwa lab. Fluorescent transcriptional reporter plasmids of two constitutive promoters (epd-icd, epdconstitutive) as well as two Crp-regulated promoters (acs, Crp reporter) were constructed as described in chapters 3 and 4 of this thesis and transferred to NQ1399 by electroporation as described previously60.

Cultivation. All experiments were performed using M9 minimal medium60, supplemented with 2g/L glucose. Cultivations were performed as follows: M9 medium batch cultures in 96 deep-well format plates (Kuehner AG, Birsfeld, Switzerland), containing the same inducer concentration as for the subsequent main culture (with the exception of slow-growing NQ393 with 10/20 mM IPTG in the main culture, which were cultivated with 30 mM IPTG in the preculture to avoid the emergence of mutations), were inoculated 1:50 from LB precultures and incubated overnight at 37°C under shaking. Subsequently, 96 deep-well plate cultures were inoculated with overnight cultures to a starting OD600 of 0.03-0.05 (total fill volume per well: 1.2 mL) and incubated at 37°C under shaking. Culture OD600’s were monitored by OD600 sampling from parallel wells on the same deep-well plate and subsequent OD600 measurements using a plate reader (TECAN infinite M200, Tecan Group Ltd, Männedorf, Switzerland). Cultivation of strains bearing transcriptional reporter plasmids, and calculation of promoter activity as the OD normalized GFP production rate was performed as described previously50,60. Steady state promoter activities in M9 minimal medium with 2 g/L glucose with varying external cyclic AMP concentrations (ranging from 0 to 10 mM) were determined during the 1.5h window during which the cultures exhibited the maximal growth rate.

Quantification of intracellular metabolite concentrations. Intracellular metabolomics samples were taken during mid-exponential phase at OD’s between 0.5 and 0.6 by fast filtration (sampling volume: 1 mL)61, and were immediately quenched in 4 mL quenching/extraction solution (40% methanol, 40% acetonitrile, 20% H2O) at -20°C31. Samples were incubated for 2h at -20°C, subsequently dried completely at 120 µbar (Christ RVC 2-33 CD centrifuge and Christ Alpha 2-4 CD freeze dryer) and stored at ~80°C until measurements. Before measurements, samples were resuspended in 100 mueL water, centrifuged for 5 min (5000g, 4°C) to remove residual particles, diluted 1:10 in water, and transferred to V-bottomed 96 well sample plates (Thermo Fisher Scientific). Samples were measured by flow-injection time-of-flight mass spectrometry with an Agilent 6550 QToF instrument operated in negative ionization mode at 4 GHz high-resolution in a range of 50–1,000 m/z as described before62. Sample processing and ion annotation was performed based on accurate mass within 0.001 Da using the KEGG
Quantification of uptake and secretion rates. Culture samples were taken at 6-8 time points together with parallel OD600 samples throughout exponential growth phase (sampling volume: 100 μL). Supernatants were separated from cells by centrifugation (5000g, 4 min, at 4°C) and transferred to V-bottomed 96 well sample plates (Thermo Fisher Scientific). Glucose and acetate concentrations in supernatants were determined by colorimetric enzymatic assays (Megazyme). All other secreted metabolites were quantified by flow-injection time-of-flight mass spectrometry as described above. Briefly, Supernatants were diluted 1:10 in water and measured with an Agilent 6550 QToF instrument operated in negative ionization mode at 4 GHz high-resolution in a range of 50–1,000 m/z. Sample processing and ion annotation was performed based on accurate mass within 0.001 Da using the KEGG eco database\textsuperscript{63} as reference and accounting for single deprotonated forms of the respective metabolite (M-H+) as described before\textsuperscript{64}. Absolute extracellular metabolite concentrations were determined using parallel dilution series of the respective metabolite in the same medium as calibration curves. Uptake and secretion rates were determined from extracellular metabolite concentrations, corresponding OD600 (conversion factor to cell dry mass: 1 OD/L = 0.413 g CDW/L), and corresponding growth rates by linear regression as described previously\textsuperscript{65}. With the exception of 2-oxoglutarate, all other quantified metabolites were secreted in minute amounts (<70 μmol/gCDW/h).

Quantification of intracellular metabolic fluxes. Intracellular central metabolic fluxes were determined by $^{13}$C flux analysis as follows. Cultivation was performed as described above, and glucose was added as the [1-$^{13}$C] isotope (> 99%; Cambridge Isotope Laboratories), or as a mixture of 20% (wt/wt) [U-$^{13}$C] (> 99%; Cambridge Isotope Laboratories) and 80% [1$^{13}$C] isotopes. Labeling samples (sampling volume 1 mL) were taken during mid-exponential phase (OD 0.5 – 0.7), cells were harvested by centrifugation (13000g, 3 min), cell pellets were washed once in cold 0.9% NaCl, and stored dry at -20°C. $^{13}$C flux ratios were determined as described previously\textsuperscript{66–67}. Briefly, cell pellets were hydrolyzed, dried, and derivatized, and labeling patterns of derivatized proteinogenic amino acids were quantified by GC-MS using a 6890 GC system combined with a 5973 Inert SL MS system (Agilent Technologies, Santa Clara, USA). Metabolic flux ratios were determined based on these labeling patterns (after correcting for naturally occurring $^{12}$C as described in\textsuperscript{58}) using the software FiatFlux\textsuperscript{34}. Two flux ratios (glyoxylate shunt and malic enzyme flux) were found to be zero in all conditions, and were discarded in subsequent analyses. Using the flux ratios, uptake/secretion rates, and the measured growth rate as inputs, absolute central metabolic fluxes were inferred using the software FiatFlux\textsuperscript{34}. 

\textsuperscript{63}
To infer metabolic fluxes beyond central carbon metabolism, Flux Balance Analysis (FBA) was performed using the Cobra toolbox\textsuperscript{35} based on an E.coli genome-scale metabolic model\textsuperscript{60}. In the first step, all measured major (glucose, acetate, 2-oxoglutarate) and minor (succinate, malate, citrate, phenylpyruvate, glycine, valine, glutamate) exchange rates, as well as the measured growth rate, were used as constraints (allowing for 5% deviation for major exchange rates and growth rate, as well as 10% deviation for minor exchange rates), and maximization of ATP production rate was used as an objective function\textsuperscript{70}. Moreover, four ratios of absolute fluxes (as determined above, namely phosphofructokinase \( \leftrightarrow \) G6P dehydrogenase; phosphoglucoisomerase \( \leftrightarrow \) phosphogluconate dehydratase; malate dehydrogenase \( \leftrightarrow \) PEP carboxylase; enolase \( \leftrightarrow \) PEP carboxykinase) were incorporated as constraints using the Cobra toolbox command addRatioReaction. Next, using the obtained ATP production rate as an additional constraint, minimization of fluxes was performed to yield the final inferred flux distribution.

**Regulation analysis.** Regulation analysis was performed as described previously\textsuperscript{9,16}. Only reactions with non-zero fluxes in at least 50% of the tested conditions were considered. Absolute fluxes were normalized to NCM3722 wild-type in the respective experiments and log-transformed. Transcriptional regulation coefficients were determined using a previously published proteomics study, which had determined relative protein concentrations in equivalent carbon and glutamate limitations\textsuperscript{21}. Relative protein concentrations were linearly interpolated to match exactly the same growth rate as in the corresponding flux measurements, normalized to NCM3722 wildtype and log-transformed. For each considered reaction, transcriptional regulation coefficients were determined separately for each limitation (by linear regression) as the slope between log-normalized fluxes and log-normalized protein concentrations. Reaction-protein pairs were used as defined in the aforementioned genome-scale metabolic model\textsuperscript{69}. For reactions that are associated to more than one protein, the final transcriptional regulation coefficients were calculated as the average regulation coefficients across all measured proteins.

Enzyme saturation regulation coefficients were determined as follows. To account for potential non-linearity of the relationship between substrate concentration and flux (e.g. due to non-Michaelis-Menten type enzyme kinetics), approximate kinetic orders (\( \alpha \)) for substrates were estimated using following equation as described previously\textsuperscript{9,16}:

\[
\min_{0 \leq \alpha \leq 4} \log(f_J) - \log(P_i) = \sum_{x \in S_I} \alpha_{ix} \cdot \log(M_x)
\]

where \( J \) denotes the normalized flux, \( P \) denotes the normalized protein concentration, and \( M \) denotes the reaction substrate(s) with corresponding kinetic order(s) \( \alpha \). \( \alpha \) was constrained to be between 0 and 4 to set a biologically realistic upper bound on the non-linear gain. Kinetic orders were estimated.
independently for each flux-enzyme pair (considering both limitations) by least square optimization using the *lsqlin* function of MATLAB. Only reactions for which all substrates had been quantified were considered. As above, for reactions that are connected to more than one protein, the final regulation coefficients were calculated as the average regulation coefficients across all measured proteins.
Figure 1. Design of this study. **Limitation:** Two orthogonal limitations, namely external glucose and internal glutamate limitation, were implemented genetically by chromosomal integration of titratable ptsG or glutamate synthase (GOGAT) into *E. coli*. **Experiment:** Titration strains were cultivated in 96-well format at varying induction levels of ptsG and GOGAT, respectively. Samples for the quantification of metabolic flux ratios and intracellular metabolite concentrations were taken during mid-exponential phase, and samples for the quantification of extracellular metabolite concentrations were taken at several time points throughout exponential growth. **Data:** Genome scale metabolic fluxes were estimated by flux balance analysis (FBA) using central metabolic flux ratios - determined by $^{13}$C flux analysis - and metabolite exchange rates - determined from the quantification of the extracellular metabolome by enzymatic assays and flow-injection time-of-flight mass spectrometry (FIA-TOF) - as constraints. Large-scale relative intracellular metabolite concentrations were quantified by FIA-TOF, and absolute concentrations of central carbon metabolites were quantified by targeted metabolomics (LC-MS/MS). Relative concentrations of about 1000 proteins (determined by shotgun proteomics) were obtained from a previous publication using equivalent limitations.
Chapter 5 – Global coordination of Escherichia coli’s metabolic response to nutrient limitation

