Mapping interactions between metabolites and transcriptional regulators at a genome-scale

Dissertation

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer. nat.)

des Fachbereichs Biologie der Philipps-Universität Marburg

Vorgelegt von

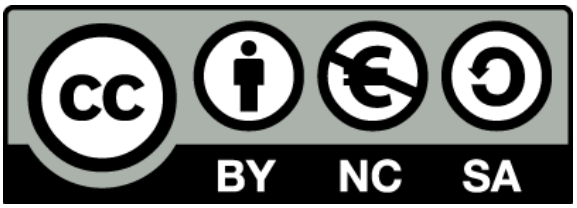
Michelle Kuntz

Master of Science, Goethe-Universität Frankfurt

Geboren in Offenbach am Main, 18. Januar 1992

Marburg, Juni 2021

Originaldokument gespeichert auf dem Publikationsserver der
Philipps-Universität Marburg
http://archiv.ub.uni-marburg.de



Dieses Werk bzw. Inhalt steht unter einer
Creative Commons
Namensnennung
Keine kommerzielle Nutzung
Weitergabe unter gleichen Bedingungen
3.0 Deutschland Lizenz.

Die vollständige Lizenz finden Sie unter:
http://creativecommons.org/licenses/by-nc-sa/3.0/de/

Die vorliegende Dissertation wurde von Juni 2018 bis Juni 2021 am Max-Planck-Institut für terrestrische Mikrobiologie/ Universität Tübingen unter Leitung von Prof. Dr. Hannes Link angefertigt.

Vom Fachbereich Biologie der Philipps-Universität Marburg (Hochschulkennziffer 1180) als Dissertation angenommen am 26.08.2021

Erstgutachter(in): Prof. Dr. Hannes Link
Zweitgutachter(in): Prof. Dr. Tobias Erb

Tag der Disputation: 07.09.2021

Table of contents

Abbreviations

ArgR	arginine repressor
ATCase	aspartate carbamoyltransferase
cAMP	cyclic adeonosine monophosphate
Cra	catabolite repressor/activator
CRISPRi	CRISPR interference
CRP	cAMP receptor protein
crRNA	CRISPR RNA
CTP	cytidine triphosphate
FACS	Fluorescence-activated Cell Sorting
ITC	isothermal titration calorimetry
LC-MS	liquid-chromatography mass spectrometry
LiP	limited proteolysis
NADPH	nicotinamide adenine dinucleotide phosphate
NGS	Next-generation sequencing
NMR	nuclear magnetic resonance spectroscopy
PAM	protospacer adjacent motif
PMI	protein-metabolite interactions
PPP	pentose-phosphate pathway
RNAi	RNA interference
RNAP	RNA polymerase
sgRNA	single guide RNA
TALE	transcription activator like effector
TF	transcription factor
tracrRNA	trans-acting RNA
TRN	transcriptonal regulatory network

Summary

The control and regulation of cellular metabolism is required to maintain the biosynthesis of building blocks and energy, but also to prevent the loss of energy and to be able to quickly adjust to changing conditions. Hence, the metabolic network and the flow of genetic information has multiple layers of regulation and information is transmitted between gene expression and metabolism. For this purpose, metabolites serve as key signals of the regulatory network to balance metabolism via the adjustment of protein levels and the activity of enzymes. Understanding these regulations and interplays of bacterial metabolism will enable us to improve the modelling and engineering of metabolic networks and ultimately to develop new antibiotics and production strains. The aim of this thesis is to investigate which regulatory mechanisms are used by the cell to respond to genetic perturbations. Moreover, we develop new methods to map protein-metabolite interactions and to prove their functionality in the cell.

After **introducing** the fundamentals of metabolic network regulation, we investigate in **chapter 1** how *Escherichia coli* (*E. coli*) reacts to genetic perturbations. We use a library of 7177 CRISPRi strains to perform a pooled fitness growth assay, demonstrating the buffering effects of metabolism. Additionally, measuring the metabolome and proteome of 30 arrayed CRISPRi strains enables us to elucidate three gene-specific buffering mechanisms.

In **chapter 2,** we use our new insights about genetic perturbations of chapter 1 to develop a method for systematically mapping interactions between metabolites and transcriptional regulators. CRISPRi leads to a knockdown of a gene and therefore induces specific changes in the metabolome and proteome of the cell. We therefore combine the pooled CRISPRi library with a fluorescent reporter for transcription factor activity and extract cells, which show a response of the reporter to the changing conditions, via FACS from the pooled library. By analyzing proteome and metabolome data, we confirm previously reported and discover new interactions.

With **chapter 3,** we provide a detailed protocol of how to work with CRISPRi libraries. We explain the design and construction of sgRNAs of arrayed as well as pooled CRISPRi strains and how to perform growth assays. Furthermore, we explain the execution and analysis of Illumina Next-generation sequencing of pooled libraries. We also explain the sorting of cells from pooled libraries via FACS.

In **chapter 4,** we show how to find new interactions between metabolites and transcription factors by external perturbations. By switching a growing *E. coli* culture between growth and glucose limitation, we provoke strong changes of metabolite levels and transcript levels. Calculating the transcription factor activity from gene expression levels and correlating them with metabolite levels, enables us to

recover known interactions but also to discover new interactions, of which we prove five in *in vitro* binding assays.

In **chapter 5,** we investigate the function of allosteric regulation of metabolic enzymes in amino acid pathways of *E. coli*. We constructed 7 mutants of allosteric enzymes to remove the allosteric feedback regulation. By metabolomics, proteomics and flux profiling analysis we show how allostery helps to adjust enzyme levels of the cell. Furthermore, using a metabolic model and the application of CRISPRi we show how well-adjusted enzyme levels make the cell more stable towards genetic perturbations.

Zusammenfassung

Die Kontrolle und Regulation des mikrobiellen Metabolismus wird benötigt für die Aufrechterhaltung der Biosynthese von Grundbausteinen und Energie, aber auch um den Verlust von Energie zu verhindern und um eine schnelle Anpassung im Falle von sich ändernden Bedingungen zu ermöglichen. Deshalb werden das metabolische Netzwerk und der Fluss genetischer Information durch mehrere Regulationsebenen reguliert und Informationen zwischen Genexpression und Metabolismus übermittelt. Zu diesem Zweck dienen Metabolite als Schlüsselsignale des regulatorischen Netzwerks, welche den Metabolismus balancieren indem sie zum einen Proteinlevel anpassen und zum anderen die Aktivität von Enzymen regulieren. Diese Regulationen und Wechselspiele zu verstehen, ermöglicht es, metabolische Modelle und Netzwerke zu verbessen und gezielt zu verändern. Außerdem bietet es die Chance zur Entwicklung neuer Antibiotika und verbesserter Produktionsstämme. Daher ist es Ziel dieser Arbeit zu untersuchen, welche regulatorischen Mechanismen der Zelle dazu dienen, genetischen Perturbationen entgegen zu wirken. Des Weiteren wird die Entwicklung neuer Methoden zur systematischen Kartierung von Protein-Metabolit Interaktionen und die Aufklärung deren Funktion innerhalb der Zelle dargelegt.

Nachdem die Regulation von metabolischen Netzwerken in der **Einleitung** besprochen wird, beschreiben wir in **Kapitel 1**, wie *Escherichia Coli* (*E. coli*) auf genetische Perturbationen reagiert. Wir nutzen eine CRISPRi Stamm-Bibliothek mit 7177 Stämmen, um einen gepoolten Fitness-/Wachstumsversuch durchzuführen, welcher die puffernden Effekte des Metabolismus aufzeigt. Des Weiteren messen wir das Metabolom und Proteom von 30 einzelnen CRISPRi Stämmen, was es uns ermöglicht drei genspezifische Mechanismen zur Kompensation von genetischen Perturbationen aufzuklären.

In **Kapitel 2** nutzen wir die gewonnen Erkenntnisse aus Kapitel 1 und entwickeln eine Methode um systematisch Interaktionen zwischen Metaboliten und transkriptionellen Regulatoren in *E. coli* zu finden. Da CRISPRi zu einer Inhibition der Transkription des Zielgens führt, induziert dies spezifische Änderungen im Metabolom als auch Proteom. Deshalb kombinieren wir die gepoolte CRISPRi Stamm-Bibliothek mit einem fluoreszenten Reporter für Transkriptionsfaktor-Aktivität und extrahieren Zellen mittels Durchflusszytometrie, welche eine Reaktion des Reporters auf die sich ändernden Bedingungen zeigen. Durch die Messung des Proteoms und Metaboloms einiger dieser Stämme validieren wir bereits bekannte als auch neu gefundene Interaktionen.

In **Kapitel 3** stellen wir ein detailliertes Protokoll zum Arbeiten mit CRISPRi Stamm-Bibliotheken zur Verfügung. Wir erklären das Design und die Konstruktion von Plasmiden für kleine Guide-RNAs,

sowohl für einzelne als auch gepoolte CRISPRi Stämme und erklären die Durchführung von Wachstumsexperimenten. Außerdem beschreiben wir die Durchführung und Analyse von Illumina Next-generation sequencing von gepoolten Stamm-Bibliotheken. Darüber hinaus erklären wir detailliert, wie einzelne Zellen aus der gepoolten Zell-Bibliothek mittels Durchflusszytometrie isoliert werden können.

In **Kapitel 4** zeigen wir eine Methode zur systematischen Kartierung von Interaktionen zwischen Metaboliten und Transkriptionsfaktoren mittels externer Perturbationen. Durch den Wechsel von einer *E. coli* Kultur zwischen optimalen Wachstumsbedingungen und der Limitation von Glucose, erzeugen wir starke Änderungen von Metabolit-Leveln und Transkriptions-Leveln. Die Berechnung von Transkriptionsfaktor-Aktivitäten mittels Transkriptions-Leveln und die Korrelation dieser mit Metabolit-Leveln, ermöglicht die Validierung von bereits bekannten Interaktionen, als auch das Finden von neuen Interaktionen, von welchen fünf mittels *in vitro* Bindeassays bestätigt werden.

In **Kapitel 5** untersuchen wir die Funktion von allosterischen Regulationen metabolischer Enzyme im Aminosäure Stoffwechsel von *E. coli*. Wir konstruieren 7 Mutanten von allosterisch-regulierten Enzymen und entfernen dadurch den Regulationsmechanismus. Durch Metabolom-Analysen, Proteomstudien und der Messung des biosynthetischen Flusses in diesen Mutanten zeigen wir, wie Allosterie die Einstellung von Enzym-Leveln in der Zelle ermöglicht. Des Weiteren wenden wir ein metabolisches Modell an und stören den Stoffwechsel der Zellen mittels CRISPRi und zeigen damit, wie gut angepasste Enzym-Level Zellen robuster gegen genetische Perturbationen machen.

Introduction

Metabolic Networks and how they are regulated

The physiological and biochemical properties of a cell are determined by all the metabolic and physical processes that build the metabolic network. The flow of genetic information within such a biological system is explained in the central dogma of molecular biology (Crick, 1970). It describes the flow of genetic information from DNA to RNA and finally into protein. Once the information is packed into a protein, it cannot be transferred to another protein or back into DNA. This dogma applies only to information, which is the precise determination of a sequence, either in the shape of a base pair sequence in nucleic acids or the order of amino acids in a protein. But how is this flow of information regulated to achieve optimal conditions of the cell's metabolism? Moreover, how are large biochemical networks of bacterial metabolism controlled, where thousands of genes, enzymes and reactions must be coordinated?

Since complete genomes of organisms can be sequenced, it is possible to reconstruct the full network of biochemical reactions in a cell. The latest genome-scale model of *Escherichia coli* K-12 MG1655 includes 1515 open reading frames, 2719 metabolic reactions and 1192 metabolites (Monk et al., 2017). Looking at those numbers, one could ask how the cell manages to carry out all reactions needed to synthesize building blocks and energy, while fine-tuning the enzyme levels to prevent the loss of energy and staying flexible in case of changing conditions. To understand these interplays, it is advantageous to focus on single metabolic reactions and pathways first, with the aim to understand how regulatory interactions guide these reactions (**Figure 1**).

A metabolite is transformed into a product in a sequence of catalyzed reactions. These reactions are carried out by enzymes which are encoded in genes and are transcribed and translated. In summary, transcription, translation and the enzyme as a final product are suitable targets of regulation. The regulation of the transcription rate takes place by attenuation or the interaction between transcription factors and metabolites, which both lead to changes in gene expression. As transcription is known to be quite slow (1 min/gene [1 kilo bp]), the cell also needs other strategies to adapt rapidly to changing conditions (Shamir et al., 2016). Thus, metabolic control is also achieved by translational regulation and post-translational protein-modifications, such as acetylation, methylation and phosphorylation (Macek et al., 2019). These covalent modifications can be either reversible or irreversible, and usually different enzymes are responsible for the forward and reverse reactions (Martin, 2014). Another mechanism to quickly adjust to new circumstances is allosteric enzyme

regulation, which modulates enzyme activities and therefore reaction rates and metabolic fluxes in seconds (Chubukov et al., 2014; Gerosa and Sauer, 2011; Link et al., 2013).

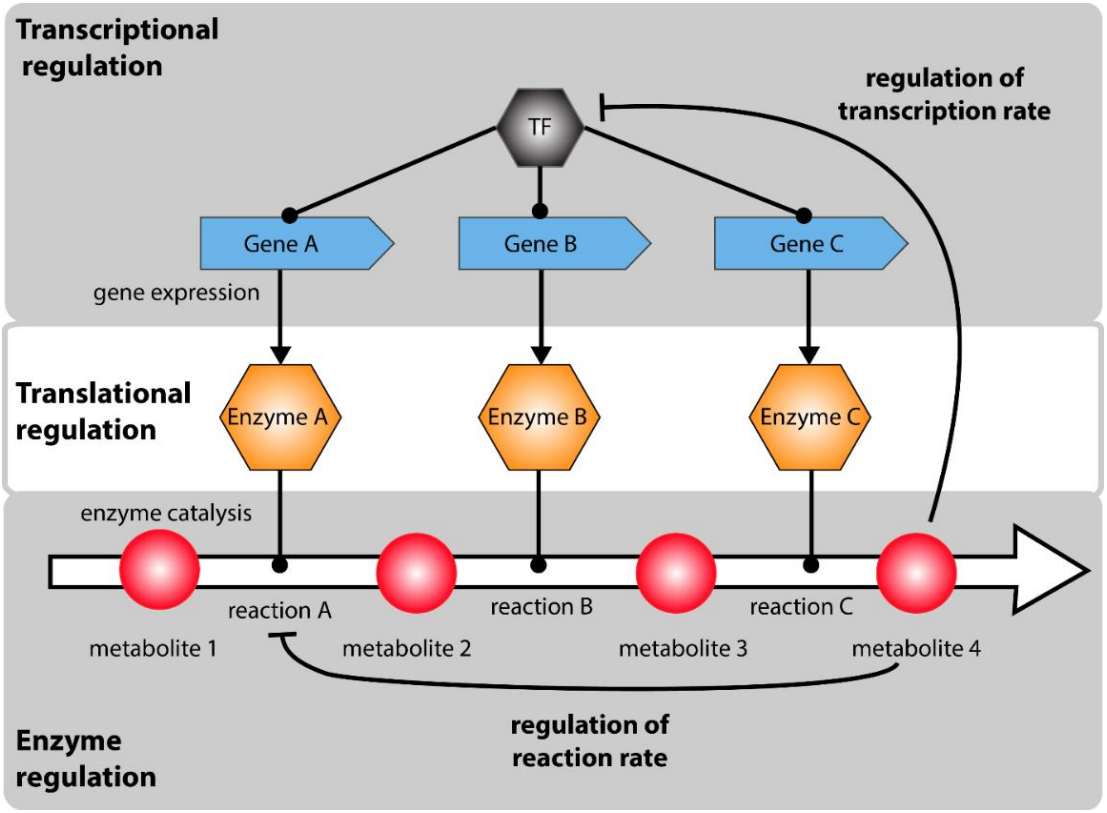


Figure 1. Layers of regulation in metabolic networks. Scheme of an exemplary metabolic pathway in amino acid biosynthesis. This includes the transcriptional regulation via metabolite-transcription factor (TF) interaction as well as a negative feedback loop of the product metabolite on the first enzyme of the metabolic pathway, which regulates the reaction rate of the enzyme (modified after Donati et al., 2018).

When looking at the metabolic pathway and all these levels of regulation, we can see that information is not only transferred in the form of nucleic acid or proteins, but metabolites also can serve as information carriers and inputs. They serve as signals for fluctuations in intracellular or extracellular conditions and lead to adaptation or maintenance of the metabolic system. Taking this into account, metabolites not only control metabolic pathways they are directly linked to but are also key signals for the whole metabolic network of a cell. Thus, the investigation of metabolites and their interactions with proteins, will help to understand the global regulation of large biochemical networks of metabolism.

Metabolites as key signals in metabolism

As technologies have developed enormously in the past decade, it is now possible to measure thousands of metabolites with high sensitivity (Johnson et al., 2016). This allowed the study of

metabolites not only as substrates and products of metabolism, but also their function as key signals and global regulators (Chubukov et al., 2014; Li and Snyder, 2011; Milanesi et al., 2020).

Previously, researchers measured metabolite profiles of 4913 strains in yeast and approximately 3800 *E. coli* strains, each strain having a different gene deletion (Fuhrer et al., 2017; Mülleder et al., 2016). They were able to not only precisely link metabolic changes to the deleted enzyme, but also uncovered novel gene-metabolite interactions. Based on their analysis, they could also predict the metabolism-related function of 72 non- annotated genes. Thus, a gene deletion leads to very specific changes of metabolite profiles, meaning that the metabolome contains information about the status of the whole cell.

Information carried by metabolites can be sensed and processed on different layers, but most of them are based on an interaction between the metabolite and a protein, the so-called protein-metabolite interactions (PMIs). As described earlier, these PMIs can be classified into i) enzyme catalysis, ii) allostery and iii) transcriptional regulation (**Figure 2A**). Allostery and transcriptional regulation will be described more specifically in the following section.

Types of protein-metabolite interactions

Allostery

The binding of a ligand to the allosteric site of an enzyme induces a conformational change in the target protein, leading to a change in the activity of the protein, whether the function of the enzyme is catalytic (metabolic enzymes) or binding (receptors). This effect was first described by the Monod-Wyman-Changeux Model (Monod et al., 1965), which defined the characteristics of allosteric regulation: i) the effector molecule must differ chemically from the substrate of the regulated reaction, ii) it binds to any other site of the enzyme other than the active one and iii) binding of the effector molecule leads to a conformational change of the protein. The term cooperativity describes the binding of ligands to proteins with multiple binding sites and therefore an interaction of binding processes (Ricard and Cornish-Bowden, 1987). The activity and function of proteins can be strongly influenced by allosteric binding of metabolites. An example of protein classes, that are modulated by the cells own metabolites, are metabolic enzymes (Gerosa and Sauer, 2011) , protein kinases (Li et al., 2010) and transcription factors (Motlagh et al., 2014).

The latest mechanistic model of the *E. coli* metabolism contains 295 allosteric interactions, showing that allosteric interactions are very common for enzymes (Khodayari and Maranas, 2016). The most prominent example of an allosterically regulated enzyme is the aspartate carbamoyltransferase

(ATCase). It catalyzes the first step of the pyrimidine biosynthesis, which is the committed step of the pathway, an irreversible reaction leading to a modification of the metabolite which after that step can only be used to synthesize the pathway's final product. With increasing concentrations the end product CTP inhibits the enzyme, preventing the synthesis of unnecessary intermediates (Gerhart, 2014). Consequently, this ensures that the pathway flux is not higher than actually needed and indicates that allosteric regulation functions as a valve controlling the fluxes of pathways. This could also be shown in a study, where *E. coli* was exposed to oxidative stress, upon which NADPH levels decreased. This led to an increasing activity of glucose-6-phosphate dehydrogenase, which is allosterically inhibited by NADPH, and therefore an increased flux through the pentose phosphate pathway (PPP), which refilled NADPH levels (Christodoulou et al., 2018).

The amino acid metabolism of *E. coli* relies heavily on allosteric regulation, as 16 out of 20 pathways are regulated by allosteric feedback inhibition (Reznik et al., 2017). This means that the amino acid as an end-product inhibits the enzyme catalyzing the committed step in the pathway. Amino acid metabolism pathways are not only regulated by allostery, but also by transcriptional regulation and attenuation. However, there are still unanswered questions about the function of allosteric regulation in these pathways and it remains challenging to investigate this phenomenon at a systemic level (for further discussion of this topic see chapter 5 of this thesis).

Transcriptional regulation

The interaction between a metabolite and a transcription factor (TF) is a special form of allosteric regulation. Because of the non-catalytic nature of the TF, the change of the activity of the protein does not result in a change of flux in a certain pathway, but rather influences the binding activity of the TF to specific DNA sequences (promoter region). Upon the binding of the metabolite, the TF can bind or release the promoter region, which in turn blocks or promotes the binding of the RNA polymerase (Latchman, 1997). This leads to increased or decreased enzyme levels of the gene regulated by the TF, depending on the TF acting as repressor or activator. Therefore, transcriptional regulation affects how much enzyme is produced by the cell in a certain pathway. One of the best studied molecular systems is the *lac* promoter of *E. coli* (Jacob and Monod, 1961). *E. coli* uses glucose as the preferred carbon source, but if it is not available it can use other carbon sources like lactose. The *lac* operon encodes for a membrane protein, which acts as a transporter for lactose. It also encodes a β-Galactosidase, which catalyzes the cleavage of lactose into galactose and glucose. In the absence of lactose, the operon is repressed by the TF *lacI*, but an increase of allolactose, the intermediate product of lactose catabolism, leads to an allosteric inactivation of *lacI,* which then releases the promoter and the *lac* operon is expressed. Simultaneously, the cell signals low glucose levels with cAMP concentrations,

which increases upon the depletion of glucose (Notley-McRobb et al., 1997). cAMP binds to CRP, which is a global transcriptional regulator having a direct influence on transcription rates. Thus, the cell senses the presence of lactose and the shortage of glucose and tightly controls the synthesis of proteins to process these carbon sources. The transcriptional regulation is, therefore, another layer of regulation aimed at preventing the extra burden of protein synthesis in lower energy conditions. This protein synthesis can lead to lower growth rates (Shachrai et al., 2010), and in this example, the induction of the *lac* operon without lactose leads to a 5% reduction of the growth rate (Novick and Weiner, 1957).

Many TFs not only have one target promoter and cannot only bind to one specific promoter sequence, but instead influence the transcription of many genes in different metabolic pathways. They are organized in so called transcriptional regulatory networks (TRNs), which are clustered into basic units, motifs, modules and networks (Babu et al., 2004). Basic units describe the interaction between a transcription factor and the target gene binding site and are organized in network motifs, which cover specific patterns of interregulation and are the most dominant in a network. These motifs are clustered into modules which are transcriptional units and their interconnecting interactions are finally compiled into a regulatory network. One of the best known TRNs exists for *E. coli*, covering 220 TFs, 3493 regulatory interactions and 4745 regulated genes (Santos-Zavaleta et al., 2019). High-throughput methods like Selex or Chip-seq enabled genome-wide quantification of TF-DNA binding (Furey, 2012; Ishihama et al., 2016). For 211 of 220 TFs a target gene could be identified. The number of regulated genes varied by a large margin for individual TFs. CRP, for example, regulates 535 genes and AlaS only one. The regulons of only 9 TFs (CRP, FNR, IHF, FIS, H-NS, ArcA, FuR, NarL and Lrp) account for 46% of all known TF-mediated regulatory interactions in *E. coli*. A library of fluorescent transcriptional reporters was used to quantify the activity of promoters in the central carbon metabolism of *E. coli* in 26 different environmental conditions. It was shown that 70% of the variance in promoter activity could be explained by global transcriptional regulation (Kochanowski et al., 2017). Furthermore, the measurements of metabolites revealed that only 3 metabolites explained most of the transcriptional regulation. Even with these methods, only ~140 interactions between 220 TFs and 2600 metabolites were found, so there is still a challenge to systematically identify functional regulatory protein-metabolite interactions on a large-scale, as many more interactions are expected (Madan Babu and Teichmann, 2003).

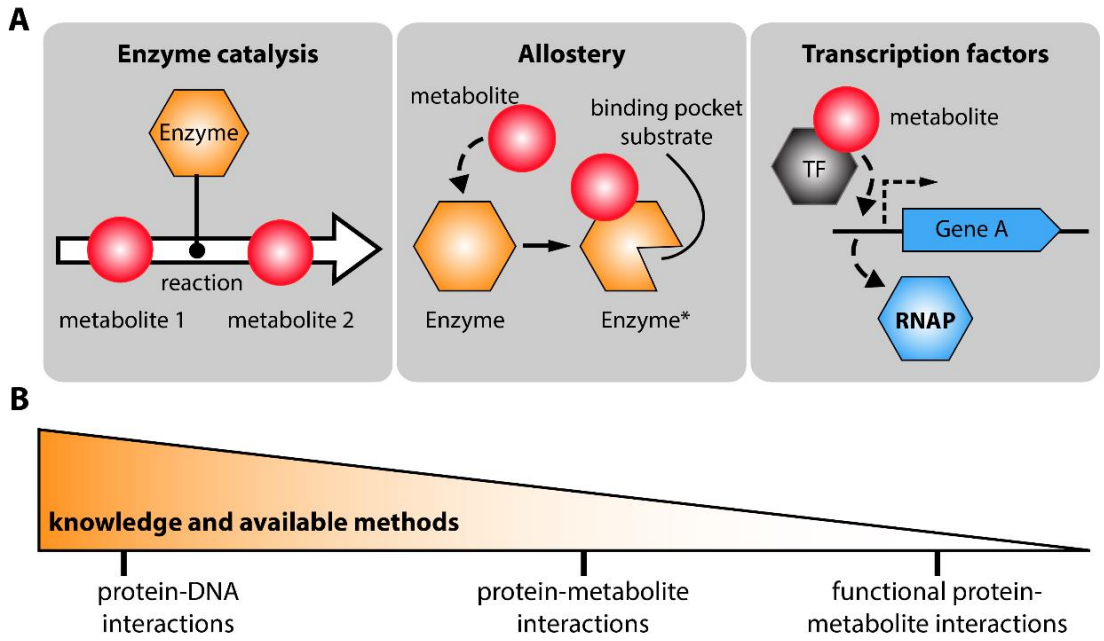


Figure 2. Types of protein-metabolite interactions. (A) Metabolites can be catalyzed by an enzyme or bind to special regions of metabolic enzymes (allostery). This leads to a conformational change, which results in the change of reaction rates of the enzyme. They can also bind transcription factors (TF) and therefore have an influence on gene expression (RNAP- RNA polymerase). (B) There are many methods available to investigate protein-DNA interactions and therefore many are known. The knowledge and availability of methods decrease when looking at protein-metabolite interactions, especially when it comes to functional protein-metabolite interactions, which were not only shown in *in vitro* assays but also had an impact on the organism in *in vivo* studies.

Methods to identify protein-metabolite interactions

To understand the regulation of metabolic networks, it is crucial not only to know which genes are targeted by TFs, but also to find the regulatory metabolite which carries the information about the status of the cell and binds to the respective protein. A well-established approach is to prove PMIs by *in vitro* binding assays, like isothermal titration calorimetry (ITC) or nuclear magnetic resonance spectroscopy (NMR) (He et al., 2015; Nikolaev et al., 2016). These methods all make use of changing properties of the protein upon the binding of a ligand. As it is only possible to test one protein against one metabolite, they only allow to prove binding events but do not support the screening for many metabolites at the same time. Hence, finding interacting metabolites, which are not directly linked to the pathway of the TFs target gene is almost impossible. Considering the numbers, there could be millions of functional PMIs, as we know that there are approximately one million protein molecules in a bacterial cell (Milo, 2013) and about 100-fold more metabolites (Bennett et al., 2009). Due to the instable nature of PMIs, only the development of multi-omics methods (proteomics, metabolomics) in the last decade enabled the invention of more systematic approaches. Consequently, it is now possible to measure at a larger scale and to test many proteins against one metabolite or the other way around. Metabolomics-based LC-MS enabled mapping PMIs by immobilizing proteins and

incubating them with a pool of metabolites. The protein-metabolite complexes were purified and the extracted metabolites were later on measured by untargeted metabolomics (Tagore et al., 2008; Vinayavekhin and Saghatelian, 2011). This combination of untargeted LC-MS and native MS was not only used to find potential interacting metabolites, but also applied to confirm new interactions (Qin et al., 2019). Furthermore, dynamically measuring metabolites and fluxes in *E. coli* during growth switches allowed a systematic identification of allosteric interactions that govern the switch between glycolysis and gluconeogenesis (Link et al., 2013). A chemo proteomic approach enabled the first systematic mapping of PMIs on a proteome-wide scale (Piazza et al., 2018). The researchers incubated whole cell-lysates with metabolites and performed a proteinase K and trypsin treatment. Hence, the proteomics measurements of the digested proteome showed different peptide patterns for proteins if the metabolite is bound to them. In summary, this method combined limited proteolysis (LiP) with MS in the presence of unmodified metabolites in a cell-like environment and enabled to identify a big network of known and previously unknown PMIs and binding sites to proteins. The authors claim that based on their data, about one quarter of the measured proteome interacted with at least one of the 20 investigated metabolites, covering all known cellular processes, and indicating that the size and impact of PMIs is in fact larger than previously considered. However, the method might miss interactions due to low MS coverage because of the treatment of the cell-lysate, and it cannot distinguish between functional und unspecific binding events (Diether and Sauer, 2017; Kochanowski et al., 2015). Altogether, the development of mapping approaches is more advanced than the evaluation of functionality, as the binding of a metabolite to a protein is ultimately dependent on the conditions the cell is facing. However, the low affinity of PMIs and their transient nature still presents a challenge (Diether et al., 2019) (**Figure 2B**). Hence, no *in vivo* method exists that detects functional interactions in a high-throughput manner and their direct effects on cellular processes, like transcription. The current challenge still remains to differentiate between functional and unspecific metabolite-protein interactions and to detect them at a large-scale (see chapter 2 and 4 in this thesis).

Utilization of a better understood protein-metabolite interaction network

PMIs do not only occur inside of isolated, linear metabolic pathways, but regulatory metabolites can also be located in other metabolic branches and transcription factors can have hundreds of target genes. Therefore, the genetic manipulation of a metabolic pathway and its regulation in a production strain can have a tremendous impact on the overall function of the metabolic network, which most often influences the maximal productivity and the robustness towards perturbations (He et al., 2016). For instance, production strains often have transcriptional feedback regulation removed to improve production rates (Nielsen and Keasling, 2016; Park et al., 2018). Several arginine overproduction

strains were investigated regarding growth and production, and it could be shown that the deletion of the transcriptional regulator ArgR led to the highest production but also had the biggest impact on growth (Sander et al., 2019). Due to protein burden caused by protein overexpression, the strain suffered of a limitation in the pyrimidine nucleotide biosynthesis and was limited in growth. To avoid these burdens for the cell, the design of production strains is often reviewed by *in silico* approaches (Ko et al., 2020; Long et al., 2015). Hence, it is crucial to provide enough experimental data which can be used for setting up computational models, which cover the genome-wide metabolic network of the strain. Therefore, for future models it is very important to i) map all existing PMIs and to ii) investigate their functionality under certain conditions, as, for instance, different carbon sources. Due to the implementation of many high-throughput methods a lot of PMIs could be mapped, nevertheless, testing their functionality remains a big challenge.

Besides of metabolic engineering of production strains, the development of new drugs is also a big field of application for PMIs. The discovery of new antibiotics is especially crucial, as many multi-resistant bacterial pathogens have emerged due to the misuse of antibiotics in the clinical and agricultural environment (Stokes et al., 2019). Most of the known antibiotics target cell envelope synthesis, DNA replication, transcription and protein biosynthesis, resulting in perturbations of metabolic homeostasis of the cell (Zampieri et al., 2017). Hence, a better understanding of the regulation of metabolism could lead to the discovery of new targets for antibiotics treatment.

CRISPR interference (CRISPRi) as a tool to investigate bacterial metabolism

Short introduction to the world of CRISPR

“Genome editing”, in other words, the ability to precisely change the base pair sequence of a DNA sequence at a predetermined location is not only crucial for studying the function of genes but also to uncover biological mechanisms. Insertion, modification or deletion of DNA sequences or even whole genes is the basis for studying effects of genetic modifications on the phenotype of a cell. Furthermore, it enables engineering of strains for biotechnology. Methods to modify the genomes of organisms have been available for many years. The discovery of restriction enzymes, which normally protect bacteria against phages, allowed the first manipulation of DNA in a test tube in the 1970s (Smith and Welcox, 1970) and was followed by many other discoveries and inventions. Recombinases enable altering the genome by homologous recombination (Capecchi, 1989; Olorunniji et al., 2016), while programmable nucleases (Chandrasegaran and Carroll, 2016), such as meganucleases and zinc

finger nucleases, induce targeted double-strand breaks. However, the efficiency of homologous recombination is heavily dependent on the organism and programmable nucleases require protein engineering expertise, as the method includes the development of DNA-binding proteins to bind to custom targets (Pickar-Oliver and Gersbach, 2019). Therefore, the discovery of CRISPR-Cas9 and its ability to bind and modify DNA sequences paved the way to a new era of genome editing.

CRISPR (clustered regularly interspaced palindromic repeats) was first described in 1987 in the genome of *E. coli* (Ishino et al., 1987), and its function remained unclear, even though it was later on identified in many different species across bacteria (40% of all species) and archaea (60%) (Makarova et al., 2011). The big mystery about the spacer sequences could only be solved when many phage genomes were sequenced at the time of the human genome project in the early twentieth century. Interestingly, the non-repeating spacers originated from viruses and other mobile genetic elements and were always located in the neighborhood of CRISPR associated (Cas) genes, which have nuclease and helicase domains (Jansen et al., 2002). Hence, it was assumed and finally shown that CRISPR serves bacteria as an adaptive defense system, using antisense RNAs as a memory and a guide to eliminate invaders (Brouns et al., 2008; Horvath and Barrangou, 2010; Makarova et al., 2006). CRISPR systems can be divided into 2 classes: class 1 systems use a complex of different proteins to degrade DNA, and class 2 systems are using one large Cas protein. Both classes are divided into 6 types (Wright et al., 2016). The Cas9 protein belongs to the class 1, type II CRISPR systems and was found in *Streptococcus pyogenes* (Deltcheva et al., 2011). It is one of the simplest CRISPR systems and consists of an endonuclease with a four-component system, that includes two small RNA molecules guiding the protein to its target DNA sequence, the mature CRISPR RNA (crRNA) and a partially complementary trans-acting RNA (tracrRNA). Jennifer Doudna and Emanuelle Charpentier re-engineered it into a two-component system and fused the two RNAs to a single guide RNA (sgRNA) (Jinek et al., 2012). Combined with the endonuclease Cas9, this artificial system can be programmed to target and cut any DNA sequence, making it a perfect tool for genome editing. This invention was so remarkable, as it could also be used for genome editing of human cells (Jinek et al., 2013; Mali et al., 2013), that Doudna and Charpentier were awarded the Nobel Prize in chemistry in 2020. Shortly after publishing the artificial system, CRISPR-Cas9 was modified to silence or activate genes and can therefore also be used as a programmable transcription factor (Bikard et al., 2013; Larson et al., 2013; Qi et al., 2013).

The CRISPR interference system (CRISPRi) makes use of a catalytically inactive (dead) Cas9 protein (dCas9) and is repurposed for RNA-guided transcription regulation. The dCas9 protein binds to a sgRNA with a 20-25 bp long sequence, which is complementary to either the template or non-template strand of the DNA target sequence and next to the protospacer adjacent motif (PAM; sequence: NGG).

The sgRNA is fused to the dCas9 handle and a terminator sequence, and it guides the dCas9 to the target region (**Figure 3A**). As the dCas9 protein is lacking endonuclease activity, it will only bind to the target region and as a result hinder the RNA polymerase (RNAP) to transcribe the DNA sequence at varying degree, depending on the design of the sgRNA. A sgRNA targeting the promoter region will block the initiation of transcription and is independent of the DNA strand, while only sgRNAs binding the non-template DNA strand show a silencing effect and hinder transcription elongation. Moreover, the position of the guide RNA plays a huge role for repression efficiency, as guides being closer to the promoter region usually show a stronger effect than the ones targeting regions inside or at the end of the gene (Larson et al., 2013; Qi et al., 2013).

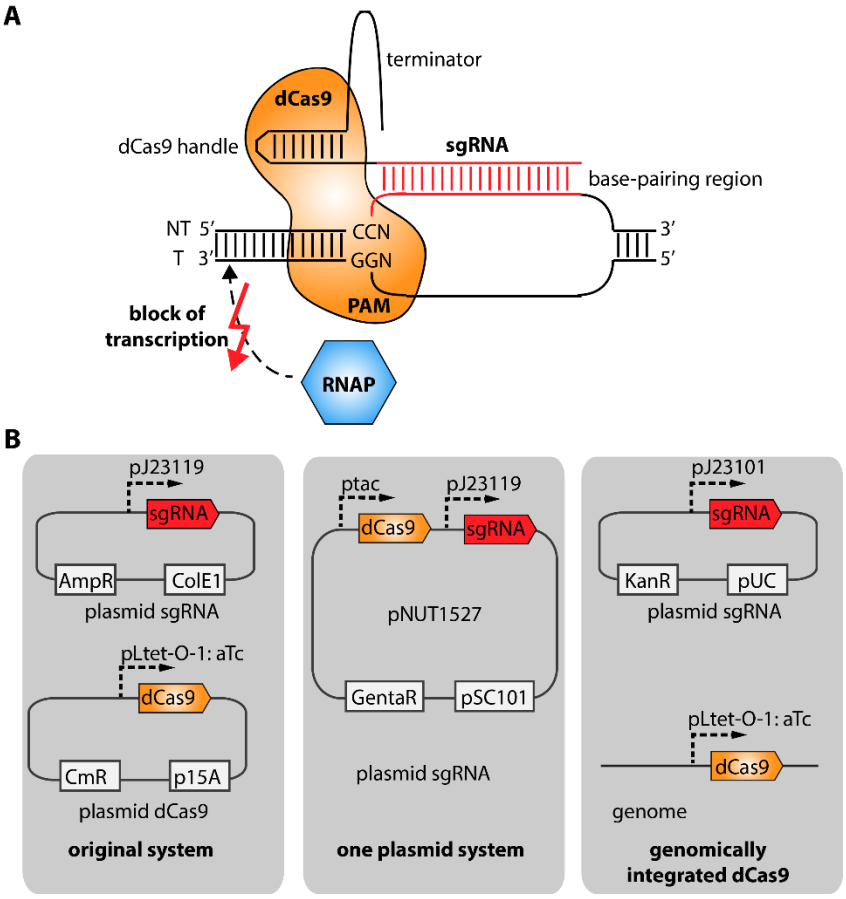


Figure 3. Functional principle of CRISPRi and available systems. (A) The dCas9 protein binds to the dCas9 handle which is fused to a sgRNA. This guides the whole complex to a genomic region, in this case on the non-template strand (NT) and pairs with the region right next to the PAM site. Therefore, the dCas9 protein blocks the binding of the RNA- polymerase (RNAP) and with this the transcription of the encoded gene. (B) (Left to right) The original system based on two plasmids (Qi et al., 2013), the system based on one plasmid expressing the sgRNA and the dCas9 enzyme (Beuter et al., 2018) and a system making use of a genomically integrated dCas9 (Lawson et al., 2017). (sgRNA- small guide RNA; AmpR- ampicillin resistance cassette; CmR- chloramphenicol resistance cassette; GentaR- gentamycin resistance cassette; KanR – kanamycin resistance cassette)

The original system consisted of one plasmid for the expression of the sgRNA while another plasmid is needed for the expression of dCas9 (Larson et al., 2013; Qi et al., 2013)(**Figure 3B**). As this is not practical when working for instance with an additional reporter plasmid, systems were generated that have the sgRNA and dCas9 protein encoded on one plasmid (Beuter et al., 2018; Kirtania et al., 2019). Often, the leakiness of promoters leads to plasmid instability and/or loss of the plasmid. As the expression of dCas9 could result in additional protein burden for the cell or the knockdown of an essential gene before induction (Huang et al., 2015), selective pressure will lead to an enrichment of cells that lose the plasmid. Hence, to prevent biases in growth before induction due to leaky expression from plasmid derived promoters, dCas9 was also genomically integrated and expressed from a regulated chromosomal promoter, which was optimized to prevent leaky expression (Lawson et al., 2017).

Taking the previous studies into account, CRISPRi-induced gene silencing is a powerful and highly efficient method to block and repress transcription with up to 1000-fold repression (Qi et al., 2013). This in turn allows i) to change the regulation of metabolic pathways and to study ii) the function of essential genes.

Application of CRISPRi in biotechnology

For the design of production strains, the repression of genes for key enzymes in metabolism is often required to redirect metabolic fluxes from growth to the desired product or to balance the production for increased titers, yields and productivity. As described before, CRISPRi leads to a highly efficient repression of transcription and therefore can be used to study the effect of genes on production rates. It was already described for many organisms that CRISPRi can be applied to generate production strains for industry and especially in *E. coli*, many pathways were optimized by using CRISPRi (Schultenkämper et al., 2020). For example, the downregulation of the transcriptional regulator ArgR led to a 2-times higher growth rate of an arginine overproducing *E. coli* strain compared to a deletion strain, while maintaining almost the same specific production (Sander et al., 2019). While CRISPRi is a powerful tool to directly influence a metabolic pathway and therefore the strain's production capacity, it can also be used to identify genes of unknown function for their possible catalytic function, and therefore the possibility of using strains or their enzymes for the production of certain compounds (Lee et al., 2018). The CRISPRi system also can be used to uncouple production from growth via inducible dCas9 expression. This means that once the desired biomass for production is reached, the expression of dCas9 can be induced which results in a dynamic repression of the desired target genes and starts the production of the compound of interest (Zheng et al., 2019).

Hence, the application of CRISPRi in biotechnology has a big potential for metabolic engineering of production strains and processes, but also for the discovery of new biochemical products.

CRISPRi library screenings

From genotype to phenotype, how can we identify genes which contribute to specific biological phenotypes, and especially, how can we do this on a large-scale? Targeted genome regulation methods, such as zinc-finger or transcription activator like effector (TALE) proteins and RNA interference (RNAi) were broadly used in the past to investigate the functions of genes. As an example, RNAi, which perturbs genes on the mRNA level with complementary RNAs, is strongly limited by off-target effects, low efficiency, toxicity, and it can only be used in organisms containing the needed enzyme machinery (Larson et al., 2013). Hence, the simpler design of CRISPRi and the ability to use it in a lot of organisms make it a more predictable and specific tool to test genes for their function.

The screening of CRISPRi libraries can be carried out in an arrayed or a pooled format (Shalem et al., 2015). While for an arrayed screen, all plasmids are cloned in separate wells, and all cells are transformed and measured in arrays, the pooled screen facilitates the whole workflow. Usually, oligonucleotides are synthesized and delivered as an oligo pool in one tube. Cloning techniques like Golden Gate and Gibson Assembly enable us to insert the pooled sgRNA sequences into the vector plasmid in one reaction (Engler et al., 2008; Gibson et al., 2009). Additionally, the transformation of the recipient strain and the screen readout are performed in a pooled fashion using methods like Next-Generation sequencing (NGS) or Fluorescence-activated Cell Sorting (FACS). Furthermore, synthesizing large DNA oligonucleotide libraries is becoming cheaper every year, as well as DNA sequencing (Wetterstrand, 2020), which allows the usage of CRISPRi in large sgRNA libraries. Therefore, pooled screenings are usually less laborious and less expensive while arrayed screens allow us to directly measure single cellular phenotypes via fluorescence, image-analysis or metabolomics and proteomics. Nevertheless, in the case of pooled libraries, a lot of experimental strategies were established to filter out single strains based on their phenotype. For instance, single-molecule fluorescence time-lapse imaging of an *E. coli* CRISPRi library identified genotypes *in situ* after providing a detailed characterization of the phenotype (Lawson et al., 2017). This makes it possible to extend live cell microscopy to the scale of pooled libraries. Similarly, the combination of a CRISPRi library with a reporter plasmid for monitoring the growth rate of all strains allowed the selective enrichment of slow-growing cells via cell sorting (Beuter et al., 2018). Hence, the choice of format mostly depends on how many targets should be screened and how the strains with a desired phenotype should be further investigated.

A lot of pooled CRISPRi screenings have been carried out in the past years and were mostly used to investigate the function of genes. For a library of 235 different CRISPR knockdowns, the position of the replication fork was monitored via live-cell microscopy for >500 cell cycles (Lawson et al., 2017). The subsequent genotyping made it possible to map each genetic perturbation to a specific phenotype and to determine the gene's function in cell cycle control. A more frequently used way to investigate CRISPRi libraries is the execution of pooled growth assays. The library is grown under certain growth conditions and the fitness of strains is calculated via the foldchange of the abundance of guides over time, which is measured via NGS. Bikard and colleagues created a library with 92000 sgRNAs targeting the whole *E. coli* genome (Cui et al., 2018). This screen not only confirmed previously reported design rules and properties of such assays but also revealed off-target effects and sequence-specific toxicity effects ("bad-seed" effects) that have not been previously reported. After removing all guides with these effects, they ended up with a library of 23000 sgRNAs, which they used to predict gene essentiality in a pooled fitness assay (Rousset et al., 2018). 78% of the genes previously annotated as essential by gene knockout studies (Baba et al., 2006) were identified, and some targets showed no tolerance to small reductions of their expression levels and could therefore be promising antibiotic targets. In another study, the researches performed fitness assays with a library consisting of 60000 sgRNAs targeting the whole *E. coli* genome (Wang et al., 2018) and showed similar results to the study mentioned before. Hence, it could be demonstrated how powerful CRISPRi is as a tool to map prokaryotic genetic networks precisely on a large-scale and that it is a good alternative to, for example, gene knockout libraries, which are not available for all organisms and often require expensive automation systems to investigate them.

Since it was shown that pooled CRISPRi library screenings could be used to investigate the genetic network of *E. coli*, it was also applied to other bacterial species and studied under various growth conditions (Todor et al., 2021). Targeting 3400 conserved genes in 3 growth media and 18 different *E. coli* strains, the study showed big variations of gene essentiality between the strains and conditions and highlighted the impact of mobile genetic elements on the essentiality of core genes within a single species (Rousset et al., 2021). In a separate study, the screening of pathogens and microbiome strains revealed the functionality of the CRISPRi system while the strains are located in the host (Liu et al., 2021; Qu et al., 2019; Shields et al., 2020). This could enable the identification of genes needed for commensalism and pathogenesis in *in vivo* screens, leading to new insights for targeted drug development and especially for the development of new antibiotics.

Objectives of this thesis

Understanding metabolic networks requires not only to identify all components of metabolism, such as genes, proteins and metabolites, but also to understand their function and how all of these components are connected. The regulation of metabolism is still a big mystery, as many regulatory mechanisms were investigated, but newer studies suggest a much higher number of interactions, especially between proteins and metabolites. Besides this, the functionality of many of those proven interactions remains a big question, as not many options exist to measure them in an *in vivo* context. Hence, in the following chapters of this thesis, we aim to understand i) which regulatory mechanisms are used by *E. coli* to react to genetic perturbations of metabolism via CRISPR interference and ii) how can we use this knowledge to find functional interactions between transcriptional regulators and metabolites.

In the first chapter, we perform a genome-wide CRISPRi library screening in *E. coli* and investigate how metabolism reacts to decreasing enzyme levels. Furthermore, we characterize the metabolome and proteome of some of these strains and show which local regulatory mechanisms buffer decreases of specific enzymes.

In the second chapter, we use the gained knowledge of chapter one to establish a method to screen for metabolite-transcription factor interactions *in vivo* on a large-scale. In CRISPRi strains, metabolite and protein levels changed remarkably specific. We combine CRISPRi with reporter plasmids to study the influence of the perturbation on transcription factor activity. This enables us to confirm known protein-metabolite interactions in an *in vivo* context and to reveal new interactions.

In chapter three, we provide a protocol on how to generate and work with CRISPRi libraries. It includes the execution of fitness growth assays and the sorting of cells from the pooled library based on their phenotype.

In chapter four, we use dynamic metabolite and transcriptome data of *E. coli* to identify known metabolite-transcription factor interactions. Furthermore, we predict and validate new interactions in central metabolism based on the dynamic data set.

In chapter five, we investigate the relevance and function of allosteric feedback inhibition *in vivo.* We use metabolomics, proteomics and flux profiling to study the interaction between allosteric feedback inhibition and transcriptional regulation in 7 mutants with allosterically dysregulated amino acid pathways. Moreover, we use CRISPR interference to show how this interplay, resulting in enzyme overabundance, makes cells more robust against genetic perturbations.

References

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular Systems Biology *2*.

Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M., and Teichmann, S.A. (2004). Structure and evolution of transcriptional regulatory networks. Current Opinion in Structural Biology *14*, 283–291.

Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J., and Rabinowitz, J.D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nature Chemical Biology *5*, 593–599.

Beuter, D., Gomes-Filho, J.V., Randau, L., Díaz-Pascual, F., Drescher, K., and Link, H. (2018). Selective Enrichment of Slow-Growing Bacteria in a Metabolism-Wide CRISPRi Library with a TIMER Protein. ACS Synthetic Biology *7*, 2775–2782.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res *41*, 7429–7437.

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. Science *321*, 960–964.

Capecchi, M. (1989). Altering the genome by homologous recombination. Science *244*, 1288–1292.

Chandrasegaran, S., and Carroll, D. (2016). Origins of Programmable Nucleases for Genome Engineering. Journal of Molecular Biology *428*, 963–989.

Christodoulou, D., Link, H., Fuhrer, T., Kochanowski, K., Gerosa, L., and Sauer, U. (2018). Reserve Flux Capacity in the Pentose Phosphate Pathway Enables Escherichia coli's Rapid Response to Oxidative Stress. Cell Systems *6*, 569-578.e7.

Chubukov, V., Gerosa, L., Kochanowski, K., and Sauer, U. (2014). Coordination of microbial metabolism. Nature Reviews Microbiology *12*, 327–340.

Crick, F. (1970). Central Dogma of Molecular Biology. Nature *227*, 561–563.

Cui, L., Vigouroux, A., Rousset, F., Varet, H., Khanna, V., and Bikard, D. (2018). A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9. Nature Communications *9*.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602–607.

Diether, M., and Sauer, U. (2017). Towards detecting regulatory protein–metabolite interactions. Current Opinion in Microbiology *39*, 16–23.

Diether, M., Nikolaev, Y., Allain, F.H., and Sauer, U. (2019). Systematic mapping of protein-metabolite interactions in central metabolism of *Escherichia coli*. Molecular Systems Biology *15*.

Donati, S., Sander, T., and Link, H. (2018). Crosstalk between transcription and metabolism: how much enzyme is enough for a cell? Crosstalk between transcription and metabolism. Wiley Interdisciplinary Reviews: Systems Biology and Medicine *10*, e1396.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A One Pot, One Step, Precision Cloning Method with High Throughput Capability. PLoS ONE *3*, e3647.

Fuhrer, T., Zampieri, M., Sévin, D.C., Sauer, U., and Zamboni, N. (2017). Genomewide landscape of gene–metabolome associations in *Escherichia coli*. Molecular Systems Biology *13*, 907.

Furey, T.S. (2012). ChIP–seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. Nature Reviews Genetics *13*, 840–852.

Gerhart, J. (2014). From feedback inhibition to allostery: the enduring example of aspartate transcarbamoylase. FEBS Journal *281*, 612–620.

Gerosa, L., and Sauer, U. (2011). Regulation and control of metabolic fluxes in microbes. Curr. Opin. Biotechnol. *22*, 566–575.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods *6*, 343–345.

He, C., Howes, B.D., Smulevich, G., Rumpel, S., Reijerse, E.J., Lubitz, W., Cox, N., and Knipp, M. (2015). Nitrite dismutase reaction mechanism: kinetic and spectroscopic investigation of the interaction between nitrophorin and nitrite. J. Am. Chem. Soc. *137*, 4141–4150.

He, F., Murabito, E., and Westerhoff, H.V. (2016). Synthetic biology and regulatory networks: where metabolic systems biology meets control engineering. Journal of The Royal Society Interface *13*, 20151046.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science *327*, 167–170.

Huang, L., Yuan, Z., Liu, P., and Zhou, T. (2015). Effects of promoter leakage on dynamics of gene expression. BMC Systems Biology *9*.

Ishihama, A., Shimada, T., and Yamazaki, Y. (2016). Transcription profile of *Escherichia coli*: genomic SELEX search for regulatory targets of transcription factors. Nucleic Acids Research *44*, 2058–2074.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J Bacteriol *169*, 5429–5433.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology *3*, 318–356.

Jansen, R., Embden, J.D.A. van, Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol *43*, 1565–1575.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science *337*, 816–821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. ELife *2*.

Johnson, C.H., Ivanisevic, J., and Siuzdak, G. (2016). Metabolomics: beyond biomarkers and towards mechanisms. Nature Reviews Molecular Cell Biology *17*, 451–459.

Khodayari, A., and Maranas, C.D. (2016). A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. Nature Communications *7*.

Kirtania, P., Hódi, B., Mallick, I., Vass, I.Z., Fehér, T., Vass, I., and Kós, P.B. (2019). A single plasmid based CRISPR interference in Synechocystis 6803 – A proof of concept. PLOS ONE *14*, e0225375.

Ko, Y.-S., Kim, J.W., Lee, J.A., Han, T., Kim, G.B., Park, J.E., and Lee, S.Y. (2020). Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. Chemical Society Reviews *49*, 4615–4636.

Kochanowski, K., Sauer, U., and Noor, E. (2015). Posttranslational regulation of microbial metabolism. Current Opinion in Microbiology *27*, 10–17.

Kochanowski, K., Gerosa, L., Brunner, S.F., Christodoulou, D., Nikolaev, Y.V., and Sauer, U. (2017). Few regulatory metabolites coordinate expression of central metabolic genes in *Escherichia coli*. Molecular Systems Biology *13*, 903.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc *8*, 2180–2196.

Latchman, D.S. (1997). Transcription factors: An overview. The International Journal of Biochemistry & Cell Biology *29*, 1305–1312.

Lawson, M.J., Camsund, D., Larsson, J., Baltekin, Ö., Fange, D., and Elf, J. (2017). *In situ* genotyping of a pooled strain library after characterizing complex phenotypes. Molecular Systems Biology *13*, 947.

Lee, S.S., Shin, H., Jo, S., Lee, S.-M., Um, Y., and Woo, H.M. (2018). Rapid identification of unknown carboxyl esterase activity in Corynebacterium glutamicum using RNA-guided CRISPR interference. Enzyme and Microbial Technology *114*, 63–68.

Li, X., and Snyder, M. (2011). Metabolites as global regulators: A new view of protein regulation: Systematic investigation of metabolite-protein interactions may help bridge the gap between genome-wide association studies and small molecule screeni. BioEssays *33*, 485–489.

Li, X., Gianoulis, T.A., Yip, K.Y., Gerstein, M., and Snyder, M. (2010). Extensive In Vivo Metabolite-Protein Interactions Revealed by Large-Scale Systematic Analyses. Cell *143*, 639–650.

Link, H., Kochanowski, K., and Sauer, U. (2013). Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. Nature Biotechnology *31*, 357–361.

Liu, X., Kimmey, J.M., Matarazzo, L., de Bakker, V., Van Maele, L., Sirard, J.-C., Nizet, V., and Veening, J.-W. (2021). Exploration of Bacterial Bottlenecks and Streptococcus pneumoniae Pathogenesis by CRISPRi-Seq. Cell Host & Microbe *29*, 107-120.e6.

Long, M.R., Ong, W.K., and Reed, J.L. (2015). Computational methods in metabolic engineering for strain design. Current Opinion in Biotechnology *34*, 135–141.

Macek, B., Forchhammer, K., Hardouin, J., Weber-Ban, E., Grangeasse, C., and Mijakovic, I. (2019). Protein post-translational modifications in bacteria. Nature Reviews Microbiology *17*, 651–664.

Madan Babu, M., and Teichmann, S.A. (2003). Evolution of transcription factors and the gene regulatory network in Escherichia coli. Nucleic Acids Res. *31*, 1234–1244.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct *1*, 7.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., et al. (2011). Evolution and classification of the CRISPR–Cas systems. Nature Reviews Microbiology *9*, 467–477.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-Guided Human Genome Engineering via Cas9. Science *339*, 823–826.

Martin, B.L. (2014). Regulation by Covalent Modification. In ELS, John Wiley & Sons Ltd, ed. (Chichester, UK: John Wiley & Sons, Ltd), p.

Milanesi, R., Coccetti, P., and Tripodi, F. (2020). The Regulatory Role of Key Metabolites in the Control of Cell Signaling. Biomolecules *10*.

Milo, R. (2013). What is the total number of protein molecules per cell volume? A call to rethink some published values. BioEssays *35*, 1050–1055.

Monk, J.M., Lloyd, C.J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., et al. (2017). iML1515, a knowledgebase that computes Escherichia coli traits. Nature Biotechnology *35*, 904–908.

Monod, J., Wyman, J., and Changeux, J.-P. (1965). On the nature of allosteric transitions: A plausible model. Journal of Molecular Biology *12*, 88–118.

Motlagh, H.N., Wrabl, J.O., Li, J., and Hilser, V.J. (2014). The ensemble nature of allostery. Nature *508*, 331–339.

Mülleder, M., Calvani, E., Alam, M.T., Wang, R.K., Eckerstorfer, F., Zelezniak, A., and Ralser, M. (2016). Functional Metabolomics Describes the Yeast Biosynthetic Regulome. Cell *167*, 553-565.e12.

Nielsen, J., and Keasling, J.D. (2016). Engineering Cellular Metabolism. Cell *164*, 1185–1197.

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

Notley-McRobb, L., Death, A., and Ferenci, T. (1997). The relationship between external glucose concentration and cAMP levels inside Escherichia coli: implications for models of phosphotransferase-mediated regulation of adenylate cyclase. Microbiology *143*, 1909–1918.