Figure 2. E. coli’s metabolic response to glucose (blue) and glutamate (red) limitation. A) Absolute metabolic fluxes in central carbon metabolism. Left: map of central carbon metabolism. Gray lines denote reactions with zero flux in all tested conditions (malic enzymes, glyoxylate shunt). Numbers correspond to right figures in right panel. Right: Absolute metabolic fluxes for selected reactions plotted against the respective growth rate as determined by $^{13}$C flux analysis. Top panel: major exchanged metabolites. Lower, numbered, panels: selected reactions. Pgi: Phosphoglucoisomerase. Pfk/Fba: flux from F6P to 2 GAP. Zwf: flux from F6P to 6PG. Pgm/Eno: flux from 3PG to PEP. CS/Acn: flux from OAA + acetyl-CoA to isocitrate. aKG to Succinate: flux through aKG dehydrogenase, succinyl-CoA synthethase and succinate oxidoreductase. Ppc: flux through PEP carboxylase. Error bars denote standard deviation of flux estimate based on error propagation of the standard deviations of flux ratios and exchange rates (three biological replicates). B) Intracellular metabolome response (of 430
annotated ions) in mid-exponential growth as determined by FIA-TOF mass spectrometry. Upper panel: heatmap of log₂ fold-changes (relative to NCM3722 wild-type) for both limitations. Annotated ions were sorted by Kmeans clustering of log₂ fold-changes, using squared Euclidian distance as metric. Annotated ions showing little variation across conditions (coefficient of variance across all conditions < 30%) were filtered out before clustering and assigned to cluster 12. Lower panel: example ions for each cluster plotted (in linear scale) against the respective growth rate. Numbers underneath metabolite names denote m/z. Error bars denote standard deviation of three biological replicates.
Figure 3. Regulation analysis relates observed flux and protein changes. Regulation analysis was used to assess the degree to which the observed changes in flux for a reaction can be explained by changes in the concentration of the respective catalyzing protein, which is quantified as the slope between log normalized flux and protein concentration, also termed transcriptional regulation coefficient ($\rho_P$). If a reaction has a regulation coefficient of 1, the observed flux changes can be fully explained by proportional changes in protein concentration. **A)** Distribution of transcriptional regulation coefficients for glucose (upper panel) and glutamate (lower panel) limitation. Transcriptional regulation coefficients were determined separately for each limitation by linear regression. For reactions that are associated to more than one protein, the final transcriptional regulation coefficients were calculated as the average regulatory coefficients across all measured proteins (total number of considered reactions: 236). The fraction of reactions whose flux changes can be well explained by the observed changes in protein concentration (regulation coefficients between 0.5 and 1.5) is highlighted by grey areas with corresponding percentages. Individual examples are shown in **B**. Blue and red lines denote the linear regression between measured flux and protein concentrations for glucose and glutamate limitation, respectively. Black lines denote full proportionality. $\rho$: corresponding regulation coefficients. Top panel: central metabolic reactions. Middle panel: biosynthetic reactions showing a shift from positive to negative $\rho_P$ for glucose and glutamate limitation, respectively. Bottom panel: biosynthetic reactions which deviate from this behavior.
Figure 4. Assessing the effect of enzyme saturation on flux. A) Distribution of regulation coefficients quantifying the contribution of changes in protein concentration (left column) and enzyme saturation (middle column) to the explanation of observed flux changes, as well as their combined effect (right column). Upper panel: glucose limitation. Lower panel: glutamate limitation. Only reactions, for which at least one protein as well as all substrates had been quantified, were considered in the analysis (total number of considered reactions: 114). $H^+$, $NH_4$, $CO_2$, and $H_2O$ were omitted as substrates. The fraction of reactions with regulation coefficients between 0.5 and 1.5 is highlighted by grey areas with corresponding percentages in brackets. Individual examples are shown in B). Blue and red lines denote the linear regression between measured flux and protein concentrations for glucose and glutamate limitation, respectively. Black lines denote full proportionality. $\rho$: corresponding regulation coefficients.
Figure 5. Relationship between normalized flux and ratio of product to substrate concentrations (Q*) in glucose (blue) and glutamate (red) limitation. X-axis: flux divided by relative protein concentration and normalized to NCM3722 wild type. Y-axis: ratio of relative product and substrate concentrations as determined by FIA-TOF. Error bars denote standard deviation of three biological replicates (calculated from standard deviations of individual metabolites by error propagation).
Figure 6. Direct and indirect global transcriptional regulation by Crp. A) Proteome fraction of catabolic (C), anabolic (A), and ribosomal (R) proteome sectors in carbon (blue) and anabolic (red) limitation. Error bars denote standard deviation of three biological replicates. Data were obtained from23. See supplementary figure S5 for details. B) Intracellular concentration of cyclic AMP (cAMP) in glucose (blue) and glutamate (red) limitation as determined by targeted metabolomics. Error bars denote standard deviation of three biological replicates. C) Maximal steady state growth rate as a function of external cAMP concentration in strain NQ1399, which lacks adenylate cyclase as well as cAMP phosphodiesterase. Error bars denote the standard deviation of 8 biological replicates. D) Steady state promoter activities of two constitutive (epd-icd, epdconstitutive) and two solely Crp-regulated (acs, Crp reporter) fluorescent transcriptional reporters in strain NQ1399 at varying external cAMP concentrations. cAMP concentrations range from 0 (white filled circle) to 10 mM (black filled circle) as shown in C). E) Proteome fraction of catabolic (C), anabolic (A), and ribosomal (R) proteome sectors in strain NQ1399 at varying external cAMP concentrations (ranging from 0 to 3 mM), normalized to respective fraction at maximal growth rate. Shotgun proteomics measurements were performed as described previously23.
Supplementary figures

Supplementary figure 1. Flux ratios in glucose (blue) and glutamate (red) limitation as determined by $^{13}$C flux analysis. (1) Ratio of glycolysis to sum of Pentose phosphate (PP) + Entner Doudoroff (ED) pathway. (2) Flux ratio of pyruvate (PYR) derived through ED pathway. (3) Upper bound of phosphoenolpyruvate (PEP) derived through transketolase. (4) Upper bound of E4P derived through transketolase. (5) Flux ratio of serine (SER) derived from glycine (GLY) through glycine cleavage pathway. (6) Flux ratio of glycine derived from serine. (7) PEP derived from oxaloacetate (OAA) through PEP carboxykinase (Pck). (8) OAA derived from PEP through PEP carboxylase (Ppc). Error bars denote standard deviation of three biological replicates. Not shown: flux ratios for malic enzymes and glyoxylate shunt, which were found to be zero in all tested conditions.
Supplementary figure 2. Metabolite exchange rates in glucose (blue) and glutamate (red) limitation.

A) Absolute secretion rates as determined by flow-injection time-of-flight (FIA-TOF) mass spectrometry. Left: maximal secretion rates across all conditions. Right: secretion rates of six minor secretion products plotted against the growth rate in the corresponding conditions. Numbers underneath metabolite names denote m/z. Secretion rates were calculated as described in the methods section. Error bars denote standard deviation of three biological replicates.

B) Rates of major exchange products plotted against the growth rate in the corresponding conditions. Glucose and acetate exchange rates were calculated based on the quantification of extracellular glucose and acetate by colorimetric enzyme assays. 2-oxoglutarate secretion rate was quantified by FIA-TOF mass spectrometry as described above. Right panel: relative secretion rate calculated as the carbon-normalized fraction of glucose that is being secreted by acetate or 2-oxoglutarate. Error bars denote standard deviation of three biological replicates. Note that all experiments were performed in M9 minimal medium with glucose as sole carbon source.
Chapter 5 – Global coordination of Escherichia coli’s metabolic response to nutrient limitation

Supplementary figure 4. Intracellular metabolome response in glucose (blue) and glutamate (red) limitation. A) KEGG metabolic map of *E. coli*. Filled circles denote metabolites that were quantified by untargeted metabolomics (flow-injection time-of-flight (FIA-TOF) mass spectrometry, annotation with 0.001 Da mass accuracy based on KEGG *E. coli* model). In total, 430 unique ions were annotated, mapping to 900 unique metabolites. B) Distribution of relative standard deviations (three biological replicates, two technical replicates per biological replicate) across all ions and conditions. The median relative standard deviation is about 10%. C) Comparison of log2 fold-changes for all metabolites that were quantified by both untargeted (FIA-TOF) and targeted (LC-MS/MS) metabolomics. R: Pearson correlation coefficient. D) Absolute intracellular concentration of central carbon metabolites as determined by LC-MS/MS. Error bars denote standard deviation of three biological replicates. In all conditions, the energy charge (defined as \([\text{ATP} + 0.5 \text{ ADP}] / [\text{ATP} + \text{ADP} + \text{AMP}]\)) was between 0.78 and 0.9, suggesting a stable physiological state.
Supplementary figure 5. Proteome response to carbon (blue) and glutamate (red) limitation as determined by proteomics. Data were obtained from\textsuperscript{23}. Carbon limitation was implemented by titration of a lactose transporter. Comparison with glucose limitation showed strong agreement in the proteome response. Glutamate limitation was implemented as described in the main text of this study. Proteins were assigned to sectors as defined previously\textsuperscript{23}, and subsequently summed up to yield the total proteome fraction. Error bars denote standard deviation of three biological replicates.
Supplementary figure 6. Regulation analysis of arginine biosynthesis pathway in glucose (blue) and glutamate (red) limitation. Left: schematic depiction of arginine biosynthesis pathway including all substrates and cofactors. In bold next to each reaction are the associated proteins. Right: relationship between observed flux changes and changes in protein (left column) or substrate concentration (middle column), as well as their combined effect (right column). Blue and red lines denote the linear regression between measured flux and protein concentrations for glucose and glutamate limitation, respectively. Black lines denote full proportionality.
Supplementary figure 7. Regulation analysis of central metabolic reactions in glucose (blue) and glutamate (red) limitation. A) Glycolysis/pentose phosphate pathway reactions. Relationship between observed flux changes and changes in protein (left column) or substrate concentration (middle column), as well as their combined effect (right column). Blue and red lines denote the linear regression between measured flux and protein concentrations for glucose and glutamate limitation, respectively. Black lines denote full proportionality. B) Anapleurotic and TCA cycle reactions.
Supplementary figure S8. Enrichment analysis of transcriptional activators (left) and repressors (right). Information on each transcription factor’s target genes was obtained from RegulonDB\textsuperscript{71}. For each transcription factor – sector combination, the number of repressed and activated target genes was determined, and enrichment was calculated as the probability that least the observed number of target genes would have been found by chance based on the total number of measured proteins, the total number of TF target genes, and the number of proteins associated to the respective sector\textsuperscript{72}. Note that ArgR and Cra, which show weak enrichment for target genes in the A-sector, regulate less than 10% of the proteins associated to this sector (as determined based on the reported transcriptional regulatory network).
Figure S9. Effect of external amino acid supply on growth in glutamate limitation. A) Upper panel: OD600 time courses of NCM3722 (wild-type) and NQ393 (IPTG-inducible GOGAT titration strain) growing in M9 glucose (2 g/L) with addition of different amino acids. Lower panel: corresponding maximal growth rate. Data show average of six biological replicates. B) $^{15}$N-labeling experiment to determine the fraction of glutamate derived from aspartate. Cells were grown in M9 glucose (2 g/L) with 10 mM 100% $[^{15}$N$]$ ammonium. Unlabeled aspartate was added at inoculation, and intracellular metabolomics samples were obtained during mid-exponential phase at OD 0.5. Lower panel: fraction of glutamate that is derived from aspartate (determined as fraction of glutamate with $^{14}$N).

Figure S10. Activity of constitutive and Crp-regulated promoters at varying cyclic AMP concentrations. Steady-state promoter activities of two constitutive (epd-icd, epd$^{{\text{constitutive}}}$) as well as two exclusively Crp-activated (acs, Crp reporter) fluorescent transcriptional reporters in strain NQ1399 (lacking endogenous cyclic AMP production) at varying external cyclic AMP concentrations.
References


71. Salgado, H. *et al.* RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory

Chapter 6 – Posttranslational regulation of microbial metabolism

This chapter has been published in:
Abstract

Fluxes in microbial metabolism are controlled by various regulatory layers, all of which alter the abundance or activity of metabolic enzymes. Recent studies suggest a division of labor between these layers: transcriptional regulation mostly controls the allocation of protein resources, passive flux regulation by enzyme saturation and thermodynamics allows rapid responses at the expense of higher protein cost, and posttranslational regulation is utilized by cells to directly take control of metabolic decisions. In this review, we present recent advances in elucidating the role of these regulatory layers, focusing on posttranslational modifications and allosteric interactions. As the systematic mapping of posttranslational regulatory events has now become possible, the next challenge is to identify those regulatory events that are functionally relevant under a given condition. Understanding precisely how a regulatory signal affects metabolic function might facilitate metabolic engineering projects by solving potential bottlenecks in non-native pathways.
Introduction

Regulation of metabolic fluxes lies at the core of many microbial processes and comprises a plethora of regulatory layers, such as transcriptional regulation, posttranslational modification, and allosteric. All of these regulatory layers ultimately act by altering the capacity of metabolic enzymes through changes in their expression or activity. Arguably, the role of transcriptional regulation is better understood, and its importance for adapting to changes in nutrient availability and coordinating catabolism and anabolism is well established. On the other hand, focusing on flux control in central metabolism paints a different picture, with so far little correspondence between flux and transcriptional changes. Moreover, recent studies have demonstrated that transcriptional regulation can be surprisingly promiscuous and often driven by global physiological parameters such as the growth rate rather than the exact nature of the perturbation. Also, the stochastic nature of transcription and translation events makes it difficult to precisely tune enzymes to their flux requirement. These findings suggest that cells do not—and arguably cannot—use transcriptional regulation to fine-tune the expression of each metabolic enzyme in accordance to changes in flux. Instead, cells use transcriptional regulation to merely set the scope of possible fluxes (Figure 1A), and apply other regulatory layers to determine the exact location inside this space.

Recent reviews have discussed the prevalence of posttranslational modifications (PTMs) in microbial metabolic enzymes. In this review, we highlight recent advances in elucidating how microbial metabolism is shaped by posttranslational regulation by PTMs as well as allosteric. We start by outlining the mechanisms which enable cells to change metabolic fluxes even in the absence of designated regulation, and close by discussing the implications for the engineering of microbial pathways.

When is posttranslational regulation really necessary?

Cells of highly organized multicellular organisms can cooperatively control their immediate extracellular environment to ensure homeostasis. Without that privilege, microorganisms must always be ready to respond quickly to unpredictably changing conditions—often by adjusting their internal fluxes accordingly. But, do they necessarily have to actively regulate all the enzymes whose flux is changing? Even in the absence of allosteric effectors or PTMs, the physical nature of enzyme kinetics allows flux to be regulated passively by changing the saturation level (substrate concentration relative to K_M value) or the thermodynamic driving force (for reversible reactions). For example, the flux through an unsaturated enzyme can be adjusted by altering the concentration of its substrate. Conversely, the flux through a reversible enzyme operating close to thermodynamic equilibrium can be changed or even reversed by moderately changing the ratio of its substrates and products (Figure 1B-C). Thus, enzyme kinetics and thermodynamics provide metabolic networks with considerable inherent flexibility to respond to environmental changes, and there is growing evidence that a large
fraction of central metabolic enzymes may rely on these passive regulatory mechanisms. For example, quantitative metabolomics in *E. coli* revealed that the concentrations of many metabolites in central metabolism are around their respective *Kₘ* value. Consequently, these enzymes operate at saturation levels at which they are sensitive to changes in substrate concentrations. Further studies have also shown that many reactions in the central metabolism of *E. coli* and *S. cerevisiae* operate close to equilibrium, in particular in lower glycolysis. The enzymes catalyzing these reactions essentially operate in fire-and-forget mode, in which no additional regulation is needed for rapid metabolic changes. However, this mode of operation does come at a high cost: for enzymes operating close to equilibrium, the majority of their capacity is wasted on exchange fluxes without contributing to the net flux. Similarly, an enzyme is at least 50% underutilized when its substrate concentration is lower than the *Kₘ*. Consequently, such passively regulated enzymes may need to be expressed at very high concentrations to support the flux required by the cell. Potentially this explains why some of the most abundant proteins in *E. coli* and other organisms are the reversible enzymes in lower glycolysis and TCA cycle.

Such passive flux control constitutes a fundamental trade-off between higher inherent robustness and more efficient protein usage. Since protein costs are the highest cellular investment, transcriptional regulation enables more efficient resource allocation – at the cost of reduced flexibility and robustness. Moreover, often it is advantageous to regulate the activity of enzymes directly. One important case occurs when two or more enzymes perform antagonistic functions, e.g. ATP wasting futile cycles in central metabolism, or the simultaneous operation of biosynthetic and catabolic reactions. In fact, these enzymes are frequently subject to extensive allosteric regulation.

Another common case requiring post-translational regulation is in controlling the relative flux between metabolic branches that utilize the same substrate pool. Even if we ignore environmental fluctuations, intrinsic noise in gene expression causes the concentration ratio between the two branching enzymes to differ among cells in a population. If the cells would rely only on this ratio to determine the relative fluxes, the metabolic outcome would be as noisy as the gene expression itself (Figure 1D).

In such cases, post-translational regulation can make the system robust to the expression noise. For example, the first step of a biosynthetic pathway which branches out from glycolysis or the TCA cycle, is typically allosterically regulated in order to prevent excessive flux from leaving central metabolism and depleting the metabolite pool needed by other pathways. Similarly, without post-translational regulation, enzymes catalyzing irreversible reactions would have to be precisely fine-tuned to not have a higher capacity than their downstream reactions – or otherwise their product would continue to accumulate indefinitely (Figure 1E). A striking example for what happens when disrupting the coordination between different sections of the same pathway is “turbo metabolism” in yeast:
mismatch in enzyme capacity between upper glycolysis (which consumes ATP) and lower glycolysis (which produces ATP) can lead to a depletion of ATP upon glucose addition which essentially slows down glycolysis to a halt \(^{23,24}\). Another example comes from Reaves et al.: using \(E.\ coli\) mutant strains which lack allosteric negative feedback loops in the pyrimidine biosynthesis pathway, the authors showed that cells secrete pyrimidine intermediates which accumulate if the pathway is imbalanced (e.g. due to the external supply of other intermediates) to ensure homeostasis of the pathway’s end product UTP \(^{25}\). The fact that microbes secrete a surprisingly diverse set of metabolites even in minimal medium suggests that similar overflow mechanisms may be quite common \(^{26}\).

**Mapping posttranslational regulatory events**

Advances in mass-spectrometry based proteomics have greatly facilitated the discovery and quantification of a multitude of different covalent protein modifications in microbes \(^{12,27,28}\), and we can now detect many protein modifications at genome scale \(^{29–31}\). These efforts have yielded comprehensive PTM maps, in which a particular modification can target hundreds of proteins \(^{32}\), and in turn proteins can be modified by multiple different PTMs \(^{33–35}\). For example, over 50% of the detected succinylation sites in \(E.\ coli\) and \(S.\ cerevisiae\) proteins are also subject to acetylation \(^{30}\).

In contrast to these advances in the identification of PTMs, the identification of allosteric protein-metabolite interactions has lagged considerably. Nevertheless, recent years have seen the advent of various novel experimental methods to systematically identify protein-metabolite interactions \(^{36}\), that directly detect metabolite binding \(^{37–40}\), or structural protein alterations caused by metabolite binding \(^{41,42}\). A particularly promising example for the latter category is limited proteolysis coupled to targeted proteomics, which identified yeast proteins that change their in vivo conformation between growth on glucose and growth on ethanol, allowing to relate a number of these conformational changes to the central carbon metabolite fructose-1,6-bisphosphate \(^{43}\). However, we note that most protein-metabolite interactions are still being identified using (often laborious) in vitro activity assays, in which few putative interactions are tested in a typical study (figure 2, data from BRENDA database, a repository for kinetic information on enzymes).

**Distinguishing regulatory signals from regulatory chatter**

Thus, recent technological developments have boosted our ability to identify proteins that are regulated by posttranslational regulation. The next step is to identify those regulatory events that are actually relevant to a particular phenotype. If recent findings from the analysis of transcriptional regulatory networks are any indication, the fraction of metabolically relevant regulatory events could be rather small: adaptation to a perturbation often triggers transcriptional responses in hundreds of genes, few of which seem to be relevant for the perturbation at hand (e.g. \(^{44}\)). Given all this potential
regulatory “chatter” under a given condition, how can we map a regulatory event to its function in the in vivo context?

The identification of PTM sites or allosteric interactions which are highly conserved across species, or which have co-evolved, can help to select regulatory events which are more likely to be functionally relevant. However, conservation of a particular PTM site (or allosteric interaction) as such does not necessarily help to reveal its function. Consequently, efforts to address this issue try to relate posttranslational regulatory events to other physiological measures, such as fluxes or metabolite concentrations. For example, a recent work sought high correlations between the phosphorylation degree of central metabolic enzymes in yeast and their respective metabolic flux to predict phospho-sites that affect enzymatic activity. Another study combined untargeted metabolomics and phosphoproteomics data from a library of over 100 yeast kinase/phosphatase knockout strains to predict functional phospho-sites in 47 enzymes which consistently affect the concentrations of neighboring metabolites. A similar work in mammalian cells monitored the dynamics of insulin signaling using phospho-proteomics and metabolomics to reconstruct the insulin-dependent signal flow through the phosphorylation network to central carbon metabolism. While such correlation-based approaches tend to work well for PTM networks where the degree of the modification can be directly measured, applying them to identify relevant allosteric interactions would be challenging since they are non-covalent. Here, mathematical models can offer some help, as was recently illustrated in a work which combined dynamic metabolomics and ensemble-modelling to systematically infer allosteric interactions in E. coli’s glycolysis that are relevant in vivo during fast carbon source transitions.

Another commonly used approach to assess the relevance and function of a posttranslational regulatory event is to genetically modify the respective site on the target protein and to monitor the outcome in vivo. Such “organism minus one interaction” experiments have long been instrumental in validating the function of posttranslational regulatory events, and allow to assess the relative role of PTMs and allosteric interactions in regulating an enzyme’s activity. For example, Xu et al. studied the posttranslational regulation of the enzyme pyruvate kinase in yeast, which is subject to both phosphorylation and allosteric activation by fructose-1,6-bisphosphate. Using different mutant forms of this enzyme, the authors showed that during carbon starvation, its in vivo allosteric regulation by fructose-1,6-bisphosphate - but not the phosphorylation of the enzyme - is used to counter the increased substrate saturation due to the higher PEP concentration. Naturally, such genetic approaches are limited by our ability to modify microbial genomes, as well as our ability to predict mutations which will abolish the interaction of interest without impairing protein activity otherwise, but recent advances in genome editing and computational protein engineering should soon make systematic large-scale studies possible.
As we have illustrated in this section, the identification of functional posttranslational regulatory events in a sea of potential regulatory chatter remains a major challenge when trying to understand the role of posttranslational regulation in metabolism. Clearly, the strategies outlined in this section are not exhaustive, and some general methods for analyzing signaling networks could be applied to study metabolism as well. For example, Chasman et al. integrated phosphoproteomics, transcriptomics and growth fitness data in a probabilistic computational framework to infer the active signaling subnetworks of the transcriptional response to salt stress in yeast. We envision that such computational methods may prove to be powerful tools in the functional annotation of PTMs and allosteric interactions.

Coarse-grained posttranslational regulation?

As mentioned above, many microbial proteins are subject to highly condition-specific posttranslational regulation. One way to achieve such extensive and condition-specific regulation is to use a large network of regulatory proteins that integrate and process regulatory inputs, such as the phosphorylation network of yeast. However, many other posttranslational modifications seem to be transferred by a surprisingly small number of regulatory proteins. For example, acetylation in *E. coli* is mostly regulated by one acetyltransferase (Pat) and one deacetylase (CopB), which have been shown to control in particular acetate metabolism. Interestingly, it has recently become clear that some posttranslational modifications, such as acetylation and succinylation, can occur non-enzymatically in vitro through inherently reactive metabolites. For example, it was recently shown that protein acetylation in *E. coli* is globally increased during carbon starvation concomitantly with an increase in acetyl-phosphate. While such an increase in the global degree of acetylation could also be caused by hitherto unknown acetyltransferases, it is tempting to speculate that cells exploit reactive metabolites such as acetyl-phosphate, succinyl-CoA or 1,3-bisphosphoglycerate to directly link their metabolic activity to the coarse-grained regulation of protein activity. Likewise, the discovery that nucleotide phosphates (such as ATP) regulate many more glycolytic enzymes than previously thought raises the possibility that ATP and other regulatory metabolites could act as global allosteric regulators.

Metabolism as a source of phenotypic variability

Although cells can use many intertwined regulatory mechanisms to ensure robust metabolic adaptation to changing environments, these mechanisms can also generate phenotypic heterogeneity in clonal populations at the level of metabolism, as was recently illustrated in a series of remarkable studies. For example, Kotte et al. demonstrated that the transition from glycolytic to gluconeogenic carbon sources in *E. coli* relies on the activation of a positive feedback loop in central metabolism. Cells that do not manage to activate this feedback loop during the transition,
presumably due to fluctuations in the basal expression of gluconeogenic enzymes, are trapped in a non-growing persister state. Currently, it is not clear whether such phenotypes are merely “regulatory bugs” that only emerge in lab environments. Nevertheless, the fact that the non-growing E.coli subpopulation was also found to be much more resistant to antibiotic treatment suggests a potential bet-hedging strategy at population level which balances between fast growth and resistance to adverse conditions. Overall, the discovery of noise-generating processes at the level of metabolism opens up exciting new avenues in the study of cellular heterogeneity.

Conclusion

Our understanding of the role of posttranslational regulation in microbial metabolic decision making is still in its infancy. Although some metabolic functions can adapt passively to external flux changes, the emerging picture is that transcription manages resource allocation, and PTMs as well as allosteric protein-metabolite interactions coordinate metabolic fluxes. Further advances in our ability to map, quantify and functionally annotate such regulatory events will certainly deepen our understanding of the different roles of transcriptional and posttranslational regulation, and might also have broader implications for understanding processes underlying health and disease and in metabolic engineering. Tuning the expression of individual enzymes is often a key step in the optimization of a production pathway. However, approaches to identify and alleviate thermodynamic bottlenecks – e.g. by altering the cofactor specificity of the respective enzyme, by using an isozyme with different cofactor specificity, or by increasing the precursor/cofactor supply – might result in a better trade-off between product yields and enzymatic cost. Moreover, by targeting a pathway’s posttranslational regulation we may be able to tackle a fundamental regulatory issue that has long haunted metabolic engineers, as was recently illustrated by Van Heerden et al.: the flux through a biosynthetic pathway is typically inhibited by the level of its end product, a negative feedback ensuring that biosynthetic rate is tuned by demand. Thus, attempts to increase the flux in such a pathway by over-expressing it are likely to be suppressed by the native regulatory mechanisms. Indeed it was shown long ago that transferring a biosynthetic pathway to a different organism can greatly increase its production rate, presumably since it lacks the native “regulatory background”, and a recent study could increase the production of the building block malonyl-CoA in yeast by abolishing feedback through Snf1-dependent phosphorylation of acetyl-CoA carboxylase, the first committed enzyme in fatty acid biosynthesis.