Novick, A., and Weiner, M. (1957). ENZYME INDUCTION AS AN ALL-OR-NONE PHENOMENON. Proceedings of the National Academy of Sciences *43*, 553–566.

Olorunniji, F.J., Rosser, S.J., and Stark, W.M. (2016). Site-specific recombinases: molecular machines for the Genetic Revolution. Biochemical Journal *473*, 673–684.

Park, S.Y., Yang, D., Ha, S.H., and Lee, S.Y. (2018). Metabolic Engineering of Microorganisms for the Production of Natural Compounds. Advanced Biosystems *2*, 1700190.

Piazza, I., Kochanowski, K., Cappelletti, V., Fuhrer, T., Noor, E., Sauer, U., and Picotti, P. (2018). A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. Cell *172*, 358-372.e23.

Pickar-Oliver, A., and Gersbach, C.A. (2019). The next generation of CRISPR–Cas technologies and applications. Nature Reviews Molecular Cell Biology *20*, 490–507.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell *152*, 1173–1183.

Qin, Q., Wang, B., Wang, J., Chang, M., Xia, T., Shi, X., and Xu, G. (2019). A comprehensive strategy for studying protein-metabolite interactions by metabolomics and native mass spectrometry. Talanta *194*, 63–72.

Qu, J., Prasad, N.K., Yu, M.A., Chen, S., Lyden, A., Herrera, N., Silvis, M.R., Crawford, E., Looney, M.R., Peters, J.M., et al. (2019). Modulating Pathogenesis with Mobile-CRISPRi. Journal of Bacteriology *201*.

Reznik, E., Christodoulou, D., Goldford, J.E., Briars, E., Sauer, U., Segrè, D., and Noor, E. (2017). Genome-Scale Architecture of Small Molecule Regulatory Networks and the Fundamental Trade-Off between Regulation and Enzymatic Activity. Cell Reports *20*, 2666–2677.

Ricard, J., and Cornish-Bowden, A. (1987). Co-operative and allosteric enzymes: 20 years on. European Journal of Biochemistry *166*, 255–272.

Rousset, F., Cui, L., Siouve, E., Becavin, C., Depardieu, F., and Bikard, D. (2018). Genome-wide CRISPR-dCas9 screens in E. coli identify essential genes and phage host factors. PLOS Genetics *14*, e1007749.

Rousset, F., Cabezas-Caballero, J., Piastra-Facon, F., Fernández-Rodríguez, J., Clermont, O., Denamur, E., Rocha, E.P.C., and Bikard, D. (2021). The impact of genetic diversity on gene essentiality within the Escherichia coli species. Nature Microbiology *6*, 301–312.

Sander, T., Wang, C.Y., Glatter, T., and Link, H. (2019). CRISPRi-Based Downregulation of Transcriptional Feedback Improves Growth and Metabolism of Arginine Overproducing *E. coli*. ACS Synthetic Biology *8*, 1983–1990.

Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeida, D., García-Sotelo, J.S., Alquicira-Hernández, K., Muñiz-Rascado, L.J., Peña-Loredo, P., et al. (2019). RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. Nucleic Acids Research *47*, D212–D220.

Schultenkämper, K., Brito, L.F., and Wendisch, V.F. (2020). Impact of CRISPR interference on strain development in biotechnology. Biotechnology and Applied Biochemistry *67*, 7–21.

Shachrai, I., Zaslaver, A., Alon, U., and Dekel, E. (2010). Cost of Unneeded Proteins in E. coli Is Reduced after Several Generations in Exponential Growth. Molecular Cell *38*, 758–767.

Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR–Cas9. Nature Reviews Genetics *16*, 299–311.

Shamir, M., Bar-On, Y., Phillips, R., and Milo, R. (2016). SnapShot: Timescales in Cell Biology. Cell *164*, 1302-1302.e1.

Shields, R.C., Walker, A.R., Maricic, N., Chakraborty, B., Underhill, S.A.M., and Burne, R.A. (2020). Repurposing the Streptococcus mutans CRISPR-Cas9 System to Understand Essential Gene Function. PLOS Pathogens *16*, e1008344.

Smith, H.O., and Welcox, K.W. (1970). A Restriction enzyme from Hemophilus influenzae. Journal of Molecular Biology *51*, 379–391.

Stokes, J.M., Lopatkin, A.J., Lobritz, M.A., and Collins, J.J. (2019). Bacterial Metabolism and Antibiotic Efficacy. Cell Metabolism *30*, 251–259.

Tagore, R., Thomas, H.R., Homan, E.A., Munawar, A., and Saghatelian, A. (2008). A global metabolite profiling approach to identify protein-metabolite interactions. J. Am. Chem. Soc. *130*, 14111–14113.

Todor, H., Silvis, M.R., Osadnik, H., and Gross, C.A. (2021). Bacterial CRISPR screens for gene function. Current Opinion in Microbiology *59*, 102–109.

Vinayavekhin, N., and Saghatelian, A. (2011). Discovery of a protein-metabolite interaction between unsaturated fatty acids and the nuclear receptor Nur77 using a metabolomics approach. J. Am. Chem. Soc. *133*, 17168–17171.

Wang, T., Guan, C., Guo, J., Liu, B., Wu, Y., Xie, Z., Zhang, C., and Xing, X.-H. (2018). Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. Nature Communications *9*.

Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available at: www.genome.gov/sequencingcostsdata. Accessed [21.06.2021, 09:30 am].

Wright, A.V., Nuñez, J.K., and Doudna, J.A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell *164*, 29–44.

Zampieri, M., Zimmermann, M., Claassen, M., and Sauer, U. (2017). Nontargeted Metabolomics Reveals the Multilevel Response to Antibiotic Perturbations. Cell Reports *19*, 1214–1228.

Zheng, Y., Su, T., and Qi, Q. (2019). Microbial CRISPRi and CRISPRa Systems for Metabolic Engineering. Biotechnology and Bioprocess Engineering *24*, 579–591.

Chapter 1

Multi-omics analysis of CRISPRi-knockdowns identifies mechanisms that buffer decreases of enzymes in *E. coli* metabolism

Michelle Kuntz[1,2,3], Stefano Donati[1,3], Vanessa Pahl[1,2], Niklas Farke[1,2], Dominik Beuter[1], Timo Glatter[1], José Vicente Gomes Filho[1], Lennart Randau[1], Chun-Ying Wang[1,2], Hannes Link[1,2,4]*

[1] Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

[2] Present address: Interfaculty Institute for Microbiology and Infection Medicine Tübingen, University of Tübingen, Auf der Morgenstelle 24, 72076 Tübingen, Germany

[3] These authors contributed equally

[4] Lead Contact

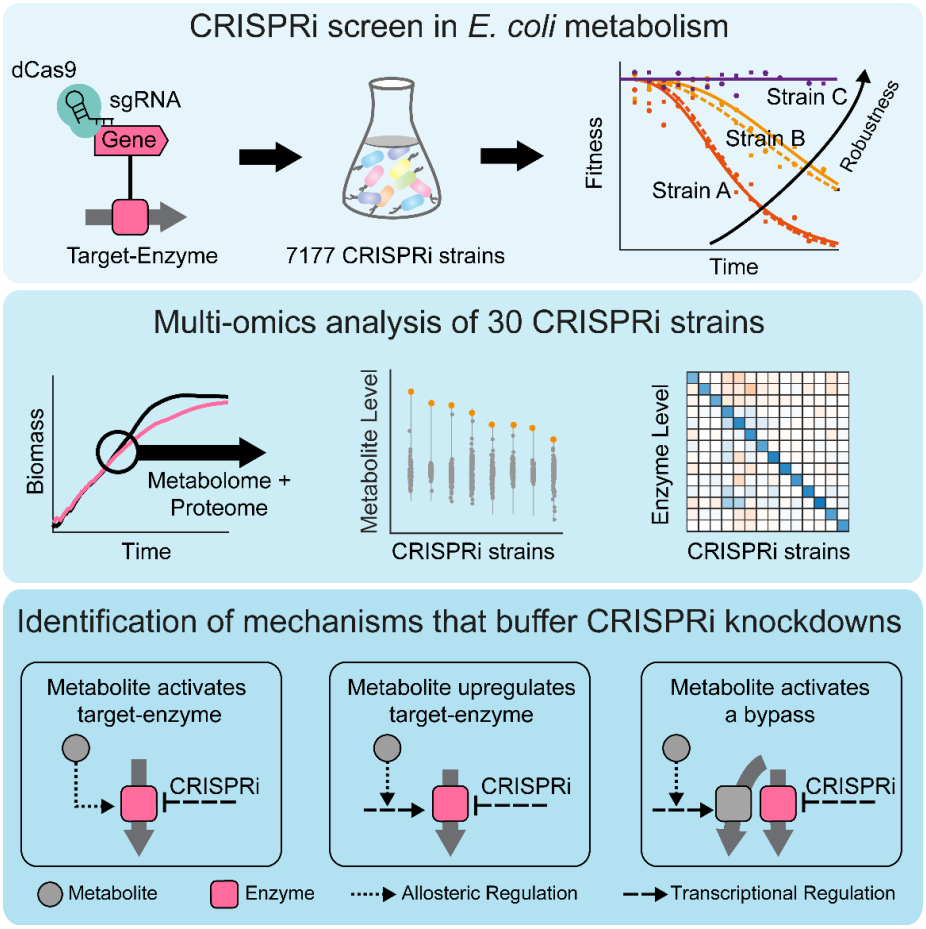
This chapter is written in manuscript style and was published in Cell Systems 2021, 12, 56–67.e1–e6. My contribution to this work included performing experiments with the pooled CRISPRi library as well as single CRISPRi strains, performing of Illumina sequencing and analyzing data from the pooled CRISPRi screen and co-writing the manuscript.

Summary

Enzymes maintain metabolism and their concentration affects cellular fitness: high enzyme levels are costly, and low enzyme-levels can limit metabolic flux. Here, we used CRISPR interference (CRISPRi) to study the consequences of decreasing *E. coli* enzymes below wild-type levels. A pooled CRISPRi screen with 7177 strains demonstrates that metabolism buffers fitness defects for hours after induction of CRISPRi. We characterized the metabolome and proteome responses in 30 CRISPRi strains and elucidated 3 gene-specific buffering mechanisms: Ornithine buffered the knockdown of carbamoyl phosphate synthetase (CarAB) by increasing CarAB activity, S-adenosylmethionine buffered the knockdown of homocysteine transmethylase (MetE) by de-repressing expression of the methionine pathway, and 6-phosphogluconate buffered the knockdown of 6-phosphogluconate dehydrogenase (Gnd) by activating a bypass. In total, this work demonstrates that CRISPRi screens can reveal global

sources of metabolic robustness and identify local regulatory mechanisms that buffer decreases of specific enzymes. A record of this paper's Transparent Peer Review process is included in the Supplemental Information.

Graphical Abstract



Introduction

Enzymes catalyze biochemical reactions that maintain metabolism and cell growth. Correspondingly, expression levels of enzymes influence cellular metabolism and fitness. Growth of *E. coli*, for instance, is affected by the abundance of single enzymes (Dekel and Alon, 2005; Li et al., 2014), as well as by the total mass of catabolic enzymes (You et al., 2013). However, it is not clear to which extent the expression of each single enzyme in a cell influences metabolism and fitness.

The consequences of perturbing enzyme-levels were investigated with knockout libraries of yeast and *E. coli* (Baba et al., 2006; Giaever et al., 2002). Studies with these libraries showed that the absence of a single enzyme has very precise and specific effects on metabolism (Fuhrer et al., 2017; Mülleder et al., 2016) and transcription (Kemmeren et al., 2014). However, knockouts are extreme cases and they are not feasible for key metabolic enzymes, which are essential for growth on single carbon sources like glucose. Moreover, knockouts are static and therefore they may reflect metabolic states that have already adapted at the level of gene expression (Hosseini and Wagner, 2018; Ishii et al., 2007) or by mutations (McCloskey et al., 2018a). Thus, it is difficult to learn about acute responses to perturbations of enzyme-levels using gene deletions.

To understand the consequences of enzyme-level perturbations around wild-type levels, a series of studies measured enzyme expression in different environmental conditions (Buescher et al., 2012; Gerosa et al., 2015; Hackett et al., 2016). These studies showed that most enzymes have different expression levels in different conditions, and that the average enzyme mass of *E. coli* cells changes more than 2-fold across conditions (Schmidt et al., 2016). However, expression changes across conditions were mainly driven by global growth-dependent regulation (Erickson et al., 2017), and delineating local regulation from these data has proven difficult (Gerosa et al., 2013; Keren et al., 2013). An approach to achieve more specific and localized changes of enzyme-levels is to delete regulatory proteins that control enzyme expression. For example, deleting protein kinases in yeast caused widespread and specific changes of enzyme levels (Zelezniak et al., 2018), and deletion of transcription factors in E. coli amino acid biosynthesis led to increases of only the enzymes that belong to the respective regulon (Sander et al., 2019). However, because deletion of regulators affects

expression of many enzymes simultaneously, it is still difficult to decipher which enzyme was responsible for certain metabolic phenotypes.

Thus, it remains an open question how cellular metabolism responds to moderate changes of a single enzyme. Such changes can occur in nature due to expression noise (Metzger et al., 2015; Newman et al., 2006) or mutations of genes that encode enzymes (Kacser and Burns, 1981). Control theory suggests that moderate changes of an enzyme have only small and local effects on metabolism, which means that local changes should not propagate globally (Levine and Hwa, 2007; Mazat et al., 1996). This robustness of metabolism is somewhat expected, but mostly theoretical studies examined the mechanisms that enable metabolic robustness (Chandra et al., 2011; Grimbs et al., 2007). The few studies that measured robustness against changes of enzyme abundance focused on specific pathways (Fendt et al., 2010; Tanner et al., 2018), but robustness was not measured at a metabolism-wide level.

Recent developments of targeted genome modification methods have advanced our ability to perturb expression of single genes with high precision and high throughput. For example, synthetic promoter libraries in yeast mapped the precise relationship between the expression level of genes and cellular fitness (Keren et al., 2016). Many of the 80 target genes in this study encoded key metabolic enzymes and their expression level had little effects on fitness of yeast. Another method to modulate gene expression is CRISPR interference (CRISPRi), which represses transcription of a target gene with a complex of deactivated Cas9 (dCas9) and a single guide RNA (sgRNA) (Qi et al., 2013). Because CRISPRi is inducible, it permits time-resolved studies (Camsund et al., 2020; Rishi et al., 2020) and functional analyses of genes that are essential and therefore not viable in knockout libraries (Peters et al., 2016; Rishi et al., 2020; Rousset et al., 2018). Many CRISPRi screens measured simple phenotypes like fitness and growth, but to our knowledge there is no comprehensive study that combined CRISPRi perturbations in metabolism with multi-omics analysis.

Here, we combined a metabolism-wide CRISPRi screen with multi-omics analysis of 30 CRISPRi strains to investigate how E. coli metabolism responds to decreases of enzyme-levels. First, we measured competitive fitness of 7177 strains in a metabolism-wide CRISPRi library. An inducible CRISPRi system enabled us to measure the time-delay between inducer addition and appearance of fitness defects. Only 7 CRISPRi strains responded within the first 4 h after induction of CRISPRi, while fitness defects of most strains appeared with a considerable time-

delay (in average 7.8 h). This provided first evidence that E. coli metabolism is robust against decreasing enzymes below wild-type levels. The metabolome and proteome of 30 CRISPRi strains showed that changes in the metabolic network were local and specific. For example, target-enzymes were always among the most downregulated proteins (in average 5-fold). At the metabolome level, we observed strong concentration changes of substrate metabolites and allosteric effectors. We show for 3 CRISPRi strains that these changes contributed to buffering the knockdown: i) increases of ornithine buffered the CarAB knockdown by allosterically activating the enzyme, ii) S-adenosylmethionine triggered a compensatory upregulation of the methionine pathway in the MetE knockdown, and iii) 6-phosphogluconate was responsible for activation of the Entner-Doudoroff pathway in the Gnd knockdown. Overall, our results highlight the central role of regulatory metabolites in maintaining robustness against ever-changing concentrations of enzymes in a cell, which occur in nature due to stochastic effects like expression noise, cell division or fluctuating environments.

Results

An inducible CRISPRi system identifies rate-limiting enzymes

For dynamic knockdowns of enzymes, we used a CRISPRi system that consisted of an aTc-inducible dCas9 on the chromosome (Lawson et al., 2017), and a constitutively expressed single guide RNA (sgRNA) on a plasmid (Qi et al., 2013) (**Figure 1A**). To evaluate dynamics of gene interference with this CRISPRi system, we targeted a YPet reporter protein inserted in the *E. coli* genome (Lawson et al., 2017). These experiments showed an exponential decrease of the YPet content per cell, indicating a constant dilution of the YPet protein by growth (**Figure 1B**). The 1-hour delay between inducer addition and decrease of YPet is probably occurring due to the time of dCas9 expression and its target search (Jones et al., 2017). Moreover, YPet expression was only repressed in the presence of the dCas9 inducer aTc, showing tight control of the CRISPRi system (**Figure 1B**). Thus, CRISPRi allowed us to dynamically decrease the abundance of proteins starting from unrepressed (wild-type) levels.

To further test the dynamics of the CRISPRi system, we targeted genes encoding enzymes in pyrimidine nucleotide biosynthesis. All pyrimidine enzymes are essential for growth of *E. coli* on glucose minimal medium. Therefore, knockdowns of pyrimidine genes should cause a

growth defect when enzyme-levels reach a critical threshold. At this threshold the target-enzyme limits biosynthesis of UMP, and eventually affects growth (**Figure 1C**). Expression of dCas9 was either induced by supplementing aTc at the start of the cultivation (induced cultures), or cells were grown without inducer (un-induced cultures). A control strain without target grew similar in induced and un-induced cultures, which means that dCas9 expression alone causes no growth burden (**Figure 1D**). Un-induced cultures of all pyrimidine knockdowns grew like the control, confirming that the CRISPRi system is tight. Induced cultures, in contrast, displayed a wide range of growth phenotypes: knockdown of the first two enzymes of the pathway (PyrB and PyrC) hardly affected growth, while the PyrE knockdown caused a strong growth defect. Knockdown of PyrF and PyrD impaired growth as well, but the effect appeared relatively late after induction of CRISPRi (around 5 hours).

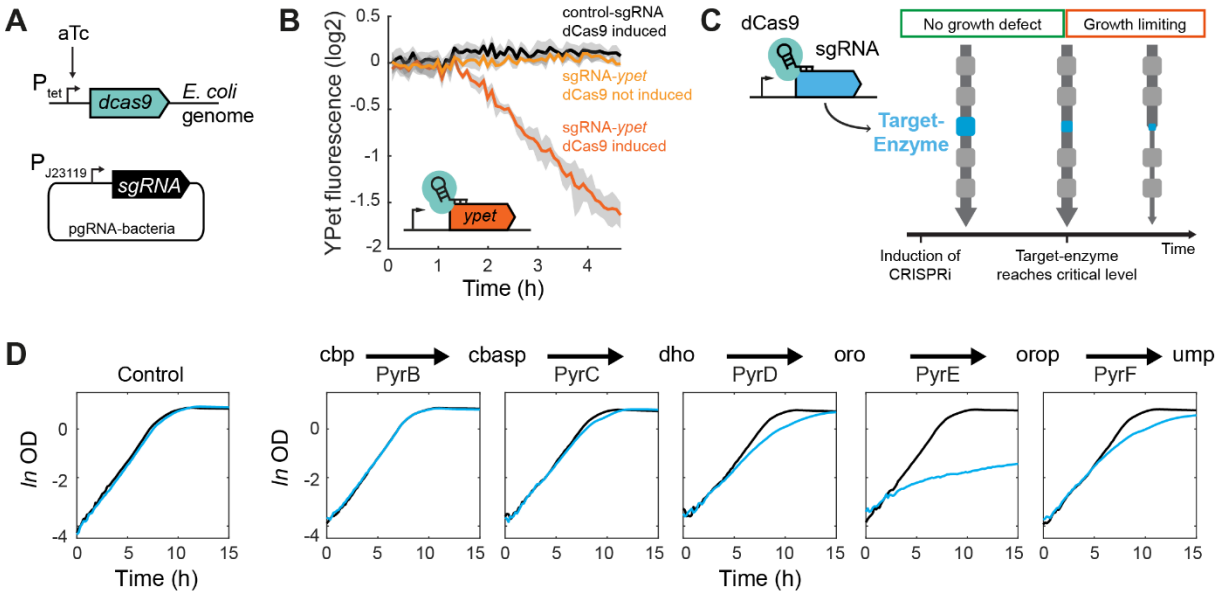


Figure 1. Dynamic knockdowns of enzymes with CRISPR interference
(A) The CRISPR interference system consisted of an *E. coli* strain (YYdCas9) that has dCas9 integrated into the genome (Lawson et al., 2017), and a single guide RNA on a plasmid (Qi et al., 2013). dCas9 is under control of an aTc inducible Ptet promoter. The sgRNA is under control of a constitutive promoter. (B) Dynamic knockdown of YPet, which is integrated into the genome of the YYdCas9 strain. YPet fluorescence is shown for cells that express either a control sgRNA (black) or a sgRNA that targets YPet (orange). YPet fluorescence per OD is normalized to an un-induced culture with the control sgRNA. The YPet knockdown was induced at time = 0 h by supplementing 200 nM of aTc. Data are represented as mean, and the grey areas are ± SD (n = 3). (C) Knockdown of an enzyme impairs growth when its concentration reaches a critical level. The target-enzyme is the enzyme, which is encoded by the gene that is repressed with CRISPRi. (D) Growth of cells expressing the control sgRNA, or sgRNAs targeting genes that encode enzymes in pyrimidine nucleotide biosynthesis. Expression of dCas9 was induced by supplementing 200 nM of aTc (blue) or dCas9 was not induced (black). Cells grew on minimal glucose medium in microtiter plates. Means of n = 3 cultures are shown.

In conclusion, CRISPRi allowed us to induce dynamic decreases of protein-levels (**Figure 1B**). The 5-hour delay between inducer addition and appearance of growth defects in the PyrF and PyrD knockdowns, suggests that the target-protein is diluted by growth until it reaches a critical level. In contrast, the early growth defect in the PyrE strain indicates that this enzyme is already expressed at a critical-level in the wild-type. This is consistent with previous reports about sub-optimal expression of PyrE in K12-derived *E. coli*, due to a frameshift mutation upstream of the *pyrE* gene (Jensen, 1993). The comparably weaker growth defects of the other pyrimidine knockdowns indicated that these enzymes do not operate at a critical level. However, an alternative explanation is that the sgRNAs for these targets are weaker or not functional. Therefore, we next targeted genes with several sgRNAs, and designed sgRNAs for all metabolism-related genes in *E. coli*.

E. coli metabolism is robust against CRISPRi-knockdowns of enzymes

The latest genome-scale model of *E. coli* metabolism, *i*ML1515, includes 1515 genes (Monk et al., 2017) and we constructed sgRNAs that target these genes using array-synthesized oligonucleotides (**Figure 2A**). Per gene we designed 4 to 6 sgRNAs that target different loci on the coding strand. The resulting sgRNAs were cloned in a pooled approach and subsequently transformed into *E. coli* that carried dCas9 on the genome (**Figure 1A**). Sequencing of the CRISPRi library showed that 7177 unique sgRNAs were present in the library and they target 1513 of the 1515 genes in the *i*ML1515 model (**Figure S1** and **Table S1**). We cultured the library for 13 h on glucose minimal medium without induction of dCas9, which hardly changed the composition of the library: the fold-change of single CRISPRi strains after 13 h was normally distributed around 1 and only 47 out 7177 strains (0.6%) showed a fold-change >2 (**Figure S1**). The stable composition of the un-induced library confirms again tight control of the CRISPRi system. Subsequently, we induced dCas9 expression and followed the library composition by next generation sequencing for 14 h in intervals of 1 h (**Figure 2A**). Every two hours, the cultures were back-diluted into fresh medium, to avoid limitations of oxygen and nutrients. To assess reproducibility, we used two independent cultivations. Fitness-scores of single CRISPRi strains were quantified as fold-change of sgRNA counts, which were reproducible between the two independent cultivations (**Figure S2**).

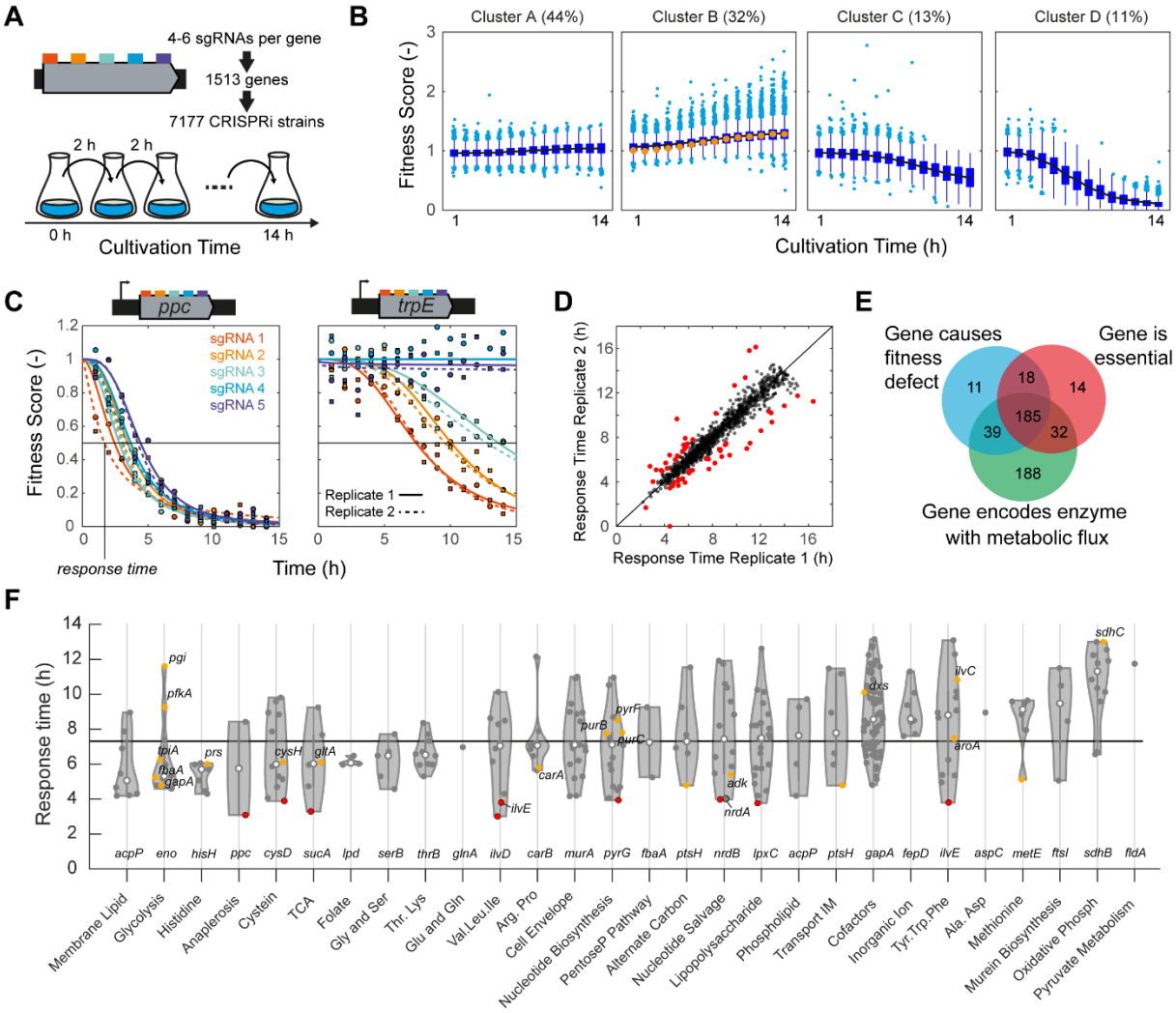


Figure 2. Dynamic knockdowns of 1513 genes in the metabolic network of *E. coli*

(A) A CRISPRi library targeting 1513 genes in the latest genome-scale reconstruction of *E. coli* metabolism (*i*ML1515). Each gene was targeted with 4-6 sgRNAs, which are equally distributed on the coding strand. sgRNAs were cloned in a pooled approach on plasmid pgRNA-bacteria and YYdCas9 was transformed with the resulting plasmid library (see also **Figure 1A**). The library was induced with 200 nM aTc at time = 0 h, and cultured for 14 h in shaking flasks. The culture was back-diluted every 2 hours into fresh medium. Samples for next generation sequencing were collected every hour. See also **Table S1**. (B) K-means clustering of fold-changes of 7177 sgRNAs. Time-course data was clustered into k = 4 clusters. Box plots represent the distribution of sgRNAs in each cluster per time point. Orange dots in Cluster B indicate a control strain that expresses a sgRNA with no target. See also **Table S2**. (C) Examples of fitness-score dynamics of CRISPRi strains (*ppc* with low variability between 5 sgRNAs, and *trpE* with high variability between 5 sgRNAs). Sigmoidal curves were fitted to the time-course of each sgRNA. The response time was defined as the time point when the fold-change of a sgRNA was 0.5. Different colors are different sgRNAs. Full and dashed lines are fits to experiment 1 (squares) and experiment 2 (circles). (D) Response times in experiment 1 and experiment 2. Shown are 1182 sgRNA that had response times below 14 h (average of exp1 and exp2). 57 sgRNAs that had more than 20% error are shown in red. See also **Table S3**. (E) Venn diagram showing the overlap between 253 “metabolic bottleneck genes” (blue), genes that are essential on glucose minimal medium (red), and genes that encode enzymes with metabolic flux (green). (F) Response times of all 253 “metabolic bottleneck genes”. See also **Table S4**. Shown is the average response time of the 2 strongest sgRNAs of each gene. Genes are grouped into metabolic categories according to the definition in *i*ML1515. The name of the most sensitive target is shown for each category. Red dots are genes with response times below 4 h. Orange dots are the 30 targets in **Figure 3**. See also **Figure S1-S4**.

To explore dynamic patterns of fitness-scores of the 7177 CRISPRi strains in the library, we performed k-means clustering with their individual time profiles (**Figure 2B**). Fitness-scores of 45% of the CRISPRi strains were constant for 14 h (cluster A). Another 30% of the strains in cluster B showed a slight increase of fitness-scores. This cluster included a control strain that expressed a sgRNA with no target (orange line in Cluster B, **Figure 2B**). This shows that increasing fitness-scores are due to a relative enrichment of strains that have wild-type like growth. However, some strains had higher fitness-scores than the control strain suggesting that these knockdowns confer a competitive advantage over the wild-type. Knockdowns of 18 genes resulted in fitness-scores >1.5 after 14 h (with at least two sgRNAs, **Table S2**), thus indicating that expression of these genes is not optimal on glucose minimal medium. Two of the sub-optimally expressed genes encoded enzymes that produce important secondary messengers in *E. coli*: cyclic-AMP (*cyaA*) and ppGpp (*relA*). This observation is consistent with a previous study that showed sub-optimal regulation by cyclic-AMP and ppGpp in *E. coli* (Towbin et al., 2017). We confirmed the fitness advantage of the *relA* strain in microtiter plate cultivations (**Figure S3**).

The remaining 25% of sgRNAs in cluster C and D caused mild and strong fitness defects, respectively. The sigmoidal dynamics of fitness-scores in cluster D suggest that the CRISPRi strains in this cluster drop out from the library, presumably because the knockdown created a metabolic bottleneck. To identify at which time point the knockdowns created a metabolic bottleneck, we estimated a “response-time” for each CRISPRi strain by fitting a sigmoidal function to the time-course of the fitness-score (**Figure 2C** and **Table S3**). The response-time was defined as the time point when the knockdown caused a 50% reduction of fitness, and response times were reproducible between the two experiments (**Figure 2D**). In total, 253 genes were targeted by at least 2 sgRNAs that caused response-times below 14 h, and we refer to these 253 targets as metabolic-bottleneck genes (**Table S4**). Most metabolic-bottleneck genes had similar response times for the 4-6 sgRNAs: 70% had responses times that varied less than +/- 20% between different sgRNAs (**Figure S4**). The different sgRNAs bind at different positions of the target genes, and therefore they should have different repression efficiencies (Qi et al., 2013). Yet, the position hardly affected response times (**Figure S4**). This result indicates that repression efficiencies have smaller effects on response times than target-specific factors.

The majority of the metabolic-bottleneck genes (203 out of 253, 80%) are essential for growth on glucose medium (**Figure 2E**). According to simulations with the *i*ML1515 model, 224 of the 253 metabolic-bottleneck genes (88%) encode enzymes that carry metabolic flux with glucose as sole carbon source. Only 11 genes (4%) are neither essential nor encode enzymes with metabolic flux (**Table S4**). For 3 of the 11 genes, the fitness defect can be explained by polar effects, because an essential or flux-carrying gene is encoded downstream of the targeted gene and in the same operon. The remaining 8 genes may have previously unrecognized functions that have strong effects on cellular fitness.

The average response time of the 252 bottleneck genes was 7.8 h, which is relatively late compared to the 7 most sensitive targets that had a response time below 4 h (red dots in **Figure 2F**). The 7 most sensitive targets were: the *ilvE/ilvD* operon, *ppc*, *sucA, lpxC*, *cysD*, *pyrG* and the *nrdA/nrdB* operon. A hypothesis is that these genes encode enzymes that are rate-limiting steps and therefore they are expressed near critical levels. For example, ribonucleoside-diphosphate reductase (NrdAB) supplies deoxyribonucleotide triphosphates (dNTPs) for DNA replication, and previous work showed that NrdAB is rate-limiting for DNA synthesis (Gon et al., 2006). Similarly, PEP carboxylase (Ppc) supplies TCA-cycle precursors for biosynthesis of 10 out of the 20 amino acids (anapleurosis). Thus, near-critical Ppc levels may limit overall protein synthesis. This hypothesis is supported by the observation that overexpression of Ppc increases the growth rate of *E. coli* (Chao and Liao, 1993).

In summary, only 7 out of 1513 metabolism-related genes had response times below 4 h. The majority of knockdowns, however, responded late to induction of CRISPRi (on average 7.8 hours). This suggests that *E. coli* is robust against reducing the abundance of most metabolic enzymes, and that few enzymes are expressed near critical levels. Next, we wondered how strongly the abundance of target-enzymes decreased and which mechanisms buffered decreases of enzymes.

CRISPRi enforces consistent decreases of target-enzymes and specific proteome responses

To probe how strongly CRISPRi downregulated the target-enzymes, we measured the proteomes of 30 CRISPRi strains (**Figure 3A** and **Table S5**). The target-enzymes included one of the most sensitive targets in the pooled CRISPRi screen: PEP carboxylase (Ppc) that converts

PEP to oxaloacetate in *E. coli*. We also included PckA which catalyzes the reverse reaction and should have no relevance for growth on glucose. Other targets were distributed over the metabolic subsystems, like glycolysis (Pts, Pgi, PfkA, PfkB, FbaA, GapA, Eno, TpiA, PykA, PykF) and the oxidative pentose phosphate pathway (Zwf and Gnd). From the TCA cycle we selected the first step catalyzed by citrate synthase (GltA), as well as the succinate dehydrogenase complex (SdhABCD). Furthermore, 8 target-enzymes were in biosynthesis pathways of amino acids (AroA, IlvC, MetE, GdhA) and nucleotides (Adk, PyrF, PurB, PurC), or both (Prs and CarAB). The remaining targets were CysH in sulfur assimilation, GlmS in amino sugar biosynthesis and Dxs in the isoprenoid pathway. We cultured these strains in microtiter plates and measured their proteomes 4.5 h after dCas9 induction. At this time-point, growth phenotypes appeared in 10 out of 30 CRISPRi strains (**Figure 3B** and **Figure S5**). Each strain was cultured in triplicates with and without induction of dCas9, resulting in a total of 180 proteome samples.

The proteome data showed that all target-enzymes decreased to a similar extent (in average 5.1-fold, **Figure 3C** and **Figure S6**). In 20 of 29 knockdowns the target-enzyme was the most downregulated protein among all 1506 measured proteins (**Figure S7**). Target-enzymes hardly decreased in un-induced cultures, showing that the CRISPRi system is tight and inducible (**Figure 3C**). We confirmed for the PfkA and MetE strain that target-enzymes were also downregulated at earlier time points (**Figure S8**), supporting our assumption that target-enzymes decrease progressively after induction of CRISPRi (similar to the YPet knockdown, **Figure 1B**).

Intuitively, stronger decreases of the target-enzyme should cause a stronger growth defect. However, there was no correlation between decreases of target-enzymes and reduction of growth (**Figure S9**). This indicates again that repression efficiencies have smaller effects on growth phenotypes than target-specific factors (e.g. overcapacities of the target-enzyme itself). However, we observed a correlation between reduction of growth and the number of significantly changed proteins (2-fold, p-test<0.05, **Figure S9**). This means that strains with a growth defect had stronger proteome changes, whereas the proteome was stable in strains without a growth defect. We then analyzed if these proteome changes were a global growth-dependent response (Scott et al., 2010) or if proteome changes were specific.

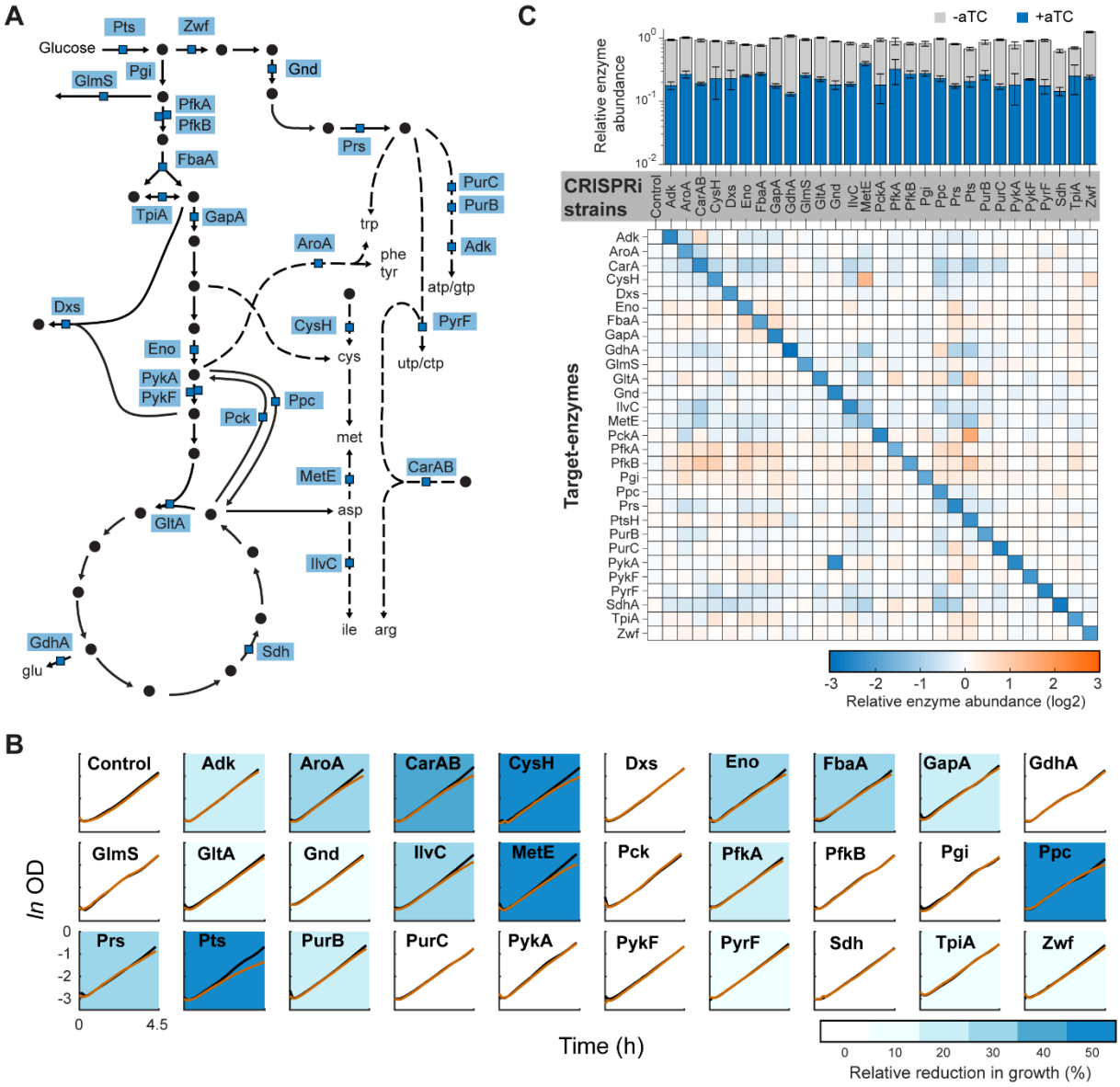


Figure 3. Growth defects and abundances of target-enzymes in 30 CRISPRi strains
Metabolic map showing the target-enzymes of 29 CRISPRi strains. The control strain expressed a sgRNA without a spacer sequence. Operon structures of the targets are shown in **Figure S6**. See also **Table S5**. (B) Growth curves of the 30 CRISPRi strains. See also **Table S6**. Un-induced cultures are shown in black. Induced cultures are shown in orange (200 nM aTc was supplemented at time = 0 h). Samples for proteomics were collected at the end of the cultivation (4.5 h). Growth curves show means of n = 3 cultures. Background colors indicate the reduction in growth rates at the time of sampling. Growth rates were estimated using linear regression with the last 4 time points of growth curves. Abbreviations of target-enzymes are described in **Table S6**. (C) The bar plot shows abundances of target-enzymes in cultures with inducer (blue) and without inducer (grey). Data is normalized to the average enzyme-level in un-induced cultures. The heatmap shows fold-changes of target-enzymes between induced and un-induced cultures. Data was calculated using the means of n = 3 samples per strain, error bars are propagated errors. See also **Table S7**. See also **Figure S5-S11**.

Because the average similarity of proteome changes between pairs of CRISPRi strains was only 6% (**Figure S10**), we concluded that each knockdown caused specific proteome changes. For example, the proteome changes affected different metabolic subsystems (**Figure S11**),

and in some CRISPRi strains (Pts, AroA, MetE, CarAB, IlvC), enzymes in the same metabolic subsystem as the target-enzymes were upregulated (green dots in **Figure S7**).

In summary, CRISPRi decreased the abundance of target-enzymes in average 5-fold (**Figure 3C**). Decreases of target-enzymes hardly affected growth of 19 CRISPRi strains, while growth rates of 10 CRISPRi strains declined ~1 h before the sampling time-point (**Figure 3B**). Thus, *E. coli* metabolism tolerates substantial decreases of enzyme-levels and we next wondered which mechanisms enable this robustness.

Substrates and allosteric effectors respond to CRISPRi-knockdowns of enzymes

To understand how *E. coli* metabolism responded to the ~5-fold decrease of target-enzymes, we measured the metabolome of the 30 CRISPRi strains. Therefore, we collected samples for metabolomics at the same time point as proteomics samples (4.5 hours), and measured 119 intracellular metabolites by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Especially substrate metabolites responded strongly and specifically to knockdowns of enzymes (**Figure 4**). In 18 out of 29 knockdowns, the substrate increased more than 2-fold and was one of the most changing metabolites. Products, in contrast, were more stable than substrates (**Figure 4**, **Figure S12**). This observation is consistent with a study in yeast, which suggested that increases of substrates can maintain fluxes and global metabolite homeostasis (Fendt et al., 2010).

In 4 strains allosteric effectors of the target enzyme responded most strongly to the knockdown (CarAB, GlmS, Ppc and Zwf, **Figure 4**). Ornithine, for example, is an allosteric activator of carbamoyl phosphate synthetase (CarAB) and ornithine increased 512-fold in the CarAB knockdown. Because ornithine levels in *E. coli* are 37-fold lower than the activation constant of CarAB (Bennett et al., 2009; Bueso et al., 1999), increases of ornithine should lead to a 92% higher activity of CarAB *in vivo* (**Figure S13**). Thus, allosteric activation of CarAB by ornithine could buffer the CarAB knockdown. Similarly, knockdown of Ppc decreased the concentration of aspartate (13-fold) and malate (16-fold), which are both allosteric inhibitors of Ppc. Absolute concentrations of malate and aspartate are above the respective inhibition constants of Ppc (Bennett et al., 2009; Gold and Smith, 1974). Therefore, decreases of aspartate and malate should relieve inhibition of Ppc, which increases its activity 4.1-fold in

the Ppc knockdown (**Figure S13**). In the GlmS and Zwf strain we observed a similar relieve from allosteric inhibition, because the respective reaction product glucoseamine-P and NADPH decreased. This is in line with previous work showing that decreases of NADPH release overcapacities of Zwf (Christodoulou et al., 2018) and that glucoseamine-P is a potent inhibitor of GlmS activity (Deng et al., 2006).

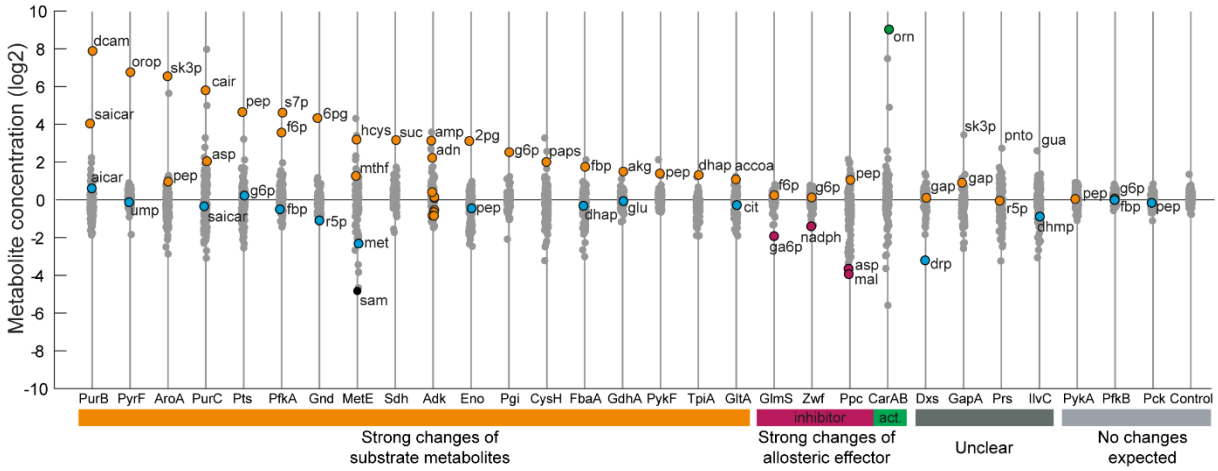


Figure 4. Metabolome changes in 30 CRISPRi strains are local and specific
(A) Intracellular concentration of 119 metabolites in the 30 CRISPRi strains. See also **Table S8**. Metabolite levels are shown as log2 fold-change between induced and un-induced cultures. Samples were collected after 4.5 hours cultivation in 12-well plates (see **Figure 3C**). Data are represented as mean (n = 2). Substrates of the target-enzymes are shown in orange, products in blue, allosteric inhibitors in magenta and allosteric activators are green. S-adenosylmethionine (sam) in the MetE strain is shown in black (related to **Figure 6**). Note that isomers were not separated: g6p and f6p is the total pool of hexose-phosphates, r5p is the total pool of pentose- phosphates, dhap and gap is the total pool dhap/gap. Abbreviations of metabolites are described in **Table S5**. See also **Figure S12**.

In conclusion, knockdowns of enzymes caused specific and localized metabolome changes: 22 CRISPRi strains showed strong concentration changes of either the substrate metabolite or a known allosteric effector of the target-enzyme. In 4 CRISPRi strains (Dxs, GapA, Prs, IlvC) we could not directly link substrates or known effectors to the target-enzyme, although in the Dxs knockdown the product metabolite (1-deoxyxylulose-5-phosphate) decreased most strongly and might be an unknown regulator of the enzyme. Three other CRISPRi strains showed no metabolome changes, which was expected because the target-enzyme is not used on glucose (Pck) or a minor isoenzyme (PfkB and PykA). The local and specific metabolome changes in the CRISPRi strains are in line with previous work on enzyme-level perturbations (Fendt et al., 2010; Kacser and Burns, 1973), which proposed that local metabolome changes

buffer against global changes. Next, we sought to obtain further evidence for a buffering function of metabolite concentration changes in three CRISPRi strains (CarAB, MetE and Gnd).

Ornithine buffers the CarAB knockdown

Despite their potential to buffer knockdowns, allosteric effectors are probably not responding to CRISPRi because they are regulators, but rather because they are located up- or downstream of the target-enzyme. For example, ornithine increases in the CarAB knockdown most likely because ornithine carbamoyltransferase (ArgF and ArgI) is limited due to low levels of the CarAB product carbamoyl phosphate (**Figure 5A**). This also implies that allosteric activation of CarAB by ornithine cannot fully compensate the CarAB knockdown. Nevertheless, ornithine could have a buffering function that alleviates the consequences of the knockdown. To understand the regulatory role of ornithine in the CarAB knockdown, we developed a small metabolic model of CarAB and the arginine-pyrimidine branch point (**Star Methods** and **Figure 5A**). Kinetic parameters of the model were randomly sampled 1,000 times from physiologically meaningful ranges based on *in vitro* parameters. With each of the 1,000 parameter sets, we simulated the CarAB knockdown using two different models: the first model included allosteric activation of CarAB by ornithine (allosteric model), and the second model did not include this regulation (non-allosteric model). The allosteric model was more robust against the CarAB knockdown than the non-allosteric model (**Figure 5B**). Especially fluxes remained relatively constant in the allosteric model: 796 of the 1000 simulations maintained 95% of the initial steady state flux. In contrast, flux in the non-allosteric model decreased continuously to about 50% of the initial steady state. Moreover, concentrations of the end-products arginine and UTP/CTP were more stable in the allosteric model than in the non-allosteric model. These model results suggest that allosteric activation of CarAB by ornithine can minimize perturbations to metabolic flux and end-products in arginine and pyrimidine nucleotide biosynthesis.

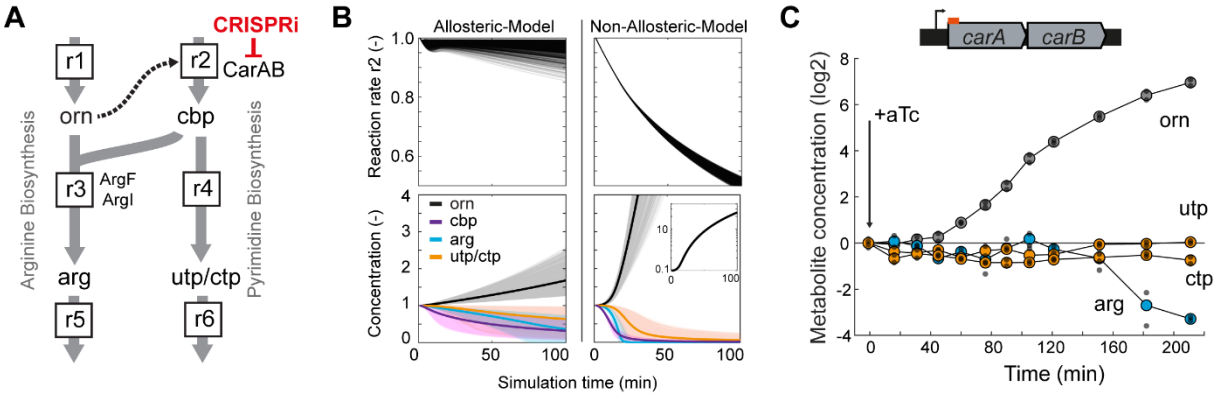


Figure 5. Ornithine buffers the CarAB knockdown
(A) Stoichiometry of the kinetic model of CarAB and the arginine-pyrimidine branch point. The dotted arrow indicates allosteric activation of CarAB by ornithine (orn). (B) Simulation results of the allosteric model and the non-allosteric model with 1000 parameter sets (thin lines). Thick lines are the average of 1000 simulations. Shown are the simulated reaction rate of r2 and metabolite dynamics of ornithine (orn, black), carbamoyl phosphate (cbp, purple), arginine (arg, blue) and utp/ctp (orange). CRISPRi was simulated by setting the expression rate of CarAB to zero at t=0 min. The insert shows the full range average ornithine levels in the non-allosteric model. (C) Measured concentration of ornithine (orn), arginine (arg), utp and ctp in the CarAB knockdown. Metabolite levels are normalized to the time point before induction. The culture was induced with aTc at t = 0 min. Small grey dots are measurements in n = 2 cultures and large colored dots are the mean. See also **Figure S13**.

To confirm the model results we measured metabolites dynamically in the CarAB knockdown (**Figure 5C**). Consistent with the simulation results ornithine increased already 40 min after induction of the CarAB knockdown, while the end-products arginine, CTP and UTP remained relatively constant for at least 2 hours. The fast response of ornithine shows that the CarAB knockdown perturbs the arginine-pyrimidine branch point early after induction of CRISPRi. However, the perturbation did not propagate into the end-products of both pathways. Thus, the combination of a metabolic model and dynamic metabolite data provides additional evidence that ornithine has the potential to buffer the CarAB knockdown.

S-adenosylmethionine causes a compensatory upregulation of the methionine pathway in the MetE knockdown

Metabolome changes can also modulate gene expression through allosteric interactions between metabolites and transcription factors. For example, S-adenosylmethionine (SAM) is an allosteric activator of MetJ, which is a transcription factor that controls genes involved in methionine and SAM biosynthesis (**Figure 6A**). SAM was the most decreased metabolite in the MetE knockdown (**Figure 4**), and correspondingly all enzymes in the methionine biosynthesis pathway were upregulated in the MetE strain (except the target MetE, **Figure**

6B). Thus, we hypothesized that low levels of SAM de-activated the transcription factor MetJ, which in turn de-repressed expression of genes that encode enzymes in methionine and SAM biosynthesis. To confirm that MetJ responded to the MetE knockdown, we expressed GFP in the MetE strain using a MetJ regulated promoter (Zaslaver et al., 2006). Indeed, GFP expression increased with a 2-hour delay after induction of CRISPRi, showing that MetJ responded to the knockdown (**Figure 6C**). Thus, we concluded that low SAM levels caused a compensatory upregulation of the methionine pathway. This hypothesis is supported by SAM levels in all 30 CRISPRi strains: only two strains had low SAM levels (MetE and Ppc strain), and methionine biosynthesis enzymes increased only in those strains (**Figure S14**).

Next, we wondered if the SAM-MetJ regulation buffered the MetE knockdown. Therefore, we supplemented SAM to the MetE strain, and expected that this prevents decreases of SAM and consequently the compensatory upregulation of the methionine pathway. The growth defect of the induced MetE strain was indeed stronger in the presence of SAM (**Figure 6D**), thus indicating that decreases of SAM buffered the knockdown. The stronger growth defect in the presence of SAM was not due to a general toxic effect, as the un-induced MetE strain was not influenced by SAM (**Figure 6D**).

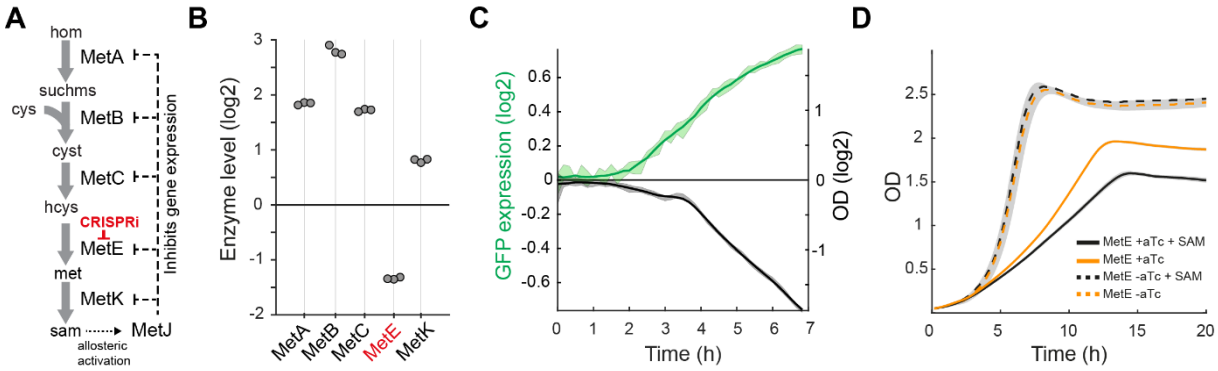


Figure 6. SAM buffers the MetE knockdown
(A) Schematic of methionine and S-adenosylmethionine (SAM) biosynthesis and regulation by the transcription factor MetJ. (B) Abundance of enzymes in methionine and SAM biosynthesis in the MetE strain. See also **Table S7**. Enzyme levels are shown as log2 fold-change between induced and un-induced cultures (n = 3 cultures). (C) The MetE knockdown was transformed with a fluorescent reporter plasmid that expressed GFP from a MetJ regulated promoter (pUA66-*metB-gfp*). The fold-change of GFP/OD between induced and un-induced cultures is shown in green. The fold-change of OD is shown in black. (D) Growth of the induced MetE strain (full lines) and the un-induced MetE strain (dashed lines), with supplementation of 1 mM SAM (black) and without (orange). Lines in (C) and (D) are means of n=3 cultures, and shadows show the standard deviation. Induced cultures were supplemented with 200 nM aTc at t = 0 h. See also **Figure S14**.

In summary, proteome and metabolome data recovered the known allosteric interaction between SAM and the transcription factor MetJ. GFP-promoter fusions confirmed that this regulation is active in the MetE knockdown, and supplementing SAM supported the hypothesis that SAM buffered the MetE knockdown. Similar to the SAM-MetJ interaction, the metabolome/proteome data recovered interactions between arginine and ArgR (active in the CarAB strain), acetyl-serine and CysB (active in the MetE and CysH strains) and transcriptional attenuation by valine (active in the IlvC strain) (**Figure S14**). This highlights the potential of CRISPRi and multi-omics data to identify regulatory metabolite-protein interactions that are functional *in vivo*.