The prevalence of post-translational regulation in nature suggests that there is much to gain by harnessing it to control engineered pathways. Novel approaches to alter a pathway’s regulatory structure, for example by engineering allosteric interactions, might be used to facilitate the growth of cells in very large bioreactors where nutrients are not uniformly distributed or when being fed a complex medium such as industrial sludge or oceanic oil-spills.
Chapter 6 – Posttranslational regulation of microbial metabolism

Figures

Figure 1. When is regulation actually needed for controlling flux? (A) The stoichiometric space of feasible flux states is defined by the set of expressed enzymes, but the condition-specific fluxes are largely determined by enzyme kinetics and all layers of regulation including post-translational ones. (B) When an enzyme is not saturated, the flux will be sensitive to the concentration of substrate, but higher levels of the enzyme will be needed to achieve the same flux as in the saturated case. (C) Similarly, for the same net flux, a reaction closer to thermodynamic equilibrium ($\Delta G' \sim 0$) will be more sensitive to changes in substrate and product levels, but will require relatively higher amounts of the enzyme due to the counter-productive backward flux. (D) Branch-points are particularly sensitive to changes in enzyme levels, especially if a precise flux-ratio is required. (E) In a pathway comprised of reversible reactions the metabolite concentrations can partially compensate for the stochasticity in enzyme expression levels. If some reactions are irreversible, however, a significant imbalance between consecutive reactions can cause a deleterious accumulation of the intermediate compound, or a severe depletion that would choke all downstream reactions.
Figure 2: Activating and inhibiting compounds registered in the BRENDA database. Left panel: number of activators and inhibitors reported every year since 1950 until 2011. Right panel: despite the recent advent of high-throughput methods to screen protein-metabolite interactions, still the vast majority of reported effectors were identified in publications which contain only few such discoveries.
References


Chapter 7 – Systematic identification of protein-metabolite interactions in complex metabolite mixtures by ligand-detected NMR

Yaroslav Nikolaev*, Karl Kochanowski*, Hannes Link, Uwe Sauer, Fred Allain

*: equal contribution

Contributions:
YN, KK, HL and US designed the study. YN and KK wrote the chapter, with contribution from all other authors. YN and KK performed all experiments and analyses.
Chapter 7 – Systematic identification of protein-metabolite interactions in complex metabolite mixtures by ligand-detected NMR

Abstract

Protein-metabolite interactions play a vital role in the regulation of numerous cellular processes. Consequently, identifying such interactions is a key requisite for understanding cellular regulation. However, the non-covalent nature of the binding between proteins and metabolites has so far hampered the development of methods to systematically protein-metabolite interactions, and the few available, largely mass-spectrometry based, approaches are restricted to specific metabolite classes, such as lipids. In this study, we address this issue and show the potential of ligand-detected NMR, a method routinely used in drug development, to systematically identify protein-metabolite interactions. As a proof-of-concept, we select four well-characterized bacterial and mammalian proteins (AroG, Eno, PfkA, BSA) and identify metabolite binders among complex mixes comprising of up to 33 metabolites. Ligand-detected NMR captures the vast majority of reported protein-metabolite interactions, and detects a number of novel ones, such as promiscuous binding of the highly negatively charged metabolites citrate, AMP, and ATP, as well as binding of L-tryptophan and L-tyrosine to AroG. Using in vitro enzyme activity assays, we further assess the functional relevance of these novel interactions in the case of AroG and show that L-tryptophan and L-tyrosine act as novel inhibitors of AroG activity. Thus, we conclude that ligand-detected NMR is suitable for the systematic identification of functionally relevant protein-metabolite interactions.
Chapter 7 – Systematic identification of protein-metabolite interactions in complex metabolite mixtures by ligand-detected NMR

Introduction

Interactions between proteins and metabolites are pivotal for the regulation of many diverse cellular processes, such as metabolism 1, gene expression 2,3, and chromatin remodeling 4, allowing cells to mount regulatory responses based on their current metabolic state. Therefore, approaches to systematically map such interactions are a key prerequisite to understanding cellular regulation 1.

However, the non-covalent nature of protein-metabolite interactions makes them notoriously difficult to detect experimentally. Consequently, compared to the plethora of available methods to detect other types of biological (i.e. protein-protein or protein-DNA 1) interactions, the development of equivalent methods for the detection of protein-metabolite interactions has so far lagged behind 5-7. Although recent advances towards this end have led to the development of few methods to identify such interactions, they are largely restricted to specific metabolite classes, such as lipids 8,9, or rely on the indirect identification of protein-metabolite interactions, for example through metabolite-induced changes in protein conformation 10 or stability 11, which are not necessarily indicative of functional protein-metabolite interactions 12. Other indirect methods rely on detecting the sequestration of free metabolites through protein binding, but require equimolar amounts of proteins and metabolites, restricting their utility to proteins which can be easily purified in large amounts and are stable at high concentrations 13. Thus, due to the lack of generally applicable systematic approaches, protein-metabolite interactions are still largely being identified using laborious in vitro activity assays 7, which are often not easily amenable to non-enzymatic proteins.

Complementary to these largely mass-spectrometry based approaches, interactions between biomolecules can be analyzed using a range of Nuclear Magnetic Resonance (NMR)-based techniques. In particular, interactions between large molecules (e.g. proteins) and small molecules (e.g. metabolites) can be directly identified by “ligand-detected NMR” methods such as saturation transfer difference (STD) NMR 14, water-ligand observed via gradient spectroscopy (water-LOGSY) 15 and T1Rho relaxation NMR 16. These methods allow the direct detection of ligand binding to a purified protein, without any labeling of either component, and without the need to perform activity based in vitro assays. Ligand-detected NMR has been primarily developed in the context of high-throughput screening of synthetic compound libraries, thereby facilitating drug discovery 17-21. However, outside of drug discovery applications, it has so far not been explored for the systematic identification of novel functional interactions of endogenous metabolites with proteins, with example studies focusing on few selected proteins or metabolites 22,23, or the protein-centric comparison of ligand-binding profiles to identify functionally related proteins 24.

Here, we use ligand-detected NMR to systematically identify interactions between well characterized bacterial and mammalian proteins and endogenous metabolites in vitro. Using two complementary
NMR methods, we recover most of the known protein-metabolite interactions from complex metabolite mixtures comprising up to 33 chemically diverse metabolites, and identify several potential novel interactions. Further, we validate the functional relevance of two of these novel interactions, namely between the protein AroG and the metabolites L-tryptophan and L-tyrosine, using enzymatic activity assays. Overall, this study demonstrates the power of NMR for the systematic identification of endogenous protein-metabolite interactions.
Results

Selection of target proteins. To assess the utility of ligand-detected NMR for identifying protein-metabolite interactions, we selected three metabolic enzymes from E. coli with known allosteric effector metabolites, namely, phosphofructokinase I (PfK), enolase (Eno), two enzymes of glycolysis, as well as 2-dehydro-3-deoxyphosphohexonate aldolase (AroG), which catalyzes the first step in chorismate biosynthesis, and purified them from a library of his-tagged overexpression strains 25. To illustrate that this approach is not restricted to enzymes, and to provide a widely available benchmark protein, we further selected a well-characterized non-enzymatic protein of comparable size, namely bovine serum albumin (BSA) (see Table 1 for full list of proteins with molecular sizes and reported interactions). A key challenge for approaches to identify protein-metabolite interactions is the vast chemical diversity of endogenous metabolites. We therefore selected seven chemically diverse metabolites (Table 2, left panel) which include known interactors such as L-phenylalanine, an allosteric inhibitor of AroG, as well as reported regulatory metabolites 1,26.

Identification of protein-metabolite interactions by two complementary ligand-detected NMR methods. To systematically identify interactions between selected proteins and metabolite mixtures, we used two commonly used 27 NMR methods, namely water-LOGSY 15 and T1rho relaxation 16. Although these methods utilize different physical mechanisms, both detect the change of certain NMR properties of a (small-sized) metabolite when it is bound to a large-sized target, a protein in our case (figure 1). Both water-LOGSY and T1rho relaxation rely on an interaction- and Kd-dependent signal amplification effect, which is strongest for interactions with Kd’s in the low-micromolar range, and which decreases at higher (up to ~10 mM) and lower Kd’s (up to ~0.1 nM) 27. More specifically, this amplification effect is strongest when a large fraction of metabolite molecules bind and dissociate from the protein target in the duration of the experiment. This condition is fulfilled for interactions with low-µM Kd’s (assuming kon is diffusion limited). If the interaction is much stronger (Kd << µM), the protein-metabolite complex does not dissociate quickly enough, so only a small fraction of metabolites bind to the protein over the experimental time. Conversely, if the interaction is much weaker, the lifetime of the protein-metabolite complexes is too short to give significant perturbation of metabolite properties, so again only a small fraction of the metabolite signal is affected. To quantify this amplification effect and to facilitate the comparison of signals obtained from different metabolites and experiments, we defined the metric ‘fractional signal intensity’, which reflects how much of the metabolite signal is altered in a water-LOGSY or T1rho experiment, when compared to a simple 1D reference experiment (see Methods and figure 1). When compared across all our experimental data, this metric indeed shows a good correlation for the same interaction hits observed independently by water-LOGSY and T1rho experiments (Supplementary Figure S4A).
Using this metric to identify metabolite-protein interactions, we tested each protein against the seven-metabolite mix. We found that both methods yielded similar interaction maps (Figure 2B and 2D), suggesting that the identified interactions are robust against method-dependent variability. Notably, all previously reported protein-metabolite interactions were recovered. Additionally, both methods detected several hitherto unreported interactions, with differences between the tested proteins: BSA and AroG exhibited significant interactions with larger number of metabolites than Eno and PfkA (5 vs 2 hits in this 7-compound mixture). Notably, citrate (CIT) and adenosine-monophosphate (AMP) showed interactions with three (CIT) and four (AMP) out of four tested proteins, suggesting that these metabolites show higher tendency for binding than the other tested metabolites.

To exclude that these observations reflect unspecific changes of the metabolite properties in presence of any protein target (e.g. due to unspecific self-aggregation of metabolites facilitated by any protein interface), we further tested recombinant green fluorescent protein (GFP), which has no reported interactions with metabolites. We found that indeed none of the tested metabolites were found to interact with GFP, with the exception of a very weak interaction with AMP in the water-LOGSY experiment (Figure S1). Note that due to its smaller molecular size (Table 1), in the employed NMR experiments GFP yields generally weaker interaction signals than the other tested proteins. Nevertheless, together with the marked differences in detected interactions between the other tested proteins, these results suggest that the detected interactions mostly constitute specific metabolite binding to the respective proteins.

**Recovering protein-metabolite interactions in complex metabolite mixtures.** A key property of many protein-metabolite interactions is that they can be often strongly affected by the presence of other metabolites, which compete for the same protein binding site, or which make the protein more sensitive for the presence of a particular metabolite 28. To assess the potential impact of such interference, we tested each protein with two more complex mixes comprising 15 and 33 central metabolites (Table 2, Supplementary Figure S2). Despite increased crowding in the resulting spectra, we found most of the previously identified interactions to be unambiguously recovered in the more complex metabolite mixes (example for AroG shown in Figure 3A, summary for all proteins in Figure 3B). In the water-LOGSY experiment, of the 13 interactions identified in the 7-metabolite mix (total sum for all four tested proteins), 4 and 5 reduced their signal below the confident detection limit (S/N ≥ 3) in mixtures of 15 and 33 metabolites, respectively. In the T1rho experiment, out of 15 detected interactions, 7 and 6 were below the confident detection limit in 15- and 33-metabolite mixes, respectively, presumably due to competition for the binding sites or spectral crowding (overlapping peaks due to very high density of signals in the given region of the spectrum). Nevertheless, similar to the initial experiments, ligand-detected NMR identified most of the reported interactions (Figure 3B)
as well as a number of novel interactions in these complex mixtures. Taken together, these results show that most of the identified interactions are robust against potential interference by other metabolites, allowing us to generate systematic interaction maps between proteins and metabolites even in complex metabolite mixtures.

**Effective affinity range of detectable protein-metabolite affinities.** An important criterion to assess the functional relevance of a protein-metabolite interaction in the cellular context is a metabolite’s affinity to protein, which allows to determine the degree of saturation for the respective interaction if the metabolite’s *in vivo* concentration is known \(^{29}\). We therefore wanted to assess the effective affinity range of the detectable protein-metabolite interactions. Comparison of the detected interactions with corresponding reported \(K_d/K_m\) values (Table S1) revealed that we could detect interactions with affinities as strong as 9 µM (AroG-PEP). To also assess the effective lower affinity limit, we further quantified the \(K_d\) value of one of the weakest interactions in our data set, namely between Pfka and FBP (supplementary figure S3), and found it to be above 1 mM (1.15±0.29 mM). This effective detectable affinity range (~1 µM to ~1mM) is well in agreement with the aforementioned sensitivity range of Water-LOGSY and T1rho relaxation experiments. Given the fact that the \(K_m\) values of enzymes \(^{30}\) as well as metabolite concentrations in bacterial central carbon metabolism \(^{29}\) are typically larger between 1 µM and 1 mM (i.e. the concentrations of the metabolites tested here), we conclude that the ligand-detected NMR approach presented here is well suited to detect most *in vivo* relevant protein-metabolite interactions.

**Validation of novel protein-metabolite interactions by in vitro activity assays.** Finally, we focused on functionally verifying the novel protein-metabolite interactions that had been identified by ligand-detected NMR. To check whether these novel interactions also affect the activity of the respective proteins, we performed *in vitro* activity assays in presence or absence of the potential interactor. In particular, we focused on the proposed interactions with AroG. Using enzymatic *in vitro* assays, we tested the effect of seven metabolites, which were identified to consistently bind AroG with both NMR methods, on AroG activity and found that several of these metabolites indeed caused a reduction in enzyme activity (Figure 4). The strongest effect on AroG activity was found for L-phenylalanine, its reported allosteric regulator. Moreover, addition of L-tryptophan or L-tyrosine caused a 20% reduction in AroG activity. In contrast, the promiscuous binding metabolites citrate and AMP did not affect AroG activity. Taken together, these results confirm that ligand-detected NMR is suitable to identify functionally relevant protein-metabolite interactions.
Discussion

The systematic identification of protein-metabolite interactions remains a key challenge in the investigation of cellular regulatory processes. In this work, we present ligand-detected NMR as a widely applicable approach to detect such interactions in vitro. As a proof-of-concept, we used water-LOGSY and T1rho relaxation NMR experiments to systematically detect binding between four well-characterized enzymatic and non-enzymatic proteins and several metabolite mixtures comprising up to 33 metabolites. We could recover all reported interactions in the basic (7-metabolite) mixtures, and a majority (8/12=75%) of the reported interactions even in the complex (33-metabolite) mixtures. Moreover, we identified a number of novel interactions, most notably promiscuous binding of nucleotide phosphates and citrate, and binding of all aromatic amino acids to the protein AroG. Finally, using in vitro activity assays, we determined the impact of these novel interactions on protein activity and found that, besides its reported allosteric effector L-phenylalanine, AroG activity is also affected by L-tryptophan and L-tyrosine.

One surprising result in this study is the promiscuous binding of the nucleotide phosphates AMP and ATP. These metabolites showed binding to all proteins tested, albeit to different extent (their signals were affected much stronger in presence of BSA, AroG and PfKA compared to eno and GFP). Interestingly, in our data sets addition of ATP reduced AMP binding to PfKA and AroG enzymes (see intensities of corresponding hits in Figure 3B), but not to BSA protein. This suggests that triphosphate is favored in the first two cases, but cannot be stably accommodated in the nucleotide binding side(s) of BSA. Notably, a recent study showed that nucleotide phosphates regulate the activity of many glycolytic enzymes in yeast 31, and nucleotide phosphates may exert similar pleiotropic effects in E.coli. The physiological relevance of promiscuous binding of citrate is less clear. One possibility is that citrate, which has several acidic oxygen moieties and thus may ‘mimic’ phosphate groups, may bind to the same phosphate-binding sites on the protein surface and thereby modulate the effect of nucleotide phosphates on protein activity. However, since neither AMP nor citrate were found to change the activity of the one tested protein, namely AroG, we cannot exclude that the identification of promiscuous binding of these metabolites may constitute technical artifacts of the employed NMR experiments and our sample conditions. For example, the relatively low ionic strength of the buffer may lead to electrostatic enhancement of $k_{\text{on}}$ rates for these strongly charged compounds, and/or cause unspecific binding of highly polar metabolites to protein surfaces. In case of citrate, the strong couplings present in its spin system could result in additional alteration of the observed signals even upon weak binding to protein, which is not taken into account in current analyses of water-LOGSY and T1rho relaxation experiments. Moreover, these promiscuous interactions may alter other protein
properties, such as complex formation or aggregation, which do not directly translate into activity changes in vitro.

Another surprising result is the binding of L-tryptophan and L-tyrosine to AroG. Although AroG activity was only weakly affected by these metabolites, *E. coli* has two additional isoenzymes of AroG, namely AroH and AroF, which exhibit strong allosteric regulation by L-tryptophan and L-tyrosine, respectively \(^{32,33}\). Given the high sequence similarity of these three iso-enzymes \(^{32}\), it is tempting to speculate that these metabolites do indeed bind all three iso-enzymes, but differ in the extent to which they affect each enzyme’s activity due to small differences in protein structure. Overall, our results suggest that systematic screens based on ligand-detected NMR may allow to shed light on the relationship between a protein’s sequence and its susceptibility to regulation by a particular allosteric effector, or even whole classes of metabolites with similar chemical properties.

To facilitate quantitative comparison of detected protein-metabolite interactions, we defined a fractional signal intensity metric – a measure of how much an individual signal (resonance) of a metabolite is altered in presence of the protein target in a water-LOGSY or T1rho relaxation experiment in comparison with a basic 1D reference experiment (see Methods). Similar parameters for specific experiment types were used in NMR studies before \(^{34,35}\), but to our knowledge no NMR experiment-independent metric has been defined in the literature. Fractional signal intensity reflects how many molecules of a specific metabolite have interacted with the (protein) target during the experimental time. Under the common sub-stoichiometric protein-metabolite concentrations, interactions with very high (~1 nM) and very weak (~10mM) affinities approach a fractional signal near zero. The maximum is observed when a large fraction of metabolite molecules bind to the protein for long enough to have their NMR properties changed, and dissociate fast enough to allow other metabolite molecules to bind and change their properties. For diffusion-limited \( \text{k}_{\text{on}} \) values (~10^8 M\(^{-1}\) s\(^{-1}\)) this is achieved for interactions with low-\( \mu \)M \( K_{d} \)’s (\( k_{\text{on}}\approx 100 \text{ s}^{-1} \), residence time \( \approx 10 \text{ ms} \)) \(^{36}\). Thus, to a rough approximation the fractional signal intensity in our setup reflects how close a protein-metabolite interaction is to the low-\( \mu \)M \( K_{d} \) range. The approximation comes from the fact that such fractional signal will also depend on other parameters, with the size and shape of the (protein) target being the most notable one, as the rate at which ligands’ NMR properties are altered upon binding to the target is higher for targets which tumble slower in the solution. Other parameters include the number of interaction sites on the target, potential allosteric interactions between these sites, and physico-chemical properties of the interactions and metabolite molecule in vicinity of the observed nucleus (this influences chemical shifts, spin relaxation rates and exchange broadening of the observed signals). For water-LOGSY also the structural topology of protein-metabolite interaction interface and properties of exchangeable protons may have strong influence. Despite the various parameters which
have to be taken into account, it seems possible that with sufficient amount of experimental data (i.e. fractional signal intensities and Kd values measured for a range of protein-metabolite interactions) the relationship between Kd and fractional signal intensity can be defined with reasonable accuracy, allowing to estimate interaction Kd’s from single measurements (e.g. in context of high-throughput screens). Notably, for simpler line-broadening NMR experiments a proof-of-principle for such estimates has already been established 35.

In this work, we have used two complementary ligand-detected NMR methods. Despite working via different mechanisms, both water-LOGSY and T1rho relaxation depend on alteration of certain NMR properties of a small-size metabolite upon its binding to a large-size target. We found that both methods showed generally good agreement regarding the detection of protein-metabolite interactions and effective detectable affinity range, suggesting that using only one method should be sufficient to detect such interactions in high-throughput applications. The choice of method may then depend on the chemical properties of the protein-metabolite interactions of interest and prior information about the chemistry and structural topology of possible interaction. Water-LOGSY depends to a certain extent on properties of structured water molecules, exchangeable protons and dipolar interactions at the binding site 36, making it more discriminative to the chemistry and topology of interaction, especially for non-polar metabolites. T1rho relaxation depends largely on the macroscopic properties of the target-ligand complex and therefore should not show such discrimination. Furthermore, not depending on H2O for magnetization transfer, T1rho relaxation experiments could potentially be performed in D2O, thereby increasing robustness of the experimental setup (no need for suppression of the strong H2O solvent signal) and improving sensitivity for signals located near the H2O frequency, which may be masked otherwise. Under the experimental conditions employed in this study, both absolute and fractional signal intensities observed for the same interactions in water-LOGSY and T1rho relaxation (Figure S4) are very similar, not allowing to favor either of the two methods based on its sensitivity. But the above considerations suggest that T1rho relaxation experiments may be more suitable in high-throughput applications.

A key advantage of ligand-detected NMR over classical activity-based assays is the possibility to test the impact of several metabolites simultaneously. The results in this study highlight the trade-off between simplicity of data interpretation and throughput: testing only few metabolites simultaneously simplifies subsequent data analysis and peak assignment at the expense of limited throughput and increased consumption of often scarce protein. Conversely, increasing the number metabolites will inevitably lead to ambiguous assignment of potential protein-metabolite interactions due to spectral overlap of chemically similar compounds. Also, compounds may compete for the same binding site on the target (e.g. protein), reducing the observed signal intensities of one another and increasing the
Chapter 7 – Systematic identification of protein-metabolite interactions in complex metabolite mixtures by ligand-detected NMR

likelihood of false-negatives. This is especially critical if tight binders are present in the mixture, which block the binding interface(s) for access of weak binders. As an illustrative example, in our measurements AMP signals were strongly attenuated in presence of ATP in experiments with PfkA and AroG (Figure 3B – compare 7-compound and 15-compound mixtures). Nevertheless, most interactions were recovered even when the complexity of the metabolite mixture was increased. Moreover, given that cells essentially constitute highly complex mixtures of thousands of different metabolites, using more complex/realistic metabolite mixtures may actually allow to distinguish robust interactions which are likely to be relevant in vivo from interactions that only occur in vitro.

One limitation of the presented ligand-detected NMR methods is their restriction to detecting protein-metabolite binding, which does not directly translate into changes in protein activity. NMR can also be used to determine a protein’s activity, but without additional labeling of substrate molecules the sensitivity of such NMR-based enzymatic assays is likely to be lower than the sensitivity of common colorimetric, mass-spectrometric or fluorescence-based assays. Therefore, we envision that ligand-detected NMR could be used to narrow down the number of potential interactions that are subsequently tested using targeted activity assays. Another limitation of the presented methods is the requirement for moderate amounts of metabolites and proteins – with common NMR instrumentation requiring minimum ~10 nmol of a compound at a concentration of around 30-40 μM (with a sample volume of 300 μl) and at least ~1 nmol of protein at minimum concentration ~1 μM to stay within reasonable experimental durations (around 1 hour per sample). This makes the investigation of protein complexes, which can often only be purified in minute amounts, more challenging. However, recent advances in genome editing, which allow the simultaneous purification of multiple proteins by addition of affinity tags, may enable the larger-scale purification of protein complexes, or proteins for which no overexpression library is currently available. Moreover, since NMR is non-destructive, the samples can be analyzed multiple times and stable sample components can be recovered. A third limitation of the ligand-detected NMR methods used here is the minimum size of the protein. Since the key characteristic differentiating free vs bound metabolites in these methods is the rate of tumbling of these compounds in solution (the protein-metabolite complex tumbles substantially slower than the free metabolite, giving rise to differences in the NMR properties of free and bound metabolite), the reasonable minimum protein size for ligand-detected NMR experiments is considered to be about 10-30 kDa. However, bacterial and eukaryotic enzymes are typically larger than 10 kDa. Moreover, for proteins with sizes below ~30 kDa a range of alternative, target-detected, NMR methods is available (e.g. SAR by NMR), which probe the same binding events by looking at changes in the protein’s NMR properties.
Taken together, the results presented in this study show that ligand-detected NMR is a suitable complementary approach to map protein-metabolite interactions on a system-level scale, and can further guide the biochemical characterization of proteins. For example, ligand-detected NMR may be used to systematically identify interactions between endogenous metabolites and regulatory proteins, such as eukaryotic protein kinases/phosphates, to unravel potential feedback from metabolism to cellular regulatory layers 42.

Acknowledgements
Authors thank the Novartis NIBR team for providing pulse-programs for water-LOGSY and T1rho experiments.
Material and Methods

Reagents and strains. Unless stated otherwise, chemicals were obtained from Sigma-Aldrich. Overexpression strains for AroG, PfkA, and Enolase proteins from *E. coli* were obtained from the ASKA library. The GFP overexpression plasmid pTrc99KK-GFP with an N-terminal His6x-tag was constructed using GFPmut2 from pUA66 as a template, and pTrc99KK as a vector (primer 1: GCCTCGAGATGCATCATCATCATCATCATATGTCTAAAGGTGAAGAATTATTC, primer 2: CCGGATCCTATTGTACAATTCCATCCAC), and transformed to the *E.coli* strain BW25113 by electroporation.