6-phosphogluconate activates the Entner-Doudoroff pathway to bypass the Gnd knockdown

Knockdown of 6-phosphogluconate dehydrogenase (Gnd) increased the concentration of 6-phosphogluconate (6PG) (**Figure 4**), and upregulated enzymes in the Entner-Doudoroff (ED) pathway (**Figure 7A** and **Figure 7B**). Thus, we wondered if increases of 6PG are linked to upregulation of the ED pathway. Transcription of the ED pathway is regulated by two transcriptional repressors: KdgR and GntR. While KdgR controls only the two genes encoding the ED enzymes (Edd and Eda), GntR controls additional genes that are involved in gluconate uptake. One enzyme encoded by these genes (GntT) was also upregulated in the Gnd knockdown (**Figure 7B**), suggesting that GntR responded to the knockdown. The activity of GntR is allosterically inhibited by gluconate (Izu et al., 1997). Therefore, we assumed that accumulation of 6PG produced small amounts of gluconate, which inhibited GntR and de-repressed transcription of *edd* and *eda*. Indeed, the intracellular concentration of gluconate increased in the Gnd knockdown (**Figure 7C**). The concentration of gluconate was 50 µM in the un-induced Gnd strain, which is close to wild-type levels (42 µM) (Bennett et al., 2009). Induction of the Gnd knockdown increased gluconate to 184 µM, which was probably sufficient to inhibit GntR and de-repress expression of *edd* and *eda*.

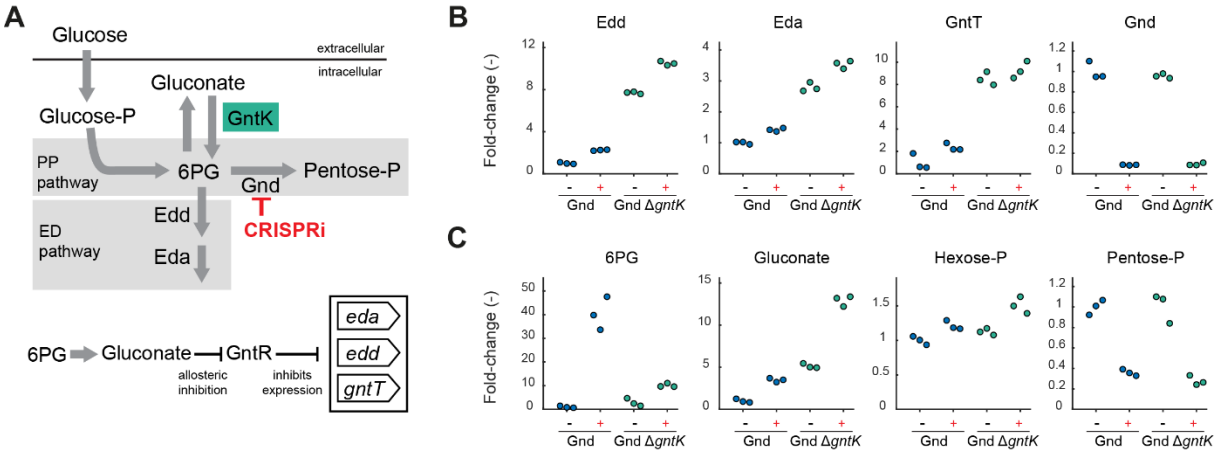


Figure 7. 6-phosphogluconate buffers the Gnd knockdown
(A) Schematic of the Entner-Doudoroff pathway (two enzymes Edd and Eda), and the oxidative Pentose-Phosphate pathway. GntK is a kinase that phosphorylates gluconate. Intracellular gluconate can derive from dephosphorylation of 6-phosphogluconate (6PG). (B) Fold-changes of the target-enzyme (Gnd), and fold-changes of all measured proteins that are regulated by the transcription factor GntR (Edd, Eda, GntT). Shown are induced (+) and un-induced (-) knockdowns of Gnd in the YYdCas9 strain (blue) and the YYdCas9-Δ*gntK* strain (green). Samples were collected after 4.5 hours cultivation in 12-well plates. Data is normalized to the un-induced Gnd strain (n = 3 cultures). (C) Same as in (B) for intracellular metabolites (6PG: 6-phosphogluconate). See also **Table S9**.

Increases of gluconate in the Gnd knockdown suggests that gluconate acts as regulatory metabolite, which does not participate in metabolism but in regulation. We expected that we can alter this regulation by disrupting the interconversion between 6-phosphogluconate and gluconate. Therefore, we deleted gluconate kinase (*gntK*) in the Gnd knockdown, which led to even higher gluconate levels in the Gnd knockdown (246 µM un-induced, and 620 µM induced, **Figure 7C**). The higher gluconate levels increased also expression of ED enzymes (**Figure 7C**). The higher abundance of ED enzymes in turn, reduced 6PG levels in the Δ*gntK*/Gnd knockdown, showing that the ED pathway is a bypass that enables overflow of excess 6PG.

In summary, the Gnd knockdown revealed a bypass function of the ED pathway, which has been observed before in the Gnd knockout (Jiao et al., 2003; McCloskey et al., 2018b). Here we discovered that expression of the ED-bypass is regulated by 6PG, which is first converted into gluconate and then interacts with the transcription factor GntR.

Discussion

Robustness is a fundamental feature of metabolism. A key requirement for metabolic robustness is that small changes in enzymes-levels have no global effects on overall metabolism. Otherwise, fluctuating enzyme-levels could decrease metabolic flux and eventually cellular fitness. Theories like Metabolic Control Analysis predicted that metabolism is insensitive to enzyme-level perturbations (Chandra et al., 2011; Grimbs et al., 2007; Kacser and Burns, 1973; Levine and Hwa, 2007), but have not measured this property at a system-level. Studies that measured flux-enzyme-metabolite relationships at a system-level examined the impact of nutritional changes on metabolism of yeast (Hackett et al., 2016) and E. coli (Gerosa et al., 2015). But how changes of enzyme-levels affect metabolism is largely unexplored.

In this study, we used CRISPRi to perturb the expression of single enzymes and found that metabolism buffers fitness defects during the initial phase after induction of CRISPRi. The opposite effect has been reported for CRISPRi-knockdowns in Bacillus subtilis (Peters et al., 2016), where a constant knockdown (~3-fold) prolonged initial lag-phases, but did not affect growth during exponential phase. Here, we observed that growth defects appeared with a time-delay and only few strains in a metabolism-wide CRISPRi library responded within the first 3-4 h (*ilvE*/*ilvD*, *ppc*, *sucA*, *lpxC*, *cysD pyrG*, *nrdA/nrdB*). Earlier studies support the high sensitivity of these targets, e.g. NrdAB seems rate-limiting for DNA synthesis (Gon et al., 2006) and overexpression of Ppc increases growth of E. coli (Chao and Liao, 1993). The high sensitivity of the ilvE/ilvD operon is probably due to the frameshift mutation upstream of ilvG, which causes suboptimal expression of these genes (Parekh and Hatfield, 1997). LpxC catalyzes the first committed step in lipid A biosynthesis and the enzyme is a drug target for antimicrobials (Löppenberg et al., 2013). Sensitive targets in glycolysis were mostly located in lower glycolysis, while upper glycolysis enzymes (PfkA and Pgi) had longer response times (**Table S2**). This observation is in line with previous reports about a thermodynamic bottleneck in lower glycolysis of E. coli (Flamholz et al., 2013).

Metabolome and proteome responses of 30 CRISPRi strains were local and specific, and together they indicated that metabolism buffers decreases of enzymes. This observation is supported by previous reports about overcapacities in metabolism, such as enzyme

overabundance (Davidi and Milo, 2017; O'Brien et al., 2016; Sander et al., 2019), reserve fluxes (Christodoulou et al., 2018) or overflow metabolism (Basan et al., 2015; Reaves et al., 2013). It has been suggested that cells control these overcapacities through metabolites that interact with enzymes directly (Christodoulou et al., 2018) or metabolites that modulate gene expression (Sander et al. 2019, Basan 2015). Substrate metabolites, for instance, can modulate enzyme activity through Michaelis-Menten relationships, and thereby buffer enzyme-level perturbations (Fendt et al., 2010) or modulate metabolic flux (Hackett et al. 2016). The strong increase of substrates in 18 knockdowns indicates that substrate-level regulation could be relevant in these strains. However, it remains open if these responses buffered lower enzyme levels, because in vivo kinetic parameters demonstrated that most of the enzymes are saturated (Bennett et al., 2009; Park et al., 2016). Yet, our three case studies (CarAB, MetE and Gnd) demonstrated that regulatory metabolites can contribute to buffering decreases of enzymes. Future studies could further probe the buffering capacity of metabolites by repressing the target gene at a lesser (or stronger) extent (Hawkins et al., 2020), and measure if this leads to milder (or stronger) metabolome and proteome changes.

In conclusion, our study shows that the metabolome responds specifically and locally to enzyme-level perturbations by CRISPRi, and that E. coli tolerates substantial decreases of enzymes. This supports the prevailing hypothesis that the abundance of single enzymes has little effects on metabolic flux and that local changes in metabolism do not propagate globally (Kacser and Burns, 1973). This mechanism may ensure a high constancy of metabolic flux despite expression noise (Newman et al., 2006; Taniguchi et al., 2010) or mutations that occur during the evolution of metabolic networks (McCloskey et al., 2018a).

Acknowledgments

We thank TJ Erb, V Sourjik and A Diepold for discussions. We thank J Elf for providing the YYdCas9 strain. This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 715650, ERC Starting Grant MapMe). SD and MK acknowledge funding from the IMPRS graduate school for environmental, cellular, and molecular microbiology from the Max Planck Society. LR acknowledges funding from DFG SPP2141.

Author Contributions

SD performed experiments with arrayed CRISPRi strains, analyzed data, performed FBA analysis and co-wrote the manuscript. MK performed experiments with the pooled CRISPRi library and co-wrote the manuscript. MK and SD analyzed data from the pooled CRISPRi screen. VP performed experiments with arrayed CRISPRi strains. NF developed and analyzed the CarAB model. DB constructed the pooled CRISPRi library. MK, JVGF and LR performed Illumina sequencing. TG measured proteomes. CYW performed experiments. HL conceived the study, analyzed data and co-wrote the manuscript.

Declaration of Interests

The authors declare no competing interests.

STAR Methods

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Hannes Link (hannes.link@synmikro.mpi-marburg.mpg.de).

Materials Availability

Plasmids and strains generated in this study are available on request from the Lead Contact, Hannes Link (hannes.link@synmikro.mpi-marburg.mpg.de).

Data and Code Availability:

Sequencing source data have been deposited at the European Nucleotide Archive (ENA) and are publicly available under the accession number: PRJEB40851. Proteome source data have been deposited at the PRIDE database and are publicly available under the accession numbers: PXD022070. Metabolome source data have been deposited at the Open Research Data Repository of the Max Planck Society (Edmond) and are publicly available at: https://edmond.mpdl.mpg.de/imeji/collection/u_8nsTTnbzAExmuZ. Original code of the CarAB model is publicly available at the Github repository: https://github.com/nfarke/Donati_Kuntz_et_al. Scripts used to generate the figures presented in this paper are not provided in this paper but are available from the Lead Contact on request. Any additional information required to reproduce this work is available from the Lead Contact.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Strains and Culture

E. coli YYdCas9 strain (Lawson et al., 2017) was the wild-type strain used in this study. NEB 5-alpha Competent *E. coli* (Cat#C2987) cells were used for cloning. All strains in this study derive from the YYdCas9 strain and are listed in the Key Resources Table.

Construction of arrayed strains

30 CRISPRi strains were created by transforming the YYdCas9 (Lawson et al., 2017) strain with pgRNA-bacteria plasmids that harbor the respective sgRNA (Addgene #44251). The spacer of sgRNAs consisted of a gene specific 20-22 base-pair region, which pairs adjacent to an NGG PAM site. The spacers were designed to bind as close as possible to the start of the coding sequence (**Supplementary Table 7**). Addgene #44251 was used as a template to prepare all plasmids, which were cloned inhouse or provided by Doulix. All plasmids were validated by sequencing. For CRISPRi of YPet, the sgRNA targeted lacZ, the first gene of the operon that includes YPet (Lawson et al., 2017). The plasmid pUA66 was used to measure promoter activity (Zaslaver et al., 2006). The Δ*gntK* mutant was constructed by P1 Phage transduction of YYdCas9 using the donor strain JW3400 (Δ*gntK*) from the KEIO collection (Baba et al., 2006). The resulting strain was cured from the kanamycin resistance gene included in the transduction cassette. The deletion of *gntK* was confirmed by sequencing. The final YYdCas9_Δ*gntK* strain was transformed with the pgRNA-*gnd* plasmid.

Construction of the CRISPRi pooled library

sgRNA guide sequences were designed with Matlab scripts by searching for 4 to 6 equally distributed NGG PAM sites on the coding strand of each gene in the iML1515 model (Monk et al., 2017). Adjacent to PAM sites, 20 nt regions were selected. 150 nt oligonucleotides were synthesized (Agilent Technologies). The 150 nt sequences contained the 20 nt sgRNA guide sequences and 65 nt flanking regions homologous to the pgRNA-bacteria backbone. Oligonucleotides were amplified with 15 cycles of PCR amplification. The pgRNA-bacteria backbone (containing the nontargeting spacer sequence 5'-AACTTTCAGTTTAGCGGTCT-3') was linearized by PCR and amplified oligonucleotides were inserted with Gibson assembly. The Gibson assembly product was purified and subsequently transformed into electrocompetent *E. coli* YYdCas9 cells. Plating on four Petri dishes with 15 cm diameter resulted in approximately 9.9×10^7 colonies. Colonies were washed from the plates, pooled and stored as glycerol stocks.

Media

Cultivations were performed with LB medium or M9 minimal medium with glucose as sole carbon source (5 g L^{-1}). M9 medium was composed by (per liter): 7.52 g Na_2HPO_4 2 H_2O, 5 g KH_2PO_4, 1.5 g $(NH_4)2_SO_4$, 0.5 g NaCl. The following components were sterilized separately and then added (per liter of final medium): 1 mL 0.1 M $CaCl_2$, 1 mL 1 M $MgSO_4$, 0.6 mL 0.1 M $FeCl_3$, 2 mL 1.4 mM thiamine-HCl and 10 mL trace salts solution. The trace salts solution contained (per liter): 180 mg $ZnSO_4$ 7 H2O, 120 mg $CuCl_2$ 2 H_2O, 120 mg $MnSO_4$ H2O, 180 mg $CoCl_2$ 6 H_2O. For strains transformed with pgRNA-bacteria plasmids, 100 μg mL^{-1} ampicillin (Amp) was added to the media. To induce expression of the dCas9 protein in the YYdCas9 strain, aTc was added to a final concentration of 200 nM. In experiments with pUA66 plasmids 34 μg mL^{-1} kanamycin was added to the medium.

METHOD DETAILS

Cultivation conditions for OD and YPet-, GFP-fluorescence measurements

Single colonies on LB+Amp agar plates were transferred into 5 mL LB+Amp liquid cultures. The LB pre-cultures were used to inoculate a second pre-culture in M9 medium that was incubated overnight in 13 mL culture tubes under shaking at 37°C. M9 pre-cultures were diluted in 150 μL M9 medium (1:50) and incubated in 96-well plates. Every strain was cultured in triplicates with and without addition of aTc to the M9 main culture (aTc was not added to pre-cultures). For YPet fluorescence measurements, 0.1 mM IPTG was added to pre-cultures and main cultures to induce YPet expression. Optical density (600 nm) and YPet fluorescence (excitation 510 nm, emission 540 nm) was measured every 5 min using a plate reader (BioTek, Synergy). For GFP measurements, GFP fluorescence (excitation 490 nm, emission 530 nm) was measured in 10 min intervals using a plate reader (Tecan, Spark).

Cultivation conditions for metabolome and proteome sampling

Single colonies were transferred into liquid 5 mL LB+Amp from fresh LB+Amp plates, and then re-inoculated in M9 medium overnight in 13 mL culture tubes under shaking at 37 °C. For metabolomics and proteomics sampling, M9 pre-cultures were adjusted to a starting OD_{600} of 0.05 into 12-well plates, with 2 mL of medium in each well. Strains were cultivated in triplicates with or without aTc, added at the beginning of the culture. Optical density at 600nm was measured every 10 min using a plate reader (Tecan, Spark) for 4.5 h. Plates were then rapidly transferred to a thermostatically controlled hood at 37 °C and kept shaking during the sampling procedure. For dynamic metabolomics, M9 pre-cultures were adjusted to a starting OD_{600} of 0.05 in a beaker containing 50 mL of medium and

a magnetic stirrer. Beakers were incubated with 400 rpm magnetic stirring in a thermostatically controlled hood at 37 °C.

Cultivation conditions of the pooled CRISPRi library

A preculture of 50 mL LB+Amp was inoculated with 500 µL of the pooled CRISPRi strain library from a glycerol stock and incubated at 37 °C for 5 hours. From the LB culture a second preculture in M9 was inoculated with a dilution of 1:10000 and incubated for 13 hours. After 13 hours the M9 preculture was in exponential phase and it was used to inoculate two main cultures with an initial OD of 0.05 in shaking flasks containing 100 mL of M9 with 200 nM of aTc to induce expression of dCas9. Every hour, OD was measured and samples for sequencing were collected. Every 2 hours, the culture was back-diluted to an OD of 0.05 with fresh and prewarmed M9 containing 200 nM of aTc. Samples were centrifuged to precipitate the cells and plasmids were extracted with the GeneJET Plasmid Miniprep Kit (ThermoFisher Scientific).

Next Generation Sequencing and Data Analysis

To generate the DNA fragments of target regions, which are compatible with Illumina sequencing, a two-step PCR approach was used. First, a 300 bp fragment including the sgRNA sequence and the flanking regions has been amplified using Q5 polymerase (New England Biolabs, USA) and specific oligonucleotides binding at the target region (NGS_F2_adapter and NGS_R2_adapter, **Supplementary Table 6**). As template, 150 ng of the purified samples were used in a 50 µL PCR reaction with the following settings: 98 °C for 30 s, 12 cycles of 98 °C for 10 s, 65 °C for 30 s and 72 °C for 15 s; final extension at 72 °C for 5 min. Afterwards, the PCR products were purified with a NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, Germany) and eluted in 20 µL water. In the second PCR, when different pairs of indexes (i5 and i7) were added to each amplicon, Phusion High- Fidelity DNA Polymerase (New England BioLabs, USA) was used with the following conditions: 98 °C for 30 s; 12 cycles of 98 °C for 10 s, 55 °C for 30 s and 72 °C for 20 s; final extension at 72 °C for 5 min. 4 ng of template was used in a final volume of 20 µL. Cleanup of the PCR products was done with AMPure XP beads (Beckman Coulter). All samples were run on a Bioanalyzer with an Agilent High Sensitivity DNA Kit (Agilent, USA) to analyze their composition. Next, 100 ng of each sample was pooled and the concentration of the pooled samples was measured using the Qubit dsDNA HS Assay on a Qubit 2.0 Fluorometer. The pooled samples were diluted, denatured and loaded on a MiniSeq High Output Cartridge following the manufacturer's instructions. To guarantee sufficient sequence diversity, 50% PhiX was spiked into the samples. Single-end reads provided sequences, which were mapped to the sgRNAs in the CRISPRi library using a Matlab Script. Read counts were calculated with single-end sequencing reads that matched to sgRNA guide sequences in the CRISPRi reference library

(**Supplementary Table 1**). Read counts per sgRNA ($reads_i$) were normalized to the total number of read counts per sample ($reads_{total}$) to obtain frequencies of sgRNAs. Frequencies were normalized to the first time point (t = 0 h) to calculate fold-changes.

Constraint-based modeling

Genes that encode enzymes with metabolic flux during growth on glucose were determined with Flux Balance Analysis (FBA). The *E. coli i*ML1515 metabolic model was downloaded from BiGG Models http://bigg.ucsd.edu/ (King et al., 2016) and FBA simulations were applied using COBRApy (Ebrahim et al., 2013) with parameters as described in Monk et al., 2017.

Kinetic modelling of the CarAB knockdown

The stoichiometry of the model is shown in Figure 5A. Mass balancing yields a system of ordinary differential equations (ODEs), F, that is a temporal function of the state variables x and the kinetic parameters p:

$$F(x,p) = \frac{dx}{dt} = \begin{cases} \frac{dorn}{dt} = r_1 - r_3 \\ \frac{dcbp}{dt} = r_2 - r_3 - r_4 \\ \frac{darg}{dt} = r_3 - \alpha_1 \cdot \mu_1 \\ \frac{dutp}{dt} = r_4 - \alpha_2 \cdot \mu_2 \\ \frac{de_2}{dt} = -\mu \cdot e_2 \end{cases} \quad \text{(Equation 1)}$$

The six reactions (r_1, r_2, r_3, r_4, r_5, r_6) are described by the following kinetic equations:

The influx into the arginine pathway r1 is constant:

$$r_1 = k_{cat1} \quad \text{(Equation 2)}$$

Allosteric activation of reaction r_2 by ornithine follows a power-law function:

$$r_2 = k_{cat,2} \cdot e_2 \cdot \left(\frac{orn}{orn_{SS}}\right)^{K_2} \quad \text{(Equation 3)}$$

,where orn_{SS} is the steady state ornithine concentration.

Reaction r_3 follows a non-ordered Bi-uni mechanism:

$$r_3 = k_{cat,3} \cdot \frac{1}{\left(1 + \frac{Km_{orn} \cdot Km_{cbp}}{orn \cdot cbp} + \frac{Km_{orn}}{orn} + \frac{Km_{cbp}}{cbp}\right)} \quad \text{(Equation 4)}$$

Reaction r4 follows simple Michaelis-Menten kinetics:

$$r_4 = k_{cat,4} \cdot \frac{cbp}{cbp + K_4} \quad \text{(Equation 5)}$$

The growth rate μ depends on r5 and r6, which follow Michaelis-Menten kinetics:

$$r_5 = \mu_{max,1} \cdot \frac{arg}{\mathrm{arg} + K_{\mu 1}} \quad \text{(Equation 6)}$$

$$r_6 = \mu_{max,2} \cdot \frac{utp}{\mathrm{utp} + K_{\mu 2}} \quad \text{(Equation 7)}$$

$$\mu = mean(\mu_1, \mu_2) \quad \text{(Equation 8)}$$

In total, the model includes 14 kinetic parameters k_{cat1}, k_{cat2}, k_{cat3}, k_{cat4}, K_2, $K_{m,orn}$, $K_{m,cbp}$, K_4, $K_{\mu 1}$, $K_{\mu 2}$, μ_{max1}, μ_{max2}, α_1 and α_2. The ensemble modelling approach (Tran et al., 2008) was used to account for uncertainties in kinetic parameters.

First, a steady flux distribution was calculated that is common for all subsequent parameter sets (r_1 = 0.958 mM min^{-1}, r_2 = 1.425 mM min^{-1}, r_3 = 0.958 mM min^{-1}, r_4 = 0.467 mM min^{-1}, r_5 = 0.958 mM min^{-1}, r6 = 0.467 mM min^{-1}). The flux distribution was estimated using flux balance analysis. Arginine and UTP efflux (r_5 and r_6) were calculated as the product of their biomass coefficients (α_1 = 95.8 mM, α_2 = 46.7 mM) and the growth rate on glucose (μ = 0.01 min^{-1}).

Binding constants (K-values) and metabolite concentrations (Ornithine = 0.01 mM, UMP = 0.50 mM, Arginine = 0.138 mM) were obtained from literature and Cbp concentration was set to 1 mM. The concentration of e2 was set to 1 mM. The binding constants were sampled 1000 times from 10-fold intervals based on literature values ($K_{m,orn}$ = 0.32 mM (argF/I, Ecocyc), $K_{m,cbp}$ = 0.36 mM (argF/I, Ecocyc), K_4 = 0.028 mM (Brenda ID: 696699), $K_{\mu 2}$ = 0.05 mM (pyrH, Ecocyc)). The power-law term K_2 was sampled between 1 and 4 in the regulated model and was set to zero in the dysregulated model. $K_{\mu 1}$ was fixed to 1×10^{-5} mM.

With the ensemble modelling approach the system is initially set into a steady state. To test stability of the steady states, eigenvalues of the Jacobian matrix were calculated, and tested if all eigenvalues are negative ($\lambda < -10^{-6}$). The procedure was repeated until 1000 stable steady states were achieved. The perturbation by CRISPRi was then simulated for all stable models by setting the expression rate of e2 to zero:

$$\frac{de2}{dt} = 0 - \mu \cdot e2 \quad \text{(Equation 9)}$$

Metabolomics measurements

Cultivations were performed as described above. Culture aliquots were vacuum-filtered on a 0.45 µm pore size filter (HVLP02500, Merck Millipore). Filters were immediately transferred into a 40:40:20 (v-%) acetonitrile/methanol/water extraction solution at -20 °C. Filters were incubated in the extraction solution for at least 30 minutes. Subsequently, metabolite extracts were centrifuged for 15 minutes at 13,000 rpm at -9 °C and the supernatant was stored at -80 °C until analysis. Metabolite extracts were mixed with a ^{13}C-labeled internal standard in a 1:1 ratio. LC-MS/MS analysis was performed with

an Agilent 6495 triple quadrupole mass spectrometer (Agilent Technologies) as described previously (Guder et al., 2017). An Agilent 1290 Infinity II UHPLC system (Agilent Technologies) was used for liquid chromatography. Temperature of the column oven was 30°C, and the injection volume was 3 µL. LC solvents in channel A were either water with 10 mM ammonium formate and 0.1% formic acid (v/v) (for acidic conditions), or water with 10 mM ammonium carbonate and 0.2% ammonium hydroxide (for basic conditions). LC solvents in channel B were either acetonitrile with 0.1% formic acid (v/v) (for acidic conditions) or acetonitrile without additive (for basic conditions). LC columns were an Acquity BEH Amide (30 x 2.1 mm, 1.7 µm) for acidic conditions, and an iHILIC-Fusion(P) (50 x 2.1 mm, 5 µm) for basic conditions. The gradient for basic and acidic conditions was: 0 min 90% B; 1.3 min 40 % B; 1.5 min 40 % B; 1.7 min 90 % B; 2 min 90 % B. The ratio of ^{12}C and ^{13}C peak heights was used to quantify metabolites. $^{12}C/^{13}C$ ratios were normalized to OD at the time point of sampling. Absolute concentrations of gluconate were determined from ^{12}C peak heights and an external calibration with an authentic standard. A specific cell volume of 2 µL mg^{-1} was used to calculate the cell volume.

Proteomics sample preparation and measurement

Cultivations were performed as described above. Culture aliquots were transferred into 2 mL reaction tubes and washed two times with PBS buffer (0.14 mM NaCl, 2.7 mM KCl, 1.5 KH_2PO_4, 8.1 Na_2HPO_4). Cell pellets were resuspended in 300 µL lysis buffer containing 100 mM ammonium bicarbonate, 0.5 % sodium laroyl sarcosinate (SLS). Cells were lysed by 5 minutes incubation at 95 °C and ultra-sonication for 10 seconds (Vial Tweeter, Hielscher). Cells were again incubated for 15 minutes with 5 mM Tris(2-carboxyethyl)phosphine (TCEP) at 90°C followed by alkylation with 10 mM iodoacetamide for 15 minutes at 25 °C. To clear the cell lysate, samples were centrifuged for 10 minutes at 15,000 rpm and the supernatant was transferred into a new tube. Protein samples were quantified using a BCA Protein Assay kit (Thermo Fisher Scientific). For each sample, 50 µg of proteins was aliquoted to new tubes, volumes were adjusted and cell lysates were digested with 1 µg trypsin (Promega) overnight at 30°C. SLS was removed by precipitation. Therefore, trifluoroacetic acid (TFA) was added to a final concentration of 1.5 % and incubated at room temperature for 10 minutes. After centrifugation (10 minutes at 10,000 rpm), the supernatant was used for C18 purification of peptides using Micro SpinColumns (Harvard Apparatus). The purified peptide solutions were dried and resuspended in 0.1 % TFA. The concentration of peptides in the samples was measured with a colorimetric peptide assay (Pierce™ Quantitative Colorimetric Peptide Assay, Thermo Fischer Scientific). Analysis of peptides was performed by with a Q-Exactive Plus mass spectrometer coupled to an Ultimate 3000 RSLC nano with a Prowflow upgrade and a nanospray flex ion source (Thermo Scientific). Peptide separation was performed on a reverse-phase HPLC column (75 µm x 42 cm) packed in-house with C18 resin (2.4 µm, Dr. Maisch GmbH, Germany). The following separating

gradient was used: 96 % solvent A (0.15% formic acid) and 4 % solvent B (99,85 % acetonitrile, 0.15 % formic acid) to 30 % solvent B over 60 minutes at a flow rate of 300 nL/min. The data acquisition mode was set to obtain one high resolution MS scan at a resolution of 70,000 full width at half maximum (at m/z 200) followed by MS/MS scans of the 10 most intense ions. To increase the efficiency of MS/MS attempts, the charged state screening modus was enabled to exclude unassigned and singly charged ions. The dynamic exclusion duration was set to 30 seconds. The ion accumulation time was set to 50 ms for MS and 50 ms at 17,500 resolution for MS/MS. The automatic gain control was set to 3x106 for MS survey scans and 1x105 for MS/MS scans. Label-free quantification (LFQ) of the data was performed using Progenesis QIP (Waters), and for MS/MS searches of aligned peptide features MASCOT (v2.5, Matrix Science) was used. The following search parameters were used: full tryptic search with two missed cleavage sites, 10ppm MS1 and 0.02 Da fragment ion tolerance. Carbamidomethylation (C) as fixed, oxidation (M) and deamidation (N,Q) as variable modification. Progenesis outputs were further processed with SafeQuant. The data was further processed with custom MATLAB scripts.

Quantification and Statistical Analysis

Statistical analysis was performed using custom Matlab scripts. The number of replicates (n) of each experiment can be found in the respective figure caption. In growth assays, n represents the number of independent microtiter plate cultures. For proteomics and metabolomics n represents the number of independent microtiter plate or shake flask cultures. Three replicates were used for metabolomics, and one of the three replicates was removed based on its Euclidean distance from the other two replicates. The remaining two replicates were used to calculate means. This removed outliers in the metabolome data set, which can occur due to the high sensitivity of the metabolome during sampling. In the proteomics dataset, proteins with an average variability between triplicates higher than 20% were removed. This left 1507 proteins that were measured in every sample. Significant proteins were defined with a two-fold cut-off and a p-value<0.05 for a two-sample t-test. Similarity of proteomes was obtained calculating the Jaccard index of significantly differentially expressed proteins.

Supplemental Information

Inventory

Supplementary Figures

Figure S1	Related to Figure 2. Library composition before and after cultivation
Figure S2	Related to Figure 2. sgRNA abundances of two replicates at all 14 timepoints
Figure S3	Related to Figure 2. Growth curve of cells expressing a sgRNA targeting *relA*
Figure S4	Related to Figure 2. Analysis of response times of different sgRNAs
Figure S5	Related to Figure 3. Growth curves of 30 arrayed CRISPRi strains
Figure S6	Related to Figure 3. Operon structure of target genes and relative enzyme abundance in knock-downs
Figure S7	Related to Figure 3. Proteome data of 30 CRISPRi strains
Figure S8	Related to Figure 3. Growth curves and proteome of m*etE* and p*fkA* knockdown
Figure S9	Related to Figure 3. Reduction of growth rates compared to target protein reduction and number of differentially expressed proteins
Figure S10	Related to Figure 3. Similarity between the differentially expressed proteins of the 30 measured proteomes
Figure S11	Related to Figure 3. Heatmaps of fold-changes of the protein-levels between induced and un-induced cultures
Figure S12	Related to Figure 4. Fold-changes of substrates and products in 30 arrayed CRISPRi strains
Figure S13	Related to Figure 5. Model of the allosteric activation of CarAB and allosteric inhibition of Ppc
Figure S14	Related to Figure 6. Fold-changes of enzymes in different amino acid biosynthesis pathways

Supplementary Tables

Table S1	Metabolism-wide CRISPRi library: 7177 CRISPRi strains plus control with read counts
Table S2	Source Data Figure 2B Fitness Scores of CRISPRi strains
Table S3	Source Data Figure 2D. Response times in the two experiments
Table S4	Source Data Figure 2F. Response times of 252 bottleneck genes
Table S5	Arrayed CRISPRi library: 30 strains with definition of target-enzymes, substrates, products
Table S6	Source Data Figure 3C. Growth data of 30 CRISPRi strains in 12-well plates
Table S7	Source Data Figure 3. Proteome of the 30 CRISPRi strains
Table S8	Source Data Figure 4. Metabolome of the 30 CRISPRi strains
Table S9	Source Data Figure 6. Metabolites and Proteins Gnd strain
Table S10	List of oligonucleotides used for cloning and Next-Generation Sequencing
Table S11	List of genotypes and spacer sequences of arrayed CRISPRi strains

Supplementary Figures

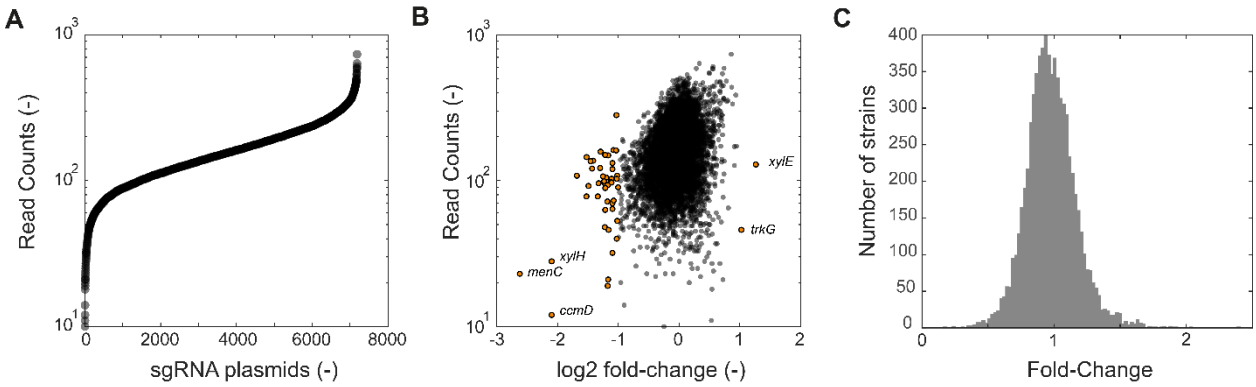


Figure S1. Related to Figure 2.

(A) Read counts of 7177 sgRNAs in the initial CRISPRi library. Shown are read counts after transformation of plasmid into *E. coli* YYdCas9 and cultivation on LB medium.

(B) Fold-change between sgRNA counts after 13 hours cultivation on M9 glucose medium (without induction), relative to the initial CRISPRi library. Fold-change is plotted against read counts of the initial library. sgRNAs with fold-change higher than 2 are shown in orange.

(C) Histogram of fold-changes between sgRNA counts after 13 hours cultivation on M9 glucose medium (without induction), relative to the initial CRISPRi library.

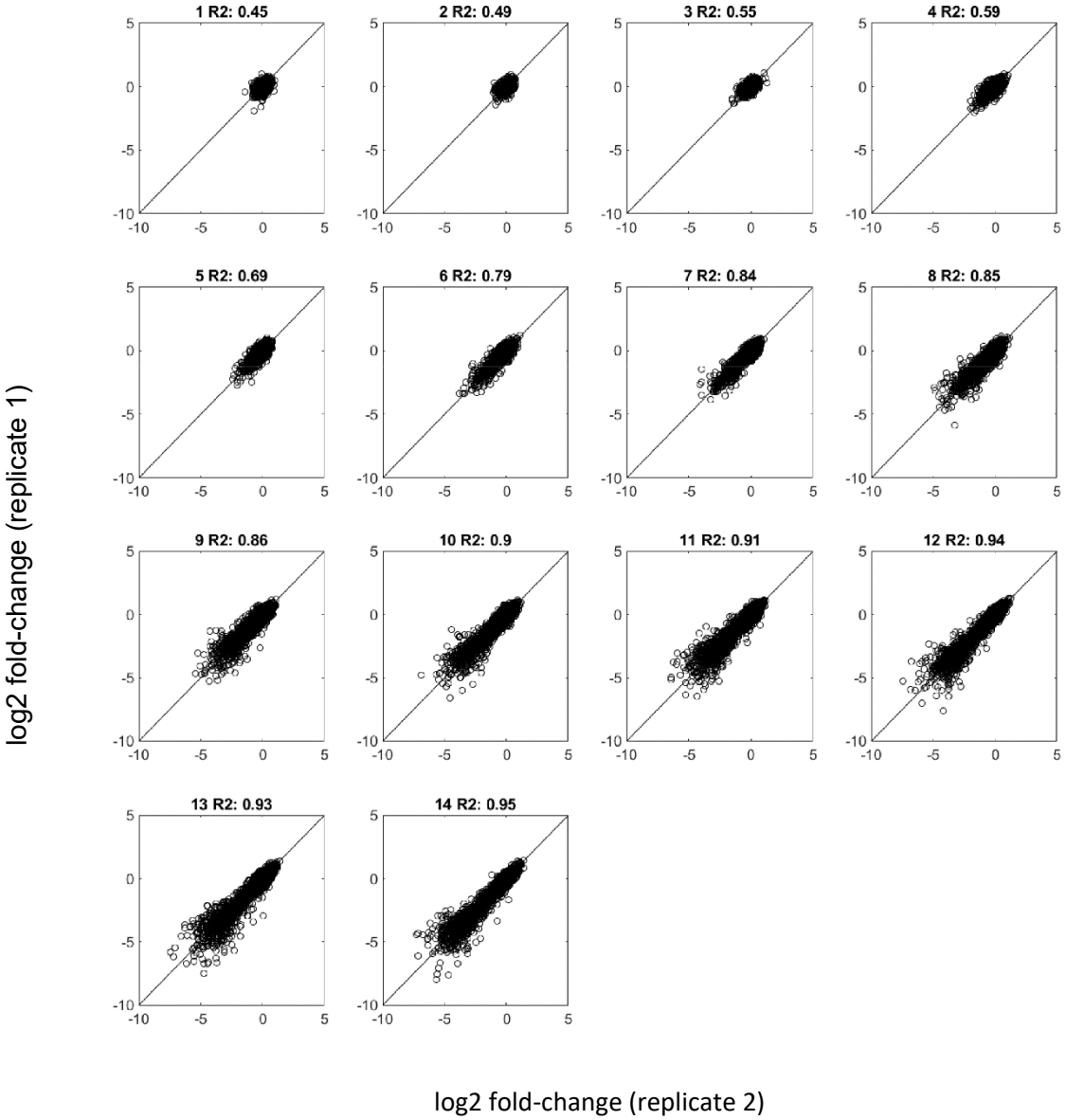


Figure S2. Related to Figure 2.

Fold-changes of sgRNA abundances in the two competition experiments. Each plot shows data for one of the 14 time points. R2 is the correlation coefficient between replicate 1 and replicate 2.

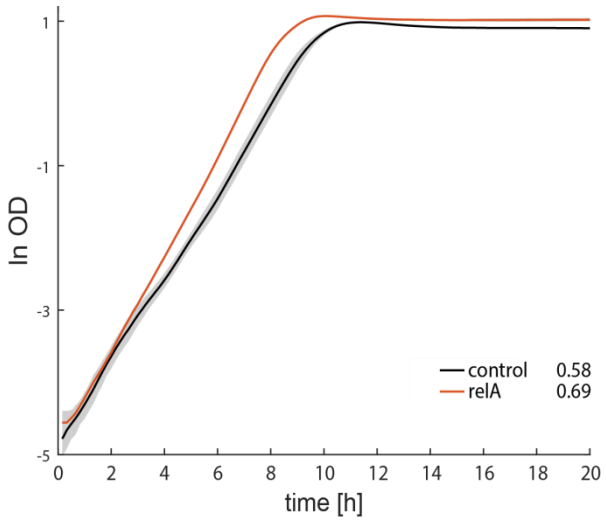


Figure S3. Related to Figure 2.

Growth of cells expressing the control sgRNA or a sgRNA targeting the gene *relA*. Expression of dCas9 was induced by adding 200 nM of aTc. Cells were grown on minimal glucose medium in microtiter plates. Means of n = 3 replicates are shown as lines, shadows are standard deviations. Maximum growth rates of the strains were calculated in the exponential phase and are shown as number in the legend (in 1/h).

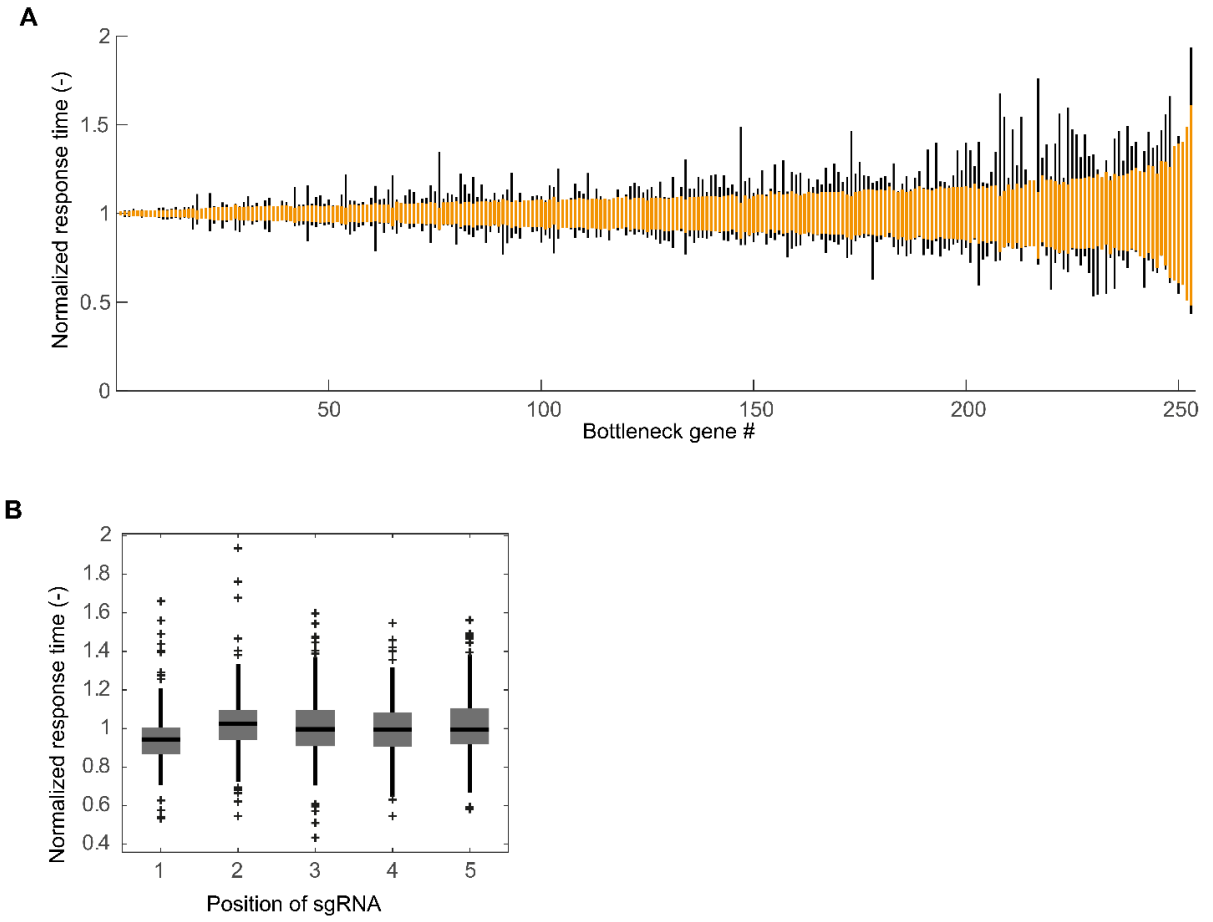


Figure S4. Related to Figure 2.

(A) Distribution of response times across 4-6 sgRNAs per target-gene. Shown are all 253 genes with response-times below 14 h (metabolic-bottleneck genes). Orange lines show the 50% percentile and black lines the 95% percentile. Response times were normalized to the average response time per target-gene.

(B) Distribution of response times across all 253 metabolic-bottleneck genes per position on the target-gene (position 1 is closest to the translation start site). Boxes show the 50% percentile and whiskers the 90% percentile. Response times were normalized to the average response time per target-gene.

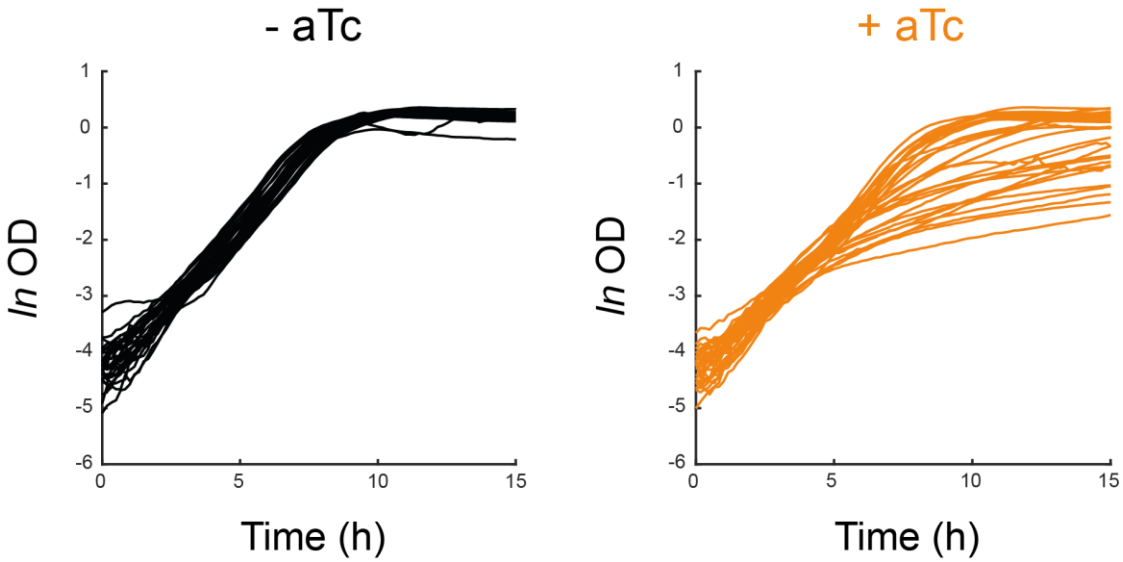


Figure S5. Related to Figure 3.

Growth of the 30 CRISPRi strains in 96-well plate cultures. Black are un-induced cultures and orange are induced cultures.

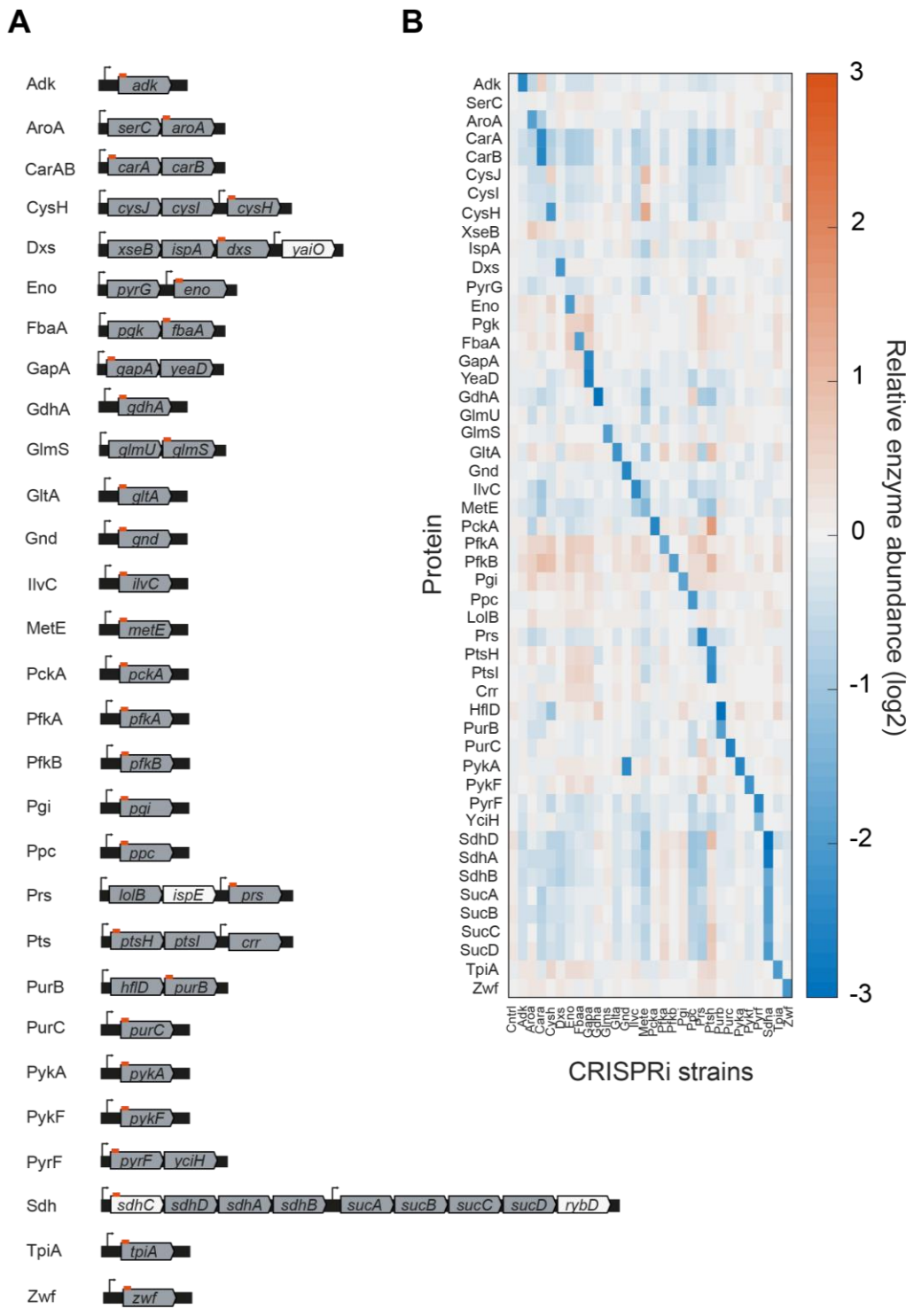


Figure S6. Related to Figure 3.

(A) Operon structure of the 29 target-genes. Red indicates the approximate loci that is targeted by sgRNAs. The genes in grey encode proteins that were measured. (B) Fold-changes of enzymes encoded by genes in the operons shown in (A). Data was calculated using the means of n = 3 samples per strain.

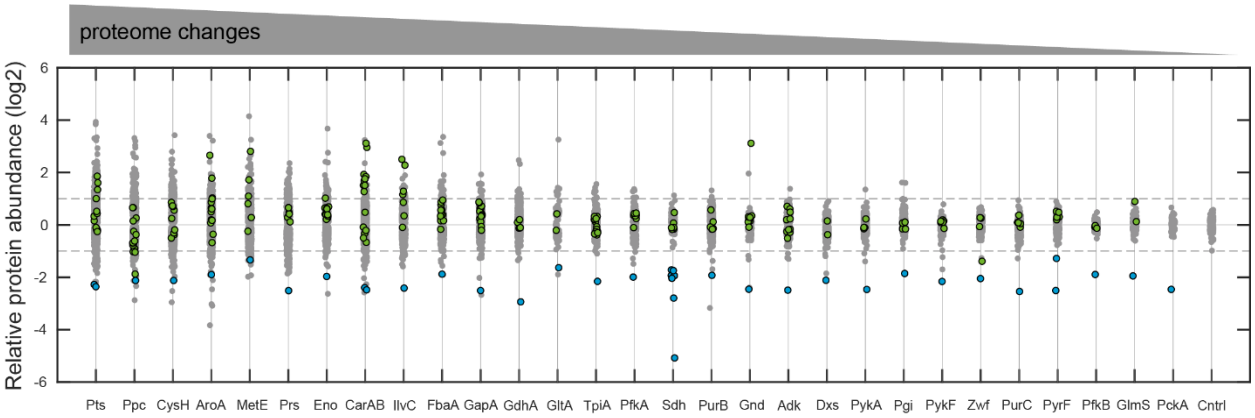


Figure S7. Related to Figure 3.

Proteome data of the 30 CRISPRi strains. Data is represented as log2 fold-change between samples from induced and un-induced cultures (n = 3). Only significantly changed proteins are shown ($p<0.05$). Strains are ordered according to the number of significantly changed proteins with fold-change > 2. Target-enzymes are shown in blue. Enzymes that are in proximity of the target-enzyme are shown in green.

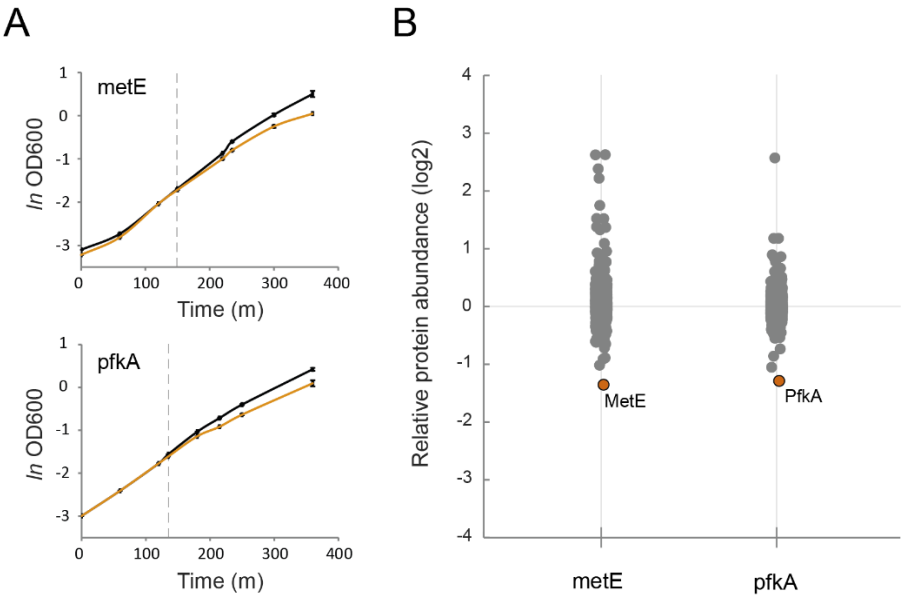


Figure S8. Related to Figure 3.

Proteome of the MetE strain and the PfkA strain at earlier time points (150 min MetE and 120 min PfkA).

(A) Growth of the MetE and PfkA strains. Un-induced cultures are black, Induced cultures are orange (n = 2 cultures). Dashed lines indicate the sampling time point.

(B) Fold change between the proteome of un-induced and induced MetE and PfkA strains (n = 2).

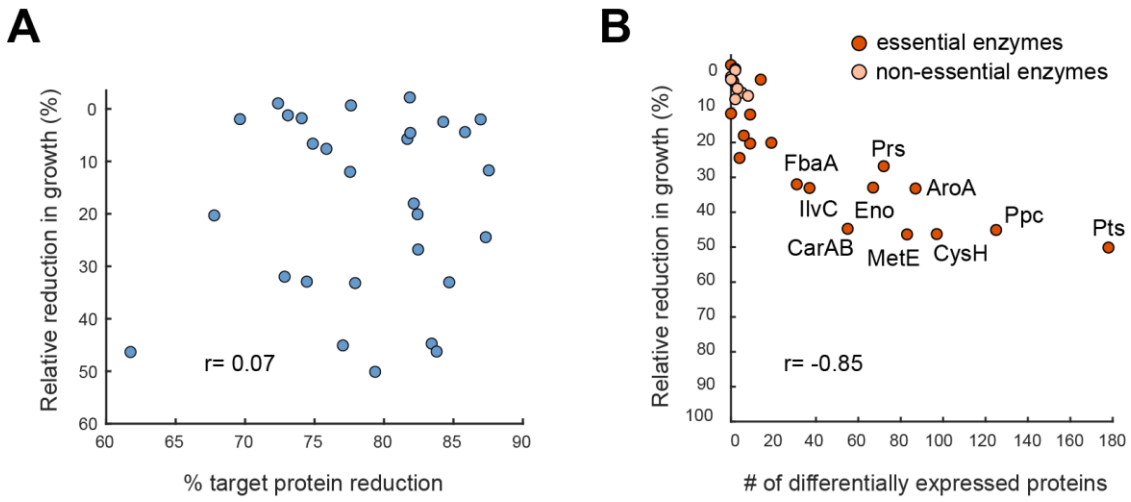


Figure S9. Related to Figure 3.

(A) Reduction in growth rates of 29 CRISPRi strains is shown on the y-axis. The reduction of the target-enzyme is shown on the x-axis.

(B) Reduction in growth rates of 29 CRISPRi strains is shown on the y-axis. The number of significant differentially expressed proteins (p-value<0.05, FC>2) is shown on the x-axis.