Cultivation. 50 mL LB shake flask cultures were inoculated 1:100 with LB precultures, and expression was induced with 0.2 mM IPTG. Cultures were incubated for 16h at 37°C with shaking (300 rpm). Cells were harvested by centrifugation (5000 g, 4°C, 15 min), washed once with 0.9% NaCl and 10 mM MgSO₄ and concentrated tenfold in lysis buffer (20 mM sodium phosphate buffer, pH 7.4, 500 mM NaCl, 20 mM imidazole, 2 mM dithiothreitol, 1 mM MgCl₂, 4 mM phenylmethylsulfonylfluorid). Cells were disrupted by passage through a French press three times at 4°C, and cell extracts were separated from cell debris by centrifugation (20000 g, 4°C, 30 min). His-tagged proteins were purified from cell extracts using nickel-sepharose gravity flow columns (GE Healthcare), and the elution buffer was replaced by the respective assay buffer (50 mM potassium phosphate buffer, pH 7.5, 10 mM MgCl₂) using filter columns with 10 kD cutoff (Millipore). Lyophilized bovine serum album was obtained from Sigma-Aldrich and resuspended in assay buffer (50 mM potassium phosphate buffer, pH 7.5, 10 mM MgCl₂).

Sample preparation and NMR measurements. Protein concentrations were measured based on their specific extinction coefficients at 280 nm immediately before NMR sample preparation. Final protein and metabolite concentrations were 30 μM and 200 μM respectively. In Kₐ determination experiments proteins were 10 μM. All samples were 425-500ul in 5mm-TA tubes (ARMAR Chemicals), and contained 5% D2O and 25 μM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). In presence of BSA protein, DSS signal appeared broad and shifted, suggesting an interaction with the protein, hence alternative chemical shift referencing was used in this case (see below).

NMR Measurements were performed on Bruker Avance III-HD 600 MHz spectrometer using CPTCI cryo-cooled probehead. A set of experiments with GFP protein (Supplementary Figure S1) was recorded on Bruker Avance III 750 MHz spectrometer with a TXI room-temperature probe. At the beginning of all experimental series temperature was calibrated to 298K using 99.8% Methanol-d4 sample. Pulse-programs for water-LOGSY and T1rho were adapted from experiments of the Novartis NIBR NMR team (Basel, Switzerland). The Polarization-Optimized WaterLOGSY experiment was used.
Water-LOGSY experiment detects changes in dipolar interactions between spins (the nuclear Overhauser enhancement, NOE), and T1rho detects an increase in magnetization decay rate (relaxation rate) in metabolites. After subtracting free protein and free metabolite spectra (details below), the interaction hits appear as positive signals in the final difference spectrum. Pulse-programs and examples of experimental datasets are in Supplementary.

**NMR processing and data analysis.** Spectra were processed in TopSpin 3.2 (Bruker) using custom-built Python scripts to process, calibrate and generate difference spectra (see Supplementary materials). The final difference spectra correspond to: \( I_{\text{metab+protein}} - I_{\text{metab}} - I_{\text{protein}} \). In T1rho individual intensities were taken as a difference of experiments with short and long spin-lock times \( I_{10\text{ms}} - I_{200\text{ms}} \). The corresponding difference in water-LOGSY experiment was implemented in the pulse-program itself and therefore did not require additional subtraction during data analysis. The phase of the spectra was set to obtain hits as positive signals in the final difference. Calibration of spectra to DSS was critical to minimize subtraction artifacts in final spectra. In case of BSA protein, due to its apparent interaction with DSS, the spectra were referenced against a ~1.2 ppm singlet peak from a trace contaminant present in all our water samples. Identification, S/N quantification, assignment and disambiguation of interaction hits was performed using custom-built Matlab scripts.

**Hit confidence cutoff based on Signal-to-Noise quantification.** Only peaks showing a signal-to-noise ratio \( \geq 3 \) after local vs global noise thresholding were considered for the analysis. Global noise was calculated as standard deviation of intensities in empty regions of the spectrum. Local noise was calculated as the standard deviation of intensities within \( \pm 4 \) peak widths around the specific signal, minus the median intensity in this region – to separate contribution of other positive signals in this region from the actual noise. Peak width was fixed at 0.005 ppm, linewidth of a singlet proton signal at half-height, after application of 1Hz-broadened exponential apodization function. The boundary between the signal and local noise regions was defined by the signal decay to zero - first point with zero or negative intensity outside of signal region (\( \pm 1 \) peak width around the signal maximum). Local noise was used for S/N calculation in cases when its value exceeded the global noise value by 3-fold (indicating that the peak is in a crowded region of the spectrum).

**Fractional signal intensity.** To get a measure of how many metabolite molecules interacted with the protein target for a significant amount of time (i.e. to approximate the combination of kinetic and equilibrium constants characterizing the given protein-metabolite interaction), the signal intensities in the final difference water-LOGSY or T1rho spectra were normalized against the corresponding maximum attainable signal intensities. The latter were estimated from the reference 1D spectrum (excitation sculpting water suppression experiment, ‘zgesgp’ in the Bruker library). The reference
spectrum was also used as a difference spectrum, to remove the contribution of protein signals. General formula for the calculation of the fractional signal intensity:

\[
\frac{\{I_{metab+protein} - I_{metab} - I_{protein}\}_{\text{wLOGSY or T1rho}}}{\{I_{metab+protein} - I_{protein}\}_{1D \text{ reference}}}
\]

Similar metrics for specific NMR experiment types were applied in other studies to facilitate compound affinity ranking\(^{34,35}\).

**In vitro enzyme assays.** Enzymatic in vitro assays for AroG were performed as described previously\(^ {33}\). Briefly, assays were performed at room temperature in reaction buffer (50 mM potassium phosphate buffer, pH 7.5, 10 mM MgCl\(_2\)) with 100 \(\mu\)M PEP and 100 \(\mu\)M E4P. Reactions were started by addition of purified AroG enzyme (final conc in assay 2.7 \(\mu\)g/mL, corresponding to a monomer concentration of 71 nM), and the decrease in PEP concentration was monitored photometrically at 232 nm every 6 seconds. Initial reaction velocities (within the first 60 seconds) were then determined by linear regression.
### Tables

**Table 1.** Overview of known protein-metabolite interactions in this study.

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Monomer size [kDa]</th>
<th>Native size [kDa]</th>
<th>Substrate/product</th>
<th>Known allosteric effectors/binding metabolites</th>
</tr>
</thead>
<tbody>
<tr>
<td>phosphofructokinase I (PfkA)</td>
<td>35</td>
<td>140 (tetramer)</td>
<td><em>F6P, ATP, FBP, ADP</em></td>
<td>PEP, <em>ADP, GDP</em></td>
</tr>
<tr>
<td>enolase (Eno)</td>
<td>46</td>
<td>92 (dimer)</td>
<td>2*-phosphoglycerate,*</td>
<td>PEP</td>
</tr>
<tr>
<td>2-dehydro-3-deoxyphosphoheptonate</td>
<td>38</td>
<td>152 (tetramer)</td>
<td>PEP, <em>erythrose-4-phosphate,</em></td>
<td>L-phenylalanine,</td>
</tr>
<tr>
<td>aldolase (AroG)</td>
<td></td>
<td></td>
<td><em>3-deoxy-7-phosphoheptonate</em></td>
<td>L-alanine</td>
</tr>
<tr>
<td>GFP</td>
<td>27.6</td>
<td>27.6 (monomer)</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>bovine serum albumin (BSA)</td>
<td>69</td>
<td>104 (monomer-dimer)</td>
<td>n/a</td>
<td>ATP, AMP, citrate,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>L-histidine, pyruvate,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>L-phenylalanine, <em>lactate</em></td>
</tr>
</tbody>
</table>

For PfkA, Eno, and AroG, information was obtained from ECOCYC \(^47\) as well as BRENDA data bases. For BSA, known metabolite binders were obtained by literature research: ATP, AMP \(^48\). Moreover, we included metabolites known to bind human serum albumin (HSA): histidine, pyruvate, phenylalanine, citrate \(^23\). *Italics*: interaction not tested.
Table 2. Metabolite mixes used in this study.

<table>
<thead>
<tr>
<th>7-metabolite mix</th>
<th>15-metabolite mix</th>
<th>33-metabolite mix</th>
</tr>
</thead>
<tbody>
<tr>
<td>pyruvate</td>
<td>pyruvate</td>
<td>pyruvate</td>
</tr>
<tr>
<td>citrate</td>
<td>citrate</td>
<td>citrate</td>
</tr>
<tr>
<td>2-oxoglutarate</td>
<td>2-oxoglutarate</td>
<td>2-oxoglutarate</td>
</tr>
<tr>
<td>FBP</td>
<td>FBP</td>
<td>FBP</td>
</tr>
<tr>
<td>shikimate</td>
<td>shikimate</td>
<td>shikimate</td>
</tr>
<tr>
<td>L-phenylalanine</td>
<td>L-phenylalanine</td>
<td>L-phenylalanine</td>
</tr>
<tr>
<td>AMP</td>
<td>AMP</td>
<td>AMP</td>
</tr>
<tr>
<td>L-glutamine</td>
<td>L-glutamine</td>
<td>L-glutamine</td>
</tr>
<tr>
<td>malate</td>
<td>malate</td>
<td>malate</td>
</tr>
<tr>
<td>6PG</td>
<td>6PG</td>
<td>6PG</td>
</tr>
<tr>
<td>G6P</td>
<td>G6P</td>
<td>G6P</td>
</tr>
<tr>
<td>PEP</td>
<td>PEP</td>
<td>PEP</td>
</tr>
<tr>
<td>phenylpyruvate</td>
<td>phenylpyruvate</td>
<td>phenylpyruvate</td>
</tr>
<tr>
<td>ATP</td>
<td>ATP</td>
<td>ATP</td>
</tr>
<tr>
<td>NAD+</td>
<td>NAD+</td>
<td>NAD+</td>
</tr>
<tr>
<td>L-tryptophan</td>
<td>L-tryptophan</td>
<td>L-tryptophan</td>
</tr>
<tr>
<td>L-tyrosine</td>
<td>L-tyrosine</td>
<td>L-tyrosine</td>
</tr>
<tr>
<td>L-histidine</td>
<td>L-histidine</td>
<td>L-histidine</td>
</tr>
<tr>
<td>L-arginine</td>
<td>L-arginine</td>
<td>L-arginine</td>
</tr>
<tr>
<td>L-lysine</td>
<td>L-lysine</td>
<td>L-lysine</td>
</tr>
<tr>
<td>L-glutamate</td>
<td>L-glutamate</td>
<td>L-glutamate</td>
</tr>
<tr>
<td>L-aspartate</td>
<td>L-aspartate</td>
<td>L-aspartate</td>
</tr>
<tr>
<td>L-asparagine</td>
<td>L-asparagine</td>
<td>L-asparagine</td>
</tr>
<tr>
<td>L-serine</td>
<td>L-serine</td>
<td>L-serine</td>
</tr>
<tr>
<td>L-threonine</td>
<td>L-threonine</td>
<td>L-threonine</td>
</tr>
<tr>
<td>L-valine</td>
<td>L-valine</td>
<td>L-valine</td>
</tr>
<tr>
<td>glycine</td>
<td>glycine</td>
<td>glycine</td>
</tr>
<tr>
<td>L-alanine</td>
<td>L-alanine</td>
<td>L-alanine</td>
</tr>
<tr>
<td>L-leucine</td>
<td>L-leucine</td>
<td>L-leucine</td>
</tr>
<tr>
<td>L-isoleucine</td>
<td>L-isoleucine</td>
<td>L-isoleucine</td>
</tr>
<tr>
<td>L-proline</td>
<td>L-proline</td>
<td>L-proline</td>
</tr>
<tr>
<td>L-methionine</td>
<td>L-methionine</td>
<td>L-methionine</td>
</tr>
<tr>
<td>L-cysteine</td>
<td>L-cysteine</td>
<td>L-cysteine</td>
</tr>
</tbody>
</table>
Chapter 7 – Systematic identification of protein-metabolite interactions in complex metabolite mixtures by ligand-detected NMR

Figures

**Figure 1. Principle of ligand-detected NMR.** The two NMR methods used in this study, namely water-LOGSY and T1rho relaxation, identify protein-metabolite interactions by detecting changes in the NMR properties of a metabolite (M) when bound to a protein (P) (see left and middle panel). The magnitude of these changes depends on the strength of the interaction and is manifested in the intensity of the subtracted signal (M+P minus M alone). Relating this subtracted signal to the 1D reference spectrum of the metabolite yields the fractional signal intensity, which serves as a proxy of the affinity between metabolite and protein: fractional signal intensities close to one designate strong (Kd in μM range) interactions, whereas fractional signal intensities close to zero designate very weak (or absent) interactions.
Figure 2. Ligand-detected NMR to study binding between four proteins and a mixture of seven chemically diverse metabolites. A) 1D1H-NMR reference spectra of the metabolites used in the mixture. B,C) WaterLOGSY and T1rho relaxation chemical shift spectra of 7-metabolite mixture in presence of four proteins (BSA, AroG, Eno, PfK). Spectra were obtained by subtraction of the spectra of free protein and free metabolites from the combined protein+metabolite sample spectrum (see Methods). Metabolites interacting with the protein are seen as positive peaks at chemical shifts (position in the frequency spectrum) characteristic for each metabolite. D,E) Interaction maps derived from waterLOGSY and T1rho relaxation experiments.Intensity of color in the boxes corresponds to the fractional signal intensity (see Methods). Black stars designate previously reported protein-metabolite interactions. Abbreviations: PYR: pyruvate, CIT: citrate, AKG: 2-oxoglutarate, FBP: fructose-1,6-bisphosphate, SKM: shikimate, PHE: L-phenylalanine, AMP: adenosine-5-monophosphate.
Figure 3. Ligand-detected NMR with metabolite mixtures of increasing complexity. A) Example of water-LOGSY and T1rho relaxation spectra of AroG with three different metabolite mixtures (mixture compositions indicated as subscripts in (B), and listed in Table 2). Light red stripes show peaks used for generation of interaction maps for 33-compound mixture in (B). B) Interaction maps resulting from water-LOGSY and T1rho experiments, for all four tested proteins, based on unambiguously assigned metabolite peaks. Intensity of color in the boxes denotes the Fractional Signal Intensity (see Methods). Black stars designate previously reported protein-metabolite interactions that were identified in the experiment, red stars indicate previously reported interactions which were not identified in the corresponding experiment. Abbreviations: PYR: pyruvate, CIT: citrate, AKG: 2-oxoglutarate, FBP: fructose-1,6-bisphosphate, SKM: shikimate, PHE: L-phenylalanine, AMP: adenosine-5-monophosphate, GLU: L-glutamate, MAL: malate, 6PG: 6-phosphogluconate, G6P: glucose-6-phosphate, PEP: phosphoenolpyruvate, PhePyr: phenylpyruvate, ATP: adenosine-5-triphosphate, NAD: nicotinamide adenine dinucleotide, TYR: L-tyrosine, TRP: L-tryptophan, HIS: L-histidine, ARG: L-arginine, ASP: L-aspartate, LYS: L-lysine, ASN: L-asparagine, GLN: L-glutamine, SER: L-serine, THR: L-threonine, ALA: L-alanine, ILE: L-isoleucine, LEU: L-leucine, MET: L-methionine, VAL: L-valine, CYS: L-cysteine, GLY: glycine, PRO: L-proline.
Figure 4. Effect of identified binding metabolites on AroG enzyme activity. In vitro enzyme activity of AroG as determined by photometric assays in presence of 100 μM of the respective metabolite. All experiments were performed in triplicate, and error bars denote the corresponding standard deviations. Metabolites denoted with *: significant change in enzyme activity (p-value < 0.02 as determined by two-tailed Student’s T-test).
Supplementary figures

Figure S1. Testing GFP as a benchmark for moderate-sized proteins in ligand-detected NMR. A) water-LOGSY and T1rho spectra of 7-metabolite mixture in presence of GFP and four larger-size proteins (BSA, AroG, Eno, PfkA). In these experiments phenylpyruvate (PhePYR) was used instead of PHE. B) 1D1H-NMR reference spectra of the metabolites used in the tested mixture. GFP shows no significant interaction with any of the tested metabolites, with the exception of a very weak interaction with AMP in water-LOGSY.
Figure S2. Reference spectra for 15-compound and 33-compound mixtures.
Figure S3. Water-LOGSY titration experiments to determine the affinity of several interaction hits observed in our datasets. Graphs represent titrations of (left-to-right) pfkA-ATP, pfkA-FBP and aroG-CIT pairs. The $K_d$-derivation was performed as described in $^{36}$, except that for the data fits the $[L]_0 \gg [P]_0$ approximation was not assumed, and the data was fit to a general solution of the one-site ligand binding equation:

$$I = I_{max} \frac{(K_D + [L] + [P]) - \sqrt{(K_D + [L] + [P])^2 - 4[P][L]}}{2[P_0]}$$

where $I$ is the observed peak integral, $I_{max}$ is the maximal integral at saturation, $[L]$ and $[P]$ are the total ligand and protein concentrations respectively.
Figure S4. Comparing fractional signal intensities in water-LOGSY and T1rho data. Individual points correspond to unambiguous metabolite:protein interaction hits observed both in water-LOGSY and T1rho experiments. (A) Fractional Signal Intensities (see Methods for definition). (B) Absolute signal intensities (normalized for the number of scans in two experiments).
### Supplementary Tables

**Table S1.** Affinities of previously reported protein-metabolite interactions tested in this study

<table>
<thead>
<tr>
<th>Protein</th>
<th>Metabolite</th>
<th>Reference [PMID]</th>
<th>Reported Km/Kd value [mM]</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSA</td>
<td>pyruvate</td>
<td>23357430</td>
<td></td>
<td>no affinity reported</td>
</tr>
<tr>
<td>BSA</td>
<td>citrate</td>
<td>23357430</td>
<td></td>
<td>no affinity reported</td>
</tr>
<tr>
<td>BSA</td>
<td>ATP</td>
<td>9474754</td>
<td>0.12</td>
<td>at pH 7.4</td>
</tr>
<tr>
<td>BSA</td>
<td>AMP</td>
<td>9474754</td>
<td>0.13</td>
<td>Ki for competitive inhibition of ATP binding</td>
</tr>
<tr>
<td>BSA</td>
<td>L-phenylalanine</td>
<td>23357430</td>
<td></td>
<td>no affinity reported</td>
</tr>
<tr>
<td>BSA</td>
<td>L-tryptophan</td>
<td>23357430</td>
<td></td>
<td>no affinity reported</td>
</tr>
<tr>
<td>BSA</td>
<td>L-histidine</td>
<td>23357430</td>
<td></td>
<td>no affinity reported</td>
</tr>
<tr>
<td>eno</td>
<td>PEP</td>
<td>7811696</td>
<td>0.135</td>
<td></td>
</tr>
<tr>
<td>pfkA</td>
<td>FBP</td>
<td>4229913</td>
<td></td>
<td>at 1 mM, no change in enzyme activity</td>
</tr>
<tr>
<td>pfkA</td>
<td>AMP</td>
<td>4229913</td>
<td></td>
<td>at 1 mM, no change in enzyme activity</td>
</tr>
<tr>
<td>pfkA</td>
<td>PEP</td>
<td>8202475</td>
<td>0.5</td>
<td>dissociation constant in absence of F6P, ATP</td>
</tr>
<tr>
<td>pfkA</td>
<td>PEP</td>
<td>8202475</td>
<td>~4</td>
<td>estimated Ki(0.5) in presence of F6P and ATP</td>
</tr>
<tr>
<td>pfkA</td>
<td>PEP</td>
<td>4229913</td>
<td>0.75</td>
<td>dissociation constant of T state of pfkA</td>
</tr>
<tr>
<td>pfkA</td>
<td>PEP</td>
<td>9335538</td>
<td>0.24</td>
<td>dissociation constant in absence of F6P, ATP</td>
</tr>
<tr>
<td>pfkA</td>
<td>ATP</td>
<td>8202475</td>
<td>0.057</td>
<td></td>
</tr>
<tr>
<td>aroG</td>
<td>PEP</td>
<td>10988284</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>aroG</td>
<td>L-phenylalanine</td>
<td>12953</td>
<td>0.075</td>
<td></td>
</tr>
</tbody>
</table>
References


Chapter 7 – Systematic identification of protein-metabolite interactions in complex metabolite mixtures by ligand-detected NMR


Chapter 8 - Concluding remarks

Karl Kochanowski
Chapter 8 - Concluding remarks

Findings
In this thesis, we aimed to identify the mechanisms underlying metabolic regulation in the model bacterium *Escherichia coli*. In particular, we focused on its transcriptional and posttranslational regulation. The key findings are:

- Two mechanisms determine *E. coli*’s transcriptional regulatory program: specific regulation by a network comprising major (hundreds of target genes) and minor (few target genes) transcription factors, as well as global regulation, which modulates gene expression globally based on the cell’s physiology. Our results demonstrate the strong contribution of global transcriptional regulation, which not only sets the maximal expression capacity of anabolic pathways (e.g. arginine biosynthesis pathway in chapter 3), but also dominates the response of many central metabolic genes (chapter 4).

- *E. coli*’s specific transcriptional program relies on few internal signals, i.e. regulatory metabolites which modulate the activity of specific transcription factors, to assess the current cellular status. For example, transcriptional regulation of *E. coli*’s central carbon metabolism in response to various environments is largely driven by only three such signals, namely F1P, FBP, and cyclic AMP (chapter 4). Similarly, the transcriptional coordination of catabolism and anabolism in response to nutrient limitation largely relies on only one signal, namely cyclic AMP (chapter 5).

- *E. coli*’s metabolic response does not depend on transcriptional regulation alone, but also requires additional posttranslational regulatory mechanisms. One such mechanism is the passive regulation of enzyme saturation through changes in substrate concentration, which adjusts *E. coli*’s approximate transcriptional response to nutrient limitation to eventually coordinate the flux through each reaction (chapter 5).

The emerging picture is that the transcriptional regulation of *E. coli*’s metabolism is governed by an approximate regulatory program that roughly partitions the proteome according to few ‘ballpark rules’ based on information from regulatory metabolites, as well as the cell’s current physiology (chapters 2 to 5). One such ballpark rule is the general coupling of gene expression to the cellular physiology: as cells encounter environments that support fast growth (e.g. external supplementation of amino acids), they respond by generally increasing each promoter’s maximal expression capacity. As exemplified for the arginine biosynthesis regulatory circuit in chapter 3, additional local transcriptional regulators may then determine how much of this increased promoter capacity is being used based on cell’s current demand for the pathway’s end product. Another ballpark rule is highlighted in chapter 5: accumulation of keto-acids is interpreted as a mismatch between catabolic and anabolic capacity (as described
Chapter 8 - Concluding remarks

previously\textsuperscript{1}) and relayed to the regulatory metabolite cyclic AMP, which in turn results in a general shift from catabolic towards anabolic protein expression.

This approximate regulatory program is implemented using only a small subset of the cell’s repertoire of transcription factors. Of particular importance is the transcription factor Crp. We found that Crp not only is a major transcriptional regulator of \textit{E. coli}'s central metabolism in a wide range of environmental conditions (chapter 4), but also mediates the aforementioned transcriptional coordination between catabolism and anabolism (chapter 5). Interestingly, our results suggest that Crp achieves this coordination through a combination of direct and indirect transcriptional regulation. Although Crp does not regulate the expression of anabolic proteins directly, Crp-dependent induction of catabolic proteins reduces the resources (i.e. free RNA polymerase and ribosomes) available for the expression of anabolic proteins. Thus, through competition for cellular resources, Crp indirectly also regulates the expression of non-target proteins. In fact, this indirect regulatory mechanism may provide a potential mechanistic explanation for the aforementioned coupling of gene expression of cellular physiology: since conditions supporting fast grow typically lead to low Crp activity\textsuperscript{1}, this may free resources for the expression of all other proteins. Evidence from literature suggests that indirect transcriptional regulation through competition for limited resources may not be restricted to Crp: a well-established example is \textit{E. coli}'s stringent response to starvation, in which activation of the sigma factor \(\sigma^5\) leads to the induction of stress-related proteins at the expense of \(\sigma^70\) dependent proteins through competition for limiting RNA polymerase\textsuperscript{2}. Another example stems from the observation that addition of the ribosome inhibitor chloramphenicol, which causes a massive induction of ribosomal proteins presumably through the transcription factor DksA\textsuperscript{3}, leads to a general reduction in the concentration of all metabolic proteins\textsuperscript{4}. Similarly, massive induction of an unnecessary synthetic protein is accompanied by a reduction of the expression of native proteins\textsuperscript{5}, suggesting that such indirect transcriptional regulation is an inherent property of the cellular expression machinery. Given that only few (chapter 4) - or even one (chapter 5) - transcription factors suffice to mount \textit{E. coli}'s transcriptional response in various conditions, what is the function of the remaining transcription factors, in particular so-called ‘general’\textsuperscript{6} or ‘global’\textsuperscript{7} regulators such as Fis, ArcA and FNR that have many reported targets in metabolism? The most parsimonious explanation is that these transcription factors respond to regulatory signals that were not triggered in the tested conditions. For example, both ArcA and FNR are known to respond to reduced oxygen availability\textsuperscript{8}, which likely did not occur in the aerated and well-mixed conditions tested in this thesis.