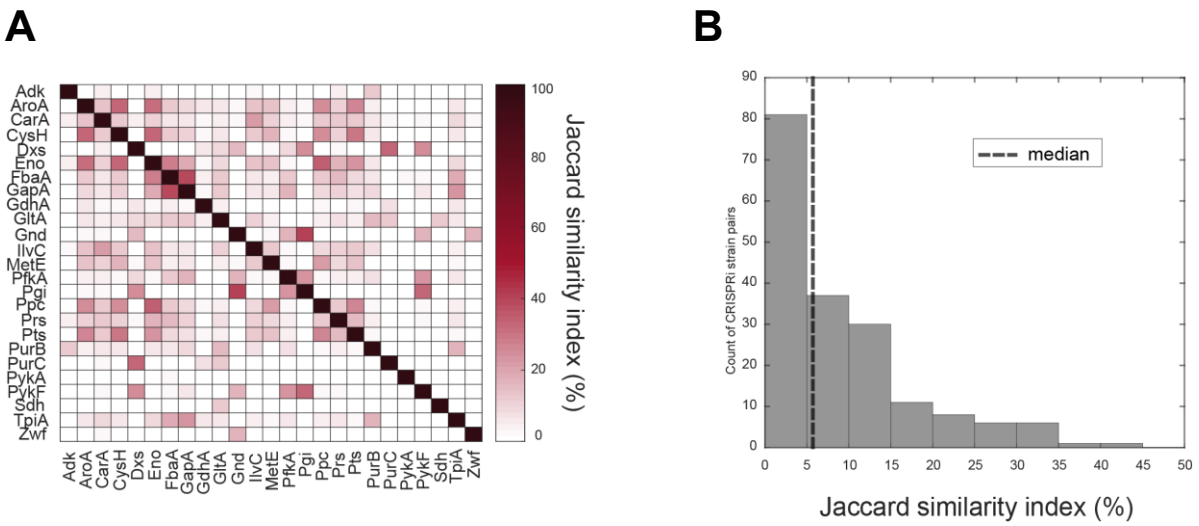


Figure S10. Related to Figure 3.

Similarity between the differentially expressed proteins of the 30 measured proteomes.

(A) Similarity matrix of differentially expressed proteins (FC=2, p-value<0.05) of the 30 measured proteomes. Similarity is defined as the Jaccard similarity index. Highest similarity was calculated for the pairs: Gnd-Pgi (40%), FbaA-GapA (38.89%), Eno-Ppc (34.27%), AroA-CysH (33.33%), CysH-Eno (32.26%), AroA-Eno (30.51%).

(B) Distribution of Jaccard similarity indexes between differentially expressed proteins in different strains. The median similarity for the distribution is 5.7%.

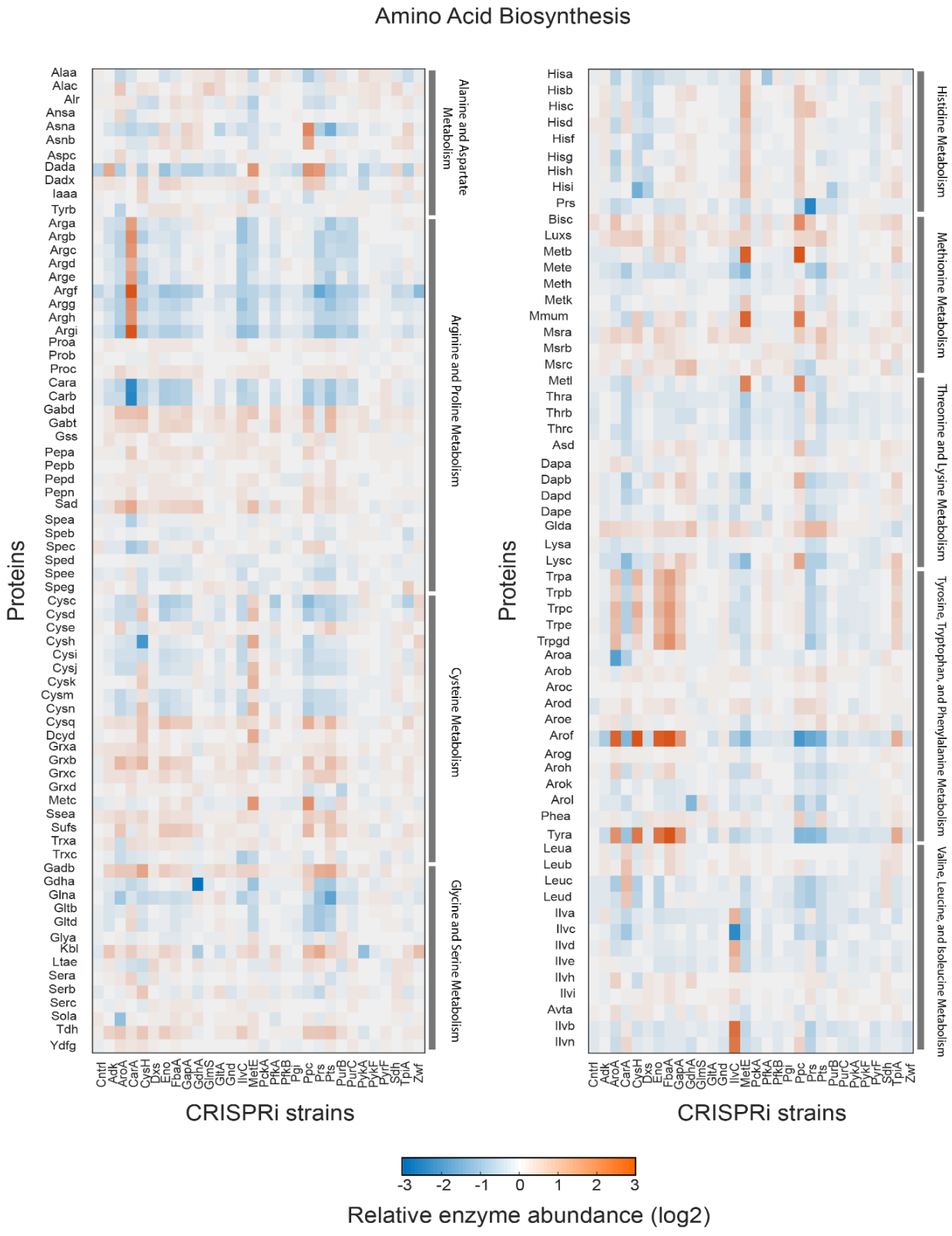


Figure S11. Related to Figure 3.

Heatmaps show log2 fold-changes of proteins between induced and un-induced cultures. Data was calculated using the means of n = 3 samples per strain. Data is organized based on metabolic subsystems in *i*ML1515.

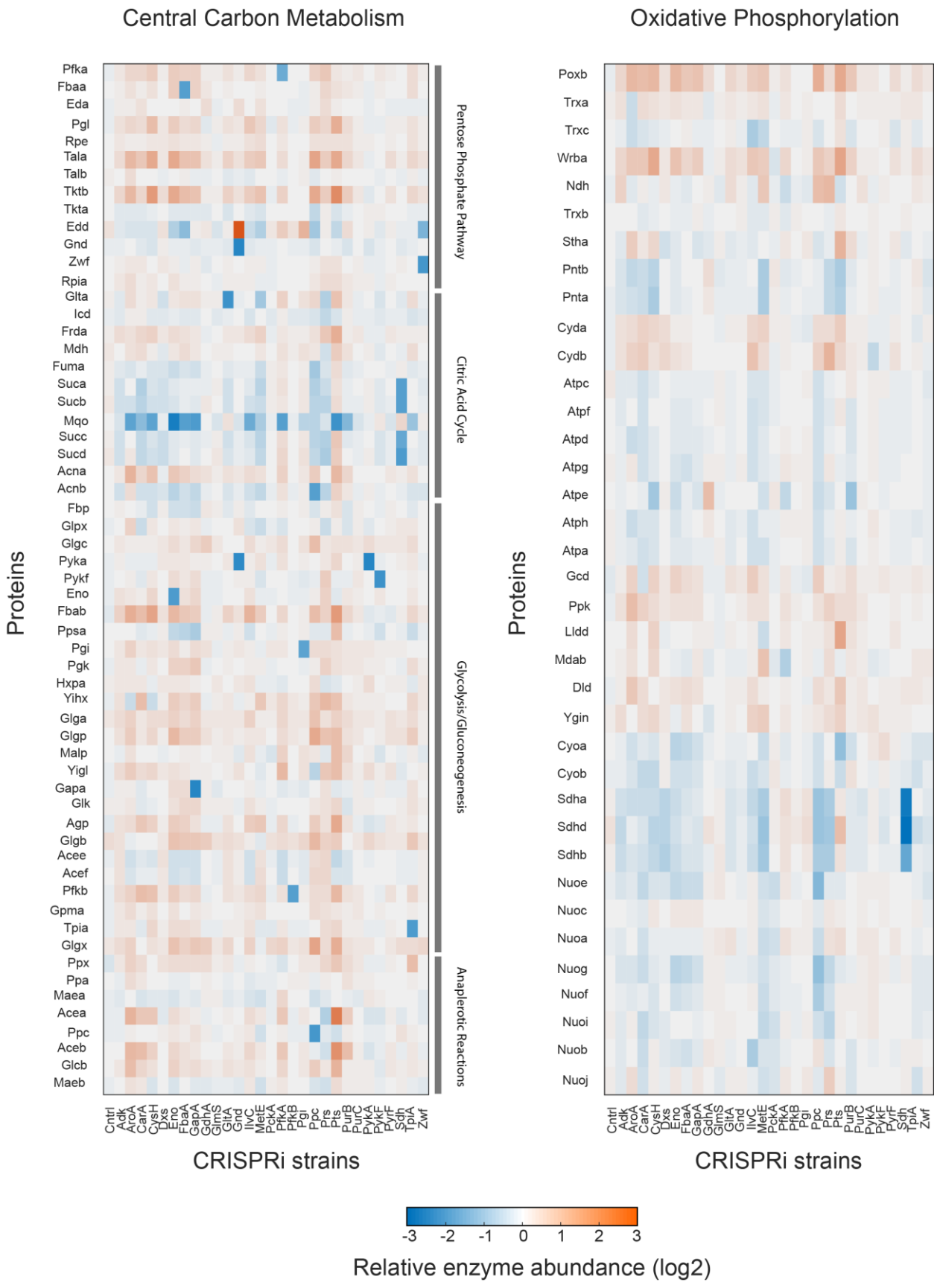


Figure S11. (continued)

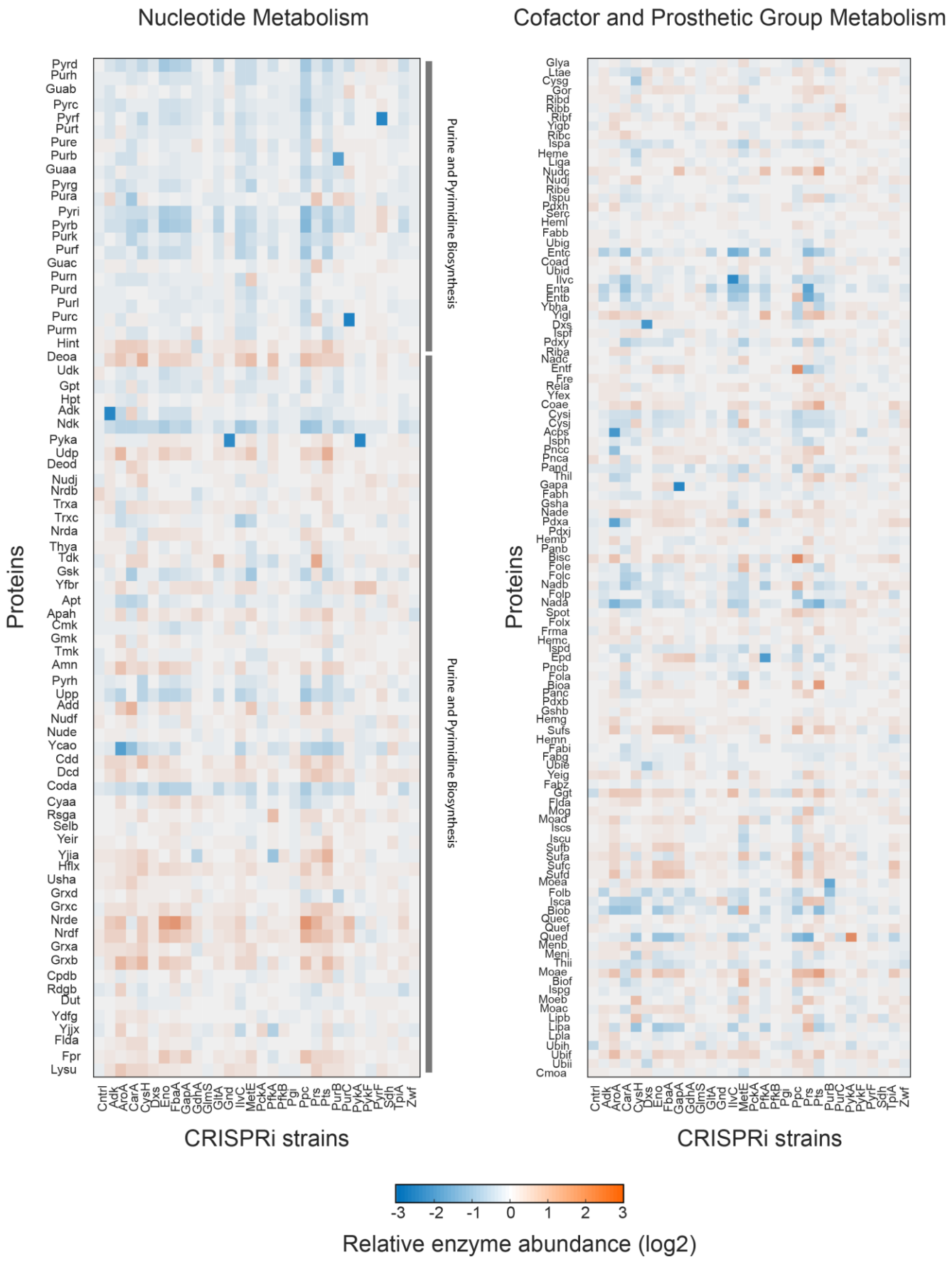


Figure S11. (continued)

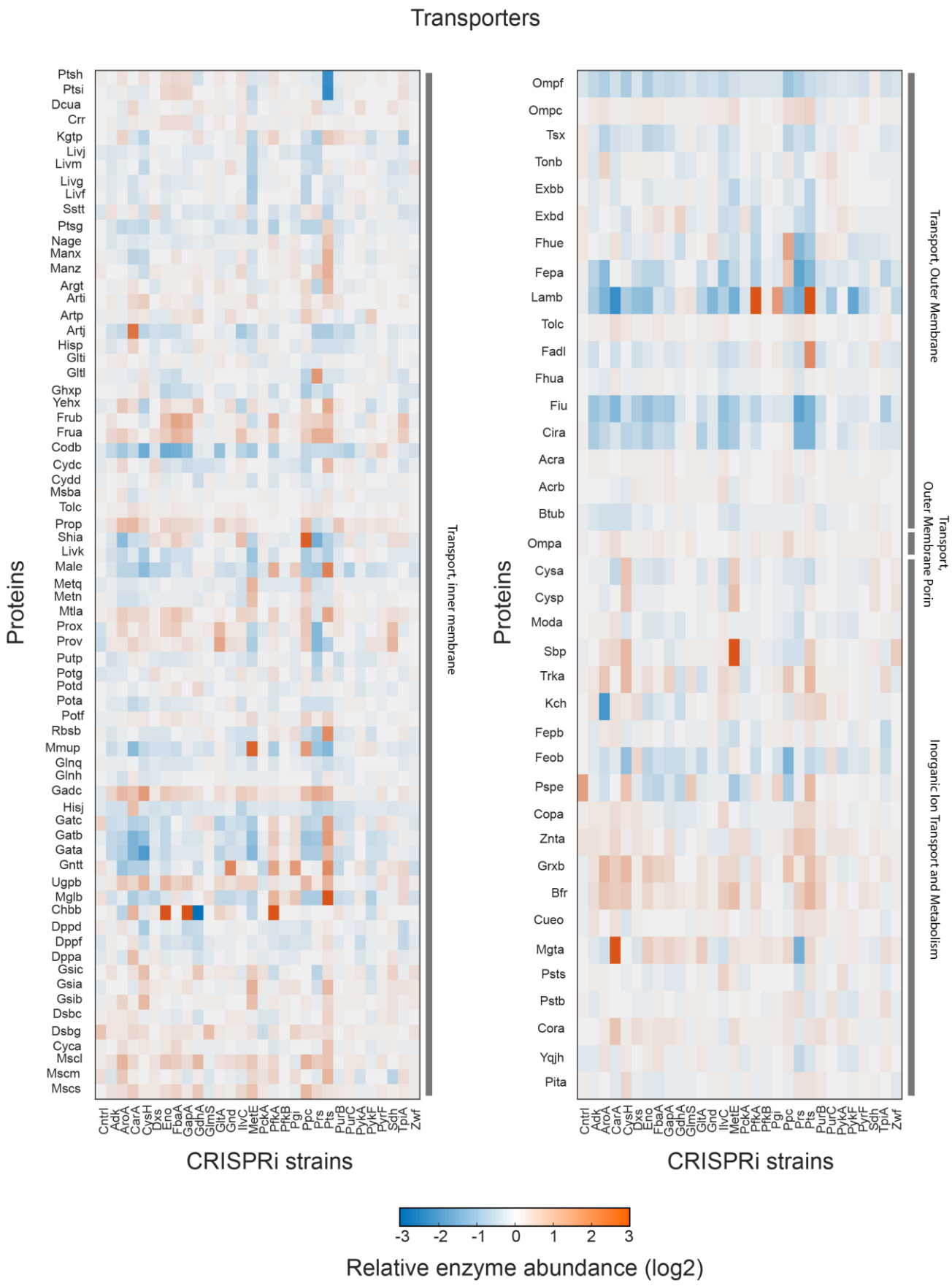


Figure S11. (continued)

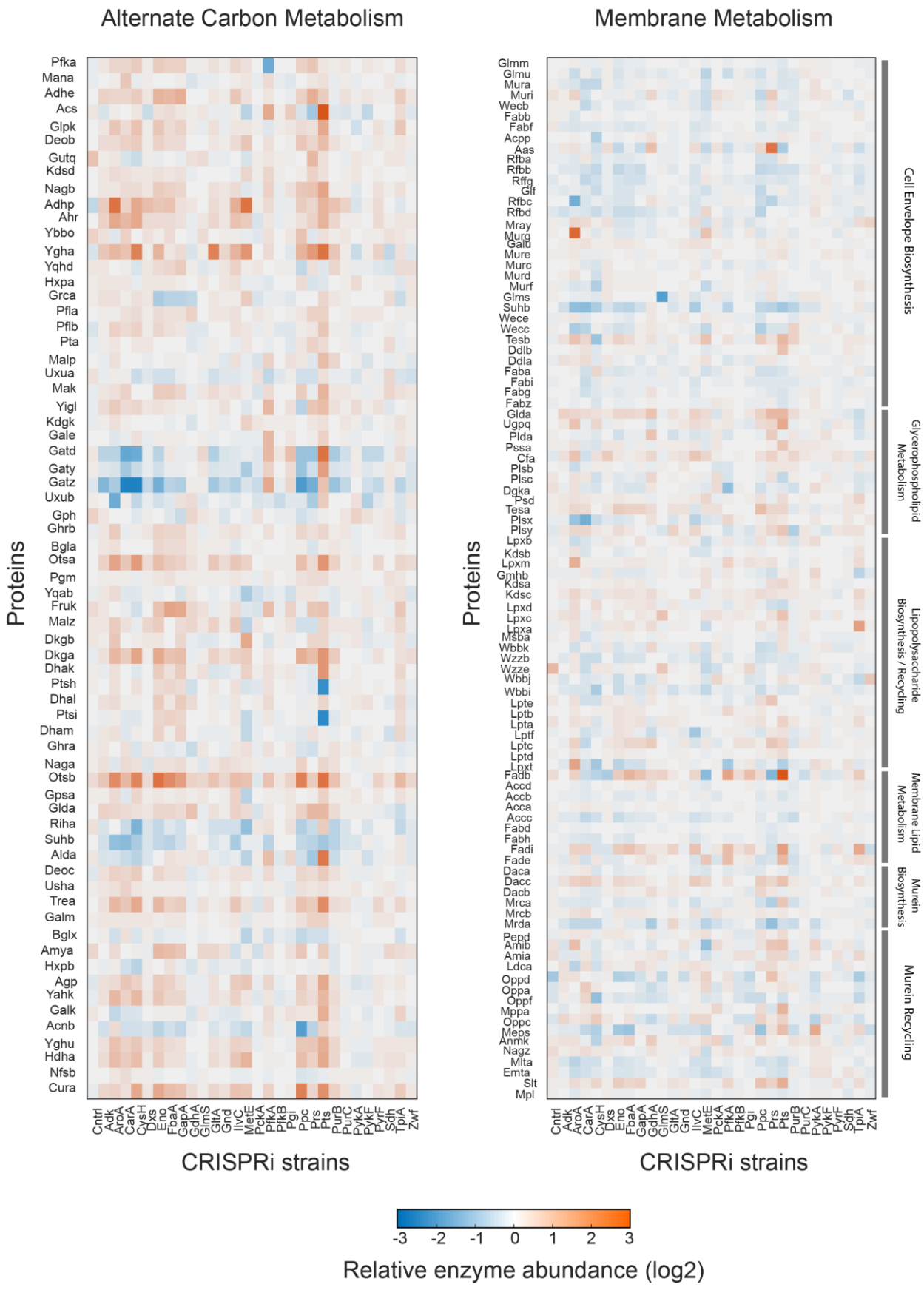


Figure S11. (continued)

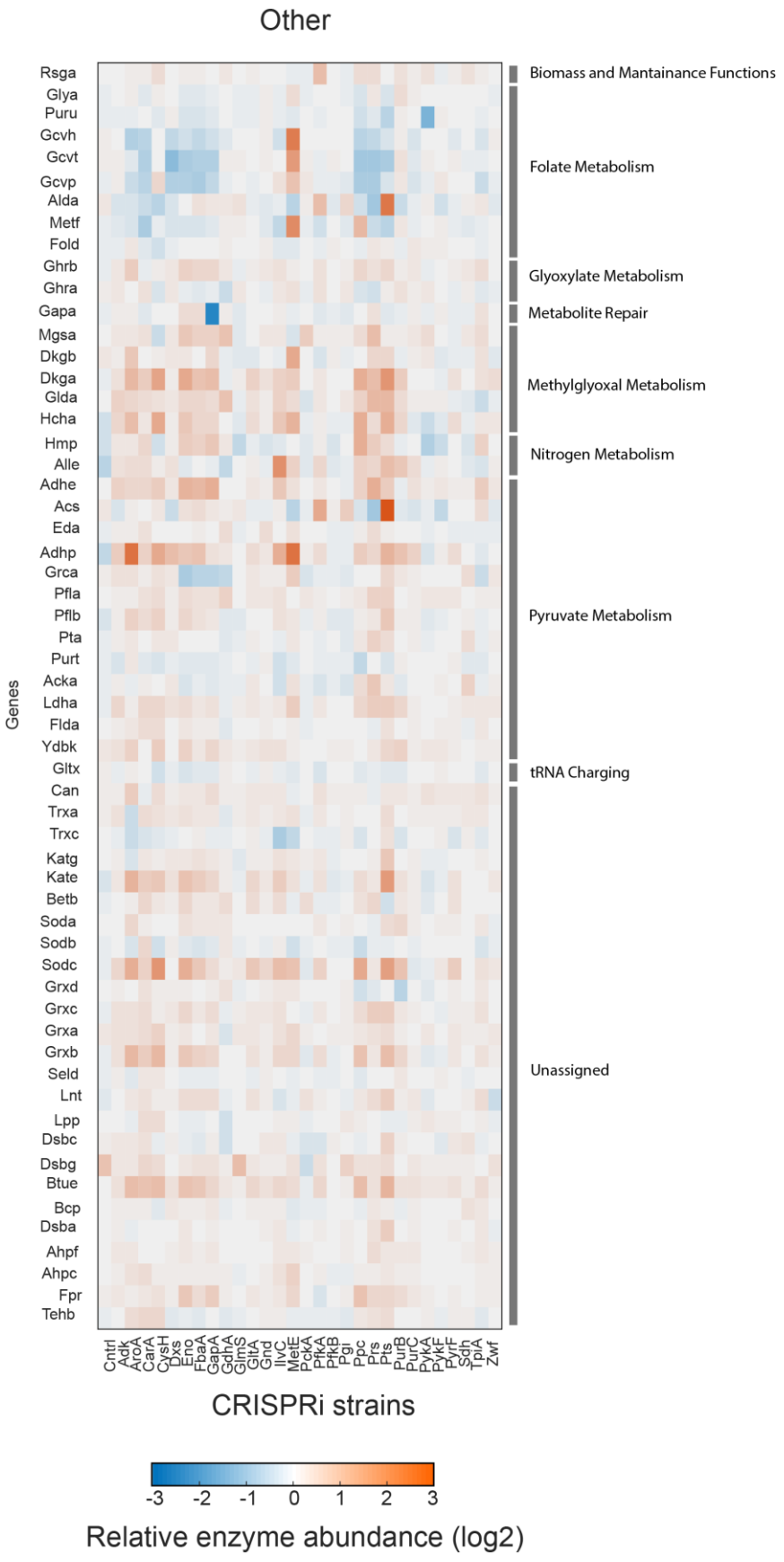


Figure S11. (continued)

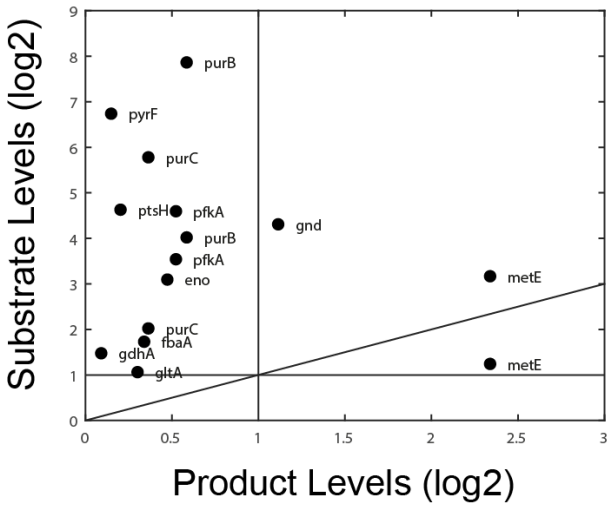


Figure S12. Related to Figure 4.

Fold-changes of substrates and products in CRISPRi strains (in which both metabolites were measured). MetE has two substrates. PurB catalyzes two reactions. Data are represented as mean (n = 2).

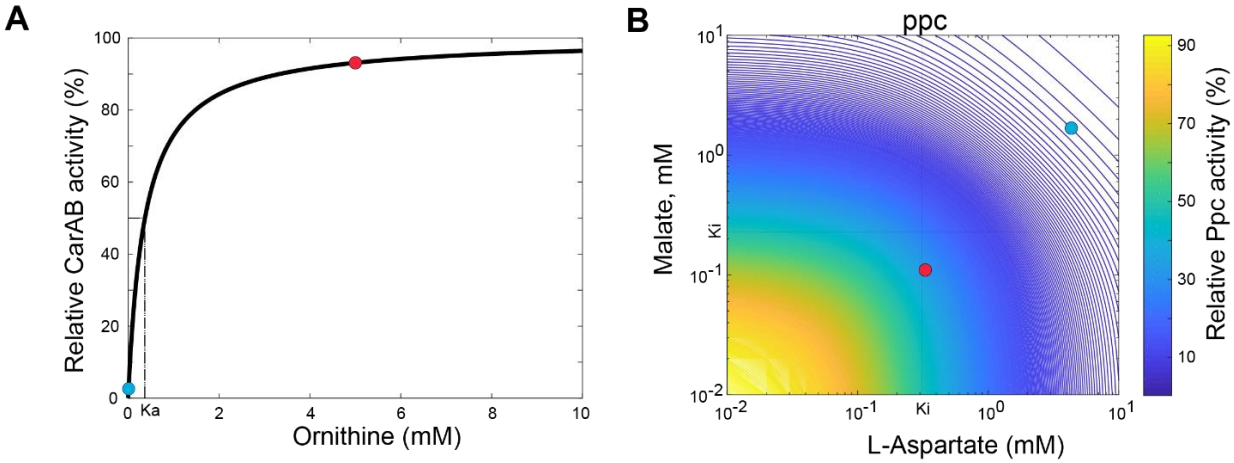


Figure S13. Related to Figure 5.

(A) Allosteric activation of CarAB by ornithine was modelled with a hyperbolic function and an activation constant (Ka) of 0.37 mM (Bueso et al., 1999). The wild-type concentration of ornithine is 0.01 mM (Bennet et al. 2009) and shown as blue dot (3% activity). The concentration of ornithine increased 512-fold in the CarAB knockdown (red dot, 94% activity).

(B) Allosteric inhibition of Ppc by malate and aspartate. Aspartate and malate bind to different sites of Ppc, and therefore inhibition was modelled as the product of two hyperbolic functions with inhibition constants (Ki) of 0.23 mM for malate (Gold and Smith, 1974), and 0.31 mM for aspartate (Gold and Smith, 1974). The wild-type concentration of malate and aspartate is 1.68 mM and 4.34 mM, respectively (Bennet et al. 2009), and shown as blue dot (8% activity). Malate and aspartate decreased 16-fold and 13-fold in the Ppc knockdown (red dot, 33% activity).

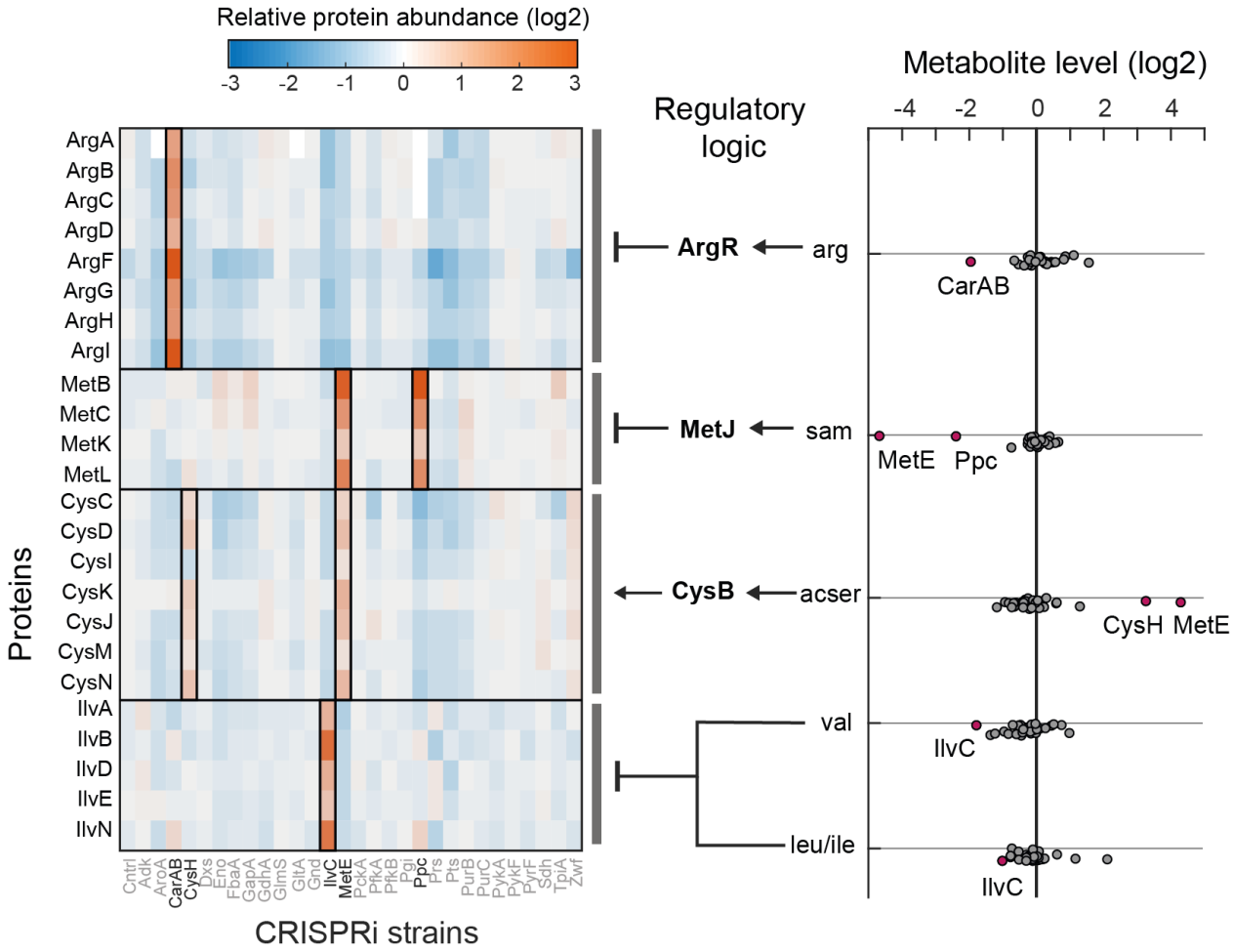


Figure S14. Related to Figure 6.

Fold changes of enzymes in pathways of arginine biosynthesis, methionine biosynthesis, sulfur assimilation (Cys), valine/isoleucine biosynthesis. CRISPRi strains with a compensatory upregulation are shown in bold. On the right panel, concentration of metabolites that are allosteric effectors of transcription factors ArgR, MetJ, CysB. Valine and isoleucine regulate enzyme expression via transcriptional attenuation. Colored dots highlight strains that showed a compensatory upregulation of the respective pathway (boxes in the heatmap).

Regulation is: arginine (arg) allosterically activates ArgR (transcriptional repressor of genes in arginine biosynthesis), S-adenosylmethionine (sam) allosterically activates MetJ (transcriptional repressor of genes in methionine and sam biosynthesis) O-acetyl-L-serine (acser) activates CysB (transcriptional activator of genes in sulfur assimilation), branched chain amino acids (valine and (iso-)leucine) inhibit expression of genes that are involved in biosynthesis of branched chain amino acids (via transcriptional attenuation).

References

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*.

Basan, M., Hui, S., Okano, H., Zhang, Z., Shen, Y., Williamson, J.R., and Hwa, T. (2015). Overflow metabolism in *E. coli* results from efficient proteome allocation. Nature *528*, 99–104.

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

Buescher, J.M., Liebermeister, W., Jules, M., Uhr, M., Muntel, J., Botella, E., Hessling, B., Kleijn, R.J., Le Chat, L., Lecointe, F., et al. (2012). Global network reorganization during dynamic adaptations of *Bacillus subtilis* metabolism. Science *335*, 1099–1103.

Bueso, J., Cervera, J., Fresquet, V., Marina, A., Lusty, C.J., and Rubio, V. (1999). Photoaffinity Labeling with the Activator IMP and Site-Directed Mutagenesis of Histidine 995 of Carbamoyl Phosphate Synthetase from Escherichia coli Demonstrate That the Binding Site for IMP Overlaps with That for the Inhibitor UMP. Biochemistry 38, 3910–3917.

Camsund, D., Lawson, M.J., Larsson, J., Jones, D., Zikrin, S., Fange, D., and Elf, J. (2020). Time-resolved imaging-based CRISPRi screening. Nat. Methods *17*, 86–92.

Chandra, F.A., Buzi, G., and Doyle, J.C. (2011). Glycolytic Oscillations and Limits on Robust Efficiency. Science *333*, 187.

Chao, Y.P., and Liao, J.C. (1993). Alteration of growth yield by overexpression of phosphoenolpyruvate carboxylase and phosphoenolpyruvate carboxykinase in *Escherichia coli*. Appl. Environ. Microbiol. *59*, 4261–4265.

Christodoulou, D., Link, H., Fuhrer, T., Kochanowski, K., Gerosa, L., and Sauer, U. (2018). Reserve Flux Capacity in the Pentose Phosphate Pathway Enables *Escherichia coli*'s Rapid Response to Oxidative Stress. Cell Syst. *6*, 569-578.e7.

Dekel, E., and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. Nature *436*, 588–592.

Deng, M.-D., Grund, A.D., Wassink, S.L., Peng, S.S., Nielsen, K.L., Huckins, B.D., Walsh, B.L., and Burlingame, R.P. (2006). Directed evolution and characterization of *Escherichia coli* glucosamine synthase. Biochimie *88*, 419–429.

Ebrahim, A., Lerman, J.A., Palsson, B.O., and Hyduke, D.R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst. Biol. *7*, 74.

Erickson, D.W., Schink, S.J., Patsalo, V., Williamson, J.R., Gerland, U., and Hwa, T. (2017). A global resource allocation strategy governs growth transition kinetics of *Escherichia coli*. Nature *551*, 119–123.

Fendt, S.-M., Buescher, J.M., Rudroff, F., Picotti, P., Zamboni, N., and Sauer, U. (2010). Tradeoff between enzyme and metabolite efficiency maintains metabolic homeostasis upon perturbations in enzyme capacity. Mol. Syst. Biol. *6*, 356.

Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W., and Milo, R. (2013). Glycolytic strategy as a tradeoff between energy yield and protein cost. Proc. Natl. Acad. Sci. *110*, 10039–10044.

Fuhrer, T., Zampieri, M., Sévin, D.C., Sauer, U., and Zamboni, N. (2017). Genomewide landscape of gene–metabolome associations in *Escherichia coli*. Mol. Syst. Biol. *13*, 907.

Gerosa, L., Kochanowski, K., Heinemann, M., and Sauer, U. (2013). Dissecting specific and global transcriptional regulation of bacterial gene expression. Mol. Syst. Biol. *9*, 658.

Gerosa, L., Haverkorn van Rijsewijk, B.R.B., Christodoulou, D., Kochanowski, K., Schmidt, T.S.B., Noor, E., and Sauer, U. (2015). Pseudo-transition analysis identifies the key regulators of dynamic metabolic adaptations from steady-state data. Cell Syst. *1*, 270–282.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. Nature *418*, 387–391.

Gold, E.W., and Smith, T.E. (1974). *Escherichia coli* phosphoenolpyruvate carboxylase: Effect of allosteric inhibitors on the kinetic parameters and sedimentation behavior. Arch. Biochem. Biophys. 164, 447–455.

Gon, S., Camara, J.E., Klungsøyr, H.K., Crooke, E., Skarstad, K., and Beckwith, J. (2006). A novel regulatory mechanism couples deoxyribonucleotide synthesis and DNA replication in *Escherichia coli*. EMBO J. 25, *488*, 1137–1147.

Grimbs, S., Selbig, J., Bulik, S., Holzhütter, H.-G., and Steuer, R. (2007). The stability and robustness of metabolic states: identifying stabilizing sites in metabolic networks. Mol. Syst. Biol. *3*, 146.

Guder, J.C., Schramm, T., Sander, T., and Link, H. (2017). Time-optimized isotope ratio LC–MS/MS for high-throughput quantification of primary metabolites. Anal. Chem. *89*, 1624–1631.

Hackett, S.R., Zanotelli, V.R.T., Xu, W., Goya, J., Park, J.O., Perlman, D.H., Gibney, P.A., Botstein, D., Storey, J.D., and Rabinowitz, J.D. (2016). Systems-level analysis of mechanisms regulating yeast metabolic flux. Science *354*, aaf2786–aaf2786.

Hawkins, J.S., Silvis, M.R., Koo, B.-M., Peters, J.M., Osadnik, H., Jost, M., Hearne, C.C., Weissman, J.S., Todor, H., and Gross, C.A. (2020). Mismatch-CRISPRi Reveals the Co-varying Expression-Fitness Relationships of Essential Genes in Escherichia coli and Bacillus subtilis. Cell Syst. https://doi.org/10.1016/j.cels.2020.09.009

Hosseini, S.-R., and Wagner, A. (2018). Genomic organization underlying deletional robustness in bacterial metabolic systems. Proc. Natl. Acad. Sci. *115*, 7075–7080.

Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., et al. (2007). Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. Science *316*, 593–597.

Izu, H., Adachi, O., and Yamada, M. (1997). Gene organization and transcriptional regulation of the gntRKU operon involved in gluconate uptake and catabolism of *Escherichia coli*. J. Mol. Biol. *267*, 778–793.

Jensen, K.F. (1993). The Escherichia coli K-12 "wild types" W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels. J. Bacteriol. *175*, 3401–3407.

Jiao, Z., Baba, T., Mori, H., and Shimizu, K. (2003). Analysis of metabolic and physiological responses to *gnd* knockout in *Escherichia coli* by using C-13 tracer experiment and enzyme activity measurement. FEMS Microbiol. Lett. *220*, 295–301.

Jones, D.L., Leroy, P., Unoson, C., Fange, D., Ćurić, V., Lawson, M.J., and Elf, J. (2017). Kinetics of dCas9 target search in *Escherichia coli*. Science *357*, 1420–1424.

Kacser, H., and Burns, J.A. (1973). The control of flux. Symp. Soc. Exp. Biol. *27*, 65–104.

Kacser, H., and Burns, J.A. (1981). The Molecular Basis of Dominance. Genetics *97*, 639–666.

Kemmeren, P., Sameith, K., van de Pasch, L.A.L., Benschop, J.J., Lenstra, T.L., Margaritis, T., O'Duibhir, E., Apweiler, E., van Wageningen, S., Ko, C.W., et al. (2014). Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. Cell *157*, 740–752.

Keren, L., Zackay, O., Lotan-Pompan, M., Barenholz, U., Dekel, E., Sasson, V., Aidelberg, G., Bren, A., Zeevi, D., Weinberger, A., et al. (2013). Promoters maintain their relative activity levels under different growth conditions. Mol. Syst. Biol. *9*, 701.

Keren, L., Hausser, J., Lotan-Pompan, M., Vainberg Slutskin, I., Alisar, H., Kaminski, S., Weinberger, A., Alon, U., Milo, R., and Segal, E. (2016). Massively parallel interrogation of the effects of gene expression levels on fitness. Cell *166*, 1282-1294.e18.

King, Z.A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A., Palsson, B.O., and Lewis, N.E. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res. *44*, D515–D522.

Lawson, M.J., Camsund, D., Larsson, J., Baltekin, Ö., Fange, D., and Elf, J. (2017). *In situ* genotyping of a pooled strain library after characterizing complex phenotypes. Mol. Syst. Biol. *13*, 947.

Levine, E., and Hwa, T. (2007). Stochastic fluctuations in metabolic pathways. Proc. Natl. Acad. Sci. *104*, 9224–9229.

Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell *157*, 624–635.

Löppenberg, M., Müller, H., Pulina, C., Oddo, A., Teese, M., Jose, J., and Holl, R. (2013). Synthesis and biological evaluation of flexible and conformationally constrained LpxC inhibitors. Org. Biomol. Chem. *11*, 6056–6070.

Mazat, J.-P., Reder, C., and Letellier, T. (1996). Why are most flux control coefficients so small? J. Theor. Biol. *182*, 253–258.

McCloskey, D., Xu, S., Sandberg, T.E., Brunk, E., Hefner, Y., Szubin, R., Feist, A.M., and Palsson, B.O. (2018a). Evolution of gene knockout strains of *E. coli* reveal regulatory architectures governed by metabolism. Nat. Commun. *9*, 3796.

McCloskey, D., Xu, S., Sandberg, T.E., Brunk, E., Hefner, Y., Szubin, R., Feist, A.M., and Palsson, B.O. (2018b). Growth adaptation of gnd and sdhCB *Escherichia coli* deletion strains diverges from a similar initial perturbation of the transcriptome. Front. Microbiol. *9*.

Metzger, B.P.H., Yuan, D.C., Gruber, J.D., Duveau, F., and Wittkopp, P.J. (2015). Selection on noise constrains variation in a eukaryotic promoter. Nature *521*, 344–347.

Monk, J.M., Lloyd, C.J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., et al. (2017). iML1515, a knowledgebase that computes *Escherichia coli* traits. Nat. Biotechnol. *35*, 904–908.

Mülleder, M., Calvani, E., Alam, M.T., Wang, R.K., Eckerstorfer, F., Zelezniak, A., and Ralser, M. (2016). Functional metabolomics describes the Yeast biosynthetic regulome. Cell *167*, 553-565.e12.

Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. Nature *441*, 840–846.

Parekh, B.S., and Hatfield, G.W. (1997). Growth rate-related regulation of the ilvGMEDA operon of *Escherichia coli* K-12 is a consequence of the polar frameshift mutation in the ilvG gene of this strain. J. Bacteriol. *179*, 2086–2088.

Park, J.O., Rubin, S.A., Xu, Y.-F., Amador-Noguez, D., Fan, J., Shlomi, T., and Rabinowitz, J.D. (2016). Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. Nat. Chem. Biol. *12*, 482–489.

Peters, J.M., Colavin, A., Shi, H., Czarny, T.L., Larson, M.H., Wong, S., Hawkins, J.S., Lu, C.H.S., Koo, B.-M., Marta, E., et al. (2016). A comprehensive, CRISPR-based functional analysis of essential genes in Bacteria. Cell *165*, 1493–1506.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Reaves, M.L., Young, B.D., Hosios, A.M., Xu, Y.-F., and Rabinowitz, J.D. (2013). Pyrimidine homeostasis is accomplished by directed overflow metabolism. Nature *500*, 237–241.

Rishi, H.S., Toro, E., Liu, H., Wang, X., Qi, L.S., and Arkin, A.P. (2020). Systematic genome-wide querying of coding and non-coding functional elements in *E. coli* using CRISPRi. Biorxiv doi: https://doi.org/10.1101/2020.03.04.975888

Rousset, F., Cui, L., Siouve, E., Becavin, C., Depardieu, F., and Bikard, D. (2018). Genome-wide CRISPR-dCas9 screens in E. coli identify essential genes and phage host factors. PLOS Genet. *14*, e1007749.

Sander, T., Farke, N., Diehl, C., Kuntz, M., Glatter, T., and Link, H. (2019). Allosteric feedback inhibition enables robust amino acid biosynthesis in *E. coli* by enforcing enzyme overabundance. Cell Syst. *8*, 66-75.e8.

Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., and Heinemann, M. (2016). The quantitative and condition-dependent *Escherichia coli* proteome. Nat. Biotechnol. *34*, 104–110.

Scott, M., Gunderson, C.W., Mateescu, E.M., Zhang, Z., and Hwa, T. (2010). Interdependence of cell growth and gene expression: origins and consequences. Science *330*, 1099–1102.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science *329*, 533–538.

Tanner, L.B., Goglia, A.G., Wei, M.H., Sehgal, T., Parsons, L.R., Park, J.O., White, E., Toettcher, J.E., and Rabinowitz, J.D. (2018). Four key steps control glycolytic flux in mammalian cells. Cell Syst. *7*, 49-62.e8.

Towbin, B.D., Korem, Y., Bren, A., Doron, S., Sorek, R., and Alon, U. (2017). Optimality and sub-optimality in a bacterial growth law. Nat. Commun. *8*, 14123.

Tran, L.M., Rizk, M.L., and Liao, J.C. (2008). Ensemble modeling of metabolic networks. Biophys. J. *95*, 5606–5617.

You, C., Okano, H., Hui, S., Zhang, Z., Kim, M., Gunderson, C.W., Wang, Y.-P., Lenz, P., Yan, D., and Hwa, T. (2013). Coordination of bacterial proteome with metabolism by cyclic AMP signalling. Nature *500*, 301–306.

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M.G., and Alon, U. (2006). A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. Nat. Methods *3*, 623–628.

Zelezniak, A., Vowinckel, J., Capuano, F., Messner, C.B., Demichev, V., Polowsky, N., Mülleder, M., Kamrad, S., Klaus, B., Keller, M.A., et al. (2018). Machine learning predicts the yeast metabolome from the quantitative proteome of kinase knockouts. Cell Syst. *7*, 269-283.e6.

Chapter 2

Mapping metabolite – transcription factor interactions at a genome-scale using CRISPRi knockdown library screenings

Michelle Kuntz[1,2], Vanessa Pahl[2], Alejandra Alvarado[2], Stefano Donati[3], Nina Odermatt[1], Timo Glatter[1], Tobias Erb[1], Hannes Link[2,4]*

[1] Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

[2] Present address: Interfaculty Institute for Microbiology and Infection Medicine Tübingen, University of Tübingen, Auf der Morgenstelle 24, 72076 Tübingen, Germany

[3] Present address: The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Kgs. Lyngby,Denmark

[4] Lead Contact

* Correspondence: hannes.link@cmfi.uni-tuebingen.de

This chapter is written in manuscript style and is in preparation for submission (work in progress). My contribution to this work included the design of experiments, performing experiments with the pooled CRISPRi library, execution of the cell-sorting, proteome sampling and measurement, analyzing data and co-writing the manuscript.

Abstract

The interaction of metabolites with transcriptional regulators is one example of how enzyme levels can be adjusted upon varying environmental and intracellular conditions a cell is exposed to. Here, we used CRISPR interference (CRISPRi) to induce changes of metabolite and protein levels *in vivo* which results in an activation of metabolic genetic circuits. We combined a library of 7177 CRISPRi knockdown strains with a fluorescent transcriptional reporter plasmid to measure the activity of a transcription factor (TF) involved in amino acid biosynthesis (ArgR) upon the induction of the knockdown. This enabled us to sort strains from the library via cell-sorting, based on their fluorescence, which were only genes located in the arginine biosynthesis pathway. Proteome as well as metabolome measurements confirmed an activation of the feedback circuit in the sorted strains. Furthermore, we could demonstrate that the knockdown of *fruK* is the only strain of our library that leads to an inactivation of Cra in the conditions we tested, which is the first time this could be shown

in vivo. In conclusion, this work demonstrates how CRISPRi can be used to identify genetic metabolic circuits *in vivo*.

Introduction

Allosteric binding of metabolites can have a strong influence on activity and function of proteins. Examples are metabolic enzymes (Gerosa and Sauer, 2011) , protein kinases (Li et al., 2010) or transcription factors (Motlagh et al., 2014), which are modulated by the cell's own metabolites. The current challenge is to distinguish between functional and unspecific metabolite-protein interactions (Kochanowski et al., 2015) and to detect them at a very large-scale. The gold standard are still in vitro binding assays, like isothermal titration calorimetry (ITC) or nuclear magnetic resonance spectroscopy (NMR) (He et al., 2015; Nikolaev et al., 2016), which are low-throughput. More systematic approaches with metabolomics-based LC-MS enabled to map the interactions between metabolites and proteins at larger scales, by immobilizing proteins and incubation with a mixture of metabolites. These were extracted from the protein-metabolite complexes and measured by untargeted metabolomics (Tagore et al., 2008; Vinayavekhin and Saghatelian, 2011). The combination of this untargeted LC-MS and native MS was applied to find potential interacting metabolites and later on to confirm the new interaction (Qin et al., 2019). Furthermore, a first systematic mapping of protein-metabolite interactions on a proteome-wide scale was enabled by a chemoproteomic approach, which combined limited proteolysis (LiP) with MS in the presence of unmodified metabolites (Piazza et al., 2018). However, no in vivo method exists that detects functional interactions in a high-throughput manner and their direct effects on cellular processes, for instance transcription.

Here, we focused on interactions between transcription factors and metabolites in *E. coli* and developed an *in vivo* based method (*MapMe*) to identify functional interactions between metabolites and transcriptional regulators (TF) at a genome scale. In earlier studies, we investigated how *E. coli* responds to decreases of single metabolic enzymes by using CRISPR interference (CRISPRi) and multi-omics analysis. We found that CRISPRi knockdowns lead to very specific and local changes of metabolite levels in the cell, which then activate a transcriptional response and induce metabolic genetic circuits (Donati et al., 2021). Because metabolite changes and the accompanying proteome changes in the CRISPRi library were remarkably specific, we reasoned that these responses could potentially inform about metabolism-transcription interactions at a large-scale. Thus, we combined a library of 7177 CRISPRi knockdowns with a fluorescent reporter plasmid to measure the activity of the TF ArgR, which is involved in amino acid metabolism. This enabled us to recover the known metabolic-genetic circuit of ArgR. Furthermore, we tested a reporter to measure the activity of the TF Cra against

all knockdowns in the CRISPRi library and found *fruK* to be the only responding target. Moreover, metabolite measurements in the *fruK* knockdown revealed that hexose-phosphate levels are not increasing upon the induction of CRISPRi, which gives an indication of a direct interaction between Cra and *fruK*.

Results

The MapMe approach

CRISPRi is making use of a catalytically inactivated Cas9 protein, which is guided by a small guide RNA to a target gene (**Figure 1A**). The transcription of this gene is then blocked by the complex, why enzyme levels will decrease and functions of this gene can be further explored (Bikard et al., 2013; Larson et al., 2013; Qi et al., 2013). For *MapMe* we made use of a pooled CRISPRi library, harboring 4-6 guides for all metabolic genes of the latest iML1515 model in *E. coli* (Beuter et al., 2018; Donati et al., 2021)(**Figure 1B**). This library is transformed with a fluorescent transcriptional reporter to measure transcription factor activity. Upon the induction of CRISPRi, metabolite levels change and lead to a deactivation of the TF and therefore a release of the TF from the promoter which should allow us to sort the cells with the highest green fluorescence via FACS after the induction of dCas9 (**Figure 1C**). Afterwards, the cells are arrayed on 96-well plates and GFP and OD are measured over time to confirm the higher green fluorescence in the induced CRISPRi strains. Those strains are then sequenced to find out which CRISPRi target they have. Also, proteome and metabolome measurements are performed with strains occurring the most, i) to exclude off-target effects and ii) to find the potential interacting metabolite of the transcription factor.

CRISPRi knockdown of CarAB leads to an inactivation of ArgR

As described in Donati and Kuntz et al. (2021), we used CRISPR interference (CRISPRi) as a tool to knockdown single enzymes in *E. coli* metabolism. We screened a library of 7177 strains for fitness defects and could see that metabolism is remarkably robust and buffers fitness defects for hours after induction of CRISPRi. Furthermore, we measured the metabolome and proteome of 30 CRISPRi strains, for example CarAB. In this strain CarAB is targeted by CRISPRi, which is an enzyme involved in pyrimidine and arginine biosynthesis (**Figure 2A**). It catalyzes the reaction of L-glutamine to Carbamoyl-phosphate, which is later used for the synthesis of arginine. If arginine accumulates in the cell, it binds to the transcriptional regulator ArgR, which then represses the arginine biosynthesis genes (**Figure 2B**).

We introduced a fluorescent reporter plasmid that expressed GFP from an ArgR regulated promoter (puA66-argE-gfp) (Zaslaver et al., 2006) into the CarAB CRISPRi strain. We cultivated the strain with and without induction of dCas9 in microtiter plates in a plate reader and measured OD and GFP levels. Indeed, GFP expression increased after 2 hours of CRISPRi induction, showing that ArgR responded to the knockdown (**Figure 2C**). Therefore, we concluded that a decrease of arginine levels in the cell led to a release of ArgR from the promoter which led to a higher transcription of *gfp* and therefore to an increased GFP fluorescence in the induced strain. Hence, the induction leads to an activation of the feedback circuit and we next sought to use this approach to search for the effects of changing metabolite levels on transcription factor activity on a large-scale.

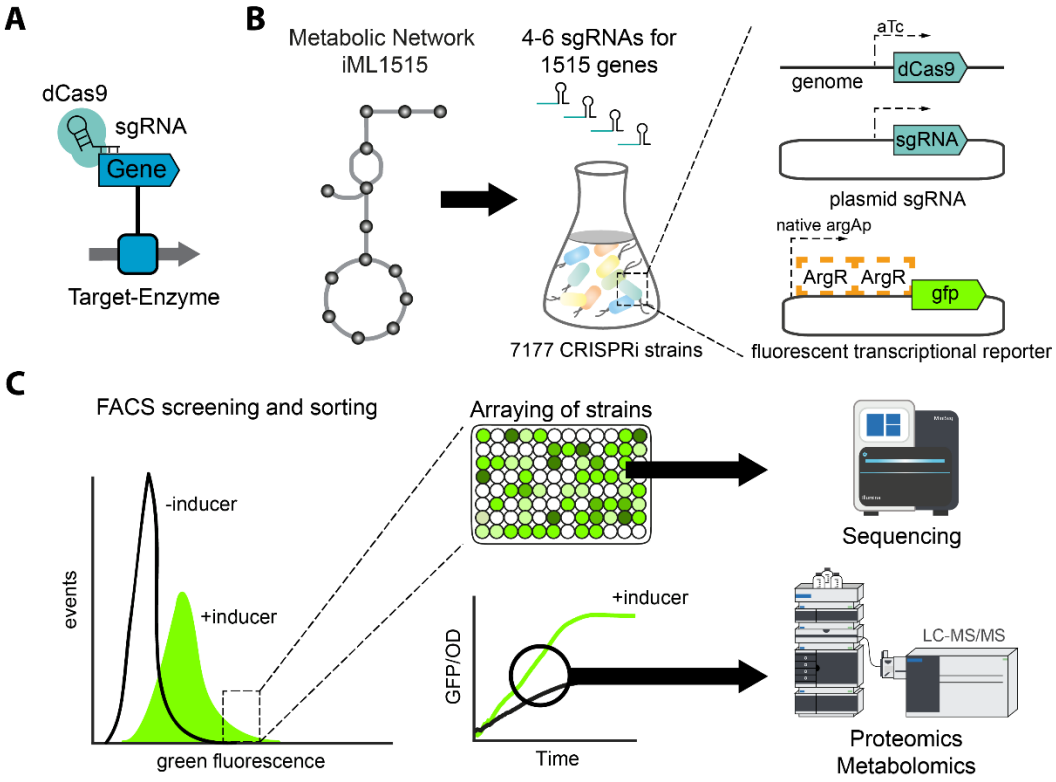


Figure 1. MapMe approach: a tool to find new metabolite-TF interactions. (A) Scheme of dCas9 and sgRNA forming a complex to repress the transcription of a target gene and therefore the abundance of the respective target enzyme. (B) Schematic of a metabolism-wide CRISPRi library which is generated in *E. coli* YYdCas9 having dCas9 encoded on the genome. The strain is first transformed with the fluorescent transcriptional reporter plasmid for TF activity (ArgR), harboring two binding sites for the transcriptional repressor ArgR and afterwards with the CRISPRi plasmid library. (C) The strain library is cultured and the cells with the highest green fluorescence are sorted by FACS (shown as dashed square). Those strains are further confirmed in plate reader experiments by comparing the growth and green fluorescence of an induced and not induced culture over time. The strains with high green fluorescence are sequenced and proteomics and metabolites are measured via LC-MS/MS.

FACS screening of CRISPRi knockdown library filters out targets involved in arginine biosynthesis

To screen for metabolite- transcription factor interactions at a large-scale, the strain library was cultured as a pool in M9 glucose with induction of dCas9. After 5 to 6 hours, cells showing the highest GFP fluorescence were sorted by FACS (**Figure 1B**). First, we inspected the GFP signal in a control strain without sgRNA (*E. coli* YYdCas9 WT). The strain showed a high variety of green fluorescence in the cells (**Figure S1**). This phenomenon was already observed by Silander et al., 2012, who were screening for promoter-mediated phenotypic noise and it leads to the conclusion that the noise of the reporter plasmids probably derived from a natural cell to cell variation. As a reference for cells that should activate ArgR, we measured GFP in a strain with a sgRNA targeting *argE*. Based on the GFP fluorescence the induced population was clearly distinguishable from the uninduced population (**Figure 2D**). Accordingly, it should be possible to filter out strains with a higher fluorescent level due to the induction of dCas9 by FACS. We subsequently measured the CRISPRi library induced and uninduced (**Figure 2E**). The induced population slightly shifted to a higher GFP fluorescence compared to the uninduced library. For cell sorting, the gate was set to sort for the top 0.25% green fluorescent cells of the overall culture. 1000 sorted cells were arrayed in 96-well plates and OD and GFP signals were measured in a plate reader. 171 out of 672 strains (25%) showed a clear increase in GFP levels in the induced compared to the uninduced state. Those strains were pooled on 96-well plates and sent for Sanger sequencing. 78% of these strains had a CRISPRi target in arginine biosynthesis, and the three top hits were ArgA (32%), CarA (12%) and CarB (9%) (**Figure 2F**). 8% of the sequenced targets were guides for the genes *glcA* and *eutD*. Those genes have no direct link to arginine levels in the cell, which led us to the question why they appeared so often in our screen. As only one out of all guides for these genes appeared, we were searching for off-target effects of those guides with the Cas-OFFinder (Bae et al., 2014). Indeed, the guide for *eutD* had an off-target for *argC*, which could explain a decrease of arginine levels in the cells and therefore a response of the reporter resulting in a higher GFP signal. Nevertheless, we could not explain *glcA* by using this approach. Hence, we decided to measure the proteome of *glcA*, *argA* as a positive control and a control strain, having a guide without a target in the *E. coli* genome. As shown in **Figure 2G**, the knockdown of *argA* led to an almost 4-fold decrease of the target gene, while dCas9 2-fold upregulated compared to the uninduced strain. Also, the knockdown induced an increase of protein levels of all ArgR targets directly linked to arginine biosynthesis, while those targets did not show any response in the control strain. We could see the same effect of the arginine biosynthesis genes for the *glcA* strain, with the only difference that *carB* was downregulated by 2-fold. Therefore, we were searching again for an off-target effect of the guide for *glcA* and found that the 11 bp right next to the PAM site matched perfectly with a sequence in the *carB* gene. This short sequence is known to be enough to provoke a downregulation of an off-target

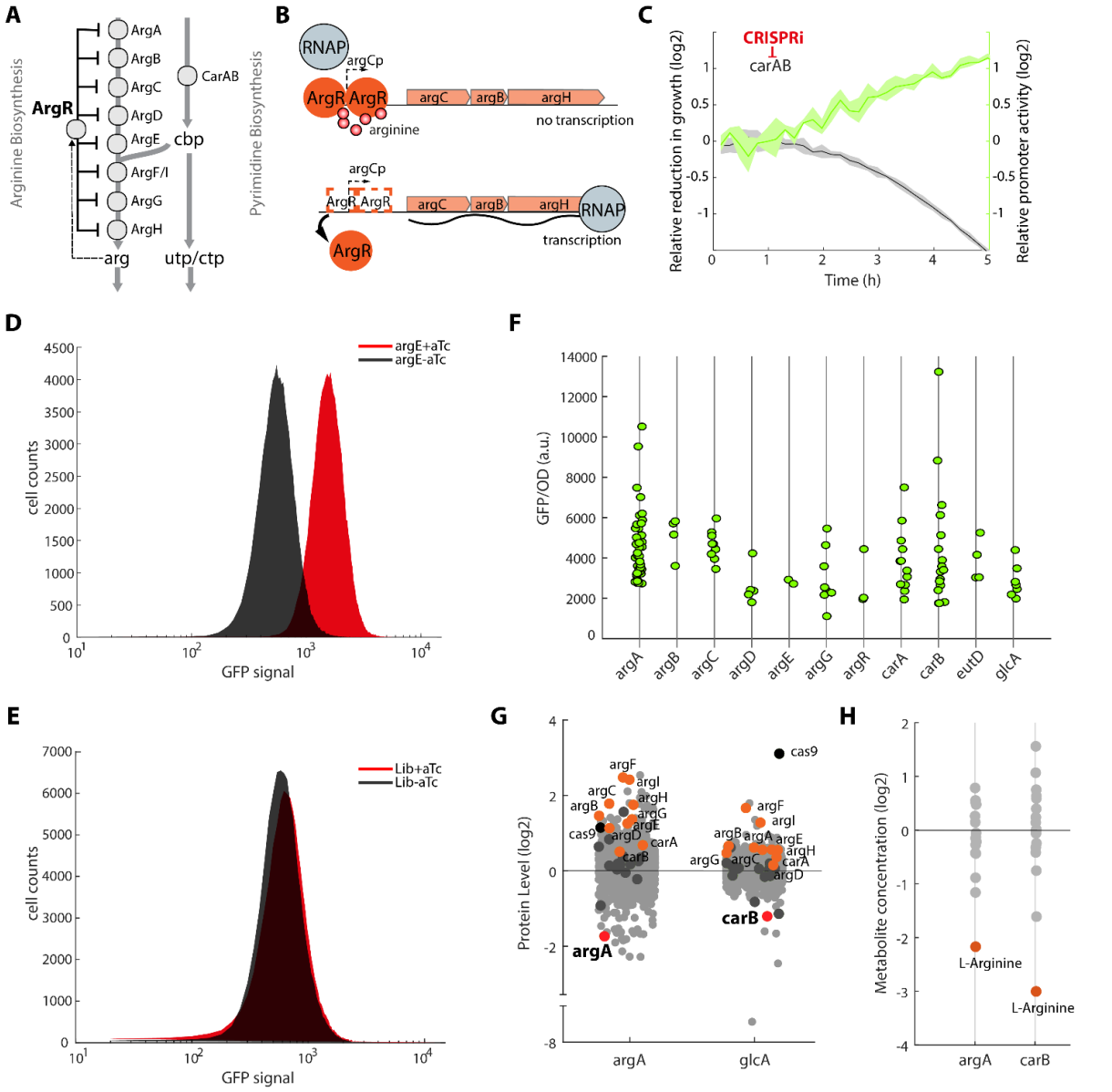


Figure 2. Screening of the CRISPRi library with a reporter for transcriptional regulator ArgR. (A) Schematic of the arginine-pyrimidine branch point. Dotted arrows indicate allosteric activation and straight line shows inhibition by the repressor ArgR. (B) Scheme of the regulatory logic of the argCBH operon, as an example for the function of ArgR as a repressor. (C) The CarAB knockdown was transformed with a fluorescent reporter plasmid that expressed GFP from an ArgR regulated promoter (puA66-*argA-gfp*). The fold-change of GFP/OD between induced and uninduced cultures is shown in green. The fold change of OD is shown in black. Both is shown as a log2-fold change. (D) The histogram shows the green fluorescent signals of 100000 cells after growth for 5.5 hours. Shown is the *argE* knockdown strain harboring a fluorescent transcriptional reporter for ArgR activity with (black) and without the induction of CRISPRi (red). (E) Shown is the CRISPRi library with 7177 guides and the fluorescent transcriptional reporter plasmid for ArgR activity not induced (black) and induced (red) (F) GFP/OD values for arrayed and sequenced strains, measured in a plate reader. GFP and OD were measured over time and shown are GFP/OD values at OD 0.5 for every strain. Green points indicate strains that were found more than once in the screening. (G) Proteome data of strains harboring a guide to target *argA* and *glcA*. Protein Level are shown in log2 foldchange compared to the uninduced strains. Dark grey dots indicate proteins related to the argR operon and orange dots are argR regulated genes belonging to the arginine biosynthesis, red dots are proteins of target genes and black dots show levels of cas9 protein. (H) Metabolite concentrations of amino acids in the *argA* and *carB* knockdown strains shown as a log2 foldchange compared to uninduced strains. Grey dots show amino acids, and orange dots show the arginine levels.

in *E.coli* (Cui et al., 2018; Rousset et al., 2018) and is not covered by the Cas-OFFinder. To support our hypothesis of an off-target effect, we designed three new guides targeting the glcA gene and measured OD and GFP levels over time. We could observe that indeed only the guide which we found in the library screening led to increased GFP levels (**Figure S2A**). Furthermore, we measure the metabolite levels of all amino acids in the argA and carB knockdown strain and could see that arginine was by far the most strongly downregulated metabolite (**Figure 2H**). Hence, with our approach we were able to find all targets that are directly linked to arginine biosynthesis due to an activation of the metabolic genetic feedback circuit. Next, we aimed to investigate a global transcription factor whose interacting metabolite could not be proven *in vivo* yet.

Cra reporter shows a strong response to knockdown of *fruK*

As shown, the method described here can be used to find known regulatory metabolites of transcription factors, we decided to screen the CRISPRi library with a reporter plasmid for the transcription factor Cra. Cra is a global regulator for carbon metabolism, and it modulates the direction of carbon flow through different metabolic pathways and is therefore involved in the expression of a large number of operons. So far it was thought to interact with fructose-1,6-bisphosphate (FBP) to regulate the glycolytic flux. Nevertheless, biochemical analysis showed that FBP neither binds nor regulates Cra activity, but fructose-1-phosphate does (Bley Folly et al., 2018). Yet, none of these interactions could be found or confirmed in *in vivo* experiments. Therefore, we transformed the YYdCas9 strain with a reporter plasmid expressing GFP from a CRA regulated promoter (puA66-*fruB-gfp,* (Zaslaver et al., 2006). It was previously shown that *fruB* is one of the strongest binding sites for Cra (Shimada et al., 2005), thus we chose this promoter to test for the binding activity of CRA in a native context. The basis strain with the reporter plasmid was then transformed with the before described CRISPRi plasmid library. The resulting strain library was again cultured as a pool in M9 glucose with induction of dCas9. After 5 to 6 hours, cells showing the highest GFP fluorescence were sorted by FACS (**Figure S3**). After comparing the GFP levels in the induced and not induced library, the gate was set to sort for the top 0.2% green fluorescent cells. The sorted cells were grown on plates overnight and arrayed on 96-well-plates. After that OD and GFP were measured in the induced and uninduced strains. 86 out of 576 strains (15%) showed a higher GFP signal in the induced state. Those strains were pooled on one plate and sent for sequencing. The most found targets were *fruB* (37%), *ulaA* (21%) and *fruK* (12%) (**Figure 3A**). *fruB* encodes for the fructose-specific PTS multiphosphoryl transfer protein FruB and phosphorylates fructose resulting in fructose-1-phosphate while transporting it into the cell, and it is located in an operon upstream of *fruK.* Consequently, a

knockdown will lead to a downregulation of the whole operon. The promoter of the operon has 2 binding sites for Cra, which blocks the transcription without the presence of the regulating metabolite. *fruK* itself is encoding for the 1-phosphofructokinase, which catalyzes the reaction of fructose-1-phosphate to fructose-1,6-bisphosphate. Hence, the knockdown of *fruK* and *fruB* both result in decreasing FruK levels and could lead to increasing fructose-1-phosphate levels in the cells and therefore an inhibition of Cra. *ulaA* as the third most appearing target in the screen, has no direct link to glycolysis metabolites and a search with Cas-OFFinder did not show any results for possible related off-targets.

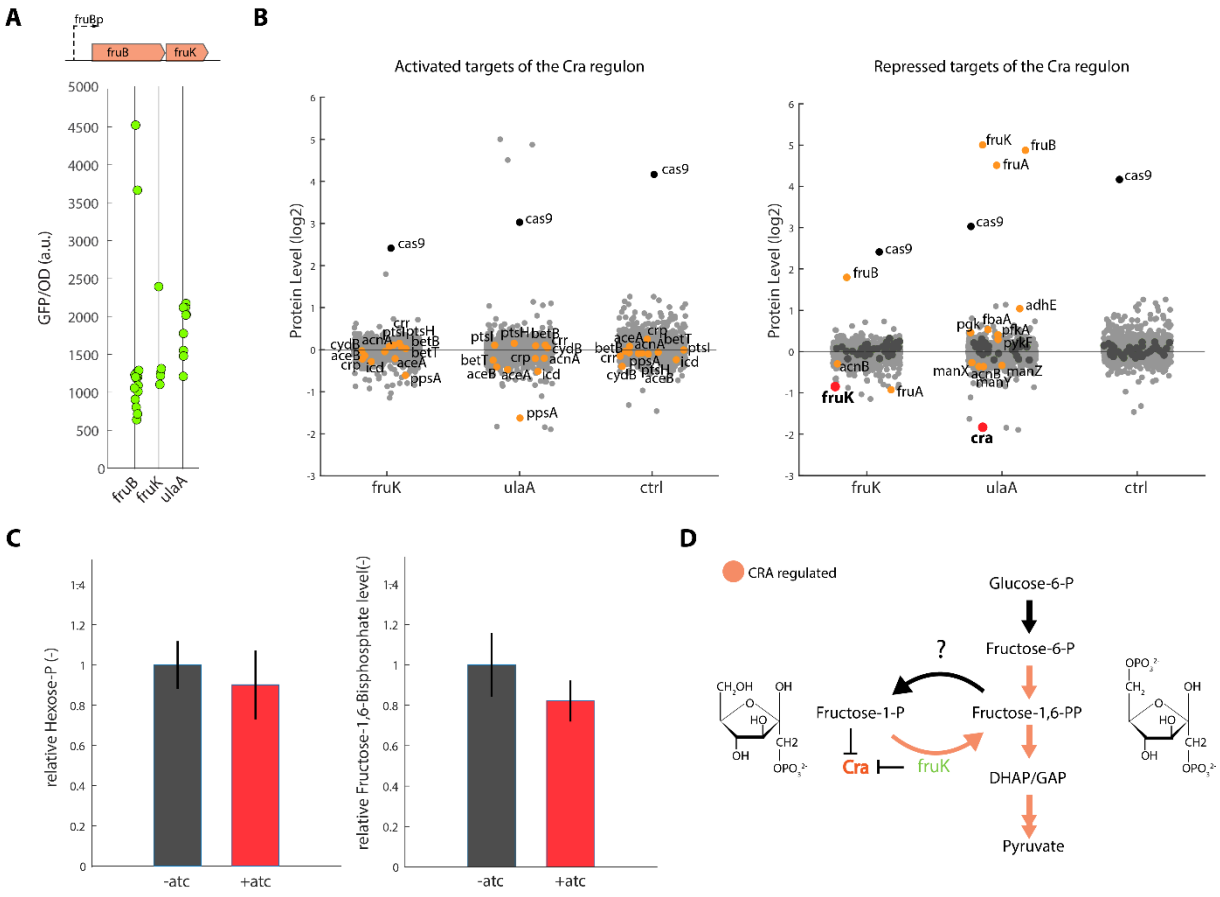


Figure 3. Screening of the CRISPRi library with a reporter for the transcriptional regulator Cra. (A)FruB operon and GFP/OD values for arrayed and sequenced strains, measured in a plate reader. GFP and OD were measured over time and shown are GFP/OD values at OD 0.5 for every strain. Green points indicate strains with guides that were found more than once in the screening. (B) Proteome data of strains harboring a guide to target fruK, ulaA and a control guide, having no target in *E. coli* genome. Protein Level are shown in log2 foldchange compared to the uninduced strains and were divided into genes belonging to the activated (left) and repressed targets (right) of Cra. Orange and dark grey dots indicate proteins whose genes are part of the Cra operon, red dots are proteins of target genes and black dots show levels of cas9 protein. (C) Metabolite data of CRISPRi strain fruK. Shown are relative Hexose-Phosphate and Fructose-1,6-bisphosphate levels. (D) Schematic of the regulation of glycolysis by the transcriptional regulator Cra. Orange arrows indicate a regulation by Cra and shown in green is fruK, which catalyzes the reaction of fructose-1-P to fructose-1,6-PP. Chemical structures of these two metabolites are shown next to their names.

To show the effect of the *fruK* knockdown on the CRA regulon and to find possible off-target effects of *ulaA*, we measured the proteome of those strains and a control strain, having no target in *E. coli* genome (**Figure 3B**). The proteome of *fruK* showed a downregulation of *fruK* and *fruA*, which is encoded after *fruK* in the operon, by almost 2-fold. On the contrary *fruB* is upregulated by 4-fold, which is located before *fruK* in the operon. Also, *ppsA*, which is a positively regulated gene of Cra, was decreased 1.2-fold. This strongly implies a response of Cra to the knockdown of *fruK*. *ulaA* caused a very strong upregulation of the *fruBKA* operon around 10-fold, while *ppsA* was downregulated 3.2-fold and *cra* 3.6-fold. This indicates that the guide for *ulaA* has an off-target in *cra* and therefore leads to a strong response of the reporter plasmid. Thus, we checked again the sequence of the guide and found that 10 bp next to the PAM site perfectly matched with the *cra* gene, and the off-target was also confirmed by measuring OD and GFP levels of strains with other guides for *ulaA* (**Figure S2B**). Accordingly, our approach enabled to find CRISPRi targets directly influencing Cra activity and furthermore giving a strong indication for fructose-1-phosphate being the main signal metabolite for CRA. We wanted to show next via metabolomics the accumulation of fructose-1-phosphate in the *fruK* knockdown.

fruK regulates CRA

Thus, we measured metabolites in the *fruK* strain and could not see an accumulation of hexose-phosphates (**Figure 3C**). As the separation of hexose-phosphates via LC-MS is extremely difficult, we first had to set-up a method with a standard mixture of fructose-1-phosphate (F1P), fructose-6-phosphate (F6P) and glucose-6-phosphate (G6P). Once we were able to detect a clear separation (**Figure S4A**), we measured these metabolites in the *fruK* knockdown strain. While we could see no change upon the induction of CRISPRi for fructose-1,6-bisphosphate levels (FBP) (**Figure 3C**), being the product of the reaction of *fruK* **(Figure 3D)**, we could not detect an accumulation of F1P in the knockdown. The F1P concentration in the knockdown was below the detection limit, and a knockdown of *fruK* did not lead to an accumulation high enough to measure F1P (**Figure S4B**). It was shown for Pseudomonas putida that F1P binds cra with a KD ~200 nm (Chavarría et al., 2011), which could be below the detection limit of our method. Furthermore, a more recent study described the direct interaction of FruK with Cra *in vivo* and that FruK also catalyzes the reverse reaction from FBP to F1P which, based on their study, regulates the affinity of the Cra/FruK complex to the Cra operators *in vitro* (Singh et al., 2017). In conclusion, we could show that *fruK* is involved in the regulation of Cra *in vivo* but it remains unclear whether this happens through a regulatory metabolite, the direct binding of *fruK* to Cra or both.

Discussion

Finding interactions between metabolites and proteins is mostly based on *in vitro* approaches (Diether and Sauer, 2017). These studies focus on the binding of metabolites to purified proteins and vice versa, and large-scale mapping approaches revealed 1000s of new interactions (Gallego et al., 2010; Lempp et al., 2019; Li et al., 2010). But the biggest challenge remains to identify and separate interactions that are functional *in vivo*, and therefore play a role in cellular metabolism from unspecific binding events, and to do this at a very large-scale.

In this study, we used CRISPRi to induce metabolic genetic circuits which enabled us to change and measure the activity of transcription factors *in vivo* with the help of fluorescent transcriptional reporters. Using a CRISPRi library with 7177 targets in *E. coli* metabolism allowed us to filter out knockdowns via cell-sorting that resulted in a response of the TF. Hence, this is an approach which only detects functional responses of TFs initiated by metabolite or protein level changes in the knockdown. Using a reporter plasmid for the TF ArgR revealed a change in its activity for the knockdowns of 9 out of 11 genes that are directly linked to the arginine biosynthesis pathway. Measuring metabolite levels showed that arginine is the regulatory metabolite of ArgR which was already shown by *in vitro* studies many years ago (Maas, 1994; Van Duyne et al., 1996). The two genes we could not detect in the library were *argH* and *argF/I*. As not every knockdown leads to the same growth defect of the cells, this can change the abundance and distribution of guides and targets in the library at different timepoints (Donati et al., 2021). Hence, the timepoint of sampling can have a big influence on which strains will be sorted from the library. Moreover, the testing conditions will have an influence on the response of the transcription factor. As many genes are only expressed under certain conditions (Nicolas et al., 2012; Schmidt et al., 2016), the knockdown of these genes will not lead to any metabolite or protein changes if the assay is carried out in a condition when the gene is not expressed. Thus, varying testing conditions and sampling timepoints might help to recover different target genes whose knockdown has an influence of the transcription factor activity. Furthermore, the manner with which the gate is set for the sorting process will have a big influence. Since in this approach we were only working with one fluorescent marker and a comparatively weaker, native promoter fused to *gfp*, the discrimination between strains with a higher green fluorescence and the rest of the population was extremely difficult. This could be improved by using a stronger promoter or a second fluorescent marker which is constitutively expressed, so the fluorescent signal of cells with a response of the TF could be easier discriminated from the background (He et al., 2019; Miao et al., 2009). Therewith, the sorting of a lot of false-positives could also be avoided. But it must

be noted that the usage of an artificial promoter will also result in a loss of the natural context of the approach.

The cryptic targets that we found after the cell sorting could be explained by off-target effects of the sgRNAs for genes that are also related to the arginine biosynthesis pathway. They were the result of a perfect match of 9 nt next to a PAM sequence in the off-target site (Cui et al., 2018). Unfortunately, this cannot be easily avoided when designing sgRNAs for CRISPRi libraries which makes it necessary to further investigate sorted targets by down-stream analysis. However, the usage of multiple sgRNAs per gene can give an idea whether the target results in a response of the TF or whether this is based on an off-target effect.

In conclusion, our study shows that the method developed here enables us to filter out strains from a CRISPRi library in which the knockdown of a target gene leads to a change of the activity of a transcriptional regulator. Further analysis via metabolomics could show that decreasing arginine levels were the cause of the response of the repressor ArgR in these strains. Hence, combining CRISPRi with fluorescent transcriptional reporters to measure protein activity and metabolomics could allow us to detect new functional interactions *in vivo* at a large-scale.

Acknowledgements

We thank A Diepold for discussions. We thank J Elf for providing the YYdCas9 strain. This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 715650, ERC Starting Grant MapMe). MK and SD acknowledge funding from the IMPRS graduate school for environmental, cellular, and molecular microbiology from the Max Planck Society.

Author contributions

MK performed experiments, analyzed data and co-wrote the manuscript. VP performed metabolite measurements. AA performed growth experiments and GFP measurement. SD analyzed data. NO and TE set-up FACS measurements. TG measured proteomes. HL conceived the study, analyzed data, and co-wrote the manuscript.

Conflict of Interests

The authors declare no competing interests.

Material and Methods

Strains and Culture

The wild-type strain used in this study was *E. coli* YYdCas9 (Lawson et al., 2017) and all strains used for this study derive from it. Strains are listed in the Strain table. For cloning NEB 5-alpha Competent *E. coli* (Cat#C2987) cells were used.

Construction of the CRISPRi pooled library

The construction of the CRISPRi library was done as described in Donati and Kuntz et al. (2021). To measure transcription factor activity the plasmid puA66 was used (Zaslaver et al., 2006).

Media

Cultivations were performed with LB medium or M9 minimal medium with glucose as sole carbon source (5 g L^{-1}). M9 medium was composed by (per liter): 7.52 g Na_2HPO_4 2 H_2O, 5 g KH_2PO_4, 1.5 g $(NH_4)2_SO_4$, 0.5 g NaCl. The following components were sterilized separately and then added (per liter of final medium): 1 mL 0.1 M $CaCl_2$, 1 mL 1 M $MgSO_4$, 0.6 mL 0.1 M $FeCl_3$, 2 mL 1.4 mM thiamine-HCl and 10 mL trace salts solution. The trace salts solution contained (per liter): 180 mg $ZnSO_4$ 7 H2O, 120 mg $CuCl_2$ 2 H_2O, 120 mg $MnSO_4$ H2O, 180 mg $CoCl_2$ 6 H_2O. For strains transformed with pgRNA-bacteria plasmids, 100 µg mL^{-1} ampicillin (Amp) was added to the media. To induce expression of the dCas9 protein in the YYdCas9 strain, aTc was added to a final concentration of 200 nM. In experiments with pUA66 plasmids 50 µg mL^{-1} kanamycin was added to the medium.

Cultivation conditions of the pooled CRISPRi library for FACS sorting

A preculture of 50 mL LB+Amp+Kan was inoculated with 500 µL of the pooled CRISPRi strain library from a glycerol stock and incubated at 37 °C for 5 hours. From the LB culture a second preculture in M9+Glc was inoculated with a dilution of 1:10000 and incubated for 13 hours. After 13 hours the M9 preculture was in exponential phase and it was used to inoculate two cultures with an initial OD of 0.02 in shaking flasks, either containing 100 mL of M9 with 200 nM of aTc to induce expression of dCas9 or without aTc. After 5-6 hours 10 ml of cells were sampled at RT and washed twice in 1x PBS. The cells were diluted in appropriate concentrations and sorted within one hour after sampling.

FACS measurements and sorting

Fluorescence activated cell sorting (FACS) was carried out on a Sony SH800SHP with all four lasers activated (405, 488, 562, 638nm), using a 100 µm nozzle sorting chip and the standard filter set. The threshold was set to 0.5% BSC with the sensor gain for FSC and BSC at 4 and 32%, respectively. Sensor gains for fluorescent channels 2 (GFP, 525/50nm) was set to 40%.

Events per second (eps) were kept below 1000, and “purity” mode was selected for sorting. The top 0.25%/ 0,2 %GFP-fluorescent cells were sorted into a 1.5 ml Eppendorf tube with 1 ml fresh M9 medium without Glucose or antibiotics. Cells were spread on big LB+Amp+Kan plates and grown overnight at 37°C. 1000 colonies were picked into deep-well plates with LB+Amp+Kan and grown for 5-6 h. Afterwards, a glycerol stock was generated.

Cultivation conditions for proteome sampling

Strains were streaked out on LB+Amp+Kan plates. Single colonies were transferred into liquid 5 mL LB+Amp+Kan, and then re-inoculated in M9+Glc medium overnight in 13 mL culture tubes under shaking at 37 °C. For proteomics sampling, M9 pre-cultures were adjusted to a starting OD_{600} of 0.02 into 12-well plates, with 2 mL of medium in each well. Strains were cultivated in triplicates with or without aTc, added at the beginning of the culture. Optical density at 600nm was measured every 10 min using a plate reader (Biotek, Synergy) for 5.5 h. Plates were then rapidly transferred to ice and 1.75 ml were sampled.

Proteomics sample preparation and measurement

Proteomics samples were prepared as described previously (Donati et al., 2021) with minor modifications. Cultivations were performed as described above. Culture aliquots were transferred into 2 mL reaction tubes and washed two times with PBS buffer (0.14 mM NaCl, 2.7 mM KCl, 1.5 KH2PO4, 8.1 Na2HPO4). Cell pellets were solubilized in 0.5 % sodium laroyl sarcosinate (SLS) and protein denaturation was carried out by applying heat (95 °C for 10min), followed by incubation with 5 mM Tris(2-carboxyethyl)phosphine (TCEP) and 10 mM iodoacetamide. 50 µg extracted protein was digested with 1 µg trypsin (Serva) overnight and peptides were purified using C18 reverse phase solid phase extraction columns (Macherey-Nagel). Peptides were analyzed using a Q-Exactive Plus mass spectrometer connected to an Ultimate 3000 RSLC nano and a nanospray flex ion source (Thermo Scientific). Peptide separation was performed on a reverse-phase HPLC column (75 µm x 42 cm) packed in-house with C18 resin (2.4 µm, Dr. Maisch GmbH, Germany). The following separating gradient was used: 96 % solvent A (0.15% formic acid) and 4 % solvent B (99,85 % acetonitrile, 0.15 % formic acid) to 30 % solvent B over 90 minutes at a flow rate of 300 nL/min. The data acquisition mode was set with the following parameters: 1 MS scan at a resolution of 70,000 with 50 ms max. ion injection fill time, MS/MS at 17,500 scans of the 10 most intense ions with 50 ms max. fill time. Label-free quantification (LFQ) of the data was performed using Progenesis QIP (Waters) and MASCOT (v2.5, Matrix Science) for spectrum/peptide identification. Progenesis outputs were further processed with SafeQuant. The data was further processed with custom MATLAB scripts.

Cultivation conditions for metabolome sampling

Strains were streaked on LB+Amp+Kan plates and single colonies were inoculated in liquid 5 mL LB+Amp+Kan for 6 hours. Different dilutions were prepared in M9 medium and incubated over night at 37°C. For strains with an ArgR reporter the M9-precultures were adjusted to a starting OD_{600} of 0.05 in 2 mL culture volume in 12-well plates using a plate reader (Tecan, Grödig, Austria, Spark). Strains were cultivated for 6.5 h in triplicates were aTc inducer was added at the beginning of the experiment. Strains with a FruB reporter were adjusted to a starting OD_{600} of 0.05 in 50 mL culture volume. Strains were cultivated in triplicates with and without inducer for 4.5 h. Optical density at 600nm was measured every 10 min using a spectrophotometer. For sampling the cultures were transferred rapidly to a thermostatically controlled hood at 37 °C while shaking them continuously during the sampling procedure.

Metabolomics measurements

Cultivations were performed as described above. Samples from cultures were vacuum-filtered on a 0.45 µm pore size filter (HVLP02500, Merck Millipore). After filtering, filter membranes were immediately transferred into a 40:40:20 (v-%) acetonitrile/methanol/water extraction solution at -20°C. Filters were incubated for 30 min in the extraction solution. Thereafter, metabolite extracts were centrifuged for 15 minutes at 13,000 rpm at -9 °C and only the supernatant was stored at -80 °C until analysis. Agilent 6495 triple quadrupole mass spectrometer (Agilent Technologies) was used for LC-MS/MS analysis as described previously (Guder et al., 2017). For liquid chromatography an Agilent 1290 Infinity II UHPLC system (Agilent Technologies) was used. Metabolite extracts were mixed with in a 1:1 ratio with prepared 13C-labeled internal standard. The ratio of 12C and 13C peaks, normalized to OD at the sampling point was used to quantify the metabolites. 3µL injection volume was used and the temperature in the column oven was 30°C. To measure amino acids the LC column Acquity BEH amide (30 x 2.1 mm, 1.7 µm) was used in acidic conditions. Used solvents were in channel A: water with 10 mM ammonium formate and 0.1% formic acid (v/v) and in channel B: acetonitrile with 0.1% formic acid (v/v). To measure hexose phosphates and fructose1,6-biphosphate the LC column iHILIC-Fusion(P) (50 x 2.1 mm, 5 µm) was used in basic conditions. Used solvents were in channel A: water with 10 mM ammonium carbonate and 0.2% ammonium hydroxide (v/v) and in channel B: acetonitrile without additive. As gradient for acidic and basic condition was used: 0 min 90% B; 1.3 min 40 % B; 1.5 min 40 % B; 1.7 min 90 % B; 2 min 90 % B. For the measurement of fructose-1-phosphate samples were processed and stored as described before. HILICpak VG-50 2D (2.0 x 150 mm) was used as LC column. The injection volume was 9 µL and the temperature of the column oven was 30°C. LC solvent in channel A was water with 10 mM ammonium formate and 0.1% formic acid (v/v) LC solvent in channel B was acetonitrile with 0.1% formic acid (v/v). Applied conditions were: 63 min 30% B.

Supplementary Figures

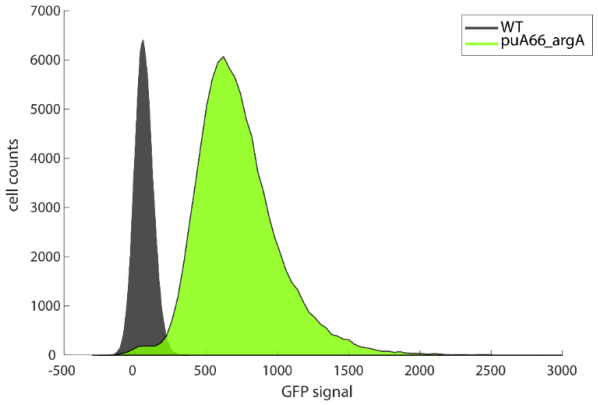


Figure S1. FACS Data control ArgR. Shown are the green fluorescent signals of 100000 cells sampled from a wildtype YYdCas9 culture (black) and the wildtype strain with the fluorescent transcriptional reporter plasmid to measure the activity of ArgR (green).

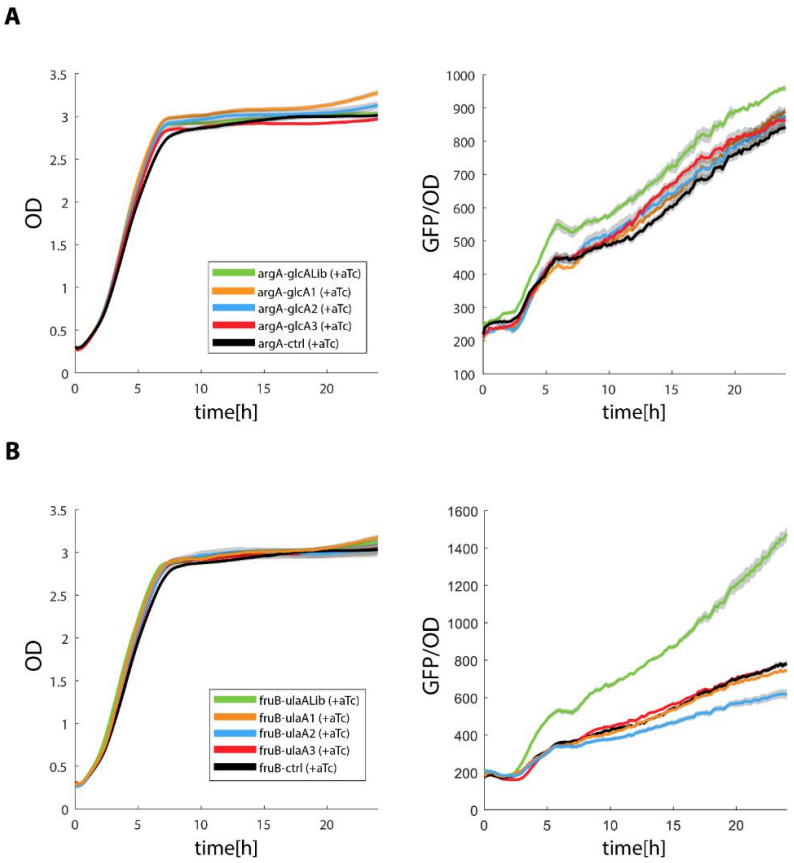


Figure S2. OD and GFP Levels of off-targets from the ArgR and Cra screening. To prove off-target activity new guides were designed (1-3) and compared to the growth and GFP levels of a control strain (ctrl) and the guide that was found in the library screening (Lib). All strains were tested with the induction of the CRISPRi system. Shown in (A) are guides targeting the gene *glcA* and in (B) guides targeting *fruB*.

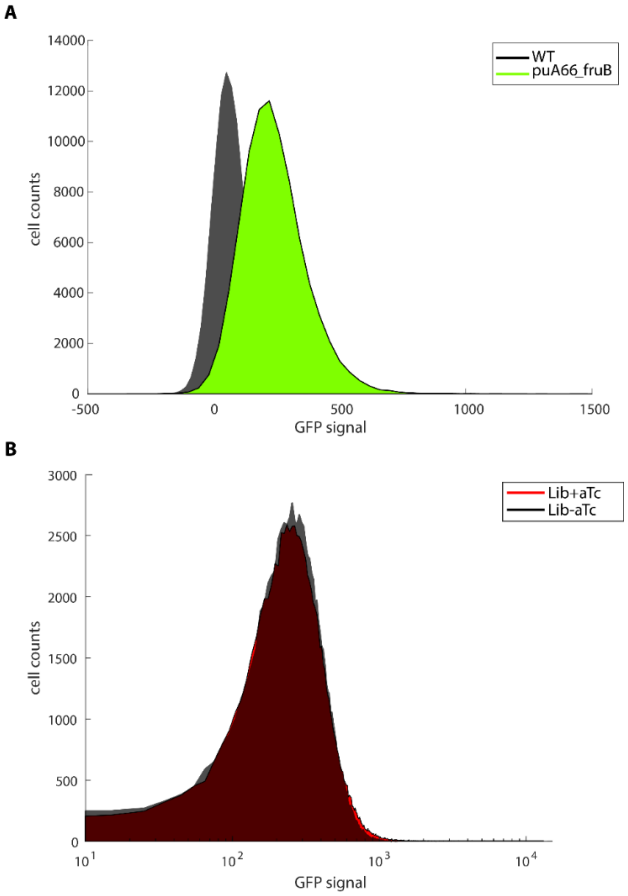


Figure S3. FACS data of the Cra library. (A) Green fluorescent signals of 100000 cells sampled from a wildtype YYdCas9 culture (black) and the wildtype strain with the fluorescent transcriptional reporter plasmid to measure the activity of Cra (green). (B) Green fluorescent signal of the Cra library with (red) and without induction of CRISPRi (black).

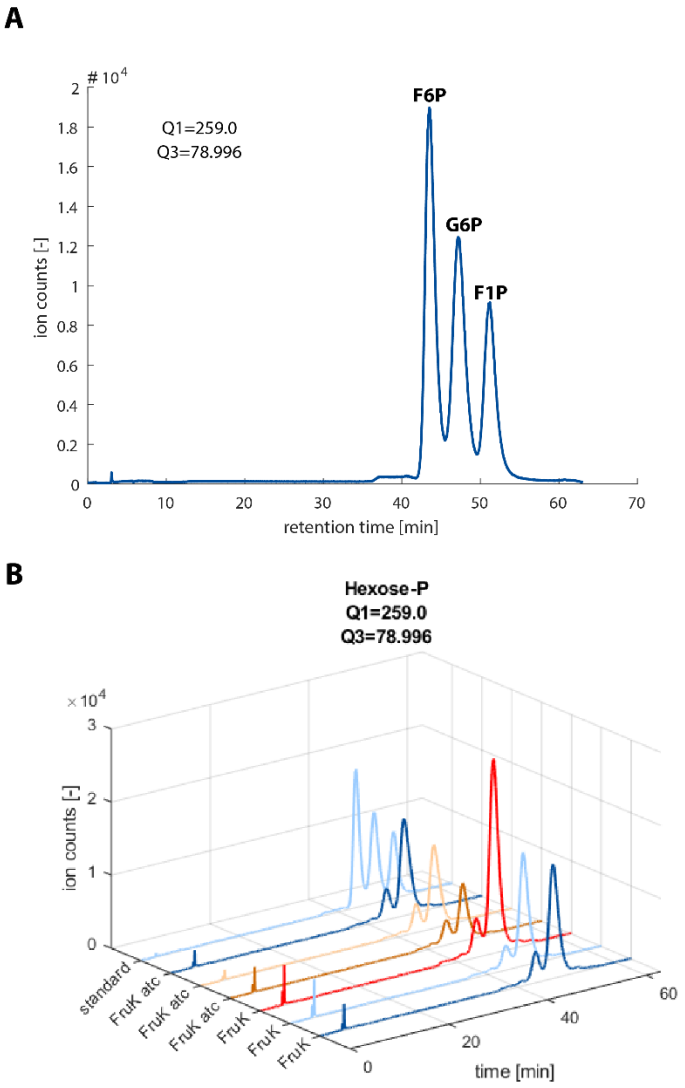


Figure S4. Measurement of Hexose-phosphates. (A) Chromatogram of the separation of fructose-1-phosphate (F1P) from glucose-6-phosphate (G6P) and fructose-6-phosphate (F6P) from a pooled standard mix, while the concentration of each metabolite in the mix was 1 µM. (B) Chromatograms of the separation of Hexose-phosphates in the standard mix and the *fruK* knockdown strain induced (atc) and not-induced.

References

Bae, S., Park, J., and Kim, J.-S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics *30*, 1473–1475.

Beuter, D., Gomes-Filho, J.V., Randau, L., Díaz-Pascual, F., Drescher, K., and Link, H. (2018). Selective Enrichment of Slow-Growing Bacteria in a Metabolism-Wide CRISPRi Library with a TIMER Protein. ACS Synthetic Biology *7*, 2775–2782.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res *41*, 7429–7437.

Bley Folly, B., Ortega, A.D., Hubmann, G., Bonsing-Vedelaar, S., Wijma, H.J., van der Meulen, P., Milias-Argeitis, A., and Heinemann, M. (2018). Assessment of the interaction between the flux-signaling metabolite fructose-1,6-bisphosphate and the bacterial transcription factors CggR and Cra: interaction between the flux-signaling metabolite fructose-1,6-bisphosphate and the bacterial transcriptional factors. Molecular Microbiology *109*, 278–290.

Chavarría, M., Santiago, C., Platero, R., Krell, T., Casasnovas, J.M., and de Lorenzo, V. (2011). Fructose 1-Phosphate Is the Preferred Effector of the Metabolic Regulator Cra of Pseudomonas putida. Journal of Biological Chemistry *286*, 9351–9359.

Cui, L., Vigouroux, A., Rousset, F., Varet, H., Khanna, V., and Bikard, D. (2018). A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9. Nature Communications *9*.

Diether, M., and Sauer, U. (2017). Towards detecting regulatory protein–metabolite interactions. Current Opinion in Microbiology *39*, 16–23.

Donati, S., Kuntz, M., Pahl, V., Farke, N., Beuter, D., Glatter, T., Gomes-Filho, J.V., Randau, L., Wang, C.-Y., and Link, H. (2021). Multi-omics Analysis of CRISPRi-Knockdowns Identifies Mechanisms that Buffer Decreases of Enzymes in E. coli Metabolism. Cell Systems *12*, 56-67.e6.

Gallego, O., Betts, M.J., Gvozdenovic-Jeremic, J., Maeda, K., Matetzki, C., Aguilar-Gurrieri, C., Beltran-Alvarez, P., Bonn, S., Fernández-Tornero, C., Jensen, L.J., et al. (2010). A systematic screen for protein–lipid interactions in *Saccharomyces cerevisiae*. Molecular Systems Biology *6*, 430.

Gerosa, L., and Sauer, U. (2011). Regulation and control of metabolic fluxes in microbes. Curr. Opin. Biotechnol. *22*, 566–575.

Guder, J.C., Schramm, T., Sander, T., and Link, H. (2017). Time-Optimized Isotope Ratio LC–MS/MS for High-Throughput Quantification of Primary Metabolites. Analytical Chemistry *89*, 1624–1631.

He, C., Howes, B.D., Smulevich, G., Rumpel, S., Reijerse, E.J., Lubitz, W., Cox, N., and Knipp, M. (2015). Nitrite dismutase reaction mechanism: kinetic and spectroscopic investigation of the interaction between nitrophorin and nitrite. J. Am. Chem. Soc. *137*, 4141–4150.

He, L., Binari, R., Huang, J., Falo-Sanjuan, J., and Perrimon, N. (2019). In vivo study of gene expression with an enhanced dual-color fluorescent transcriptional timer. ELife *8*.

Kochanowski, K., Sauer, U., and Noor, E. (2015). Posttranslational regulation of microbial metabolism. Current Opinion in Microbiology *27*, 10–17.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc *8*, 2180–2196.

Lempp, M., Farke, N., Kuntz, M., Freibert, S.A., Lill, R., and Link, H. (2019). Systematic identification of metabolites controlling gene expression in E. coli. Nature Communications *10*.

Li, X., Gianoulis, T.A., Yip, K.Y., Gerstein, M., and Snyder, M. (2010). Extensive In Vivo Metabolite-Protein Interactions Revealed by Large-Scale Systematic Analyses. Cell *143*, 639–650.

Maas, W.K. (1994). The arginine repressor of Escherichia coli. Microbiological Reviews *58*, 631–640.

Miao, H., Ratnasingam, S., Pu, C.S., Desai, M.M., and Sze, C.C. (2009). Dual fluorescence system for flow cytometric analysis of Escherichia coli transcriptional response in multi-species context. Journal of Microbiological Methods *76*, 109–119.

Motlagh, H.N., Wrabl, J.O., Li, J., and Hilser, V.J. (2014). The ensemble nature of allostery. Nature *508*, 331–339.

Nicolas, P., Mader, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M., Aymerich, S., et al. (2012). Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in Bacillus subtilis. Science *335*, 1103–1106.

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

Piazza, I., Kochanowski, K., Cappelletti, V., Fuhrer, T., Noor, E., Sauer, U., and Picotti, P. (2018). A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. Cell *172*, 358-372.e23.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell *152*, 1173–1183.

Qin, Q., Wang, B., Wang, J., Chang, M., Xia, T., Shi, X., and Xu, G. (2019). A comprehensive strategy for studying protein-metabolite interactions by metabolomics and native mass spectrometry. Talanta *194*, 63–72.

Rousset, F., Cui, L., Siouve, E., Becavin, C., Depardieu, F., and Bikard, D. (2018). Genome-wide CRISPR-dCas9 screens in E. coli identify essential genes and phage host factors. PLOS Genetics *14*, e1007749.

Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., and Heinemann, M. (2016). The quantitative and condition-dependent Escherichia coli proteome. Nature Biotechnology *34*, 104–110.

Shimada, T., Fujita, N., Maeda, M., and Ishihama, A. (2005). Systematic search for the Cra-binding promoters using genomic SELEX system. Genes to Cells *10*, 907–918.

Silander, O.K., Nikolic, N., Zaslaver, A., Bren, A., Kikoin, I., Alon, U., and Ackermann, M. (2012). A Genome-Wide Analysis of Promoter-Mediated Phenotypic Noise in Escherichia coli. PLoS Genetics *8*, e1002443.

Singh, D., Fairlamb, M.S., Harrison, K.S., Weeramange, C., Meinhardt, S., Tungtur, S., Rau, B.F., Hefty, P.S., Fenton, A.W., and Swint-Kruse, L. (2017). Protein-protein interactions with fructose-1-kinase alter function of the central *Escherichia coli* transcription regulator, Cra (Microbiology).

Tagore, R., Thomas, H.R., Homan, E.A., Munawar, A., and Saghatelian, A. (2008). A global metabolite profiling approach to identify protein-metabolite interactions. J. Am. Chem. Soc. *130*, 14111–14113.

Van Duyne, G.D., Ghosh, G., Maas, W.K., and Sigler, P.B. (1996). Structure of the Oligomerization andL-Arginine Binding Domain of the Arginine Repressor ofEscherichia coli. Journal of Molecular Biology *256*, 377–391.

Vinayavekhin, N., and Saghatelian, A. (2011). Discovery of a protein-metabolite interaction between unsaturated fatty acids and the nuclear receptor Nur77 using a metabolomics approach. J. Am. Chem. Soc. *133*, 17168–17171.

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M.G., and Alon, U. (2006). A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. Nature Methods *3*, 623–628.

Chapter 3

Creating a CRISPRi library and performing genome-wide fitness screenings in *E. coli* or selective enrichments of strains by FACS

Michelle Kuntz[1,2], Thorben Schramm[2], Dominik Beuter[1], José Vicente Gomes-Filho[1], Lennart Randau[1], Hannes Link[2,3]

[1]Max Planck Institute for Terrestrial Microbiology, Marburg 35043, Germany

[2] Present address: Interfaculty Institute for Microbiology and Infection Medicine Tübingen, University of Tübingen, Auf der Morgenstelle 24, 72076 Tübingen, Germany

[3]Lead Contact

This chapter is written in manuscript style and is in preparation (work in progress). This manuscript includes a detailed protocol for all working steps used in chapter 1 and 2 of this thesis. My contribution to this work included the development, testing and implementation of all working steps, and co-writing the manuscript.

Abstract

Pooled CRISPRi library screenings are a well-established and powerful tool to investigate the function of essential and non-essential genes. Therefore, it is very important to establish high-throughput readouts, which is mostly achieved by the combination of fitness growth assays with NGS. Furthermore, the combination of CRISPRi with fluorescent reporters allows to screen for and select strains which show a certain phenotype, from big libraries. The protocol provided here explains how to design and construct sgRNA libraries for an inducible CRISPRi system in *E. coli* with a genomically integrated dCas9 protein, the transformation of sgRNA libraries into the strain of interest, a detailed timeline how to perform fitness growth assays and NGS, data analysis of NGS and alternatively how to combine a CRISPRi library with a fluorescent reporter and cell sorting via FACS. The protocol enables to design and construct the library in 6-7 days, performing a fitness assay, NGS and the data analysis in 5-7 days or to execute a selective enrichment via FACS in 4 days.

Introduction

The sequencing of full genomes made it possible to identify the location of genes or homologues in organisms. However, for many of them the *in vivo* function could not be revealed by systematic approaches yet. In the past, targeted genome regulation methods like zinc-finger or transcription activator like effector (TALE) proteins, RNA interference (RNAi) and sRNAs were used to investigate the function of genes (Garg et al., 2012; Hannon, 2002; Segal and Barbas, 2001; Yoo et al., 2013). Nevertheless, these methods often show limitations because of many off-target effects, toxicity, low efficiency and can only be used if the investigated organism holds the appropriate enzyme machinery (Larson et al., 2013). Moreover, big strain libraries were generated, providing a knockout of every gene in the genome of a certain organism (Baba et al., 2006). Screenings of these strains enable the exploration of the specific phenotypes of the gene knockouts and helps to assign a gene to a function. Yet, these screenings have limitations in exploring functions of essential genes as the knockouts can lead to auxotrophies, which means the strains are not able to grow anymore, and their function cannot be investigated under certain conditions. In conclusion, an inducible system is required which can target every location of a gene and is transferable to many organisms. Hence, the screening of big CRISPRi libraries was used widely to study the function of genes in many bacteria (Cui et al., 2018; Peters et al., 2019; Rousset et al., 2018; Todor et al., 2021; Wang et al., 2018).

The CRISPRi system contains a dead (inactive) Cas9 protein, which can be used to silence or activate genes. For this reason, CRISPRi can be used for RNA-guided transcription regulation (Bikard et al., 2013; Larson et al., 2013; Qi et al., 2013). The dCas9 forms a complex with a 20-25 bp long single guide RNA (sgRNA), whose sequence is complementary to the template or non-template strand of the DNA target and is located next to the protospacer adjacent motif (PAM) (**Figure 1A**). The dCas9 protein does not have endonuclease activity, hence it will bind the target sequence and will interfere with transcription by sterically hindering the binding of the RNA polymerase (RNAP). Thus, every gene can be silenced by CRISPRi if its sequence contains a PAM site. The strength of the silencing can be controlled, for example by introducing mismatches or choosing different target positions inside the gene (Larson et al., 2013; Qi et al., 2013). Moreover, the expression of dCas9 or of the sgRNAs can be induced (Beuter et al., 2018; Larson et al., 2013; Lawson et al., 2017; Qi et al., 2013).

The biggest advantage of CRISPRi is that it can be used for pooled library screenings, which are less laborious and cost-intensive as experiments with single strains (Shalem et al., 2015). All steps can be performed in a pooled manner, from cloning the library to the final phenotyping. As sequencing gets cheaper every year, it not only allows to order big oligonucleotide libraries, but also to combine pooled growth assays with Next-Generation sequencing (NGS) to define the fitness of gene knockdowns.

(Donati et al., 2021; Rousset et al., 2018, 2021; Todor et al., 2021). Furthermore, the combination of a CRISPRi library with a fluorescent reporter plasmid enables the enrichment of strains with a certain phenotype via Fluorescence-activated Cell Sorting (FACS) (Beuter et al., 2018).

Development of the protocol

The CRISPRi interference library described here was initially created to select slow-growing strains from a pooled library of *E. coli* NCM3722 with the help of a TIMER protein, indicating the growth-rate of bacteria (Beuter et al., 2018). We also cloned the sgRNA library into the plasmid pgRNA of the original CRISPRi system (Qi et al., 2013) and used it in *E. coli* YYdCas9, which was created by Lawson et al. (2017). This strain has the gene for dCas9 introduced into the genome and therefore shows a more tightly controlled expression of dCas9, compared to the plasmid-based system. When performing fitness growth assays with the metabolism-wide library, we investigated dynamically how the knockdown of certain metabolic genes influences the growth of *E. coli* over time (Donati et al., 2021). Furthermore, by measuring the metabolome and proteome of 30 single CRISPRi strains, we could show which mechanisms are used by *E. coli* to buffer knockdowns. Currently, the CRISPRi library is combined with a fluorescent reporter for transcription factor activity (Kuntz et al., in preparation). In this protocol, we describe in detail how to construct and clone a CRISPRi library and how to track the abundance of library members during dynamic fitness assays by next-generation amplicon sequencing. Furthermore, we show which other methods and techniques can be used to investigate CRISPRi libraries and perform selective enrichments by FACS sorting.

Applications of the method

Our designed CRISPRi library can be used to study the roles of all metabolic genes of the latest iML1515 model of *E. coli* (addgene accession number pending). It contains 7177 strains with 4-6 sgRNAs per gene. By comparing the effect on the knockdown between the different guides, off-target effects can be identified when, for example, only one of the guides leads to a specific effect. Furthermore, the strength of guides can be measured directly. CRISPRi is not limited to *E. coli* and can also be used in other organisms like *Bacillus subtilis* or *Vibrio cholerae* (Todor et al., 2021). Various studies were done that investigated fitness effects based on CRISPRi knockdowns, effects of antibiotics, as well as comparison studies between knockout strains and the wildtype (Liu et al., 2021; Qu et al., 2019; Rousset et al., 2018; Shields et al., 2020). The here described library can easily be re-cloned into other vectors (pNUT1527), which makes it possible to apply it for other systems and organisms. Furthermore, the Matlab code provided here makes it possible to design guides for single genes or to get a list of sgRNAs for other organisms if the correct genomic sequence is loaded. Therefore, it is

possible to design oligos for single strains as well as oligo pools for the cloning of CRISPRi strain libraries, as it is described here.

This protocol mainly focuses on fitness screenings of CRISPRi libraries, but we also show which other methods can be combined instead of NGS to obtain faster readouts and for example single strains out of a pool of strains for direct investigations without re-cloning.

Comparison with other methods

To determine the role of a gene, it is necessary to disrupt its functionality inside the cell. Hence, methods like screening knockout libraries provide a good insight on how a certain gene contributes to specific phenotypes which can finally reveal its function in metabolism (Baba et al., 2006; Giaever et al., 2002). But the creation of these libraries also has its limitations as knockouts of essential genes cannot be created and the function of the gene cannot be studied because the cells cannot grow anymore. Furthermore, the construction of such arrayed libraries is very laborious and has therefore only been done for a handful of organisms. Besides this, the screening of knockout libraries is mostly performed with single cells and not in a pooled manner, as it is very difficult to map the deletion of a gene with any high-throughput technology. For example, the usage of DNA microarrays only enables the comparison of the gene expression level of the WT and one gene knockout strain (Capaldi, 2010). The strains are grown to a certain OD, the DNA or RNA is extracted and labelled with different fluorescent dyes. The microarray is produced by the annealing of cDNA or oligonucleotides to a solid surface, which are later incubated with the labelled samples and washed. Thus, the examination of DNA microarray experiments is very laborious, and one needs to precisely follow every little detail of the protocol to prevent high variations between the samples.

The application of transposon insertion sequencing (Tn-seq) has allowed to identify the function of genes at a genome-scale (van Opijnen et al., 2009; Québatte et al., 2017). The approach is based on the usage of a saturated transposon insertion library, with transposons as mobile genetic elements, inserting into every non-essential gene and therefore disrupting the function of the gene. The library is grown under a certain experimental condition and all disrupted genes, which are necessary for growth under this condition, will decrease in frequency from the whole population. The fitness of every mutant is mapped by the massively parallel sequencing of the transposon flanking region and therefore reveals the abundance and location of each mutation. However, this method also has various limitations as it cannot be applied to essential genes, the genes of interest cannot be directly targeted due to the random insertion of the transposon and thousands of mutants are needed to reach full saturation.

In contrast, the protocol provided here for pooled CRISPRi library screenings enables the precise targeting of genes of interest, leading to smaller library sizes. Moreover, CRISPRi is inducible and allows the dynamic control over the repression strength (Hawkins et al., 2020; Qi et al., 2013). Furthermore, the knockdown of an essential gene only leads to a growth phenotype after a certain time, allowing to investigate its function under different growth conditions. We show here that CRISPRi libraries can be easily screened using a fluorescent transcriptional reporter. By FACS, single cells of a CRISPRi library that have a certain phenotype can be isolated. These strains can then be further investigated, as it is known for arrayed CRISPRi library screenings (Shalem et al., 2015). Therefore, the hereby provided protocol combines the high efficiency of pooled CRISPRi screenings with the precision of experiments done with arrayed strains.

Experimental design (Figure 3)

Design and construction of the CRISPRi library

For each gene in the iML1515 model (Monk et al., 2017), 4-6 NGG PAM sites that target the coding strand and are equally distributed over the open reading frame were filtered with customized Matlab scripts. In total, 7184 sites were considered. Next to the chosen PAMs, 20 nucleotides (nt) were selected as the protospacer sequence for a small guide RNA (sgRNA). These sgRNAs were paired with 65 nt flanking regions, homologous to the pgRNA backbone, which was used as a parental vector for this study (**Figure 1B**). This resulted in 150 nt long oligonucleotides, which were synthesized by Agilent Technologies. The resulting oligonucleotide library was amplified with 15 cycles of PCR. The pgRNA vector was linearized by PCR and the before amplified oligonucleotides were inserted by Gibson assembly. The resulting product was purified and then transformed into *E. coli* YYdCas9 by electroporation. The transformants were plated on 15 cm diameter Petri dishes, which resulted in approximately 9.9×10^7 colonies. LB medium was spread on plates and all colonies were washed from the plate. The resulting cell suspension was pooled, glycerol was added and stored as a glycerol stock at -80°C.

Fitness assays with the pooled CRISPRi library

The library was grown in duplicates in shake flasks on M9 minimal medium with glucose with a starting OD of 0.05. The pre-culture was kept in exponential growth phase to allow a direct induction of the CRISPR interference system in the main culture with 200 nM anhydrotetracycline (aTc). After the induction, the cultures were grown for 14 hours, while every two hours they were back-diluted to OD 0.05. Samples were taken every hour. Cells were centrifuged and the pellet stored at -20°C.

Illumina Miniseq

After finishing the growth experiment and collecting samples for 14 hours, plasmid DNA of all samples was extracted and concentrations were measured. A 300 bp fragment of the plasmids, including the sgRNA sequence, was amplified and Illumina adapters added for performing NGS. The sequence of the adapters was dependent on the combination of samples that were sequenced in one run. For the sequencing of this library, the Illumina Mini-Seq High Output Reagent Kit (150 cycles) was used, which allows to get up to 25M reads. For a library of ~7200 sequences, we wanted to have at least a coverage of 100 reads per guide per run which results in 720000 reads per sample. Due to the high similarity of the amplicons, 50% phiX was used. To ensure sufficient coverage, we calculated with ~2M reads per sample, which allowed us to sequence 10 samples at the same time. Thus, the samples were sequenced in 3 different batches, making sure that the replicates of each time point were sequenced in the same run to not introduce any batch effects. Instead of splitting the samples, the Illumina system also provides machines and kits with a higher read output (Hiseq).

Data analysis

Demultiplexing was carried out by the MiSeq system and resulted in .fastq files. These files were loaded into our custom-scripted Matlab code and the 20 bp sgRNA sequence was trimmed from the read. The sgRNA sequences were then aligned with the sequences of the initial sgRNA library and only perfect matching sequences were counted. These counts were divided by the number of all reads per run/sample. sgRNAs with less than 10 reads were excluded from further analysis. The $\log_2$ fold-change was calculated by the division of reads of each sgRNA of each time point with the starting read count. The response time was determined for every target gene, by fitting a sigmoidal function to the time course of the fitness score. The response time is defined as the time when the knockdown reduced the fitness of the strain by 50%.

Optional screening methods for a CRISPRi library

FACS

Instead of performing an NGS screening of samples taken from a fitness assay, it is also possible to combine the pooled library with a fluorescent reporter plasmid and to sort a fraction with the desired fluorescent response from the pooled strain library. This was used either to enrich slow-growing cells by cell sorting via the monitoring of the growth rate of the strains (Beuter et al., 2018) or to find targets, whose knockdown influenced the activity of a transcription factor (Kuntz et al., in preparation). Both approaches make it possible to find and array cells, in which the knockdown of a target gene provokes a certain response of the reporter plasmid. Hence, these strains are directly available for further experiments and it is not necessary to reclone them.

The library was transformed with a fluorescent reporter plasmid and grown on M9 glucose with inducer until the desired time point. 10 ml of the culture were sampled and washed with PBS. The cells were loaded into the FACS and gates were set according to the desired fluorescence. Sorted cells were grown on plates or in a shake flask and then further investigated either by NGS or plate reader experiments.

Expertise needed to implement the protocol

For the design of sgRNAs to target the *E. coli* genome and the analysis of data sets customized Matlab scripts were used. For the customization and application in other organisms, basic programming skills are needed. When evaluating CRISPRi fitness screens with the protocol described here, expertise in handling Illumina MiniSeq system is needed. Alternatively, sequencing services can be used. Furthermore, when sorting cells from the library via FACS, the user must be familiar with handling a sorting machine or the service of a core facility is needed.

Limitations

Targeting genes with CRISPRi also has its limitations. Genes without PAM sites or PAM sites located distant to the promoter region, will lead to no or mild repression of the target. Therefore, not every gene can be repressed by CRISPRi. Furthermore, targeting genes in operons will lead to a repression of all genes located downstream of the target. Hence, the knockdown of genes in operons cannot be performed without effecting the rest of the operon and cannot be used for gene-level resolution.

The design of guides used in this protocol does not include the search for bad-seed effects or off-targets and on-target activity evaluation. Therefore, each gene is targeted by several guides and users should always exclude off-targets by calculating means of several guides or only conclude a specific phenotype if at least two guides are showing the same effect.

The execution of fitness assays with CRISPRi libraries can also be influenced by competitive effects, as some knockdowns can lead to an overproduction of supplements, which are secreted to medium or the knockdown of certain genes can lead to higher growth rates, as for example the knockdown of the gene *relA* (Donati et al., 2021; Towbin et al., 2017).

In conclusion, CRISPRi should be mainly used as a screening method and certain phenotypes should be validated by follow-up experiments of single CRISPRi strains.

Materials

Reagents

Reagent or Resource	Source	Identifier
Anhydrotetracycline (final conc. 200 nM)	Sigma-Aldrich	Cat#1035708-25MG
IPTG	Roth	Cat#CN08.2
Ampicillin (100 µg mL-1)	Roth	Cat#K029.2
Kanamycin (50 µg mL^{-1})	Roth	Cat#T832.3

Kits

- Bioanalyzer HS DNA Sensitivity Kit (cat. no. 5067-4626)
- Illumina MiniSeq High Output Reagent Kit (300-cycles; cat. no. FC-420-1003)
  MiniSeq High Output Reagent Kit (150-cycles; cat. no. FC-420-1002)
- AMPure XP beads (Beckman Coulter; cat. no. 10453438)
- NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, Germany; cat. no. 740609.50)
- DNA Clean and Concentrator-5 (Zymo Research; cat. no. D4014)

Equipments

Hardware

- Electroporator
- Bioanalyzer Agilent
- Illumina Mini-Seq
- Qubit 2.0 Fluorometer
- FACS: Sony SH800SHP
- Magnetic stand for Eppendorf tubes
- Big petri dishes (15 cm)

Software

Reagent or Resource	Source	Identifier
Matlab R2018b (9.5.0.944444) for analysis of experimental data	mathworks.com	N/A

Plasmids

Reagent or Resource	Source	Identifier
pgRNA-bacteria	Qi et al. 2013	Addgene plasmid #44251

Strains

Reagent or Resource	Source	Identifier
NEB® 5-alpha Competent E. coli: fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	New England Biolabs	Cat#C2987
YYdCas9: BW25993 intC::tetR-dcas9-aadA lacY::ypet-cat	Lawson et al. 2017	N/A

Oligonucleotides

Oligonucleotide	Sequence (5'-3')	Description
psgRNAamp-F	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Amplification of pgRNA for Gibson Assembly with amplified spacer oligonucleotides
psgRNAamp-R	ACTAGTATTATACCTAGGACTGAGCTAGC	Amplification of pgRNA for Gibson Assembly with amplified spacer oligonucleotides
protoamp-F	TTGACAGCTAGCTCAGTCCTAGGTATAATACTAGT	Amplification of spacer oligonucleotide

protoamp-R	GCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC	Amplification of spacer oligonucleotide
FWD_cassette_seq1	CCGAGTTGCTCTTGCC	Sequencing of cloned CRISPRi plasmids
Rev_pkD-pgRNA_seq2	GACTCGAGTAAGGATCCAGTTC	Sequencing of cloned CRISPRi plasmids
OH_amp_fwd	TAAGGATGATTTCTGGAATTCTAAAG	Amplification of pooled oligonucleotides
OH_amp_rev	GTGCCACTTTTTCAAGTTGATAAC	Amplification of pooled oligonucleotides
EcF_forward	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Amplification of the pgRNA backbone for Gibson Assembly with amplified pooled oligonucleotides (Qi et al., 2013)
EcF_reverse	ACTAGTATTATACCTAGGACTGAGCTAGC	Amplification of the pgRNA backbone for Gibson Assembly with amplified pooled oligonucleotides (Qi et al., 2013)
NGS_F2_adapter	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGCAATAGGCGTATCACGAGG	Amplification of a 300 bp fragment of pgRNA including the sgRNA
NGS_R2_adapter	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGACGGCGCTATTCAGATCC	Amplification of a 300 bp fragment of pgRNA including the sgRNA
Custom_N705	CAAGCAGAAGACGGCATACGAGATGGACTCCTGTCTCGTGGGCTCGG	I7 oligo
Custom_N706	CAAGCAGAAGACGGCATACGAGATTAGGCATGGTCTCGTGGGCTCGG	I7 oligo
Custom_N721	CAAGCAGAAGACGGCATACGAGATTACGCTGCGTCTCGTGGGCTCGG	I7 oligo

Custom_N503	AATGATACGGCGACCACCGAGATCTACACAGAG GATATCGTCGGCAGCGTC	I5 oligo
Custom_N504	AATGATACGGCGACCACCGAGATCTACACAGAG TAGATCGTCGGCAGCGTC	I5 oligo
Custom_N511	AATGATACGGCGACCACCGAGATCTACACCGGA GAGATCGTCGGCAGCGTC	I5 oligo
Custom_N513	AATGATACGGCGACCACCGAGATCTACACCTAGT CGATCGTCGGCAGCGTC	I5 oligo

Supplementary Protocols from companies

- NEB: Making your own electrocompetent cells
- Illumina Miniseq System Guide
- Illumina: MiniSeq System-Denature and Dilute Libraries Guide
- Illumina Adapter Sequences
- NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1)

Media and buffers

Reagent or Resource	**Source**	**Identifier**
LB Medium	Roth	Art. No. X968.2
M9 Minimal medium		
5 x base salt solution		
Na_2HPO_4 2 H_2O (30 g)	Roth	P030.2
KH_2PO_4 (15 g)	Roth	3904.1
$(NH_4)2_SO_4$ (7.5 g)	Sigma-Aldrich	A3920
NaCl (2.5 g)	Roth	9265.1
-add ddH2O to obtain 1 L		
100 x trace element solution		
ZnSO4 · 7 H2O (0.18 g)	Sigma-Aldrich	Z4750
CuCl2 · 2 H2O (0.12 g)	Sigma-Aldrich	C3279
MnSO4 · H2O (0.12 g)	Sigma-Aldrich	M7899

CoCl2 · 6 H2O (0.18 g)	Sigma-Aldrich	255599
- add ddH2O to obtain 1 L		
500 x Thiamine-HCl (Vit. B1)		
Thiamine-HCl	Sigma-Aldrich	T4625-25G
- add ddH2O to obtain 50 ml - store at 4°C		
1000 x MgSO4 solution		
MgSO4 · 7 H2O (24.6 g)	Sigma-Aldrich	63138-250G
- add ddH_2O to obtain 100 ml - store at 4°C		
1000 x $CaCl_2$ solution		
CaCl2 · 2 H2O (1.47 g)	Sigma-Aldrich	C8106-500G
- add ddH_2O to obtain 100 ml - store at 4°C		
1666 x $FeCl_3$ solution		
$FeCl_3$ · 6 H2O (1.35 g)	Sigma-Aldrich	31232-250G-D
- add ddH_2O to obtain 50 ml - store covered with tin foil (light sensitive!!!) - store at 4 °C		
Carbon source solution (200 g/L)		
Glucose (8 g)	Roth	X997.1
- add ddH_2O to obtain 40 mL (use Falcon tube) - sterile filtrate		
SOB medium		
Tryptone (2 %)	Sigma-Aldrich	T9410

Yeast extract (0.5 %)	Sigma-Aldrich	Y1625
NaCl (10mM)	Roth	9265.1
KCl (2.5 mM)	Sigma-Aldrich	P4504
$MgCl_2$ (10mM)	Roth	KK36.1
$MgSO_4$ (10mM)	Sigma-Aldrich	63138-250G
SOC medium		
SOB medium + Glucose (20 mM)		

For 1 L M9 minimal liquid medium:

Fill a Schott Bottle in the following order (mix by shaking after each step!)

1. 200 ml 5 x base salts
2. fill with app. 600 ml ddH_2O
3. 10 ml trace element solution
4. 1 ml $MgSO_4$ solution
5. 1 ml $CaCl_2$ solution
6. 0.6 ml $FeCl_3$ solution
7. 2 ml Thiamine solution
8. use ddH2O to fill up to 1 L (use mark of the bottle)

Sterile filtrate and store at room temperature.

Add carbon source before experiment.

Procedure

sgRNA design and planning of amplicons TIMING 1 day

1. Run “CRISPRi_Design_Tool.m” (**Figure 2**)
2. Add the target genes by typing in e.g. “argA” and pressing “Add target gene” or load a list of targets from an Excel sheet (in “.xls” format). The folder also contains a file “amino acid related genes.xls”, which gives an example for such loading of target genes
3. Select the number of sgRNAs per gene target and the DNA strand to be targeted. By default, the non-template (coding) strand is preselected because targeting this strand usually gives stronger repression

4. Press “Create sgRNA” to generate the protospacers (recognition sites) of the targets

Beware! The tool does only design the protospacers (recognition sites). Any handles (like the dCas9 handle) or overlaps (such as the promotor region) must be added!

5. By clicking on the entries of the “Found genes”-list box you can select which gene you want to review
6. The right table gives some metadata related to the selected gene targets (based on gene ontology consortium database, 12/04/2017)
7. By pressing the “Export to .xls” button all designed protospacers (recognition sites) are saved to an spreadsheet
8. The finished list of sgRNAs has to be loaded with another customized Matlab script and overhangs are added to perform the Gibson assembly (step 15)
9. The finished list of oligonucleotides is uploaded to the ordering platform of a company providing the service of synthesizing pooled oligo libraries (Twist, Agilent)

Cloning of a CRISPRi library TIMING 1-2 days

10. For the amplification of guide RNAs in the oligonucleotide pool, use the following amounts of reagents:

    10 µL dNTPs
    2.5 µL each primer (OH_amp_fwd + rev)
    1 µL guide RNA stock (solubilized according to supplier’s instructions)
    20 µL Q5 reaction buffer
    1 µL Q5 polymerase
    63 µL H_2O

11. Split into 2 50 µL tubes and perform a PCR amplification based on supplier’s instructions with an annealing temperature of 62°C, elongation length 20 sec for 15 cycles
12. Perform a DNA clean up with NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, Germany)
13. For the amplification of the pgRNA acceptor plasmid backbone mix the following reagents:

    2 µL dNTPs
    5 µL each primer (EcF forward, EcF reverse)
    1 µL Template DNA (pgRNA, max 100 µg)
    20 µL Q5 reaction buffer
    1 µL Q5 polymerase
    66 µL H_2O

14. Split into 2 50 µL and perform a PCR amplification based on suppliers’ instructions with an annealing temperature of 66°C, elongation length 1:30 min for 30 cycles tubes
15. Perform a gel extraction from an agarose gel of the linearized plasmid

TROUBLESHOOTING

PAUSE point: amplified vector and inserts can be stored at -20°C

16. Perform a Gibson Assembly with the amplified oligos and the linearized vector:
    - Mix 100 ng plasmid, 29 ng insert (1:5 ratio), 10 µL Gibson Assembly Master Mix, ad 20 µL H_2O
    - Incubate at 50°C, 15 min
17. Perform a NucleoSpin Gel and PCR Clean-up, elute everything in H_2O
18. Do a DpnI digest according to the supplier's instructions (this step is important to get finally rid of the backbone plasmid)

CRITICAL step: if there is too much backbone left in the control step after transforming the library, search for advice in the troubleshooting table

TROUBLESHOOTING

19. afterwards clean up with a DNA Clean-up Kit (Zymo Research) and elute in 6 µl H_2O

Preparation of electrocompetent YYdCas9 cells TIMING 3 days

(protocol modified after NEB protocol, for source check section supplementary protocols from companies)

Day 1

20. Prepare the following buffers and media:
    - SOB and SOC
    - Sterile 10% glycerol (can be autoclaved) is needed for the washes. The volume of 10% glycerol needed is 2X the culture volume (for example, a 500 ml culture requires 1L of 10% glycerol)
21. Streak out the strain that should be made electrocompetent on fresh plates: streak *E. coli* YYdCas9 on LB plates without antibiotics

PAUSE point: cells have to be incubated over night

Day 2

22. Procedure for 2* 250 ml cultures (500 ml total): Inoculate 1 colony from a fresh plate of the strain to be made electrocompetent into 10 ml of SOB in a 125 ml flask and incubate for 16-18 hours at 37°C and 250 rpm

PAUSE point: pre-culture has to be inoculated over night

Day 3

23. Have ready 2, 1 L flasks containing 250 ml each of SOB pre-warmed to 37°C. Add two drops of the overnight culture to each of the flasks
24. Shake at 37°C and 250 rpm until the cultures reach an OD600 of 0.5-0.7. Be sure to turn on centrifuge and cool rotor to 4°C well in advance of harvesting cells. Be sure to place 1 L of 10% glycerol on ice well in advance of harvesting cells
25. Place cultures on ice for 15 minutes. From this point on the cultures must be kept ice cold. Pour each 250 ml culture into chilled 500 ml (or 1000 ml) centrifuge bottles
26. Centrifuge at 5000 rpm for 10 min. Pour off the supernatant and aspirate any residual broth
27. Add 250 ml of glycerol to each of the centrifuge bottles and completely suspend the cells by pipetting up and down
28. Centrifuge at 5000 rpm for 10 min. Pour off the supernatant, it is not necessary to aspirate. Completely suspend the cells in 250 ml glycerol and re-centrifuge
29. Pour off the supernatant and suspend the cells in the residual glycerol by pipetting up and down

PAUSE point:

30. At this point you can electroporate or freeze the cells away. To freeze, add 100 microliters of the culture to microcentrifuge tubes on ice. Once you have used all of the culture, transfer the tubes to dry ice for 10 minutes or into liquid nitrogen. Once the cultures are frozen, transfer them to a -80°C freezer. The cultures should be stable for >6 months

Transformation of electrocompetent YYdCas9 cells with a CRISPRi library TIMING 2 days

31. Place Cuvettes at -20°C, turn on the electroporator and set it to 1.7-2.5 kV (optimize for strain), 200 ohms and 25 µF
32. Place recovery SOC in 37°C water bath or incubator
33. Pre-warm LB-antibiotic plates containing Amp at 37°C

34. Thaw cells on ice for 10 min or use freshly made cells
35. Place appropriate number of microcentrifuge tubes and 1 mm-electroporation cuvettes on ice
36. Flick the tube containing cells a few times to mix and add 20 µl to the microcentrifuge tubes
37. Add 1.5 µl of the prepared Gibson assembly mix (eluted in H_2O) to the cells in the microcentrifuge tube
38. Put 1 ml recovery medium in a 2 ml tube, take it with you to the electroporator
39. Transfer the DNA-cell mixture to the cold cuvette, tap on countertop 2X, wipe water from exterior of cuvette and place in the electroporation module and press pulse
40. Immediately add 980 µl of 37°C SOC, mix by pipetting up and down once and transfer back to 2 ml tube.
41. Rotate in a 37°C incubator for 1 h
42. Plate 3*300 µl on big petri dishes and dilute the remaining 100 µl in a dilution series from (1:10-1:100000) and plate on small petri dishes
43. Incubate overnight at 37°C

PAUSE point: cells have to be incubated over night

Day 4

44. Count cells on the plates of the dilution series and calculate approximate number to check the coverage of the transformation
45. Pick at least 1/1000 of the overall number of sgRNAs of your library and send to Sanger sequencing to check whether the cloning was successful

CRITICAL step: if all picked colonies have a different guide, it should be sufficient to continue with the fitness assay and NGS, if not steps 10-45 have to be repeated

TROUBLESHOOTING

46. Pool Library: put 10 ml LB on one plate and scratch colonies off carefully with spatula
47. Pool everything into a 50 ml Falcon by using a pipette
48. Mix the pooled cells carefully and prepare cryo stocks

PAUSE point: cryo stocks can be at -80°C for years

Performing a fitness assay with a pooled CRISPRi library TIMING 2-3 days

49. Grow a pre-culture of the CRISPRi strain library from a glycerol stock: add 500 µL of the strain library to 50 ml LB+Amp, incubate for 5 hours at 37°C

50. Prepare a second pre-culture in M9+Glucose+Amp, dilute LB pre-culture 1:10000 and incubate for 13 hours
51. Use the M9 pre-culture which should be in exponential phase and inoculate at least 2 main cultures with an initial OD of 0.5 in shaking flasks (100 ml M9+Glucose+Amp, 200nM aTc)
52. Measure OD every hour and collect samples for plasmid extraction and sequencing. Collect sample which equals OD 3 (if cells are at OD 1, collect 3 ml). Centrifuge samples for plasmid preparation (RT, max speed, 5 min) and store at -20°C until plasmid extraction
53. Dilute cultures back every 2 hours to OD 0.05 with fresh and pre-warmed M9+Glucose+Amp + 200 nM aTc
54. Grow cultures for the desired time (14 hours in our case, which equals to 14 doublings)
55. After all samples were collected, extract plasmids with GeneJET Plasmid Miniprep Kit (ThermoFisher Scientific)

PAUSE point: plasmid DNA can be stored at -20°C for at least one year

Next-generation Illumina sequencing TIMING 2-3 days

Amplicon amplification and addition of indices

56. Amplify the 300 bp fragment including the sgRNA sequence and flanking regions using Q5 polymerase (New England Biolabs, USA):
    - Template: 150 ng purified library plasmids as template
    - 50 µL PCR reaction with following settings: 98 °C for 10 s, 12 cycles of 98 °C for 10 s, 65 °C for 30 s and 72 °C for 15 s; final extension at 72 °C for 5 min
    - Oligos: NGS_F1 and NGS_R1
57. Purification of the PCR products with NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, Germany)

PAUSE point: amplified DNA can be stored at -20°C for at least one year

58. Amplification of the 300 bp fragment and addition of different pairs of indexes (i5 and i7) to each amplicon using Phusion High- Fidelity DNA Polymerase (New England BioLabs, USA)
    - Template: 4 ng of purified library from step 46.
    - 20 µL PCR reaction with following settings: 98 °C for 30 s; 12 cycles of 98 °C for 10 s, 55 °C for 30 s and 72 °C for 20 s; final extension at 72 °C for 5 min
59. The PCR products were cleaned up using AMPure XP beads (Beckman Coulter): Add x µL of H2O to each well of the PCR plate or strip to have a total volume of 50 µl
60. Add 50 µl of Ampure XP beads to each well of a 96-well plate or Eppendorf cup

61. Transfer 50 µl PCR product to the plate/ Eppendorf cup and pipette to mix
62. Incubate 10 minutes at RT
63. Plate the plate/ Eppendorf cups on a magnetic stand for 5 min (liquid must become clear)
64. Discard the supernatant by pipetting
65. Wash samples by adding 190 µl of freshly prepared 80% EtOH and wait 30 sec on the magnetic stand
66. Discard all supernatant and repeat the washing step for a total of 2 washes
67. Remove plate/ Eppendorf cups from the magnetic stand and let it air dry for 2-3 min (make sure there is no EtOH left, don not over dry it)
68. Remove the plate/ Eppendorf cups from the magnetic stand and resuspend in 22 µl of H2O by mixing it thoroughly
69. Incubate for 10 min at RT
70. Place it back on the magnetic stand and wait 5 min
71. Transfer 20 µl to a new plate/ Eppendorf cup

PAUSE point: amplified DNA can be stored at -20°C for at least one year

Library preparation for NGS run

(Protocol modified after Protocol A: Standard normalization method of Illumina: MiniSeq System-Denature and Dilute Libraries Guide)

Preparation of reagents:

72. Prepare a fresh dilution of 0.1 N NaOH, combine the following reagents in a microcentrifuge tube and invert the tube to mix (use within 12 hours)
    - 900 µl ddH_2O
    - 100 µl NaOH (1.0 N)
73. Prepare hybridization buffer (comes with Illumina MiniSeq High Output Reagent Kit):
    - Remove the tube from freezer and thaw at RT
    - Once it is thawed, store at 2°C to 8°C
    - Vortex briefly before usage
74. Prepare RSB:
    - 10 mM Tris-HCl, pH 8.5
    - 0.1% Tween 20
    - Store at 2°C to 8°C

Creation of a normalized library pool:

75. Measure all samples in a Qubit 2.0 Fluorometer using Qubit dsDNA BR Assay
76. Mix 10 ng of each sample that should be sequenced in one tube
77. Dilute the pooled libraries to 10nM using RSB buffer
78. Afterwards, dilute the library pool to 1 nM in a fresh tube using RSB buffer
79. Vortex briefly and centrifuge at 280 × g for 1 min

PAUSE point: pooled library can be stored at -20°C for at least one year

Denature the library:

80. Combine the following volumes in one tube:
    - 5 µl library (1nM)
    - 5 µl NaOH (0.1 N)
81. Vortex briefly and centrifuge at 280 × g for 1 min
82. Incubate at RT for 5 min
83. Add 5 µl 200 mM Tris-HCl, pH 7.0
84. Vortex briefly and centrifuge at 280 × g for 1 min

Dilution of library to loading concentration:

85. Add 985 µl of prechilled Hybridization buffer to the denatured library (total vol. 1 ml at 5 pM)
86. Vortex briefly and centrifuge at 280 × g for 1 min
87. Transfer 180 µl diluted library to a new microcentrifuge tube
88. Add 320 µl prechilled Hybridization buffer (total vol. 500 µl at 1.8 pM)
89. Vortex briefly and centrifuge at 280 × g for 1 min

Denature and dilute PhiX (modified after Denature and Dilute PhiX Control in Illumina guide):

90. Thaw a tube of 10 nM PhiX stock
91. Combine 15 µl RSB and 10 µl of PhiX (10nM) in one tube (final vol. 25 µl at 4 nM)
92. Vortex briefly and pulse centrifuge (can be stored at -25°C to -15°C for up to 3 months)
93. Combine in a microcentrifuge tube:
    - 5 µl PhiX (4 nM)
    - 5 µl NaOH (0.1 N)
94. Vortex briefly and pulse centrifuge
95. Incubate at RT for 5 min
96. Add 5 µl 200 mM Tris-HCl, pH 7.0
97. Vortex briefly and centrifuge at 280 × g for 1 min

98. Add 985 µl prechilled Hybridization buffer to the tube of denatured PhiX (total vol. 1 ml at 20 pM)
99. Combine in a microcentrifuge tube (total vol. 500 µl at 1.8 pM)
    - 45 µl of denatured PhiX
    - 455 µl of prechilled hybridization buffer
100. Invert to mix and centrifuge at 280 × g for 1 min (can be stored at -25°C to -15°C for up to 2 weeks)

Combine library and PhiX control:

101. Mix in a microcentrifuge tube:
    - 250 µl PhiX (1.8 pM)
    - 500 µl denatured library (step 78.)
    - Results in a 50% PhiX spike-in
102. Loading the MiniSeq machine with the finished library:
    - See Illumina Miniseq System Guide
    - After reagent cartridge is thawed, 500 µl of the prepared library-PhiX mix are loaded

Data Analysis of pooled CRISPRi screening TIMING 1 day

103. Fastq data file of Single-end reads were read out by custom Matlab script
104. Single-end reads are matched to sgRNA guide sequences of the CRISPRi reference library and read counts are calculated
105. Read counts of each sgRNA are normalized to the total number of read counts per sample to achieve frequencies of sgRNAs
106. To calculate fold-changes, frequencies are normalized to first time point (t= 0 h)

Optional experiments

FACS screening with CRISPRi library TIMING 4 days

- Before performing step 20, transform your base train (YYdCas9) with a suitable fluorescent reporter plasmid
- Then continue the procedure until step 48

49a. Grow a pre-culture of the CRISPRi strain library from a glycerol stock: add 500 µL of the strain library (cryo stock) to 50 ml LB+Amp, incubate for 5 hours at 37°C

50a. Prepare a second pre-culture in M9+Glucose+Amp, prepare various dilutions of the LB pre-culture to have an exponential culture on the next day and incubate for 13 hours

Dilution series of the library:

1:5.000 33 uL (library) in 165 ml (medium M9+Glc+antibiotics)

1:10.000 65 ml (of the dilution before) + 65 ml (fresh medium)

1:50.000 30 ml (of the dilution before) + 120 ml (fresh medium)

1:100.000 50 ml (of the dilution before) + 50 ml (fresh medium)

PAUSE point: cells have to be incubated for 13 hours

51a. Use the M9 pre-culture which should be in exponential phase and inoculate a main culture with an initial OD of 0.02 in a shaking flask (100 ml M9+Glucose+antibiotics, 200nM aTc in a one litre shake flask), additionally grow the library in a lower volume without inducer (control for FACS)

CRITICAL step: we start with a low OD of 0.02, so we do not have to back-dilute the library if the incubation time does not exceed 6-7 hours. Dependent on the experiment and incubation time, the library must be diluted to keep it in exponential phase!

52a. Start the FACS machine according to the supplier's instructions to minimize the time the cells are stored in PBS

53a. After an appropriate incubation time, centrifuge and wash 10 ml of the libraries 2 times with the same volume PBS buffer (RT, 4000rpm, 10 min), dilute 10-fold if OD~0.5

54a. Measure the library in the FACS machine induced and not induced, measure the fluorescence with suitable laser settings and screen at least 100000 cells

55a. Perform a cell-sorting after setting gates appropriately for what you want to sort (highest GFP levels)

- Fill the tube for the sorted cells with 1 ml M9 (without glucose to prevent cell growth as cells are kept at RT)

TROUBLESHOOTING

Some FACS systems have the option to directly sort into wells of for example 96-well plates, however many cells will not survive the sorting and wells will therefore stay empty.

56a. Streak out different dilutions on LB plates with antibiotics 100 µl, the remaining 900 µl are equally distributed and plated on 3 big petri dishes

PAUSE point: cells have to be incubated over night

57a. On the next day, count the cells to calculate the survival rate and pick single colonies for further investigations, for example into 96-well plates

TROUBLESHOOTING

- Further experiments: sequencing, plate reader measurements (fluorescence and OD), metabolomics, proteomics, microscopy

Timing

All times mentioned in the table below only cover the here described workflow in *E. coli* (**Figure 3**). Depending, on the organism, systems and machines used, knowledge of the experimenter, growth conditions and number of samples, the times can differ.

- sgRNA design (step 1-9) 1 day
- cloning CRISPRi library (step 10-19) 1-2 days
- transformation of the CRISPRi library (step 20-48) 4 days
- fitness assay (step 49-55) 2-3 days
- NGS (step 56-102) 2-3 days
- Data analysis (step 103-106) 1 day
- FACS screening of CRISPRi library (step 49a-57a) 4 days

Troubleshooting

Step	Problem	Solution
15.	The amplification of the backbone might lead to smear on the gel	having the plasmid linearized with a restriction enzyme as a control also on the gel so that you can identify the correct band
18. and 45.	There is still backbone left in your sample	perform an overnight DpnI digest of the backbone following the supplier's instructions and redo the cloning
45.	No colonies grew over night	1) Test your electrocompetent cells with another plasmid, which you know works 2) vary the molar ratios of the Gibson assembly mix 3) increase the number of Gibson Assembly reactions and transformations
55a.	It is hard to discriminate between induced and not induced cultures	1) plot the measured cultures in different ways, for example use a scatter plot instead of a histogram to see distribution of single cells 2) use different laser settings 3) try to find an appropriate positive control, which enables to set a get where cells of interest should be located 4) choose another reporter plasmid, which has an increased signal upon induction
57a.	Low survival rate after cell sorting	1) try to decrease the time that cells are stored in PBS 2) try to decrease sorting time by changing gates

Figures and legends

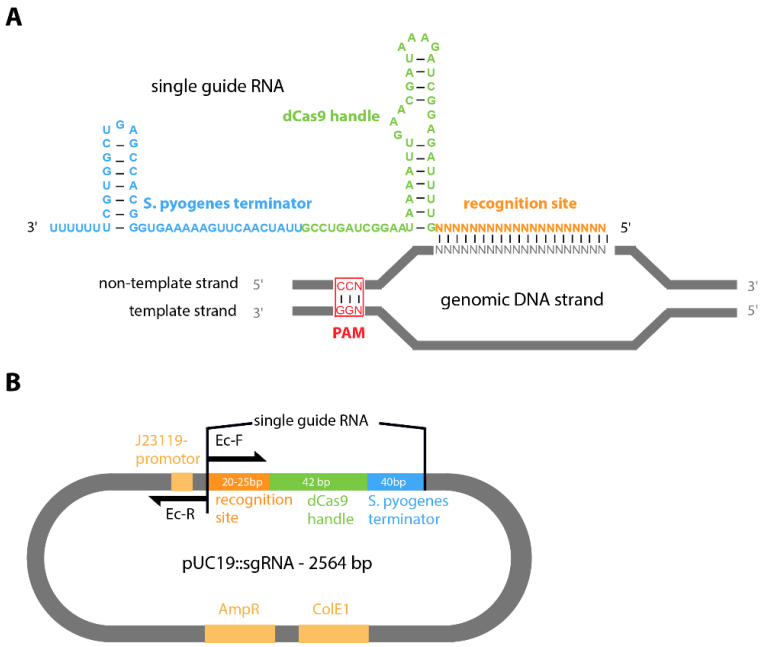


Figure 1. Schematic of the sgRNA and the pgRNA plasmid. (A) Sequence of the sgRNA used in this protocol to target a gene in the genome of *E. coli* on the non-template strand. The sgRNA includes the recognition site (20-25 bp), the dCas9 handle to bind the dCas9 protein and the terminator derived from *S. pyogenes*. Also shown is the PAM site (NGG), located upstream of the recognition site of the sgRNA. (B) pgRNA plasmid (Addgene #44251), harboring the sequence for the sgRNA constitutively expressed by promotor J23119. The vector carries an ampicillin resistance cassette (AmpR) and ColE1 origin. The whole vector can be amplified with the oligonucleotides Ec-F and Ec-R, while Ec-F can be used to exchange the sgRNA sequence (modified after Schramm, 2015; unpublished).

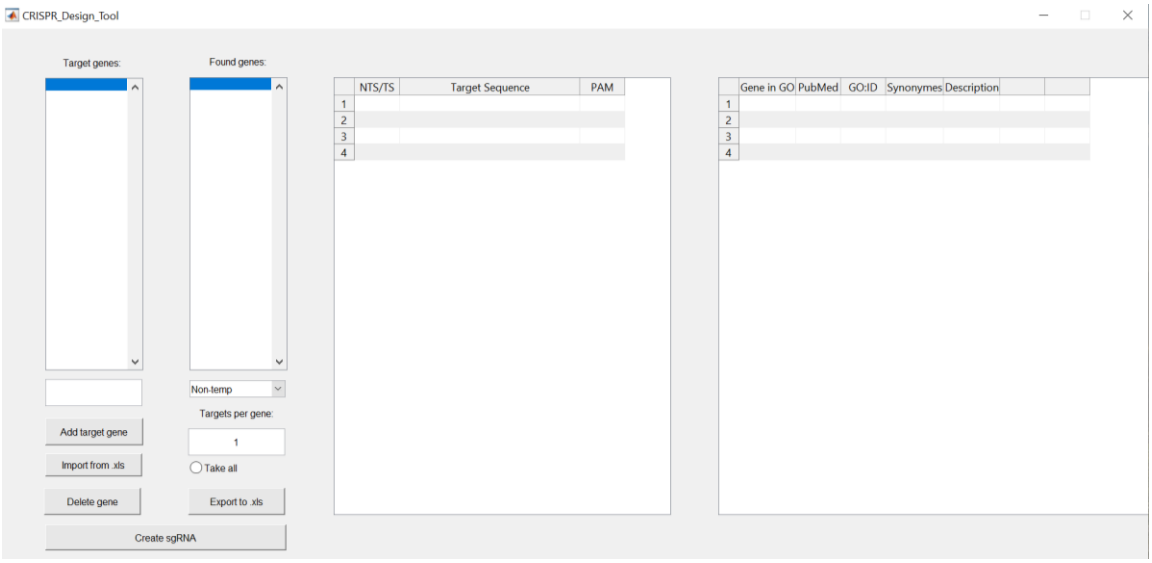


Figure 2. Window of sgRNA design tool. The CRISPR design tool provides a surface to design sgRNA for targeting genes in *E. coli*. Genes can either be selected by pressing „add target gene“ or by uploading a table „import from .xls“. The targeting strand can be selected and also how many targets per gene should be shown. “Create sgRNA” will show a list of possible sgRNA sequences on the right side of the window and the list can be exported by pressing “export to .xls”.

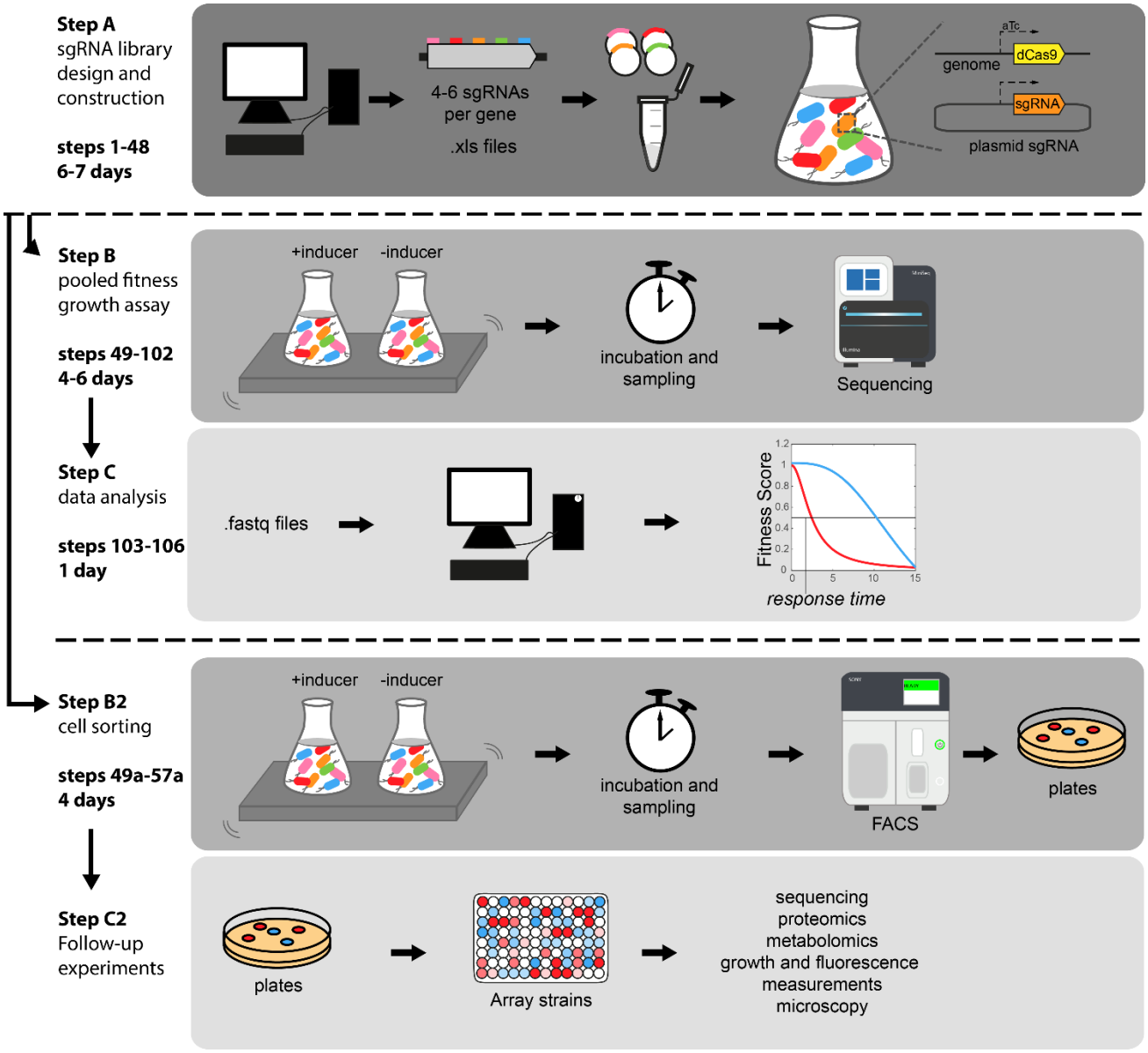


Figure 3. Outline of the workflow and timeline to work with a CRISPRi library. The here described steps can be used to either perform a fitness assay with a CRISPRi library (step A-B-C) or to enrich single cells, which show a certain phenotype, from a CRISPRi library via FACS (step A-B2-C2). Both options include step A: the design and construction of a CRISPRi library in the YYdCas9 strain, having dCas9 encoded on the genome and the sgRNA on the pgRNA plasmid. This step is either followed by step B and C, performing a fitness assay, NGS and data analysis or step B2 and C2, performing growth and sampling of the CRISPRi library, FACS sorting, plating and arraying of strains, followed by additional experiments. All steps indicate which steps of the protocol have to be followed and how long each step takes.

Author Contributions

MK performed experiments with the pooled CRISPRi library, analyzed data and co-wrote the manuscript. TS set-up Matlab scripts to design sgRNAs and co-wrote the manuscript. DB constructed the pooled CRISPRi library. MK, JVGF and LR performed Illumina sequencing. HL conceived the study, analyzed data, wrote Matlab scripts and co-wrote the manuscript.

Acknowledgements

We thank J Elf for providing the YYdCas9 strain. This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 715650, ERC Starting Grant MapMe). MK acknowledges funding from the IMPRS graduate school for environmental, cellular, and molecular microbiology from the Max Planck Society. LR acknowledges funding from DFG SPP2141.

Competing Interests

The authors declare no competing interests.

References

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular Systems Biology *2*.

Beuter, D., Gomes-Filho, J.V., Randau, L., Díaz-Pascual, F., Drescher, K., and Link, H. (2018). Selective Enrichment of Slow-Growing Bacteria in a Metabolism-Wide CRISPRi Library with a TIMER Protein. ACS Synthetic Biology *7*, 2775–2782.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res *41*, 7429–7437.

Capaldi, A.P. (2010). Analysis of Gene Function Using DNA Microarrays. In Methods in Enzymology, (Elsevier), pp. 3–17.

Cui, L., Vigouroux, A., Rousset, F., Varet, H., Khanna, V., and Bikard, D. (2018). A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9. Nature Communications *9*.

Donati, S., Kuntz, M., Pahl, V., Farke, N., Beuter, D., Glatter, T., Gomes-Filho, J.V., Randau, L., Wang, C.-Y., and Link, H. (2021). Multi-omics Analysis of CRISPRi-Knockdowns Identifies Mechanisms that Buffer Decreases of Enzymes in E. coli Metabolism. Cell Systems *12*, 56-67.e6.

Garg, A., Lohmueller, J.J., Silver, P.A., and Armel, T.Z. (2012). Engineering synthetic TAL effectors with orthogonal target sites. Nucleic Acids Research *40*, 7584–7595.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387–391.

Hannon, G.J. (2002). RNA interference. Nature *418*, 244–251.

Hawkins, J.S., Silvis, M.R., Koo, B.-M., Peters, J.M., Osadnik, H., Jost, M., Hearne, C.C., Weissman, J.S., Todor, H., and Gross, C.A. (2020). Mismatch-CRISPRi Reveals the Co-varying Expression-Fitness Relationships of Essential Genes in Escherichia coli and Bacillus subtilis. Cell Systems *11*, 523-535.e9.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc *8*, 2180–2196.

Lawson, M.J., Camsund, D., Larsson, J., Baltekin, Ö., Fange, D., and Elf, J. (2017). *In situ* genotyping of a pooled strain library after characterizing complex phenotypes. Molecular Systems Biology *13*, 947.

Liu, X., Kimmey, J.M., Matarazzo, L., de Bakker, V., Van Maele, L., Sirard, J.-C., Nizet, V., and Veening, J.-W. (2021). Exploration of Bacterial Bottlenecks and Streptococcus pneumoniae Pathogenesis by CRISPRi-Seq. Cell Host & Microbe *29*, 107-120.e6.

Monk, J.M., Lloyd, C.J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., et al. (2017). iML1515, a knowledgebase that computes Escherichia coli traits. Nature Biotechnology *35*, 904–908.

van Opijnen, T., Bodi, K.L., and Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nature Methods *6*, 767–772.

Peters, J.M., Koo, B.-M., Patino, R., Heussler, G.E., Hearne, C.C., Qu, J., Inclan, Y.F., Hawkins, J.S., Lu, C.H.S., Silvis, M.R., et al. (2019). Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. Nature Microbiology *4*, 244–250.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell *152*, 1173–1183.

Qu, J., Prasad, N.K., Yu, M.A., Chen, S., Lyden, A., Herrera, N., Silvis, M.R., Crawford, E., Looney, M.R., Peters, J.M., et al. (2019). Modulating Pathogenesis with Mobile-CRISPRi. Journal of Bacteriology *201*.

Québatte, M., Christen, M., Harms, A., Körner, J., Christen, B., and Dehio, C. (2017). Gene Transfer Agent Promotes Evolvability within the Fittest Subpopulation of a Bacterial Pathogen. Cell Systems *4*, 611-621.e6.

Rousset, F., Cui, L., Siouve, E., Becavin, C., Depardieu, F., and Bikard, D. (2018). Genome-wide CRISPR-dCas9 screens in E. coli identify essential genes and phage host factors. PLOS Genetics *14*, e1007749.

Rousset, F., Cabezas-Caballero, J., Piastra-Facon, F., Fernández-Rodríguez, J., Clermont, O., Denamur, E., Rocha, E.P.C., and Bikard, D. (2021). The impact of genetic diversity on gene essentiality within the Escherichia coli species. Nature Microbiology *6*, 301–312.

Segal, D.J., and Barbas, C.F. (2001). Custom DNA-binding proteins come of age: polydactyl zinc-finger proteins. Current Opinion in Biotechnology *12*, 632–637.

Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR–Cas9. Nature Reviews Genetics *16*, 299–311.

Shields, R.C., Walker, A.R., Maricic, N., Chakraborty, B., Underhill, S.A.M., and Burne, R.A. (2020). Repurposing the Streptococcus mutans CRISPR-Cas9 System to Understand Essential Gene Function. PLOS Pathogens *16*, e1008344.

Todor, H., Silvis, M.R., Osadnik, H., and Gross, C.A. (2021). Bacterial CRISPR screens for gene function. Current Opinion in Microbiology *59*, 102–109.

Towbin, B.D., Korem, Y., Bren, A., Doron, S., Sorek, R., and Alon, U. (2017). Optimality and sub-optimality in a bacterial growth law. Nature Communications *8*.

Wang, T., Guan, C., Guo, J., Liu, B., Wu, Y., Xie, Z., Zhang, C., and Xing, X.-H. (2018). Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. Nature Communications *9*.

Yoo, S.M., Na, D., and Lee, S.Y. (2013). Design and use of synthetic regulatory small RNAs to control gene expression in Escherichia coli. Nature Protocols *8*, 1694–1707.

Chapter 4

Systematic identification of metabolites regulating gene expression in *E. coli*

Martin Lempp[1], Niklas Farke[1], **Michelle Kuntz[1]**, Sven Freibert[2], Roland Lill[2], Hannes Link[1]

[1] Max Planck Institute for Terrestrial Microbiology, Marburg 35043, Germany

[2] Institut für Zytobiologie und Zytopathologie, Philipps-Universität Marburg, 35033 Marburg, Germany

This chapter is written in manuscript style and was published in Nature Communications 10, 4463 (2019). My contribution to this work included the implementation of a purification protocol for proteins and their purification for the execution of binding assays.

Abstract

Metabolism controls gene expression through allosteric interactions between metabolites and transcription factors. These interactions are usually measured with *in vitro* assays, but there are no methods to identify them at a genome-scale *in vivo*. Here we show that dynamic transcriptome and metabolome data identify metabolites that control transcription factors in *E. coli*. By switching an *E. coli* culture between starvation and growth, we induce strong metabolite concentration changes and gene expression changes. Using Network Component Analysis, we calculate the activities of 209 transcriptional regulators and correlate them with metabolites. This approach captures, for instance, the *in vivo* kinetics of CRP regulation by cyclic-AMP. By testing correlations between all pairs of transcription factors and metabolites, we predict putative effectors of 71 transcription factors, and validate five interactions *in vitro*. These results show that combining transcriptomics and metabolomics generates hypotheses about metabolism-transcription interactions that drive transitions between physiological states.

Introduction

Transcriptional regulation of metabolism is well characterized regarding the canonical flow of genetic information, which considers how transcription modulates the abundance of enzymes, and thereby metabolic flux and metabolites[1–4]. In reverse, metabolites convey information back to the transcription network by directly or indirectly interacting with a transcription factor (TF) [5–9] (**Fig. 1a**). In *Escherichia coli*, for example, the amino acid arginine allosterically regulates the activity of ArgR, which is a TF that controls genes involved in arginine biosynthesis, but the total regulon includes more than 400 genes[10]. Allosteric TF regulation allows a cell to tune gene expression depending on its metabolic state and theory shows that this regulation is robust and predictable by models[11]. An important consequence of allosteric TF regulation is that metabolites are not just biomass building blocks but they serve as internal signals with the potential to actively drive transitions between different physiological states.

It is largely unexplored which of the many intracellular metabolites interact with TFs[12,13], yet many transcriptional regulators are expected to bind a small molecule[5]. Currently, a major limitation to fill this gap of knowledge is the lack of methods to identify the most functionally relevant metabolite-TF interactions that control gene expression *in vivo*. Detection of physical interactions between metabolites with transcriptional regulators is mainly based on *in vitro* assays, which are low-throughput, feasible for only certain compounds and combinatorial effects cannot be assayed[14]. An alternative approach is to probe protein structural changes with proteomics, which can detect binding of a single metabolite across thousands of proteins in cell extracts, but this approach cannot decipher unspecific binding from interactions that are functional *in vivo*[15]. An *in vivo* approach has been proposed, which searches for correlations in metabolomics data and data from fluorescent transcriptional reporters. This method could indeed recover few of the known metabolites that are relevant for gene regulation of central carbon metabolism in *E. coli*[16].

Here, we measure the *E. coli* transcriptome and metabolome changes during a 20 h dynamic transition, and show that integrating these two data-types generates hypotheses about metabolite-TF interactions that may have functional relevance *in vivo*. We also construct a metabolite-TF network for *E. coli* from the literature and databases, and show that our approach recovers more than 50% of the interactions in this network that were covered by our data. Moreover, we validate five predicted interactions with *in vitro* binding assays, i.e. lysine-ArgR, tyrosine-TrpR, glutamate-SgrR, tryptophan-SoxR and dihydroxyacetone phosphate-DhaR, showing that our methodology generates physiologically meaningful results.

Results

Switching *E. coli* between growth and carbon starvation

We used a 1 Liter bioreactor to switch the culture conditions of *E. coli* between 6 h growth, 12 h carbon starvation and 2 h growth resumption. First, cells grew on minimal medium with glucose and when the culture reached an optical density (OD) of 2, we transferred cells to minimal medium without carbon source. This rapid medium exchange caused an immediate growth arrest and cells starved for a period of 12 h (**Fig. 1b**). After 12 h we added again glucose to the culture and within 10 min cells resumed growing exponentially (**Fig. 1b**). Apart from the fast growth resumption, also oxygen uptake and CO_2-production increased rapidly upon glucose addition and reached the same rate as before starvation (**Supplementary Fig. 1**). Thus, physiological parameters like growth and respiration change in a fast and reversible fashion when *E. coli* cells enter and exit carbon starvation. Next, we investigated metabolism and transcription during the growth-starvation-growth switch, and measured the concentration of 123 metabolites by LC-MS/MS (**Supplementary Fig. 2**) and 4242 transcripts by RNA sequencing (**Fig. 1c** and **Supplementary Data 1, Supplementary Data 2**). In total, we collected transcriptomics samples at 29 and metabolomics samples at 35 different time points in duplicates from a single bioreactor (**Supplementary Data 3**), with average errors of 18% for metabolites and 16% for transcripts. Only 8% of metabolites and 17% of the transcripts did not change significantly in either phase. To explore global dynamics of the metabolome and transcriptome data, we grouped each data set into four clusters (hierarchical clustering, z-score normalized). The clusters showed that the largest group of metabolites (63%) and transcripts (68%) decreased during the starvation phase and increased during the exit-phase (Cluster A, **Supplementary Fig. 3**). This group included intermediates in glycolysis like fructose-bisphosphate, dihydroxyacetone phosphate and acetyl-CoA, as well as the nucleotide-triphosphates ATP and GTP (**Fig. 2**). Another group of metabolites and transcripts accumulated during the first 4-6 hours into starvation, such as the amino acids lysine and phenylalanine that originate from degradation of proteins[17] (Cluster C, **Supplementary Fig. 3**). Similarly, accumulation of nucleotide derivatives like hypoxanthine was presumably a consequence of RNA degradation (**Fig. 2**). These data indicate that starving *E. coli* cells catabolize RNA and proteins during the early phase of starvation, an interpretation that is consistent with the relatively high production of CO_2 in this phase (**Supplementary Fig. 1**), and also with the expression of genes involved in RNA, protein and glycogen degradation processes (**Supplementary Fig. 4**). Notably, expression of genes in glycogen degradation preceded the expression of genes in RNA and protein degradation (**Supplementary Fig. 4**), confirming that glycogen functions as short-term energy storage. After switching cells back to glucose, 95% of the metabolites and 78% of transcripts

reached the same steady state levels that they had before the starvation phase. However, for many metabolites and transcripts it took over 1 h until they reached a steady state, thus indicating extensive regulation during the exit phase.

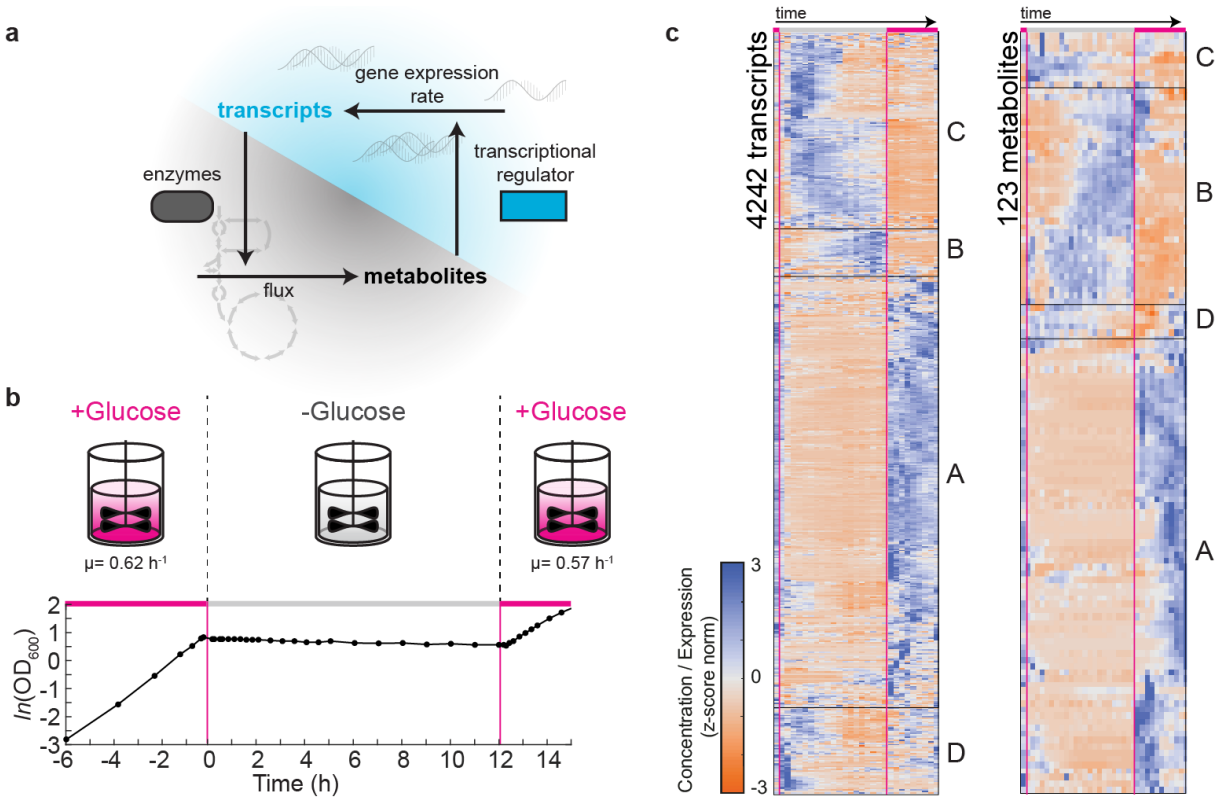


Figure 1. Dynamic metabolomics and transcriptomics during the growth-starvation switch in *E. coli*. a) Schematic of the mutual feedback between metabolism and transcription. Transcription regulates enzyme levels, which affect flux and metabolite concentrations. Metabolite concentrations regulate gene expression by allosteric interactions with transcriptional regulators. **b)** Growth of *E. coli* during the switch between growth, carbon starvation and back to growth. Cells were cultivated in a 1 L bioreactor on glucose minimal medium to an OD of 2 and then the medium was exchanged to minimal medium without carbon source. After 12 h carbon starvation glucose was added to the culture. μ is the growth rate calculated by linear regression in the first and second growth phase. **c)** Dynamic transcriptomics and metabolomics data measured at 29 and 35 time points, respectively. The first measurement was before the switch to starvation, 19 samples were collected during starvation and 9 samples during exit from starvation. Shown are z-score normalized transcript levels (in TPM) of 4242 genes and the z-score normalized concentration of 123 metabolites. Blue indicates high expression/concentration; orange indicates low expression/concentration. Data is grouped by hierarchical clustering and the four largest clusters are indicated as clusters A-D (average cluster dynamics are shown in **Supplementary Fig. 2**). Source data are provided as a Source Data file.

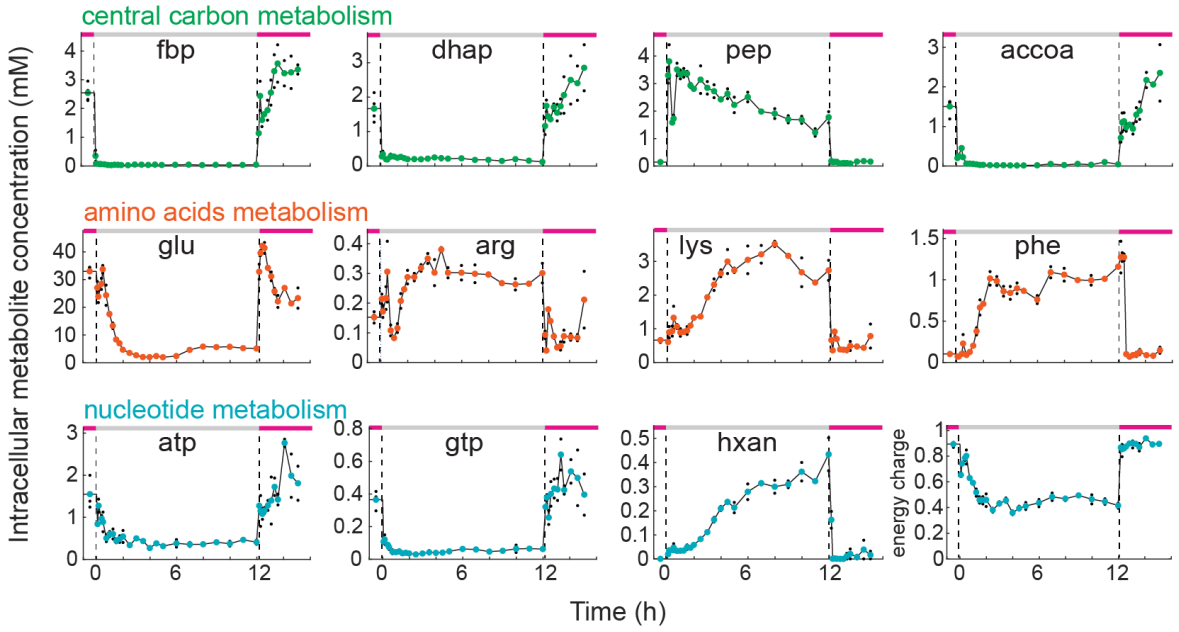


Figure 2. Examples of metabolite concentration changes during the growth-starvation-growth switch. Absolute concentration of metabolites in central carbon metabolism (green), amino acid metabolism (orange) and nucleotide metabolism (blue). The dashed lines indicate the starvation and growth phases. Black dots show concentrations of two replicates per time point (four at the first time point), colored dots are the mean. The energy charge is calculated from the concentration of ATP, ADP and AMP. (Abbreviations: fbp – fructose-1,6-bisphosphate, dhap – dihydroxyacetone phosphate, pep – phosphoenolpyruvate, accoa – acetyl-coenzyme A, glu – glutamate, arg – arginine, lys – lysine, phe – phenylalanine, atp – adenosine triphosphate, gtp – guanosine triphosphate, hxan – hypoxanthine) Source data are provided as a Source Data file.

Integrating metabolomics and transcriptomics data

To identify metabolites that are potential regulators of gene expression during the growth-starvation-growth switch, we searched for correlations between dynamics of metabolites and transcripts. Because metabolites modulate transcription through allosteric TF regulation, we sought to determine the activity of TFs and other regulators like σ^{70} and σ^{S}. The relationship between transcriptional regulators and their target genes is well-characterized in *E. coli*, in the form of a transcription regulation network[18]. A well-mapped transcription regulation network allowed us to infer activities of transcriptional regulators from measured gene expression profiles using algorithms like Network Component Analysis (NCA)[19,20]. The NCA algorithm estimates activity profiles of transcriptional regulators, which minimize the error between theoretical and measured gene expression profiles (for 2167 genes that are mapped to 209 transcriptional regulators in the *E. coli* transcription regulation network). In total we performed 100 searches with the NCA algorithm, each with a different randomized initial condition, such that we obtained means and confidence intervals for activity profiles of the 209 transcriptional regulators (**Fig. 3a, Supplementary Data 5**). These 209 activity profiles were able to reproduce 75% of the transcript dynamics and were consistent with the expected responses of transcriptional regulators during starvation and growth[21]. For example, σ^{70}, the major

sigma factor during exponential growth, was deactivated upon entry to starvation, and the stress response regulator σ^S was immediately activated (**Supplementary Fig. 5**).

Allosteric regulation of a TF by a metabolite is often described by Hill-type kinetics[11], which assumes a sigmoidal relationship between TF activity and the concentration of an effector metabolite. In a canonical example of this regulation, the secondary messenger cyclic-AMP activates CRP, which is a global TF in *E. coli*[22,23]. On the basis of Hill kinetics, we tested how well the measured cyclic-AMP concentration predicts the activity profile of CRP (**Fig. 3b**). Cyclic-AMP and CRP activity revealed indeed a Hill-type relationship with an activation constant (K_H) of 39 µM, which is very close to the *in vitro* determined value of 27 µM[23]. Thus *in vivo* metabolite and transcript data identifies the existence of the known interaction between cyclic-AMP and CRP, and additionally captures the underlying kinetics of allosteric TF regulation. Another well-known metabolite-TF pair is tryptophan and the repressor of the tryptophan operon (TrpR), which also showed Hill-type kinetics, and the *in vivo* K_H of 355 µM was again relatively close to the *in vitro* value of 160 µM[24] (**Fig 3c**).

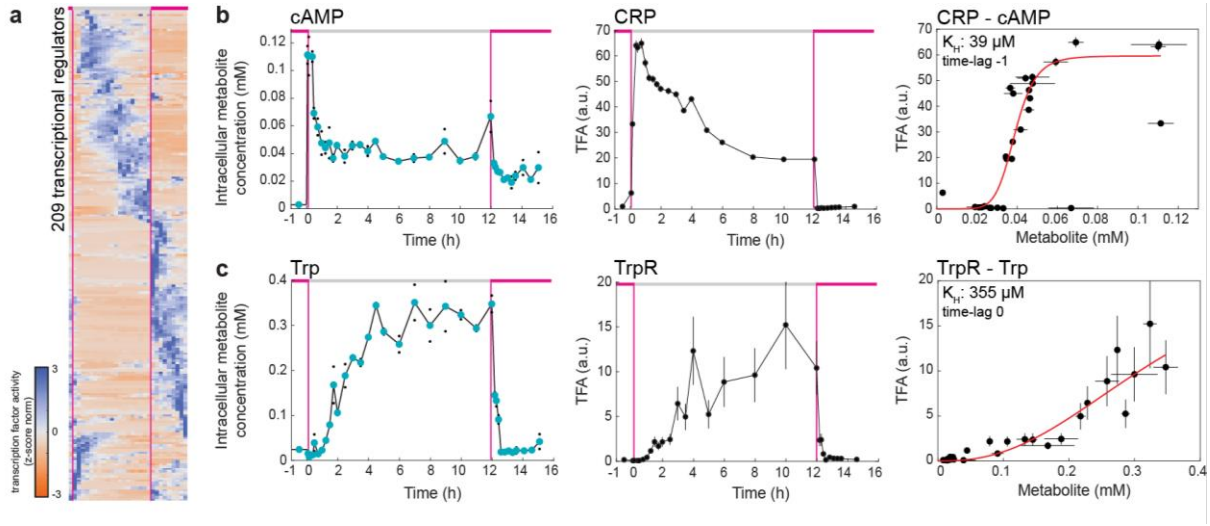


Figure 3. Metabolite levels and transcription factor activities recover regulatory interactions. a) Z-score normalized activity of 209 transcriptional regulators. Blue indicates high activity; orange indicates low activity. Activity was estimated from the transcriptome data with Network Component Analysis (NCA). **b)** Dynamics of cyclic-AMP concentration (left) and CRP activity (middle) during the growth-starvation-growth switch. Correlation of cyclic-AMP and CRP activity (right). The red line is the best fit of a Hill function (K_H = 39 µM). The correlation is shown with a time-lag of one data point to account for a time-delay between activity changes of CRP and changes in gene expression. **c)** Same correlation analysis as in **b)** for tryptophan (Trp) and the activity of TrpR, the repressor of the tryptophan operon. For cAMP and tryptophan, black dots show two replicates per time point, and colored dots are the mean. Error bars of transcription factor activity indicate the 95% confidence interval of n = 100 randomized estimates with NCA. Source data are provided as a Source Data file.

Next, we wondered how many of the known metabolite-TF interactions are covered by our data, and whether they show a Hill-type relationship. Therefore, we first constructed a "literature network" of known metabolite-TF interactions by mining RegulonDB[18], the EcoCyc database[25] and the

Allosteric Database[26]. This literature metabolite-TF network included in total 134 interactions between 87 TFs and 106 metabolites **(Supplementary Fig. 6**). 41% of the interactions are activating, 38% inhibiting and for 21% it is not known whether the metabolite inhibits or activates the TF. Our data covered interactions for 21 out of the 87 TFs, and 12 of them correlated with at least one of the known regulatory metabolites (**Fig. 4a,** Pearson's correlation coefficient $R^2 > 0.75$). Thus, our data recovered known interactions in more than 50% of the cases, and in each of these cases the correlation correctly reflected, whether the metabolite activates or inhibits the TF. In case of NadR and ExuR, our data suggests that they are inhibited by ATP and lysine, respectively.

Mapping metabolism-transcription interactions systematically

A problem of the correlation analysis was that several metabolites correlated with the activity of a TF, resulting in many false positives (**Fig. 4b**). The large number of false positives is mainly caused by metabolites that have similar dynamics. The same problem was previously reported for a multi-omics analysis of yeast metabolism, which searched for correlations between metabolites and fluxes[27]. In this study, correlations between metabolites caused also many false positives, and including prior knowledge about metabolic flux regulation solved the problem. Here, we could not adapt such an approach, due to the limited information about allosteric TF regulation. Instead, we reduced the number of putative interactions by using a distance criterion for metabolites and TFs: metabolite-TF pairs were only considered, if at least one target-gene of the TF encodes an enzyme that participates in the same metabolic subsystem as the metabolite or if the metabolite is a substrate or a product. The hypothesis behind this distance criterion is that metabolites are more likely to regulate genes that are involved in their own biosynthesis. This assumption is supported by a recent study in cancer cells, which showed that metabolite-gene pairs have a higher correlation when they are close in the metabolic network[28]. We observed a similar proximity of metabolite-TF interactions in our literature network, because more than 80% of these interactions have a small distance in the *E. coli* genome-scale metabolic model[29] (**Supplementary Fig. 6**). We then applied the distance criterion to our data and only considered metabolite-TF pairs that fulfilled the distance criterion (black dots in **Fig. 4b**). For the 12 known metabolite TF-interactions that showed a Hill-type relationship, 11 fulfilled the distance criterion, and only the interaction between ExuR and lysine was rejected. The advantage of the distance filter was that it reduced the number of highly correlating metabolites from an average of 34 metabolites per TF to an average of 9 (**Fig. 4b**).

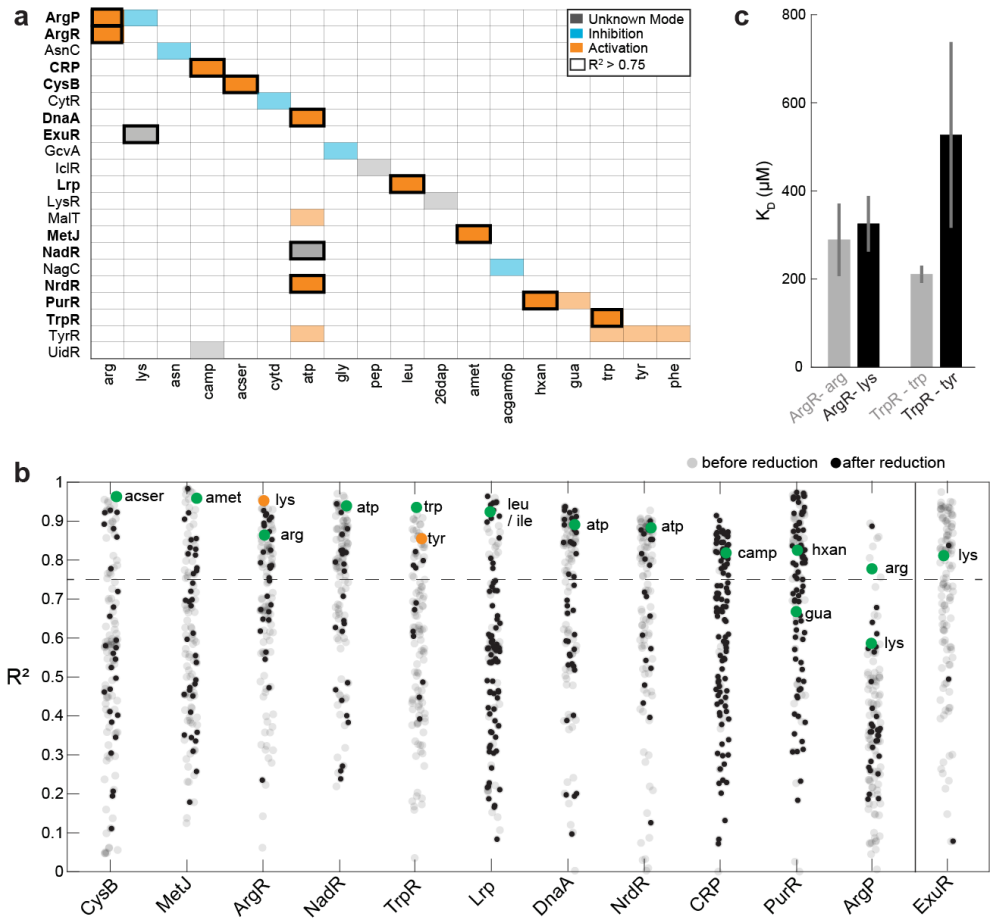


Figure 4. Identification and validation of known metabolite-transcription factor interactions. a) Metabolite-TF interactions that are described in the literature and covered by the metabolome data in **Fig. 1c** and the TF activities in **Fig. 3a**. Shown are 21 transcription factors (rows) and their respective effector metabolites (columns). Orange indicates an activation of the TF by the metabolite, blue indicates an inhibition, and grey indicates that the mode is unknown. Metabolite-TF pairs that show Hill-type kinetics in the metabolome and transcriptome data ($R^2 > 0.75$) are indicated with a box. **b)** Correlation coefficients of 12 transcription factors, which had activity profiles that correlated with at least one of the known effector metabolites (green dots). Grey and black dots are correlation coefficients with all other 123 measured metabolites. Black dots are metabolites that fulfill the distance criterion (same metabolic subsystem or product/substrate of a target-gene). Grey dots are metabolites that are rejected by the distance criterion. For ExuR the distance criterion excluded the known effector lysine. Lysine and tyrosine are indicated in orange for ArgR and TrpR. **c)** *In vitro* measured dissociation constants (K_D) of ArgR and TrpR with the known effectors (arginine and tryptophan) and the predicted additional effectors (lysine and tyrosine). Binding was measured with His-tag purified ArgR and TrpR using micro scale thermophoresis (MST). Error bars show the 95% confidence interval of K_D estimates, which are based on fitting n = 9 MST assays (proteins purified three times, each measured in three MST assays). MST data are shown Supplementary Figure 9. (Abbreviations: arg – arginine, lys – lysine, asn – aspartate, acser – O-acetyl-Serine, cytd – Cytidine, atp – adenosine triphosphate, gly – glycine, pep – phosphoenolpyruvate, leu – leucine, 26dap – diaminopimelic acid, amet – S-adenosylmethionine, acgam6p – N-acetyl-glucosamine phosphate, hxan – hypoxanthine, gua – guanine, trp – tryptophan, tyr – tyrosine, phe – phenylalanine) Source data are provided as a Source Data file.

Among the false positives that remained after the distance filter were lysine-ArgR and tyrosine-TrpR (orange dots in **Fig. 4b**). Because lysine and tyrosine share structural similarity with the known allosteric effectors (arginine for ArgR and tryptophan for TrpR), we tested if lysine and tyrosine are additional and previously unidentified regulators of ArgR and TrpR. Therefore, we purified the two TFs and tested binding of lysine and tyrosine *in vitro* using micro-scale thermophoresis (MST). The *in vitro* MST assays showed indeed binding of lysine and tyrosine to ArgR and TrpR, respectively, thus

validating the *in vivo* prediction (**Fig. 4c, Supplementary Fig. 9**). The *in vitro* assays also confirmed the known arginine-ArgR and tryptophan-TrpR interactions (**Fig. 4c, Supplementary Fig. 9**). Because ArgR regulates essential steps in lysine biosynthesis as well as two lysine transporters, the physiological function of the lysine-ArgR interaction is presumably a metabolic feedback that inhibits lysine production and import when lysine is abundant[30,31]. In case of TrpR, previous studies showed that deletion of TrpR, affects expression of *tyrA* in the tyrosine biosynthesis pathway[32]. Here, we show that also tyrosine is linked to TrpR, and the crosstalk between the two aromatic amino acids could potentially coordinate their biosynthesis.

Finally, we tested if we can generate hypotheses about the existence of metabolite-TF interactions in an unbiased fashion by fitting Hill functions to all pairs of metabolites and TFs. We first reduced the number of TFs from 209 to 125 by excluding: i) TFs that followed simple on-off-on dynamics, ii) TFs with poor estimates of activity profiles (confidence interval >100%), and iii) TFs that are part of two-component systems (these regulators are more likely modulated by external signals rather than internal metabolites). The remaining 125 TFs and 123 metabolites resulted in 15,375 metabolite-TF pairs, for which we tested if they show a Hill-type relationship. A total of 3067 metabolite-TF pairs (20%) showed a Hill type relationship ($R^2 > 0.75$ **Supplementary Data 6**), and by applying again the distance criterion we reduced this number to 513, which we considered as putative metabolite-TF interactions (**Supplementary Fig. 7, Supplementary Fig. 8, Supplementary Data 7**).

The putative 513 interactions included 71 TFs, and we focused on the 30 TFs that correlated only with one or two metabolites (**Supplementary Data 7**). The resulting network shows mostly interactions of TFs with metabolites from amino acid and nucleotide metabolism but also with intermediates in carbon and cofactor metabolism (**Fig. 5a**). We purified three of the identified TFs to test if they bind the predicted metabolite. *In vitro* MST assays validated that SoxR binds tryptophan, SgrR binds glutamate and DhaR binds the glycolysis intermediate dihydroxyacetone phosphate (DHAP) (**Fig. 5b, Supplementary Fig. 9**). SoxR is known to activate expression of *aroF* and *tyrA*, which encode enzymes catalyzing the first step in the biosynthesis of all aromatic amino acids (*aroF*) and the tyrosine branch (*tyrA*)[33]. By binding tryptophan, SoxR could be part of a feedback regulation circuit in aromatic amino acid biosynthesis, which reduces expression of *aroF* and *tyrA* when tryptophan levels are high. SgrR activates *alaC* that encodes a transaminase that converts glutamate and pyruvate to alpha-ketoglutarate and alanine[34]. This transaminase accounts, together with a corresponding isoenzyme, for 90% of the catalytic activity for biosynthesis of alanine in *E. coli*[35]. As our *in vivo* data shows an inhibition of SgrR by glutamate, low glutamate levels would upregulate *alaC*. Because low glutamate level brings the transamination reaction closer to thermodynamic equilibrium, an accompanying upregulation of *alaC* might provide the necessary enzymatic capacity[36]. The last new interaction is

DhaR and DHAP, a regulator of dihydroxyacetone kinases, which seems to activate in response to increasing DHAP levels[37]. As DhaR activates the dihydroxyacetone kinases, the interaction could be part of a positive feedback loop.

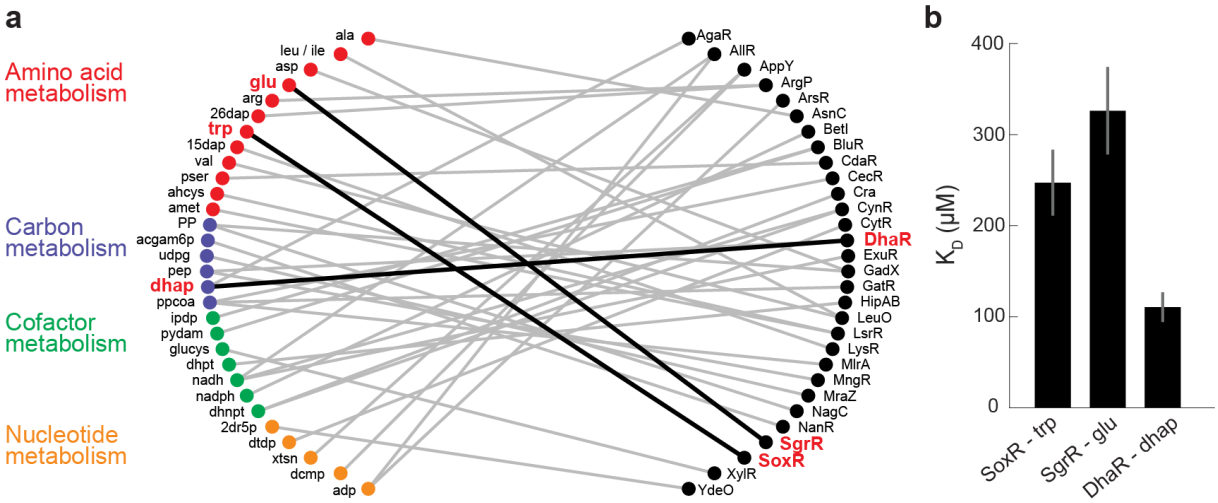


Figure 5. Identification and validation of new metabolite-transcription factor interactions. a) Network of potentially new interactions between transcription factors and metabolites. Shown are 30 transcription factors that correlated with one or maximal two metabolites (only metabolites that fulfill the distance criterion). Metabolites are colored according to the subsystems of amino acid metabolism (red), carbon metabolism (purple), cofactor metabolism (green) or nucleotide metabolism (yellow). Connections in bold black highlight new interactions that are validated *in vitro*. **b)** *In vitro* measured K_D values of the new interactions indicated in bold in **a**). Binding was measured with His-tag purified SoxR, SgrR and DhaR using micro scale thermophoresis (MST). Error bars show the 95% confidence intervals of K_D estimates, which are based on fitting n = 9 MST assays (proteins purified three times, each measured in three MST assays). MST data are shown Supplementary **Figure 9**. (Abbreviations: ala – alanine, leu/ile – leucine/isoleucine, asp – aspartate, glu – glutamate, arg – arginine, 26dap – diaminopimelic acid, trp – tryptophan, 15-dap – 1,5-diaminopimelate, val – valine, pser – 3-phosphoserine, ahcys - S-adenosylhomocysteine, amet – S-adenosylmethionine, PP – pentose phosphate, acgam6p – N-Acetyl-D-Glucosamine-P, udpg – UDP-glucose, pep – phosphoenolpyruvate, dhap – dihydroxyacetone phosphate, ppcoa – propionyl-coenzyme A, ipdp – isopentenyl diphosphate, pydam – pyridoxamine, glucys – gamma-glutamyl-cysteine, dhpt – dihydropteroate, nadh – nicotinamide adenine dinucleotide (reduced), nadph – nicotinamide adenine dinucleotide phosphate (reduced), dhnpt – dihydroneopterin, 2dr5p – deoxyribose phosphate, dtdp – thymidine diphosphate, xtsn – xanthosine, dcmp – deoxycytidine monophosphate, adp – adenosine diphosphate) Source data are provided as a Source Data file.

Discussion

In conclusion, data of the *E. coli* transcriptome and metabolome during a 20 h starvation-growth-starvation switch generated hypotheses about potential interactions between metabolites and TFs. The scale of this approach is the biggest advantage, because it allows pair-wise testing of all TFs against all metabolites. Here, we provided a first proof of principle that the combination of transcriptomics and metabolomics has a great potential to identify metabolite-TF interactions at a metabolism-wide scale. To this end we showed that many of the known metabolite-TF interactions were reflected by our data (e.g. cyclic AMP-CRP), and therefore that metabolite and gene expression data contain the information that is necessary to reconstruct metabolic-genetic networks. Moreover, we could validate five of the predicted metabolite-TF interactions with *in vitro* assays (lysine-ArgR, tyrosine-TrpR, glutamate-SgrR, tryptophan-SoxR and dihydroxyacetone phosphate-DhaR).

In our analysis, we excluded two-component systems, because they are likely responsive to external metabolites. By measuring the exo-metabolome it should be possible to identify effectors of two-component systems with the method proposed in this study. We also excluded TFs with poor estimates of activity profiles, and to include these TFs one could probe their activities with fluorescent transcriptional reporters as recently suggested[16]. Accurate information about TF activities was important for our approach because it allowed pairwise testing of Hill-type relationships between TF activities and metabolites. Here, we inferred TF activities with the NCA algorithm that requires a well-mapped transcription regulation network. While the transcription regulation network is known in *E. coli*, it is unknown for most other organisms. To overcome the need for a known transcription regulation network, the TF activities could be inferred from the transcriptome data directly without using prior knowledge about the transcription regulation network. Previous studies showed for example that machine learning methods can infer TF activities in *E. coli* based on transcriptomics data[38], and inference of regulatory metabolites with such methods was also suggested[39]. Future approaches could even consider determining TF activities and regulatory metabolites simultaneously.

The main limitation in our study was that many metabolites showed similar dynamics, which in turn caused false positive predictions of metabolite-TF interactions. The high correlation among metabolites could be a general problem in metabolomics-based inference approaches[27]. A solution for this problem is to enforce more specific metabolite concentration changes by localized perturbations of metabolism, for example by disturbing single enzymes. We anticipate that the transcriptome and metabolome of hundreds of locally perturbed metabolic states would provide sufficient information to faithfully map metabolite-TF interactions of an organism. An effective perturbation method is CRISPR interference, because of its potential to interfere with the expression of every enzyme of an organism.

A complete map of metabolite-TF interactions would advance our knowledge about the dynamic nature of metabolic regulation and enable the construction of dynamic metabolic models. Here, we focused mainly on interactions that are part of metabolic-genetic feedback circuits, because we considered the distance between TFs and metabolites. However, metabolites will not only affect transcription of genes encoding enzymes, but also affect genes involved in various other physiological processes. Understanding these long-ranging metabolite-TF interactions would dramatically increase our understanding about how metabolism drives physiological responses, e.g. to oxidative stress[40] or antibiotics[41]. Finally, there is the possibility to exploit the knowledge about metabolite-TF interactions to engineer better strains for biotechnology, e.g. by designing genetic-metabolic feedback that acts as valves in production strains[42] or growth switches[43].

Material & Methods

Strains and cultivation

E. coli BW25113 (parent strain for the Keio Collection, CGSC#: 7636) was cultivated in 1 L bioreactor with 500 mL of M9 minimal medium containing 5 g*L^-1 glucose to an optical density at 600 nm (OD) of 2. Then the culture was centrifuged at 37 °C and 1800 x g for 5 min. Pelleted cells were resuspended in M9 medium at 37°C without glucose and transferred back to the bioreactor. After 12 h, the culture was supplemented glucose to a final concertation of 5 g*L^-1 glucose. The M9 minimal medium consisted of the following components (per liter): 6 g $Na_2HPO_4 \cdot 2\ H_2O$, 3 g KH_2PO_4, 1.5 g $(NH_4)_2SO_4$, 0.5 g NaCl. The following components were sterilized separately and then added to the medium (final concentrations): 0.1 mM $CaCl_2$, 1 mM $MgSO_4$, 60 µM $FeCl_3$, 2.8 µM thiamine-HCl, and 10 mL trace salt solution. The trace salt solution contained (per liter) 180 mg $ZnSO_4 \cdot 7\ H_2O$, 120 mg $CuCl_2 \cdot 2\ H_2O$, 120 mg $MnSO_4 \cdot H_2O$, 180 mg $CoCl_2 \cdot 6\ H_2O$. The dissolved oxygen in the bioreactor was kept at 30% and pH 7 was controlled with 5 M NH_4OH and 20% H_3PO_4. The bioreactor was a BioFlo115 bioprocess system (Eppendorf, Hamburg, Germany), equipped with a pH-sensor (Mettler Toledo, Colombus, OH) and a DO-sensor (Mettler Toledo, Colombus, OH). Exhaust gas of the cultivation was analyzed by a DASGIP GasAnalyser (Eppendorf, Hamburg, Germany). The GasAnalyser was calibrated with two-point-calibration prior to the cultivation. The bioreactor cultivation was monitored with the BioCommand-Software (Eppendorf, Hamburg, Germany).

Metabolomics

For metabolomics 1 mL culture aliquots were vacuum-filtered on a 0.45 µm pore size filter (HVLP02500, Merck Millipore). Filters were immediately transferred into 40:40:20 (v-%) acetonitrile/methanol/water at -20°C for extraction. Extracts were centrifuged for 15 minutes at 11 000 xg at -9 °C. Centrifuged extracts were mixed with ^{13}C-labeled internal standard. Chromatographic separations were performed on an Agilent 1290 Infinity II LC System (Agilent Technologies) equipped with an Acquity UPLC BEH Amide column (2.1 x 30 mm, particle size 1.8 µm, Waters) for acidic conditions and an iHilic-Fusion (P) HPLC column (2.1 x 50 mm, particle size 5 µm, Hilicon) for basic conditions. We were applying the following binary gradients at a flow rate of 400 µl*min^-1: Acidic condition) 0-1.3 min: isocratic 10% A (water/formic acid, 99.9/0.1 (v/v), 10 mM ammonium formate), 90% B (acetonitrile/formic acid, 99.9/0.1 (v/v)); 1.3-1.5 min linear from 90% to 40% B; 1.5-1.7 min linear from 40% to 90% B, 1.7-2 min isocratic 90% B. Basic condition) 0-1.3 min: isocratic 10% A (water/formic acid, 99.8/0.2 (v/v), 10 mM ammonium carbonate), 90% B (acetonitrile); 1.3-1.5 min linear from 90% to 40% B; 1.5-1.7 min linear from 40% to 90% B, 1.7-2 min isocratic 90% B The injection volume was 3.0 µl (full loop injection). Eluting compounds were detected using an

Agilent 6495 triple quadrupole mass spectrometer (Agilent Technologies) equipped with an Agilent Jet Stream electrospray ion source in positive and negative ion mode. Source gas temperature was set to 200 °C, with 14 L*min^-1 drying gas and a nebulizer pressure of 24 psi. Sheath gas temperature was set to 300 °C and flow to 11 L*min^-1. Electrospray nozzle and capillary voltages were set to 500 and 2500 V, respectively. Metabolites were identified by multiple reaction monitoring (MRM), and MRM parameters were optimized and validated with authentic standards[44]. Metabolites were measured in ^{12}C- and ^{13}C isoforms, and data was analyzed with published Matlab code[44]. Metabolites were sampled four times at the first time point t_0; and two samples were collected at the remaining time points (see also reporting standards in **Supplementary Data 10**). Metabolomics metadata is accessible under the MetaboLights accession number MTBLS1044.

Transcriptomics

For transcriptomics 0.5 mL culture was transferred into reaction tubes and centrifuged at 11 000 xg for 2 min, and the pellet was frozen in liquid nitrogen. The total RNA of the cells was isolated using the Total RNA Isolation Mini Kit (Agilent, Santa Clara, CA). The integrity of the RNA was measured using the BioAnalyzer Pico-Kit (Agilent, Santa Clara, CA). RNA sequencing was performed by the Max Planck-Genome-Centre Cologne, Germany (https://mpgc.mpipz.mpg.de/home/). The sequencing reads were analyzed and mapped using the CLC Software (QIAGEN, Venlo, NL). For normalization, gene expression was calculated as transcripts per kilobase million (TPMs). RNA was sampled four times at the first time point t_0; and two samples were collected at the remaining time points. For the time points t_{13}, t_{15}, t_{19} and t_{24} one of the two replicates was excluded due to low quality of the sampled RNA. Transcriptomics metadata is accessible under the GEO number GSE131992.

Network component analysis (NCA)

NCA was performed by iteratively optimizing connectivity strength and TF-activity by using the connectivity matrix of the transcription regulation network and the measured gene expression. The optimization is a least square optimization between the gene expression and the product of connectivity and TF-activity:

$$(1)\quad \min_{A,P} \|E - AP\|^2$$

Where *E* is the log10 transformed gene expression data (in TPMs) (**Supplementary Data 9**), *A* the connectivity matrix of the transcription regulation network (matrix with regulator-gene interactions **Supplementary Data 8**) and *P* the TF-activity[19]. To generate the connectivity matrix, a matrix of transcription regulator – gene interactions was generated by combining the matrixes of TF – gene interactions and sigma factor – gene interactions of RegulonDB[18]. Additional regulation that was

added was the (p)ppGpp regulon and transcriptional attenuation, as described in the EcoCyc database[25]. To account for basal expression of every gene by the RNA polymerase we added a global regulator, which was connected to all genes in the connectivity matrix. Randomized starting points were used for each calculation cycle of the algorithm. A calculation cycle was aborted if the summed squared 2-norm of the residuals did not change by more than 1%.

Correlations between metabolites and TF activities

Metabolite concentrations and TF activities were first correlated linearly. In case of a positive linear correlation we used activating Hill kinetics as the basis for a non-linear fit. In case of a negative linear correlation we used inhibition kinetics:

(2) Activation kinetics: $y = y_{\max} * \frac{x^{\mathrm{h}}}{x^{\mathrm{h}} + {K_{\mathrm{H}}}^{\mathrm{h}}}$

(3) Inhibition kinetics: $y = y_{\max} * \frac{{K_{\mathrm{H}}}^{\mathrm{h}}}{x^{\mathrm{h}} + {K_{\mathrm{H}}}^{\mathrm{h}}}$

Where *y* is the TF activity, and *x* the metabolite concentration. K_H is the activation constant, *h* the Hill coefficient and y_{max} is the maximal TF activity, which was assumed to be constant over time. Parameters of the Hill equations (K_H and *h*) were estimated in total 50 times per metabolite-TF pair. The Hill coefficient *h* was constrained to an upper value of 10. For each pair of metabolite and TF, we tested if a negative time-shift of the TF activity by one time point would improve the parameter estimation. This accounts for the fact that TF activities are derived from gene expression data, which could potentially succeed changes of metabolite levels (**Supplementary Fig. 10**). The correlation coefficient R^2 was calculated between the measured TF activity and the transformed metabolite levels using the estimated Hill parameters.

Distances of metabolite-TF interactions

First, we remove all cofactors, as well as periplasmatic and extracellular metabolites from the stoichiometric matrix of the iJO1366 metabolic genome scale model of *E. coli*. Next, we create a metabolite-gene adjacency matrix, F, by calculating the inner product of the modified stoichiometric matrix, N, and the reaction-gene matrix, G. We finish by computing the Boolean of F, F'. Next, we transform F' it into an undirected, bipartite graph, nodes denoting metabolites and genes, respectively. For this graph, we calculate a distance matrix, D, containing all pairwise distances between metabolites and genes in F. For known metabolite-TF interactions, we look for the distances between the regulating metabolite and each of the target genes of the TF and take the smallest distance. In case a regulating metabolite is not part of the iJO1366, we omit the distance calculation[45]. The distance criterion for correlating metabolite-TF pairs (**Fig. 4b**) was also based on the genome scale model iJO1366[29]. Pairs of metabolites and TFs were only considered if at least one of the two criteria

were fulfilled. Criteria 1: the metabolite is a product or a substrate of an enzyme that is encoded by a target-gene of the TF. Criteria 2: the metabolite is listed in the same metabolic subsystem as an enzyme that is encoded by a target-gene of the TF. Subsystems of TFs were defined as the metabolic pathways controlled by the TF in the genome scale model. Subsystems of metabolites were defined according to the **Supplementary Data 1**.

Protein overexpression and purification

TFs were purified from the *E. coli* ASKA strains[46]. Cells were grown in 200 mL TB medium containing 30 µg*mL^-1 Chloramphenicol at 37 °C. When cells reached OD 0.6 we added 0.5 mM IPTG. Cells were incubated at 37 °C for 3 h more and harvested by centrifugation. Proteins were purified from the pellets using Protino™ Ni-TED-IDA 1000 Kit (Macherey-Nagel, Düren Germany). Protein purity was confirmed by SDS-PAGE and concentrations were determined by the Pierce protein BCA Assay (Thermo Fischer Scientific, Waltham, MA).

Quantitation of interactions by Microscale Thermophoresis

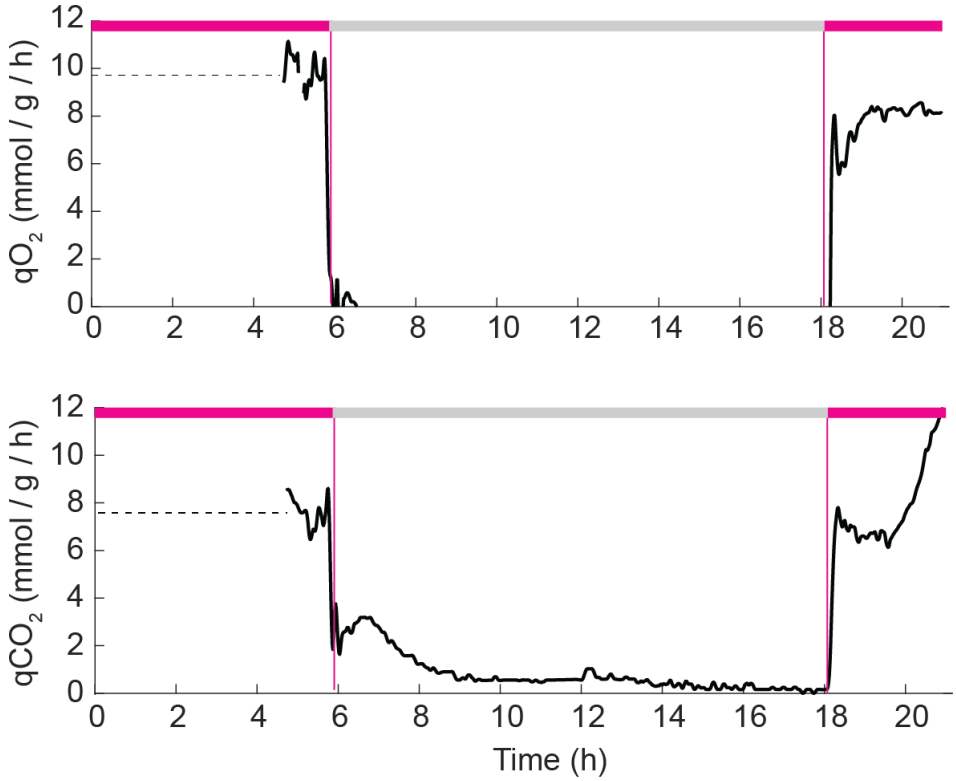
Microscale Thermophoresis (MST)[47] was performed on a Monolith NT.115 (Nano Temper Technologies GmbH, Munich, Germany) at 21°C (red LED power was set to 75% and infrared laser power to 80%). 50 nM of the respective protein was labeled with the dye Monolith His-Tag Labeling Kit RED-tris-NTA 2nd Generation (MO-L018) supplied by NanoTemper Technologies. Labeled proteins were titrated as indicated with the respective metabolite in buffer T (50 mM NaH_2PO_4, 500 mM NaCl, pH 5.7). At least nine independent MST experiments (three technical replicates of three biological replicates) were performed at 680 nm and processed by Nano Temper Analysis package 1.2.009 and Origin8 (OriginLab, Northampton, MA).

Data & Code availability

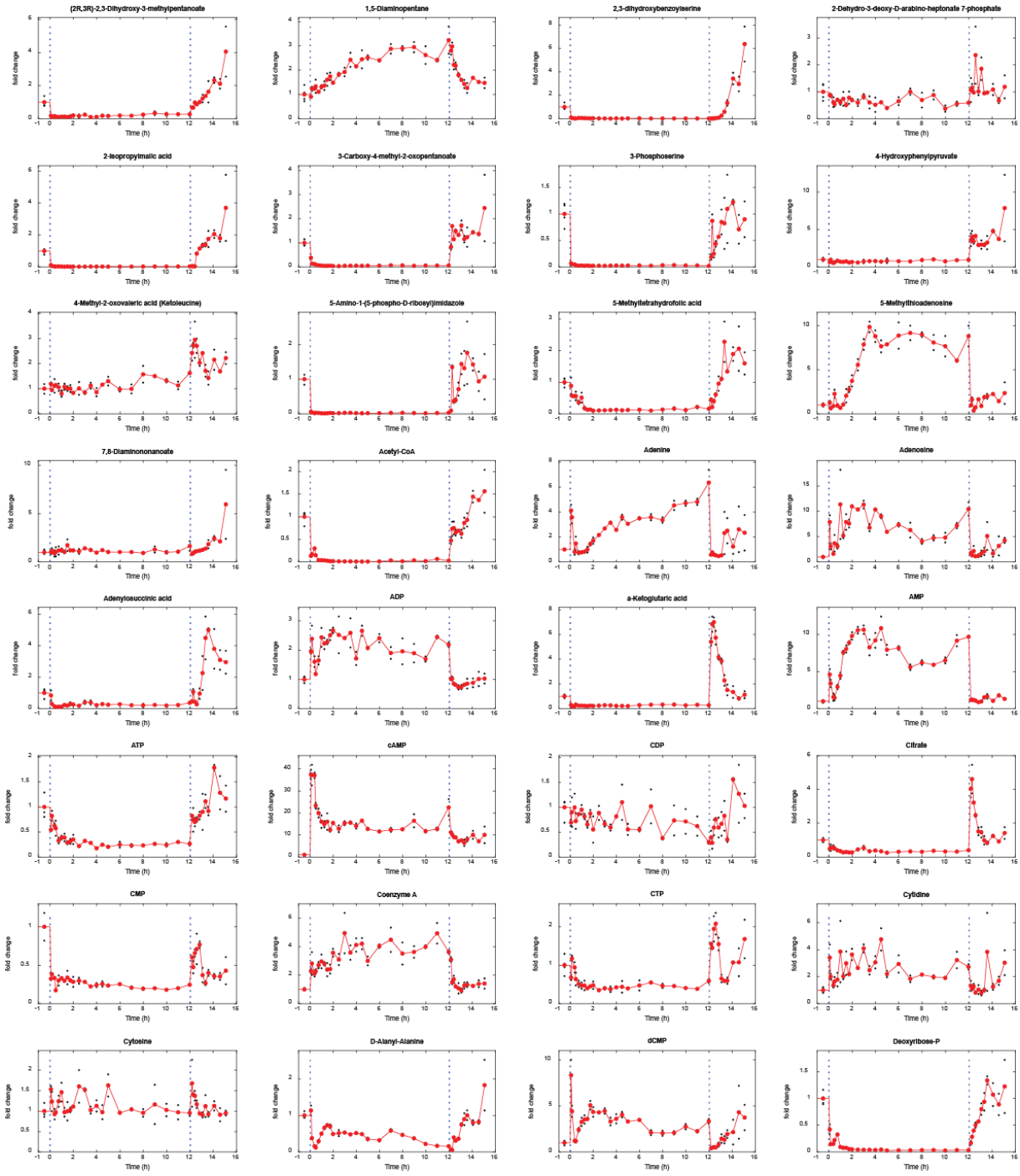
Gene expression data that support the findings of this study have been deposited in NCBI's Gene Expression Omnibus with the accession code GSE131992 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131992]. Metabolomics data that support the findings of this study have been deposited in MetaboLights database with the accession codes MTBLS1044[https://www.ebi.ac.uk/metabolights/MTBLS1044]. The source data of Figures 1a-b, 2, 3a-c, 4b-c and 5b and Supplementary Figures 1, 3, 4, 5, 7 and 9 are provided as a Source Data file. All other data are available from the corresponding author on reasonable request. Supplementary data tables can be accessed from the GitHub repository via https://github.com/mlempp/PhD_Thesis.

Matlab code to perform Network Component Analysis and Kinetic correlations can be accessed from the GitHub repository via https://github.com/nfarke/Lempp_Metabolite_TF_interaction_Ecoli.

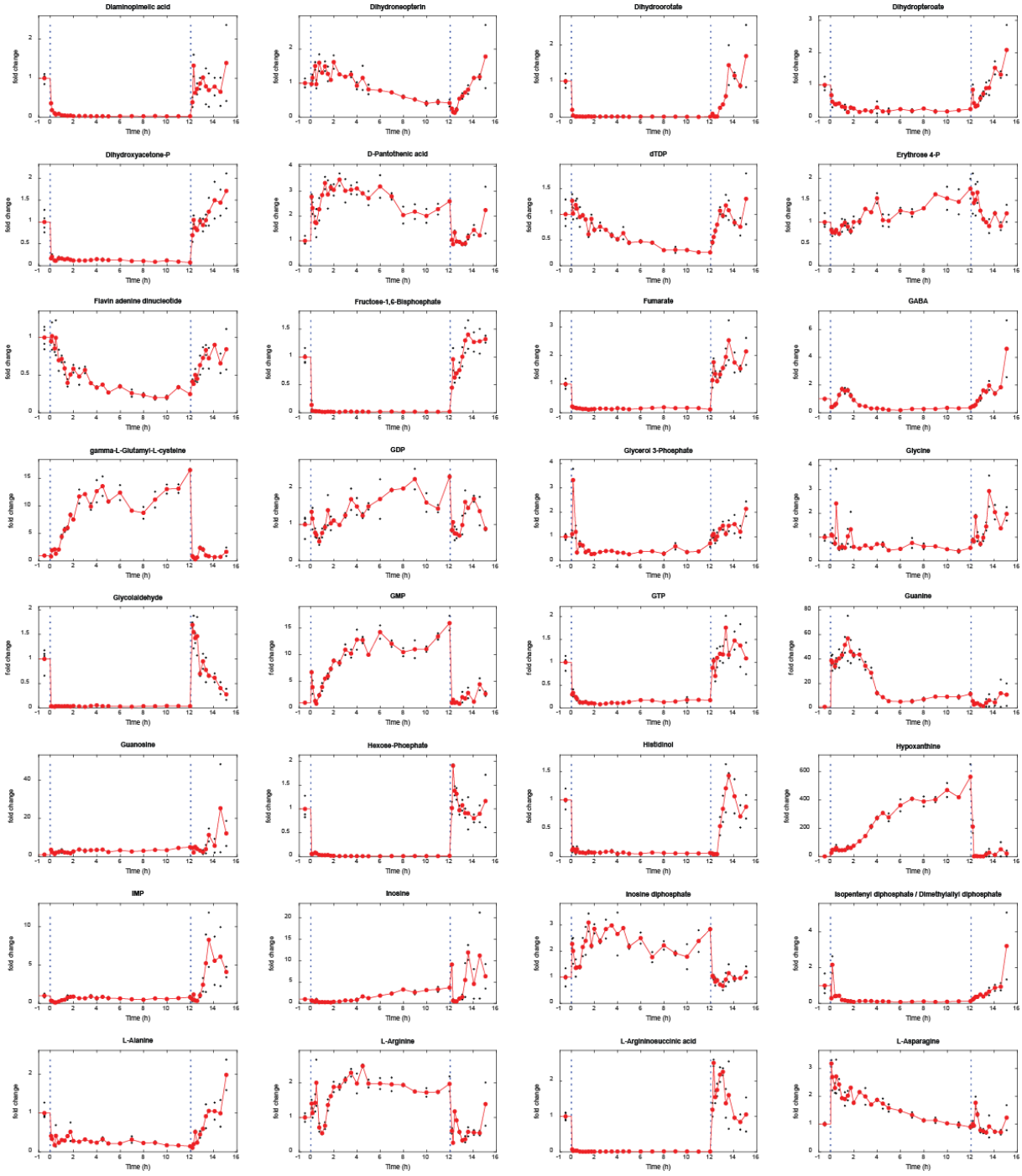
Supplemental Information



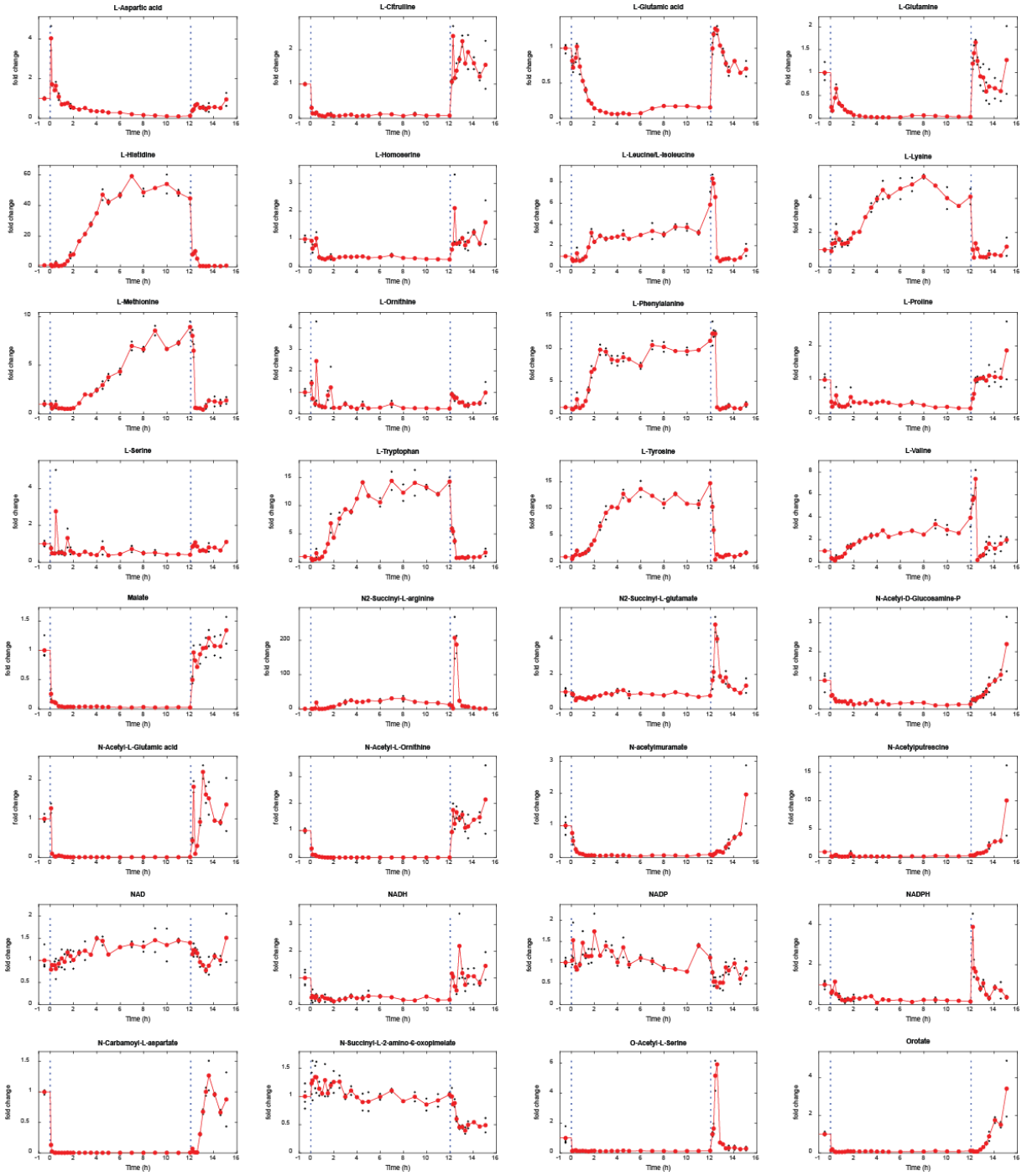
Supplementary Figure 1. Specific oxygen consumption (qO_2) and carbon production rates (qCO_2) during the growth-starvation-growth switch. The dashed line indicates the phase when biomass was too low to determine rates. Pink indicates growth phases and grey the starvation phase. (Source data are provided as a Source Data file.)



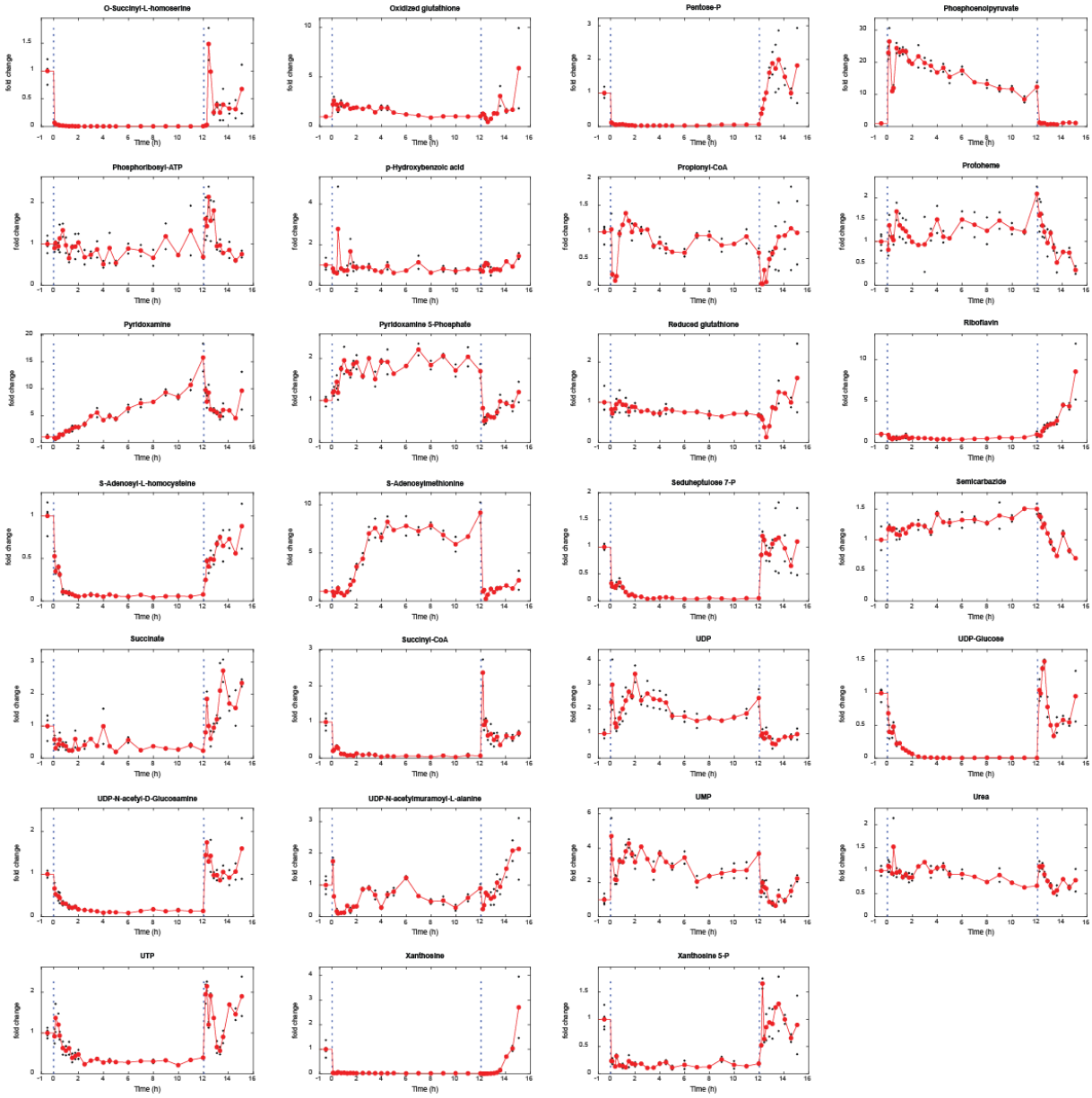
Supplementary Figure 2. Metabolite levels during the switch between starvation and growth. Relative concentrations are shown as fold change relative to the first data point. Black dots show levels of two replicates per time point (four at the first time point), red dots are the mean. Dashed vertical lines indicate the growth-starvation-growth switch.



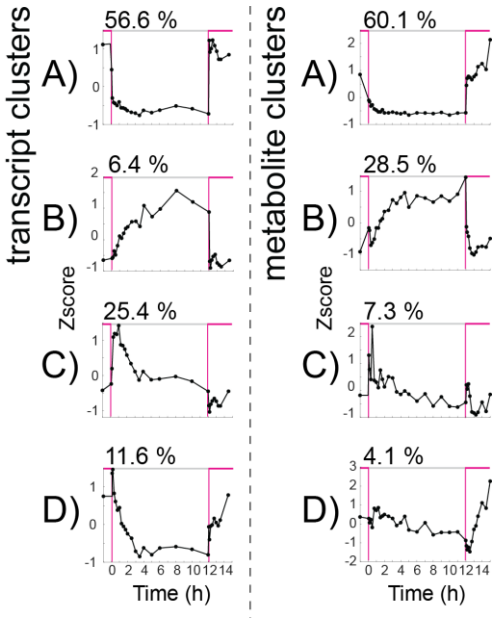
Supplementary Figure 2. (continued)



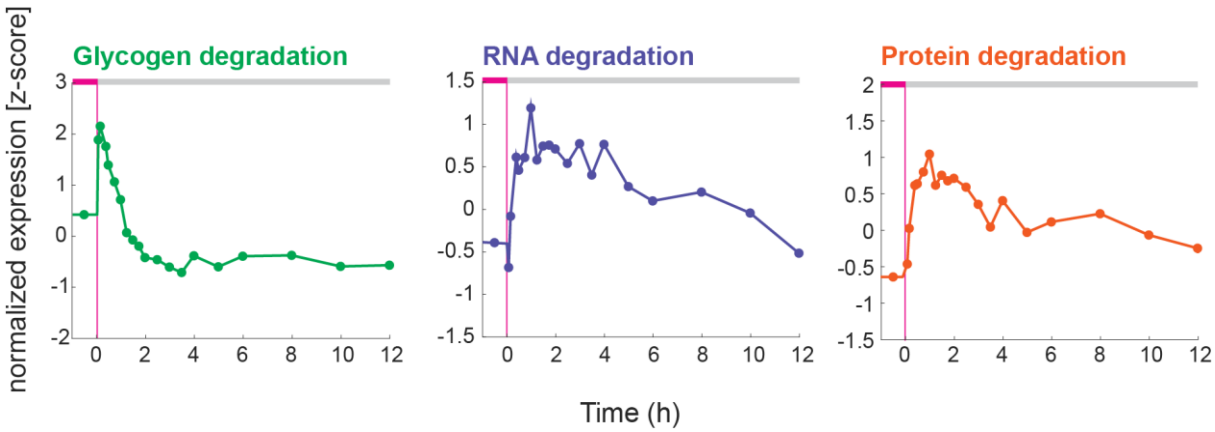
Supplementary Figure 2. (continued)



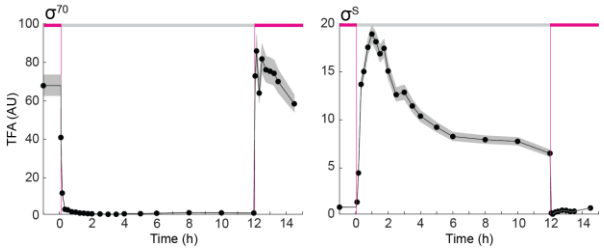
Supplementary Figure 2. (continued)



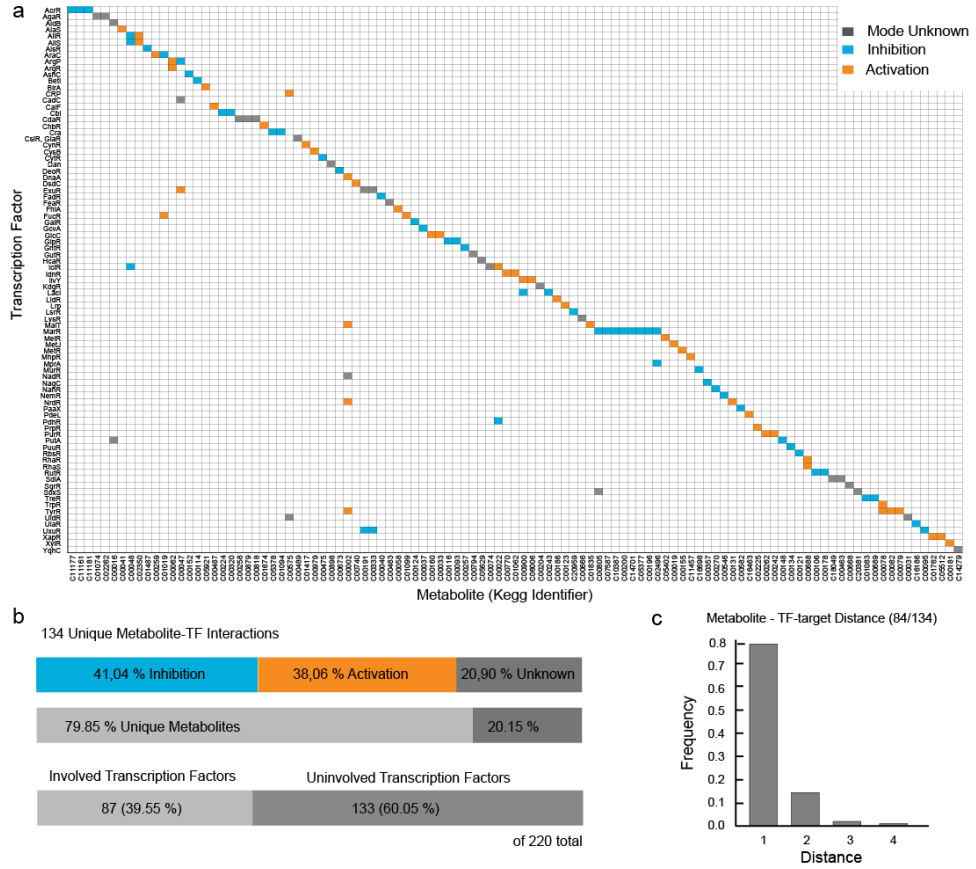
Supplementary Figure 3. Average dynamics of the hierarchical clusters A - D of transcripts and metabolites. Percentages show the number of metabolites/transcripts in cluster relative to the total number of measured metabolites/transcripts. Pink indicates growth phases and grey the starvation phase. (Source data are provided as a Source Data file.)



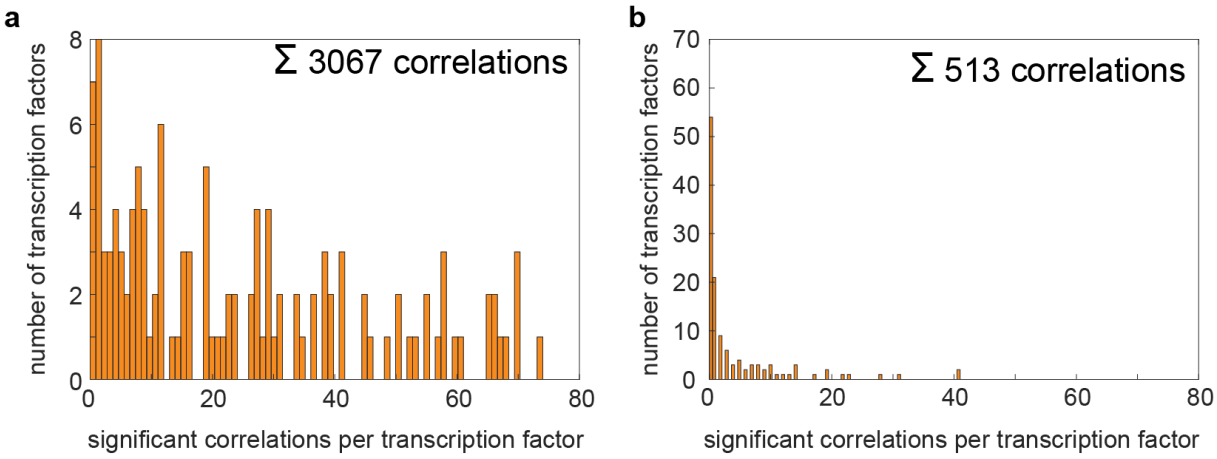
Supplementary Figure 4. Average expression of genes in degradation pathways of glycogen (green), RNA (purple) and proteins (orange) during the starvation phase. Pink indicates the growth phase and grey the starvation phase. Expression profiles include only genes that increase during the starvation phase (Glycogen degradation: 4 of 6 genes; RNA degradation: 13 of 35 genes; Protein degradation: 42 of 106 genes) (Source data are provided as a Source Data file.)



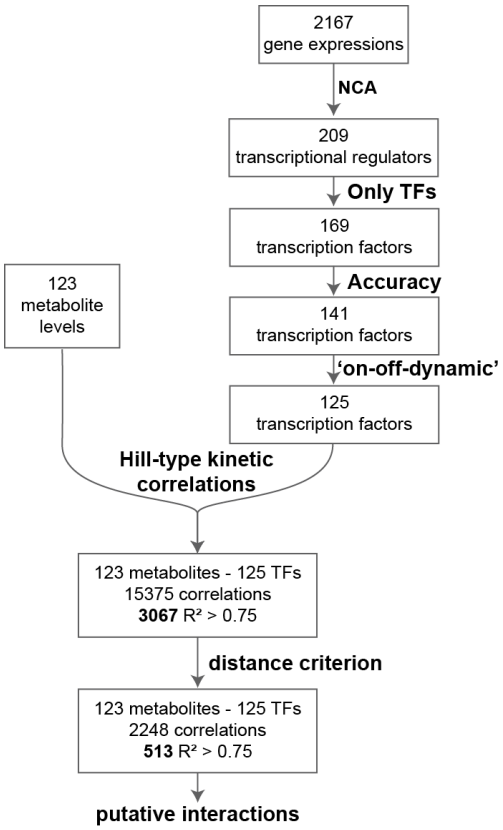
Supplementary Figure 5. Activity of the sigma factors σ^{70} and σ^{S} during the switch between starvation and growth. The grey area indicates the 95% confidence interval of n = 100 randomized estimations with Network Component Analysis. Pink indicates growth phases and grey the starvation phase. (Source data are provided as a Source Data file.)



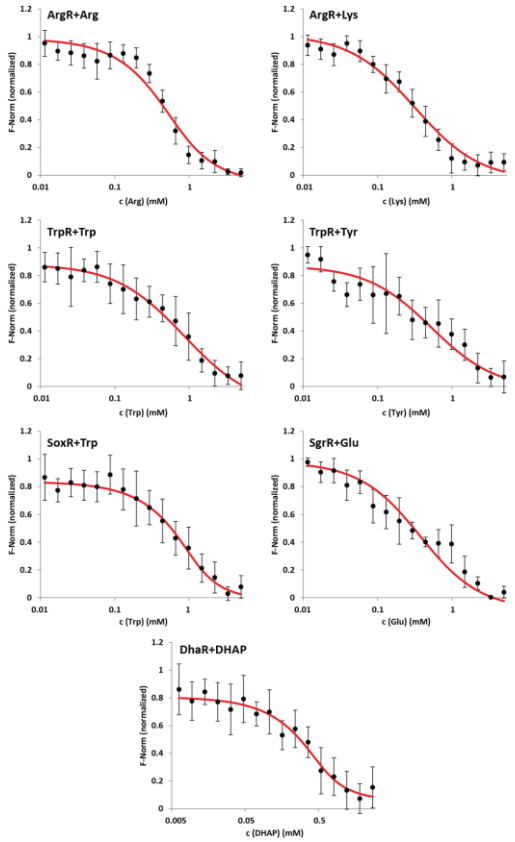
Supplementary Figure 6. a) Metabolite-transcription factor interactions that are described in the literature and databases. Shown are 87 transcription factors (rows) and their respective effector metabolites (columns). Orange indicates an activation of the TF by the metabolite, blue indicates an inhibition, and grey indicates that the mode is unknown. **b)** Mode of metabolite-transcription factor interactions; fraction of unique metabolites in the 134 different interactions; number of transcription factors for which an interacting metabolite is known. **c)** Distance between a metabolite and the target genes of the interacting transcription factor. The distance d was transformed by the following equation to account only for genes: Distance=(d+1)/2.



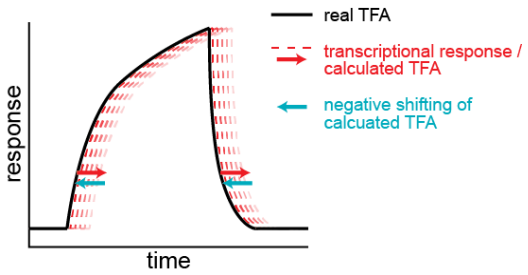
Supplementary Figure 7. a) Number of significant kinetic correlations (R^2>0.75) per transcription factor without distance filter. **b)** Number of significant kinetic correlations (R^2>0.75) per transcription factor with the distance filter. (Source data are provided as a Source Data file.)



Supplementary Figure 8. Steps to reduce the number of correlating metabolite-transcription factor pairs.



Supplementary Figure 9. *In vitro* binding assays with micro-scale thermophoresis (MST). Normalized fluorescence (black dots) was fitted (solid red line) to calculate the binding affinity of each interaction. The respective transcription factor (50 nm) was titrated with increasing amount of the putative effector metabolite. Shown are the mean and standard deviation of n = 9 MST assays (three technical replicates of three biological replicates). (Source data are provided as a Source Data file.)



Supplementary Figure 10. Schematic of the potential time lag between real and calculated transcription factor activity (TFA). Real TFA follows immediately the change in effector metabolite level (black). Because of a delay in gene expression, the actual transcriptional response has a negative time-lag (red). Therefore, we allow a negative time-lag for TFA in the correlation analysis with metabolites (blue arrows).

References

1. Buescher, J. M. *et al.* Global network reorganization during dynamic adaptations of Bacillus subtilis metabolism. *Science* **335**, 1099–103 (2012).

2. Kresnowati, M. T. A. P. *et al.* When transcriptome meets metabolome: Fast cellular responses of yeast to sudden relief of glucose limitation. *Molecular Systems Biology* **2**, (2006).

3. Bradley, P. H., Brauer, M. J., Rabinowitz, J. D. & Troyanskaya, O. G. Coordinated Concentration Changes of Transcripts and Metabolites in Saccharomyces cerevisiae. *PLOS Computational Biology* **5**, (2009).

4. Redestig, H. & Costa, I. G. Detection and interpretation of metabolite–transcript coresponses using combined profiling data. *Bioinformatics* **27**, i357–i365 (2011).

5. Chubukov, V., Gerosa, L., Kochanowski, K. & Sauer, U. Coordination of microbial metabolism. *Nature Reviews Microbiology* **12**, 327–340 (2014).

6. Browning, D. F. & Busby, S. J. W. Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology* **14**, 638–650 (2016).

7. Donati, S., Sander, T. & Link, H. Crosstalk between transcription and metabolism: how much enzyme is enough for a cell? *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **10**, (2018).

8. Jozefczuk, S. *et al.* Metabolomic and transcriptomic stress response of Escherichia coli. *Molecular Systems Biology* **6**, 1–16 (2010).

9. Geiger, R. *et al.* L-Arginine Modulates T Cell Metabolism and Enhances Survival and Anti-tumor Activity. *Cell* **167**, 829-842.e13 (2016).

10. Cho, B.-K., Federowicz, S., Park, Y.-S., Zengler, K. & Palsson, B. Ø. Deciphering the transcriptional regulatory logic of amino acid metabolism. *Nature Chemical Biology* **8**, 65–71 (2012).

11. Razo-Mejia, M. *et al.* Tuning Transcriptional Regulation through Signaling: A Predictive Theory of Allosteric Induction. *Cell systems* **6**, 456-469.e10 (2018).

12. Rinschen, M. M., Ivanisevic, J., Giera, M. & Siuzdak, G. Identification of bioactive metabolites using activity metabolomics. *Nature Reviews Molecular Cell Biology* **20**, 353–367 (2019).

13. Yugi, K. & Kuroda, S. Metabolism-Centric Trans-Omics. *Cell Systems* **4**, 19–20 (2017).

14. Folly, B. B. *et al.* Assessment of the interaction between the flux-signaling metabolite fructose-1,6-bisphosphate and the bacterial transcription factors CggR and Cra. *Molecular Microbiology* **109**, 278–290 (2018).

15. Piazza, I. *et al.* A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* **172**, 358-372.e23 (2018).

16. Kochanowski, K. *et al.* Few regulatory metabolites coordinate expression of central metabolic genes in Escherichia coli. *Molecular systems biology* **13**, 903–903 (2017).

17. Link, H., Fuhrer, T., Gerosa, L., Zamboni, N. & Sauer, U. Real-time metabolome profiling of the metabolic switch between starvation and growth. *Nature methods* **12**, 1091–1097 (2015).

18. Gama-Castro, S. *et al.* RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Research* **44**, D133–D143 (2016).

19. Liao, J. C. *et al.* Network component analysis: Reconstruction of regulatory signals in biological systems. *Proceedings of the National Academy of Sciences* **100**, 15522–15527 (2003).

20. Kao, K. C. *et al.* Transcriptome-based determination of multiple transcription regulator activities in by using network component analysis Escherichia coli. *Proceedings of the National Academy of Sciences* **101**, 641–646 (2004).

21. Sharma, U. K. & Chatterji, D. Transcriptional switching in Escherichia coli during stress and starvation by modulation of sigma activity. *FEMS Microbiology Reviews* **34**, 646–657 (2010).

22. Kolb, A., Busby, S., Buc, H., Garges, S. & Adhya, S. Transcriptional Regulation by cAMP and its Receptor Protein. *Annual Review of Biochemistry* **62**, 749–797 (1993).

23. Małecki, Jij., Polit, A. & Wasylewski, Z. Kinetic Studies of cAMP-induced Allosteric Changes in Cyclic AMP Receptor Protein from Escherichia coli *Journal of Biological Chemistry* **275**, 8480–8486 (2000).

24. Arvidson, D. N., Bruce, C. & Gunsalus, R. P. Interaction of the Escherichia coli trp aporepressor with its ligand, L-tryptophan. *Journal of Biological Chemistry.* **261**, 238–243 (1986).

25. Keseler, I. M. *et al.* The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Research* **45**, D543–D550 (2017).

26. Shen, Q. *et al.* ASD v3.0: unraveling allosteric regulation with structural mechanisms and biological networks. *Nucleic Acids Research* **44**, D527–D535 (2016).

27. Hackett, S. R. *et al.* Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* **354**, (2016).

28. Ortmayr, K., Dubuis, S. & Zampieri, M. Metabolic profiling of cancer cells reveals genome-wide crosstalk between transcriptional regulators and metabolism. *Nature Communications* **10**, 1841 (2019).

29. Orth, J. D. *et al.* A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011. *Molecular Systems Biology* **7**, (2011).

30. Pathania, A. & Sardesai, A. A. Distinct Paths for Basic Amino Acid Export in Escherichia coli: YbjE (LysO) Mediates Export of L-Lysine. *Journal of Bacteriology.* **197**, 2036–2047 (2015).

31. Peterkofsky, B. & Gilvarg, C. N-Succinyl-l-diaminopimelic-glutamic Transaminase. *Journal of Biological Chemistry* **236**, 1432–1438 (1961).

32. Sander, T. *et al.* Allosteric Feedback Inhibition Enables Robust Amino Acid Biosynthesis in E. coli by Enforcing Enzyme Overabundance. *Cell Systems* **8**, 66-75.e8 (2019).

33. Seo, S. W., Kim, D., Szubin, R. & Palsson, B. O. Genome-wide Reconstruction of OxyR and SoxRS Transcriptional Regulatory Networks under Oxidative Stress in Escherichia coli K-12 MG1655. *Cell Reports* **12**, 1289–1299 (2015).

34. Vanderpool, C. K. & Gottesman, S. The Novel Transcription Factor SgrR Coordinates the Response to Glucose-Phosphate Stress. *Journal of Bacteriology* **189**, 2238–2248 (2007).

35. Kim, S. H., Schneider, B. L. & Reitzer, L. Genetics and Regulation of the Major Enzymes of Alanine Synthesis in Escherichia coli. *Journal of Bacteriology* **192**, 5304–5311 (2010).

36. Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W. & Milo, R. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences* **110**, 10039–10044 (2013).

37. Bächler, C., Schneider, P., Bähler, P., Lustig, A. & Erni, B. Escherichia coli dihydroxyacetone kinase controls gene expression by binding to transcription factor DhaR. *The EMBO Journal* **24**, 283–293 (2005).

38. Fang, X. *et al.* Global transcriptional regulatory network for Escherichia coli robustly connects gene expression to transcription factor activities. *Proceedings of the National Academy of Sciences* **114**, 10286–10291 (2017).

39. Noor, E., Cherkaoui, S. & Sauer, U. Biological insights through omics data integration. *Current Opinion in Systems Biology* (2019).

40. Campbell, K., Vowinckel, J., Keller, M. A. & Ralser, M. Methionine Metabolism Alters Oxidative Stress Resistance via the Pentose Phosphate Pathway. *Antioxidants & Redox Signaling* **24**, 543–547 (2016).

41. Campos, A. I. & Zampieri, M. Metabolomics-Driven Exploration of the Chemical Drug Space to Predict Combination Antimicrobial Therapies. *Molecular Cell* **74**, 1291-1303.e6 (2019).

42. Gupta, A., Brockman Reizman, I. M., Reisch, C. R. & Prather, K. L. J. Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nature Biotechnology* **35**, 273–279 (2017).

43. Burg, J. M. *et al.* Large-scale bioprocess competitiveness: the potential of dynamic metabolic control in two-stage fermentations. *Current Opinion in Chemical Engineering* **14**, 121–136 (2016).

44. Guder, J. C., Schramm, T., Sander, T. & Link, H. Time-Optimized Isotope Ratio LC–MS/MS for High-Throughput Quantification of Primary Metabolites. *Analytical Chemistry* **89**, 1624–1631 (2017).

45. Reznik, E. *et al.* Genome-Scale Architecture of Small Molecule Regulatory Networks and the Fundamental Trade-Off between Regulation and Enzymatic Activity. *Cell Reports* **20**, 2666–2677 (2017).

46. Kitagawa, M. *et al.* Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA Research* **12**, 291–299 (2005).

47. Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S. Molecular Interaction Studies Using Microscale Thermophoresis. *Assay and Drug Development Technologies* **9**, 342–353 (2011).

Chapter 5

Allosteric Feedback Inhibition Enables Robust Amino Acid Biosynthesis in *E. coli* by Enforcing Enzyme Overabundance

Timur Sander[1], Niklas Farke[1], Christoph Diehl[1], **Michelle Kuntz[1]**, Timo Glatter[1], Hannes Link[1,2]

[1]Max Planck Institute for Terrestrial Microbiology, Marburg 35043, Germany

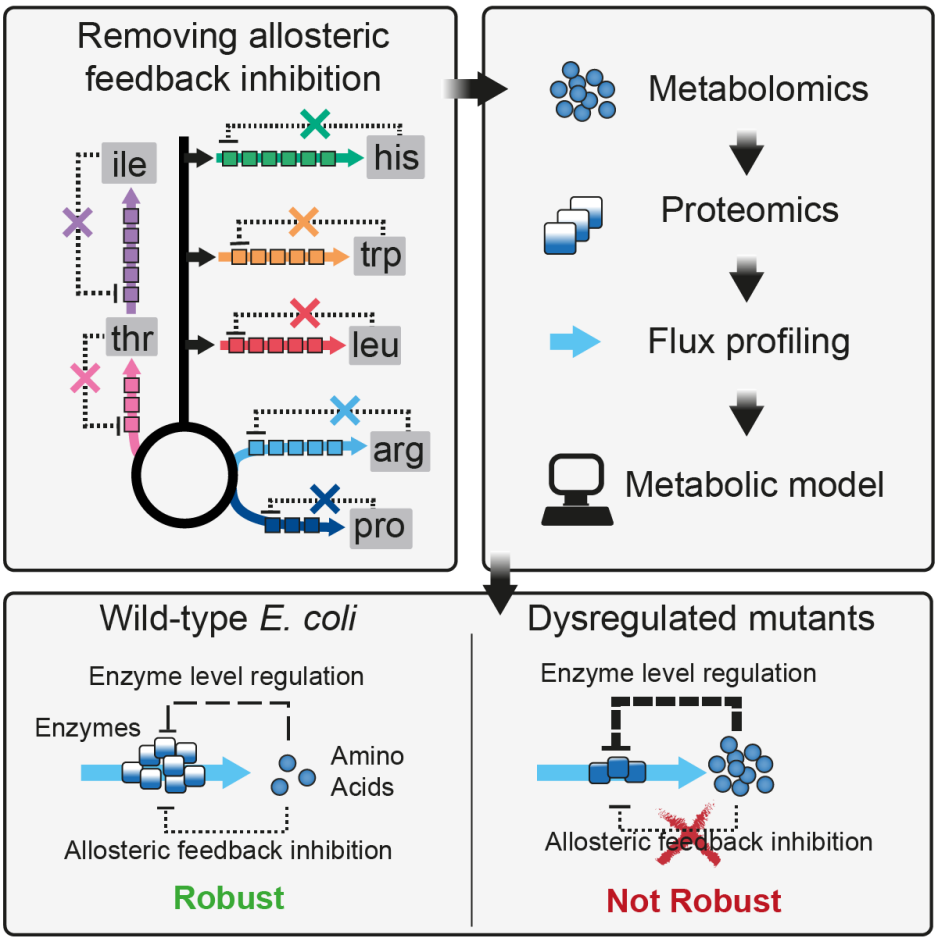
[2]Lead Contact

This chapter is written in manuscript style and was published in Cell Systems 2019, 8 (1), 66–75.e8. My contribution to this work included cloning plasmids with small guide RNAs for applying CRISPR interference in allosteric feedback mutants and performing growth experiments with the respective strains.

Summary

Microbes must ensure robust amino acid metabolism in the face of external and internal perturbations. This robustness is thought to emerge from regulatory interactions in metabolic and genetic networks. Here, we explored the consequences of removing allosteric feedback inhibition in seven amino acid biosynthesis pathways in *Escherichia coli* (arginine, histidine, tryptophan, leucine, isoleucine, threonine and proline). Proteome data revealed that enzyme levels decreased in five of the seven dysregulated pathways. Despite lower enzyme levels flux through the dysregulated pathways was not limited, indicating that enzyme levels are higher than absolutely needed in wild-type cells. We show that such enzyme overabundance renders the arginine, histidine and tryptophan pathway robust against perturbations of gene expression, using a metabolic model and CRISPR interference experiments. The results suggest a sensitive interaction between allosteric feedback inhibition and enzyme level regulation that ensures robust yet efficient biosynthesis of histidine, arginine and tryptophan in *E. coli*.

Graphical Abstract



Introduction

Regulation of microbial metabolism involves a wide range of mechanisms that act on different cellular layers and together control the abundance and activity of enzymes (Chubukov et al., 2014). An example is end-product inhibition of amino acid biosynthesis in *Escherichia coli*, which can act on enzyme abundance through transcriptional regulatory cues, and enzyme activities through allosteric feedback inhibition. However, since metabolic reaction rates are determined by both enzyme abundance and enzyme activity it has been difficult to disentangle the specific roles of the two regulatory layers, and to understand how they interact to control metabolism (Chubukov et al., 2013; Daran-Lapujade et al., 2007; Kuile and Westerhoff, 2001).

Allosteric feedback inhibition of the committed step in biosynthetic pathways is thought to maintain homeostasis of end-products (Umbarger, 1956), and 16 out of 20 amino acids in *E. coli* feedback inhibit enzymes of their own biosynthesis pathway (Reznik et al., 2017). The consequences of dysregulating these enzymes were mainly studied *in vitro* (Schomburg et al., 2013), or in the context of biotechnological overproduction strains (Hirasawa and Shimizu, 2016). For the case of nucleotide biosynthesis in *E. coli*, a detailed *in vivo* study showed that removing allosteric feedback inhibition did not perturb nucleotide homeostasis (Reaves et al., 2013). In the absence of allosteric feedback inhibition, additional regulatory mechanisms accomplished proper control of the pathway by channeling the excess of nucleotides into degradation pathways (so-called directed overflow). Theoretical analyses, in contrast, suggest a key role of allosteric feedback inhibition in achieving end-product homeostasis (Hofmeyr and Cornish-Bowden, 2000), metabolic robustness (Grimbs et al., 2007), flux control (Kacser and Burns, 1973; Schuster and Heinrich, 1987) and optimal growth (Goyal et al., 2010).

The abundance of enzymes in *E. coli* amino acid metabolism is mainly regulated at the layer of transcription, either by transcriptional attenuation (Yanofsky, 1981) or transcription factors (Cho et al., 2008, 2012). For example, a set of four transcription factors (ArgR, TrpR, TryR and Lrp) control expression of 19 out of 20 amino acid pathways, by sensing the availability of amino acids via allosteric binding (Cho et al., 2012). This regulation ensures that enzymes in amino acid pathways are only made when they are needed (Schmidt et al., 2016; Zaslaver et al., 2004). As a consequence of such need-based enzyme level regulation, one would expect that enzyme levels are not higher than absolutely needed for amino acid biosynthesis. However, recent data suggest that cells express the majority of enzymes at higher levels than necessary to fulfill biosynthetic demands, and that such enzyme overabundance provides a benefit in changing environments (Davidi and Milo, 2017; O'Brien et al., 2016). For example, enzyme overabundance enables a quick activation of the pentose phosphate

pathway upon stresses (Christodoulou et al., 2018), and similar benefits were attributed to overabundant ribosomes (Mori et al., 2017) and coenzymes (Hartl et al., 2017).

Here we constructed seven *E. coli* mutants, each with a different feedback-dysregulated amino acid biosynthesis pathway (arginine, histidine, tryptophan, leucine, isoleucine, threonine and proline), and measured proteins, metabolites, fluxes and growth of the mutants. In all seven feedback-dysregulated pathways the concentration of amino acid end-products increased, and in five pathways we measured lower enzyme levels. Despite the lower enzyme levels biosynthetic flux was not limited, indicating that these enzymes are not operating at maximal capacity in wild-type cells. By combining theoretical and experimental analysis, we show that this enzyme overabundance provides a robustness benefit against genetic perturbations, in case of the arginine, tryptophan and histidine pathway.

Results

Dysregulating Allosteric Enzymes Changes Levels of Specific Amino Acids in *E. coli*

To explore the function of allosteric feedback inhibition in the arginine, histidine, tryptophan, leucine, isoleucine, threonine and proline biosynthesis pathways, we first created a panel of seven allosterically dysregulated *E. coli* mutants (**Figure 1A** and **Table S1**). Using a scarless CRISPR method (Reisch and Prather, 2015), we introduced point mutations into genes encoding the allosteric enzyme that catalyzes the committed reaction in each pathway (*argA*, *hisG*, *trpE*, *leuA*, *ilvA*, *thrA* and *proB*). These mutations have been shown previously to abolish the allosteric interaction while not affecting enzyme activity, thereby allowing us to study regulation of the pathway in the absence of allosteric feedback (Caligiuri and Bauerle, 1991; Csonka et al., 1988; Doroshenko et al., 2013; Gusyatiner et al., 2005; LaRossa et al., 1987; Lee et al., 2003; Rajagopal et al., 1998). For N-acetylglutamate synthase (ArgA), we confirmed with *in vitro* assays that the mutation does not affect enzymatic activity, and abolishes inhibition by arginine (**Figure S1**). To analyze metabolism of the mutants we quantified intracellular metabolites during exponential growth on glucose by LC-MS/MS (Guder et al., 2017). Stronger metabolic changes were restricted to amino acid biosynthesis, with specific increases between 2- and 16-fold of only the amino acid products of the dysregulated pathways (**Figure 1B**). Despite these changes within the dysregulated pathways, the remaining amino acid concentrations as well as the global metabolite profile remained relatively stable (**Figure 1B** and **S2**). Thus, dysregulating allosteric enzymes in *E. coli* amino acid biosynthesis elevated the intracellular concentration of the corresponding amino acid product.

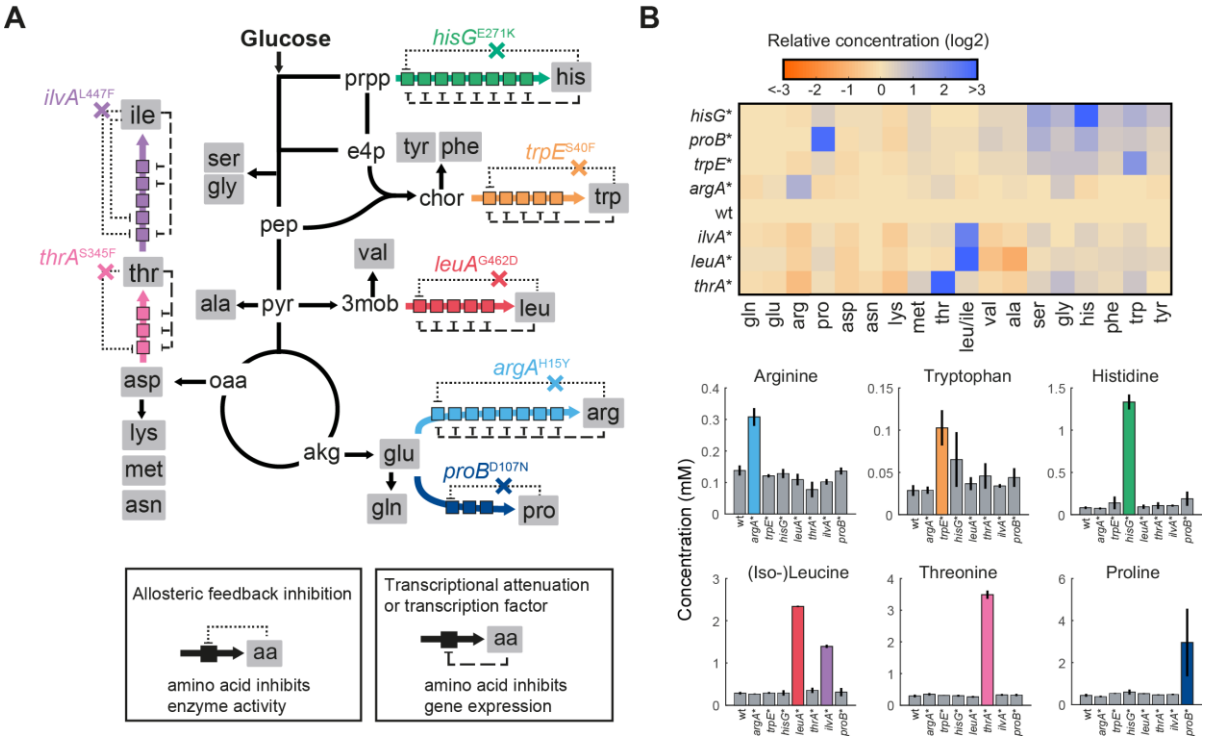


Figure 1. Amino Acid Profile of Feedback-dysregulated *E. coli* Mutants
(A) Seven amino acid pathways were dysregulated by genomic point mutations in the indicated genes. See also **Table S1**. Negative allosteric feedbacks of amino acids on enzymes in the biosynthetic pathways are shown as dotted lines. Negative transcriptional feedbacks of amino acids are shown as dashed lines. Boxes indicate enzymes in the biosynthesis pathways. (B) Relative concentrations of intracellular amino acids in wild-type *E. coli* and the seven dysregulated mutants. Bar plots show absolute concentrations of the amino acid in the dysregulated pathways. See also **Figure S2.** Data are represented as mean, and error bars are ± SD (n = 3).

Lower Expression of Enzymes in Feedback-Dysregulated Pathways

With the exception of proline biosynthesis, all of the dysregulated pathways are additionally controlled at the layer of enzyme abundance, either via transcription factors or transcriptional attenuation. To probe if elevated amino acid concentrations in our mutants affected enzyme levels in the corresponding pathways, we measured their proteomes (**Figure 2A**). The data covered relative abundances of 173 out of the 204 enzymes annotated to amino acid metabolism in the latest *E. coli* metabolic model (Monk et al., 2017). Enzyme expression was indeed lower in five of the seven dysregulated pathways (*argA**, *trpE**, *hisG**, *leuA**, *thrA**), indicating that the elevated amino acid concentrations caused a compensatory downregulation of their associated pathway (**Figure 2A**). Enzyme levels did not change in the *proB** and *ilvA** mutant, which is expected because proline biosynthesis lacks enzyme level regulation and isoleucine biosynthesis is subject to a second allosteric feedback that was not removed (**Figure 1A** and **2A**). The *leuA** mutant showed more global changes in enzyme levels than the other mutants. The high leucine concentration in this strain likely activates the leucine responsive transcription factor Lrp, which acts on many genes in amino acid metabolism

(Cho et al., 2008). In the *argA** mutant we observed an expected accompanying decrease in histidine biosynthesis enzymes, which are additional targets of the transcription factor ArgR (Gama-Castro et al., 2016). Apart from the compensatory downregulation of biosynthetic enzymes, enzymes in dedicated amino acid degradation pathways were upregulated in three mutants (AstC in the arginine mutant, TnaA in the tryptophan mutant and PutA in the proline mutant, **Figure 2A**). This likely constitutes an additional compensatory mechanism similar to the directed overflow reported for nucleotides (Reaves et al., 2013).

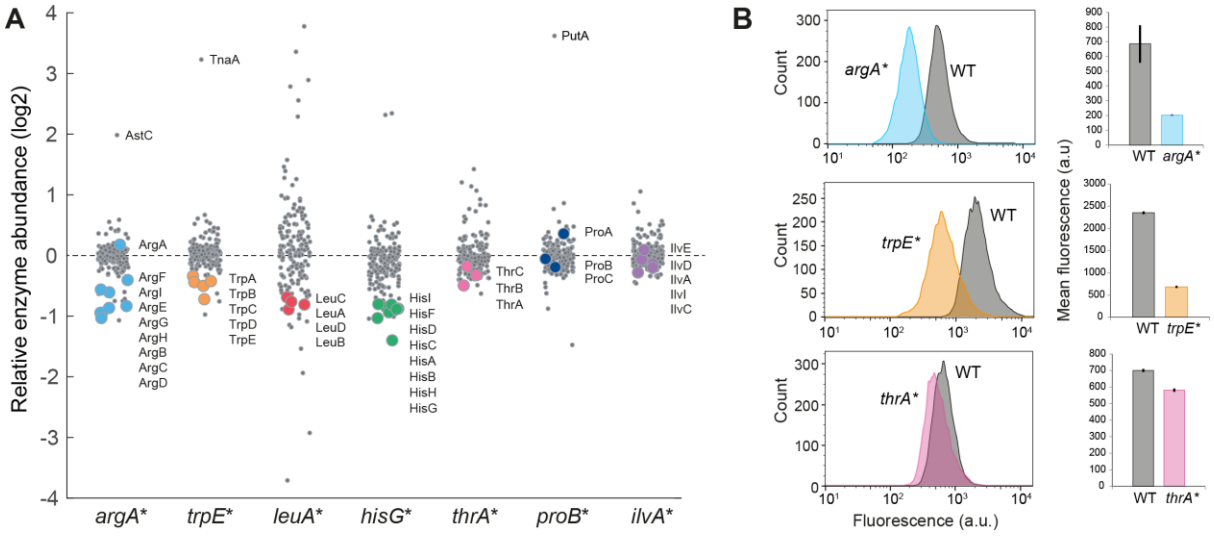


Figure 2. Expression of Enzymes in Feedback-dysregulated Pathways
(A) Abundance of 173 enzymes in amino acid metabolism (out of 204 enzymes in total), relative to the level in the wild-type. Data are represented as mean (n = 3). For each strain the enzymes in the dysregulated pathway are shown as colored dots. Enzymes in degradation pathways of arginine, tryptophan and proline are indicated by their names. (B) GFP-fluorescence measured by flow cytometry. GFP-promoter fusions were transformed in wild-type cells and the indicated mutant. Upper panel: pP*argA-gfp*; middle panel: pP*trpL-gfp;* lower panel: pP*thrL-gfp*. Histograms represent fluorescence of 10,000 single cells. Mean fluorescence was calculated from 10,000 single cells of n = 3 independent cultures. See also **Figure S3**.

To obtain additional evidence for lower enzyme levels in the dysregulated pathways, we used GFP-promoter fusions and measured fluorescence in single cells (**Figure 2B**). GFP expression from an ArgR-regulated promoter was indeed ~3-fold lower in the *argA** mutant compared to the wild-type. Similarly, a TrpR-regulated promoter was ~3-fold stronger repressed in the *trpE** mutant. The cell-to-cell variation in GFP content was similar in wild-type cells and the mutants, thus indicating that all cells in the population of allosteric feedback mutants have lower enzyme levels in the dysregulated pathway. A GFP reporter with the *thrL* leader peptide was only 17% repressed in the *thrA** mutant compared to the wild-type, which is consistent with the small decrease of enzymes levels in the dysregulated threonine pathway (**Figure 2A** and **2B**). We also fused GFP to the *hisL* and *leuL* leader peptides, but they did not report repression by amino acids even when they were added to the

medium (**Figure S3**). Probably transcriptional attenuation by *hisL* and *leuL* requires the genomic context, and cannot function on plasmids. In summary, proteome data revealed a lower expression of enzymes for five of the seven dysregulated pathways (*argA**, *trpE**, *hisG**, *leuA** and *thrA**). GFP-promoter fusions confirm this enzyme level regulation at the single cell level, and indicate that downregulation of enzymes in the *argA**, *trpE** and *thrA** mutant occurs at the transcriptional layer.

Allosteric Feedback Inhibition Enforces Enzyme Overabundance

Next, we wondered if lower expression of enzymes limits the biosynthetic capacity of the mutants. First, we tested steady state growth on glucose minimal medium and seven other carbon sources (**Figure S4**). All mutants showed wild-type like growth, except the *leuA** mutant, which grew in average 10% slower than the wild-type. To test if lower enzyme levels affect biosynthetic capacity in dynamics shifts, we starved cells for carbon and measured growth resumption on glucose minimal medium (**Figure 3A**). During the initial phase of growth resumption all mutants had the same growth rate as the wild-type. Only the *leuA**, *ilvA** and *thrA** mutants reached lower growth rates than the wild-type during the subsequent 4 hours. The three strains had also lower ODs after 20 hours starvation. Similarly, nutritional upshifts from galactose to glucose did not affect growth of the mutants (**Figure S5**). The three strains with highest reduction in enzyme levels (*argA**, *trpE** and *hisG**) grew like the wild-type in all tested conditions, indicating that biosynthetic capacity is not limited by lower enzyme level. The advantage of lower protein costs in these pathways was either too subtle to be detected by growth assays, or counterbalanced by negative effects of feedback-dysregulation.

To directly probe biosynthetic capacity, we traced intracellular fluxes of amino acids with ^{15}N labeling experiments (**Figure 3B**). Labeling of arginine, tryptophan and proline was similar in the respective mutant and the wild-type, whereas histidine, (iso)-leucine and threonine labeled slower in the mutants. However, it is important to consider that labeling rates depend on fluxes, and also on absolute pool sizes of amino acids. Because amino acid pools were higher in the mutants we used a method for quantitative analysis of the labeling profiles to estimate fluxes (Yuan et al., 2008). To account for unknown labeling profiles of upstream nitrogen precursors, we calculated fluxes for a wide range of precursor labeling rates in the literature (Yuan et al., 2006). The flux estimates show that none of the mutants had lower flux through the dysregulated pathways than the wild-type (**Figure 3B**). In most cases biosynthetic flux was even higher, indicating that downregulation of enzyme levels could not fully compensate the loss of allosteric feedback inhibition in some of the mutants. This might

be the reason for the growth-phenotype of the *leuA**, *ilvA** and *thrA** mutants in dynamic growth experiments (**Figure 3A**).

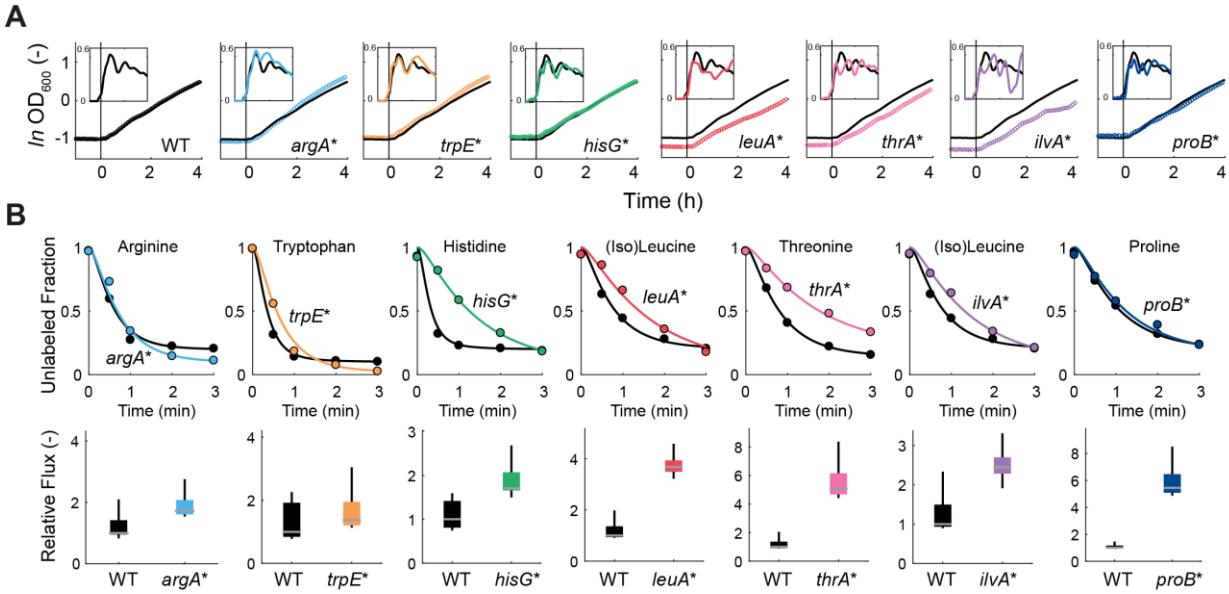


Figure 3. Growth and Biosynthetic Flux of Feedback-dysregulated *E. coli* Mutants
(A) Growth resumption after 20 hours carbon starvation of wild-type *E. coli* and the seven dysregulated mutants. Cells were starved in minimal medium and glucose was added at t = 0 h. OD was measured in 5 minute intervals in a plate reader. Shown are means of n = 3 cultures. Inserts show the specific growth rate in h^{-1} during the same time period. Growth rates were estimated by linear regression over a moving 30 minute window. The same wild-type growth curve and growth rate is shown in each graph in black as a reference. See also **Figure S4** and **S5**. (B) Decay of unlabeled amino acids in the wild-type *E. coli* (black) and the seven dysregulated mutants (color). The measured amino acid is indicated above each graph. Cells were loaded from shake flasks onto filters and perfused with ^{15}N-medium for different lengths of time (0, 30, 60, 120 and 180 seconds). Dots are means of n = 2 samples for each time point. Lines are means of 1000 fits of decay rates based on equations for kinetic flux profiling. Box plots show fluxes based on the 1000 fits, relative to the median flux estimate in the wild-type. Boxes contain 50% and whiskers 99% of the flux estimates.

In conclusion, the feedback-dysregulated mutants showed the same or higher flux through the dysregulated amino acid pathways than wild-type cells, although in five mutants (*argA**, *trpE**, *hisG**, *leuA** and *thrA**) enzyme levels in the dysregulated pathway were lower. Especially, the *argA**, *trpE** and *hisG** mutant had ~2-fold lower enzyme levels in the dysregulated pathways compared to the wild-type, while fluxes were 1-2-fold higher and growth was unaffected. This indicates that these enzymes are not operating at maximal capacity in wild-type *E. coli* during growth on glucose. We then hypothesized that this enzyme overabundance emerges from allosteric feedback inhibition by maintaining low concentration of end-products, which in turn increases production of enzymes (e.g. by de-repression of transcription). Next, we explored this interplay between control of enzyme activity and enzyme abundance and its relevance for cellular metabolism.

Interdependence of Allosteric Feedback Inhibition and Enzyme Level Regulation

To obtain a better mechanistic understanding of the interplay between allosteric feedback inhibition and enzyme level regulation, we developed a kinetic model of metabolism and enzyme expression (**Figure 4A**). Briefly, the model includes two enzymes e_1 and e_2, and two metabolites m_1 and m_2 in a two-step pathway. The end-product m_2 represents an amino acid, which is consumed in the last reaction for protein synthesis and growth. The end-product m_2 feedback inhibits the expression of both enzymes, as well as the activity of the first enzyme. The first reaction and the expression of both enzymes follow simple inhibition kinetics, whereas the second reaction follows Michaelis-Menten kinetics (**Figure 4A**). As such this model is a simplified representation of an amino acid biosynthesis pathway that is controlled at two layers (**Figure 1A**).

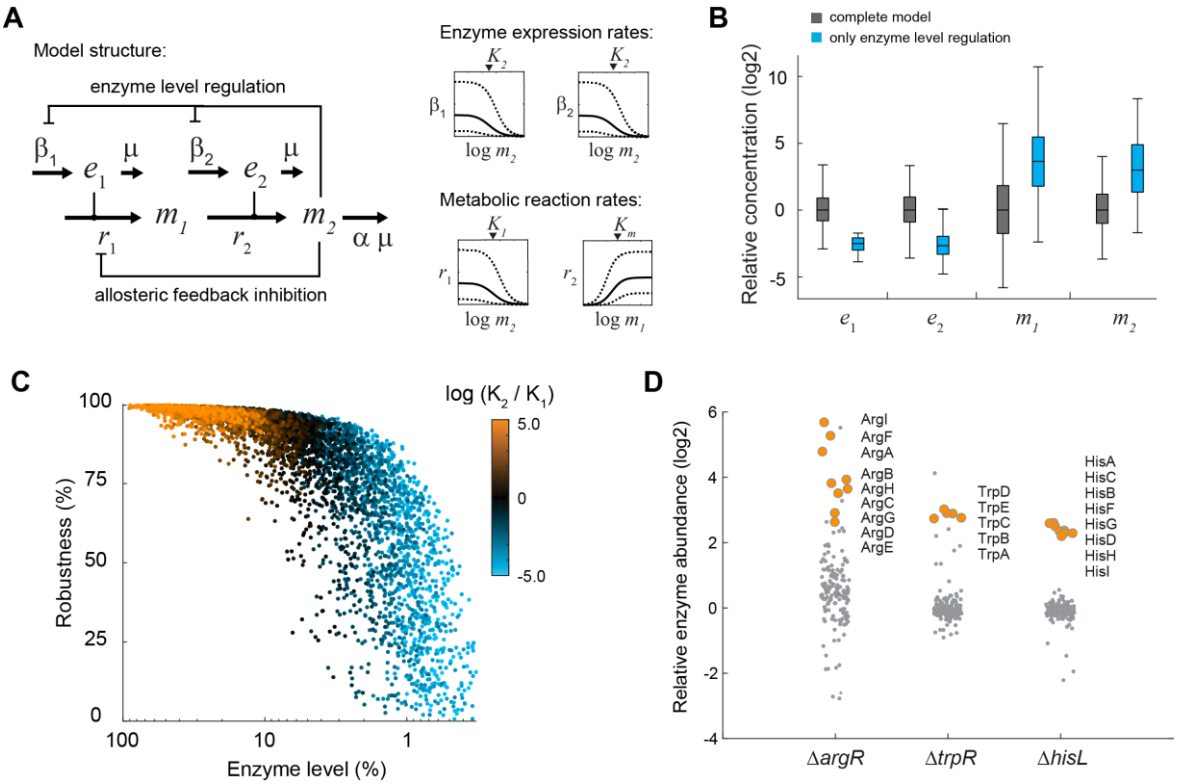


Figure 4. A Kinetic Model Predicts a Robustness-Efficiency Tradeoff
(A) Stoichiometry and structure of the kinetic model. m_1 and m_2 are metabolites, e_1 and e_2 are enzymes. Kinetics of the enzyme catalyzed reactions r_1 and r_2, as well as kinetics of enzyme expression rates β_1 and β_2 are sampled in the indicated intervals. (B) Steady state concentrations of e_1, e_2, m_1 and m_2 calculated with 5000 random parameter sets for the complete model (grey), and the model with only enzyme level regulation (blue). Boxes contain 50% and whiskers 99% of the simulated concentrations. All concentrations are normalized to the median concentrations of the complete model. See also **Figure S6** and **S7**. (C) Enzyme levels (sum of e_1 and e_2) and robustness against perturbations of $\beta_{2,max}$, for 5000 simulations of the complete model (dots). The color of each dot shows the ratio of inhibition constants for allosteric feedback inhibition (K_1) and enzyme level regulation (K_2) in the respective model. Robustness corresponds to the percentage downregulation of $\beta_{2,max}$ that was tolerated by each model. 100% enzyme abundance corresponds to the maximum theoretical enzyme concentration in the model.

As a starting point for the model analysis, we fixed the flux in the pathway to the amino acid requirement given by the growth rate of *E. coli* on glucose. We randomly sampled seven model parameters (maximal rates and binding constants) 5000 times from physiologically meaningful ranges based on literature values (Davidi and Milo, 2017; Li et al., 2014; Milo et al., 2010). For each of the thus derived 5000 parameter sets we calculated concentrations of e_1, e_2, m_1 and m_2, for a model including feedback on enzyme activity and enzyme abundance (complete model, grey in **Figure 4B**), and also for a model including only feedback on enzyme abundance (single feedback model, blue in **Figure 4B**). The simulated concentrations of e_1, e_2, m_1 and m_2 matched qualitatively the measured protein and metabolite data: the two enzymes decreased in the single feedback model (**Figure 2A**), whereas the end-product m_2 increased (**Figure 1B**). Also, the simulated concentration of the intermediate m_1 matched the measured increase of intermediates in amino acid pathways (**Figure S6**). Thus, a simple model confirms our hypothesis that allosteric feedback inhibition enforces enzyme overabundance. In theory, other types of enzyme inhibition could cause a similar increase in enzyme expression. To test this, we replaced the allosteric feedback in the model with competitive product inhibition of the second reaction (**Figure S7**). However, removing competitive product inhibition was compensated by lower substrate concentrations (m_1), and not by lower enzyme levels. This model result indicates that enzyme overabundance does not emerge from all types of enzyme inhibition.

The Interplay of two Feedbacks solves a Robustness-Efficiency Tradeoff

Next, we set out to investigate the function that emerges from the interplay between feedback on enzyme activity and enzyme abundance. While low enzyme levels are obviously advantageous due to lowering protein cost, high enzyme levels could provide a cellular benefit by improving robustness against perturbations in enzyme expression. To test this with the model, we made use of a numerical parameter continuation method to quantify robustness (Lee et al., 2014). This method iteratively decreases a model parameter until instabilities occur in the model. Robustness can then be defined as the percentage change of this parameter that was tolerated. Using this method, we calculated robustness against perturbations of the maximal expression rate of the second enzyme ($\beta_{2,max}$) in the complete model with 5000 randomly sampled parameter sets (**Figure 4C**). Changing $\beta_{2,max}$ reflects genetic or environmental perturbations of gene expression that can lead to a bottleneck in the pathway. Consistent with our expectations, models with high enzyme levels showed increased robustness, while models with lower enzyme levels were more sensitive to perturbations of enzyme expression (**Figure 4C**). However, robustness was not proportional to the enzyme level: a relatively small increase of enzyme levels already conferred a large robustness benefit. Very high enzyme levels,

in comparison, did not increase robustness substantially over more subtle changes in enzyme abundance. Our model thus reveals a tradeoff between protein costs and robustness, which can be solved by sensitively balancing enzyme levels.

The optimal balance of enzyme levels occurs in models occupying the middle of the tradeoff frontier, those models with equally strong feedback on enzyme activity and enzyme abundance (indicated by similar inhibition constants *Ki*, black dots in **Figure 4C**). We then wondered if amino acid biosynthesis in *E. coli* operates in the middle of the tradeoff frontier, meaning that both feedbacks are simultaneously active. In particular enzyme levels in the *argA*, trpE* and hisG** mutant demonstrated that wild-type *E. coli* does not operate with minimal enzyme levels in these pathways (blue dots in **Figure 4C**). To test if enzymes in these pathways are maximally expressed (orange dots in **Figure 4C**), we removed their transcriptional regulation, which functions by different mechanisms: a transcription factor (arginine), transcriptional attenuation (histidine), or both (tryptophan). In the arginine and tryptophan pathway we deleted the respective transcription factor (Δ*argR* and Δ*trpR*), and in histidine biosynthesis we removed the leader peptide *hisL*. Removing transcriptional regulation of all three pathways resulted in higher expression of enzymes in the respective pathway (**Figure 4D**): arginine enzymes increased between 5 and 60-fold, histidine enzymes about 6-fold, and tryptophan enzymes about 8-fold. This shows that *E. coli* does not operate at maximal expression of arginine, tryptophan and histidine enzymes, but rather in the middle of the tradeoff frontier. Previous studies that support this observation showed that ArgR binds to promoters of arginine genes more than 80% of the time when *E. coli* grows on glucose (Gerosa et al., 2013). Deletion of ArgR caused more global changes of amino acid enzymes than removing TrpR or HisL. This reflects the potential of ArgR to control metabolism of almost all amino acid pathways (Cho et al., 2012).

Taken together, both model and dysregulated mutants indicate a regulatory interplay in the arginine, tryptophan and histidine pathway: removing transcriptional regulation increased enzyme levels (**Figure 4D**), whereas removing allosteric regulation decreased enzyme levels (**Figure 2A**). The model shows that if feedback on enzyme activity and enzyme abundance are simultaneously active, inhibition constant of the two feedbacks must have similar values (black dots in **Figure 4C**). Inhibition constants and binding affinities in the literature show that feedbacks on enzyme activity and enzyme abundance are indeed equally strong for many amino acids (**Table S5**), corroborating the existence of a two-pronged regulation strategy.

Enzyme Overabundance Provides Robustness against Genetic Perturbations

To test if arginine, tryptophan and histidine biosynthesis are more robust against perturbations of gene expression in wild-type cells than in the feedback-dysregulated mutants, we used CRISPR interference (CRISPRi) (Larson et al., 2013). We designed single guide RNAs (sgRNA) targeting the genes *argE* in arginine biosynthesis, *hisB* in histidine biosynthesis and *trpA* in tryptophan biosynthesis. The sgRNAs were cloned on a plasmid, which harbors an inducible dCas9 and the constitutively expressed sgRNA. The three CRISPRi plasmids and a control without target sequence were transformed into the wild-type, and also into the *argA**, *trpE** and *hisG** mutants. This resulted in 16 strains with all combinations of genetic perturbations and dysregulation of the three pathways (**Figure 5A**). All strains expressing the control sgRNA without target sequence grew almost identically and also induction of dCas9 did not affect growth (**Figure 5B**).

Induction of dCas9 in strains with sgRNAs targeting *argE*, *hisB* and *trp*A reduced growth of all strains by more than 50% (**Figure 5C**). However, we observed the strongest growth defect when perturbing a gene in a dysregulated pathway. For example, CRISPRi of *argE* reduced growth of the *argA** mutant more than twice as much as the other strains. Similarly, the *hisG** and *trpE** mutant were most sensitive to perturbations of expression of *hisB* and *trpA*, respectively. The *argA**mutant was also sensitive to a perturbation of *hisB*, which matches the lower expression of histidine enzymes in this mutant (**Figure 2A**). These data confirm that feedback-dysregulated mutants are indeed more sensitive to a perturbation of gene expression. Notably, the mutants were only more sensitive to a perturbation within pathways that had lower enzyme levels, and they did not lack a general robustness.

While these data support the hypothesis that high enzyme levels render arginine, histidine and tryptophan biosynthesis more robust against perturbations of gene expression, bacteria would hardly face such strong perturbations in nature. Therefore, we designed the sgRNAs in such a way that the wild-type showed only a small growth defect without induction of dCas9 (**Figure 5D**). The mild perturbations in un-induced cultures still affected the respective mutants stronger than the other strains, causing instable growth and lower growth rates (**Figure 5D**). Thus, feedback-dysregulation renders the arginine, tryptophan and histidine pathways more sensitive against perturbations of gene expression, which may arise in nature due to the stochasticity of gene expression.