Thus, the findings in this thesis suggest that the regulation of \textit{E. coli}'s metabolism is the result of a regulatory division-of-labor between transcriptional and posttranslational regulation. \textit{E. coli}'s approximate transcriptional regulatory program sets each enzyme’s maximal capacity, and thereby
determines the boundaries of what the cell can do to mount a metabolic response. However, this transcriptional regulatory program rarely controls metabolic reactions alone (chapters 2 to 5), and therefore requires further posttranslational regulatory mechanisms (chapters 5 to 7). One such mechanism is passive regulation of enzyme activity through changes in enzyme saturation and thermodynamic driving force (chapter 6). While such passive regulation can account for a considerable fraction of observed flux changes (e.g. up to 50% in chapter 5), it is clear that the majority of flux changes cannot be explained by transcriptional and passive regulation alone, as was also shown previously for central metabolism\textsuperscript{9,10}.

Which additional mechanisms could fill this gap? An obvious candidate mechanism is allosteric regulation of enzyme activity. As we have recently demonstrated for \textit{E.coli} glycolysis, allosteric regulation is vital for the fast adaptation of metabolic pathways to environmental changes\textsuperscript{11}, and several lines of evidence suggest that it is also important in steady state (e.g. \textsuperscript{12,13}). However, the lack of methods to systematically identify allosteric interactions currently imposes severe constraints on our ability to assess the role of allostery beyond small-scale examples (chapter 6). In chapter 7, we have begun to tackle this issue using ligand-detected NMR to identify metabolite binders of three well characterized \textit{E. coli} enzymes. Already in this proof-of-principle study, we could detect and functionally validate several novel protein-metabolite interactions, and extrapolation from this small sample points towards a large uncharted allosteric interactome even in well-studied organisms such as \textit{E. coli}. Importantly, such allosteric regulation does not necessarily need to be highly specific, as exemplified by the prevalent binding of nucleotide phosphates in chapter 7, which is in line with recently observed pleiotropic low-affinity regulation of glycolytic enzymes in yeast\textsuperscript{14}. Thus, in analogy to the transcriptional program outlined above, a combination of specific and global allosteric interactions may shape \textit{E. coli}'s posttranslational regulatory response. Whether other posttranslational regulatory mechanisms, i.e. covalent protein-modifications, play an important role, is currently unclear. At least in \textit{E. coli}, examples of \textit{in vivo} functional protein-modifications of metabolic enzymes are few and far between\textsuperscript{15,16}. Nevertheless, since some of these modifications also occur non-enzymatically through reactive metabolites such as acetyl-phosphate and succinyl-CoA\textsuperscript{17,18}, they may provide a direct link from metabolism to the coarse-grained regulation of enzyme activity.

Many of the findings discussed above were enabled by combined computational-experimental approaches developed in this thesis (chapters 3 to 5). Although these approaches differ in their exact implementation, they do follow a recurrent theme with three key steps. The first step is the selection of experimental perturbations which yield a wide range of physiologies, e.g. various carbon sources (chapters 3 and 4) or gradual titration of nutrient uptake (chapter 5), and subsequent quantification of the cellular steady state response to these perturbations. The second step is to develop a
mathematical description of the process of interest that captures its essential properties, e.g. promoter activity as function of global and specific transcriptional regulation. Finally, step three combines data and mathematical description to systematically dissect and quantify the different regulatory inputs by computational inference. These approaches allowed us to dissect the contribution of global and specific transcriptional regulation in biosynthesis and central carbon metabolism (chapters 3 and 4), and to systematically identify metabolites serving as regulatory signals for central metabolic genes (chapter 4). Importantly, these data-driven approaches are not restricted to E. coli, but are widely applicable to other organisms, as well as other regulatory layers.

Outlook

From the foundational work laid out in this thesis, several key questions emerge that could be addressed in future efforts:

**How prevalent are allosteric protein-metabolite interactions in E. coli, and what is their in vivo function?** As discussed above, posttranslational regulation of enzymes activity, i.e. through allostery, is likely to play a vital role in establishing the cell’s metabolic response. In chapter 7, we have developed a NMR-based approach to systematically identify protein-metabolite interactions and applied it to a small set of well-characterized E. coli enzymes. Future studies could extend this approach to comprehensively map out protein-metabolite interactions in complete metabolic pathways. Another possibility is to include alternative methods to detect such interactions. For example, a recently developed mass spectrometry method uses limited proteolysis to systematically identify proteins which undergo conformational changes *in vitro* upon addition of a metabolite\textsuperscript{19}, and we are currently exploring its potential to identify novel target proteins of various regulatory metabolites. Complementary to these purely experimental approaches, the computational inference method developed in this thesis (chapter 4) could readily be extended to identify *in vivo* relevant transcription factor-metabolite interactions also at genome scale, if genome-scale expression and metabolomics data in various steady state conditions become available. Once a more comprehensive picture of E. coli’s protein-metabolite interactome becomes available, the next challenge is to elucidate the role of such interactions in the *in vivo* context. Towards this end, computational approaches similar to the one we used to identify *in vivo* relevant allosteric interactions in E. coli’s glycolysis\textsuperscript{11} may allow to test whether the metabolic response of a given pathway can be better explained by addition of a newly identified interaction.

**Do internal regulatory metabolites play a role outside E. coli, i.e. in eukaryotes?** A key finding of this thesis is that the regulation of E. coli’s metabolism relies on intracellular metabolites which provide a representation of its current metabolic status. Consequently, a potential research avenue is to assess the prevalence of such internal regulatory metabolites in other organisms. Recent studies have
provided some examples in yeast and mammalian cells, such as the regulation of TOR complex activity by amino acids\textsuperscript{20,21} and 2-oxoglutarate\textsuperscript{22}, glycolytic flux-dependent regulation of respiration\textsuperscript{23} by fructose-1,6-bisphosphate\textsuperscript{24,25}, acetyl-CoA driven differentiation of embryonic stem cells\textsuperscript{26}, and regulation of few yeast transcription factors by metabolites\textsuperscript{27}. However, in many cases the flow of information between metabolism and the different regulatory layers is only poorly understood\textsuperscript{28}, and for the vast majority of eukaryotic transcription factors and signaling kinases/phosphatases it is not clear whether they are actually affected by regulatory metabolites. In principle, the approaches developed in this thesis can be applied to address this issue. For example, by applying the approach developed in \textit{chapter 4} to matched phosphoproteomics and metabolomics data obtained from either time course or steady state experiments, we may be able to identify metabolites that serve as potential regulatory signals of phosphorylation events.

\textbf{Can we exploit the regulatory program outlined here to guide the metabolic engineering of microorganisms?} Current efforts to optimize a production pathway typically focus on adjusting the expression of its individual enzymes\textsuperscript{29,30}, but often neglect the cell’s native regulatory program\textsuperscript{31}. Consequently, a potential research avenue is to explicitly take the cell’s regulatory program into account, and to couple it to the production pathway. For example, by putting the production pathway under the control of a native regulatory metabolite such as 2-oxoglutarate or fructose-1,6-bisphosphate, its activity could be coupled to a specific physiological state without the need for external regulatory inputs. Conversely, by introducing synthetic regulatory feedback mechanisms from the production pathway to the cell’s global regulatory program we may be able to redirect cellular resources toward product formation in a more targeted way. One way how this may be achieved is by using synthetic regulators that detect the pathway’s end product\textsuperscript{32}. An alternative is allosteric engineering to introduce novel allosteric interactions to a protein of interest, or modulate existing ones\textsuperscript{33,34}. Towards this end, a key challenge will be the development of methods to rationally design such interactions, and to rapidly test different designs experimentally.

\textbf{Do regulatory metabolites play a role in the communication between cells?} In nature, microorganisms rarely live in isolation. Instead, they frequently coordinate their behavior both within as well as between species, for example to form biofilms\textsuperscript{35,36}. To establish this coordination, cells rely on the exchange of diverse small molecule signals\textsuperscript{37–40}. Beyond ‘designated’, often chemically complex\textsuperscript{40}, signals (e.g. bacterial quorum sensing molecules such as the ‘autoinducer’ 3-oxo-hexanoyl-L-homoserine lactone\textsuperscript{38}), there is growing evidence that cells secrete various metabolic intermediates\textsuperscript{39,41–44}, some of which may double as communication signals\textsuperscript{45,46}. For example, biofilm formation in \textit{S. typhimurium} was shown to depend on the sensing of external L-arginine\textsuperscript{46}, and numerous studies have reported the secretion and uptake of the global internal regulatory metabolite
cyclic AMP in *E. coli*\(^{47,48}\). Conversely, recent works have demonstrated the interplay between metabolism and collective cellular behavior\(^{45,49–52}\), such as the dependence of biofilm formation in *E. coli* on catabolite repression by Crp\(^{51,52}\). Consequently, it is tempting to speculate that regulatory metabolites may not only inform individual cells about their internal status, but also on the presence (and metabolic status) of their neighbouring cells, e.g. to determine the position within a colony and coordinate metabolic operation\(^{45,49}\). Future studies of native and synthetic communities may allow to identify metabolites that serve as dual internal/external regulatory signals.
References

Acknowledgements
This thesis concludes one of the most stimulating and rewarding periods in my life, and it would not have been possible without the support from many people. In particular, I would like to thank...

...my advisor Uwe Sauer for his trust and support over the years, and for knowing when to give me a gentle push in the right direction.

...my co-examiners Matthias Heinemann and Martin Ackermann for providing critical input during committee meetings. A special ‘thank you’ goes to Matthias for having given me the first taste of microbial systems biology over six years ago.

...my collaborators within the Sauer lab: Benjamin Volkmer, who taught me everything I know about experimental work and is probably an even bigger TV nerd than me; Luca Gerosa and Hannes Link, who showed me how to computationally formalize my hand waving ideas and never flinched at my butchered code; Dimitris Christodoulou, who earned his first experimental stripes while performing time course experiments with me and did not complain even once (or only very quietly); Victor Chubukov and Elad Noor for teaching me how to turn my rambling writing style into concise and engaging sentences.

...my collaborators outside the Sauer lab: Terence Hwa for his work on E. coli’s response to nutrient limitations, and for the many stimulating discussions during his sabbatical in our lab; Yaroslav Nikolaev for his work on ligand-detected NMR, and for his patience every time I got lost again in the intricacies of NMR; Ilaria Piazza and Paola Picotti for their work on limited proteolysis, and for not strangling me every time I suggested a few more ‘small control experiments’ involving two weeks of measurements.

...my Semester- and Masterstudents for their contributions to many of the projects included in this thesis and beyond: in particular Simon Brunner, Oliver Dressler, Raffael Inglin, Roland Brunner, and Dario Cerletti.

...all past and present members of the Sauer lab for creating an intellectually rigorous and fun atmosphere, and for enduring my countless questioning during group meetings. A special ‘thank you’ goes to my lab mates Dan Zheng, Laura Falter, Marieke Buffing, Daniel Sévin, Maren Diether, Balazs Szappanos, and Flavia Tschan. Sorry for all the foot tapping, air drumming, and humming, interrupted by occasional quiet cursing.

...all IMSB- and IBC-members for the many fun coffee breaks, runs through the forest, Happy Hours, Ho-Ho’s,...

...my family and friends for their love and support, and for keeping me reasonably sane.
Curriculum Vitae

Name: Karl Andreas Kochanowski
Address: Nordstrasse 278
          8037 Zurich, Switzerland
E-mail: kochanowski@imsb.biol.ethz.ch
Date of birth: 24.10.1983
Nationality: German

Education

05/2010 – present  PhD student, IMSB at ETH Zurich
          Supervisor: Prof. Dr. Uwe Sauer
03/2010  Diploma in biology, RWTH Aachen
          Final grade: 1 (“sehr gut”)
          “Identification of a putative metabolite-based flux sensor in Escherichia coli”
          Supervisors: Dr. Matthias Heinemann and Prof. Dr. Rainer Fischer
10/2006  Intermediate diploma in biology, RWTH Aachen
          Grade: 1.2
10/2004 – 03/2010  Biology (Diploma), RWTH Aachen
10/2003 – 09/2004  Civil Service in Elisabeth-hospital Essen
08/1994 – 08/2003  Viktoriaeschule Essen (secondary school)
          Final examination grade: 1.4
08/1990 – 07/1994  Winfried-Grundschule Essen (primary school)
          Grades: 1 = best, 5 = worst
Publications

The following works were published during the completion of this thesis:


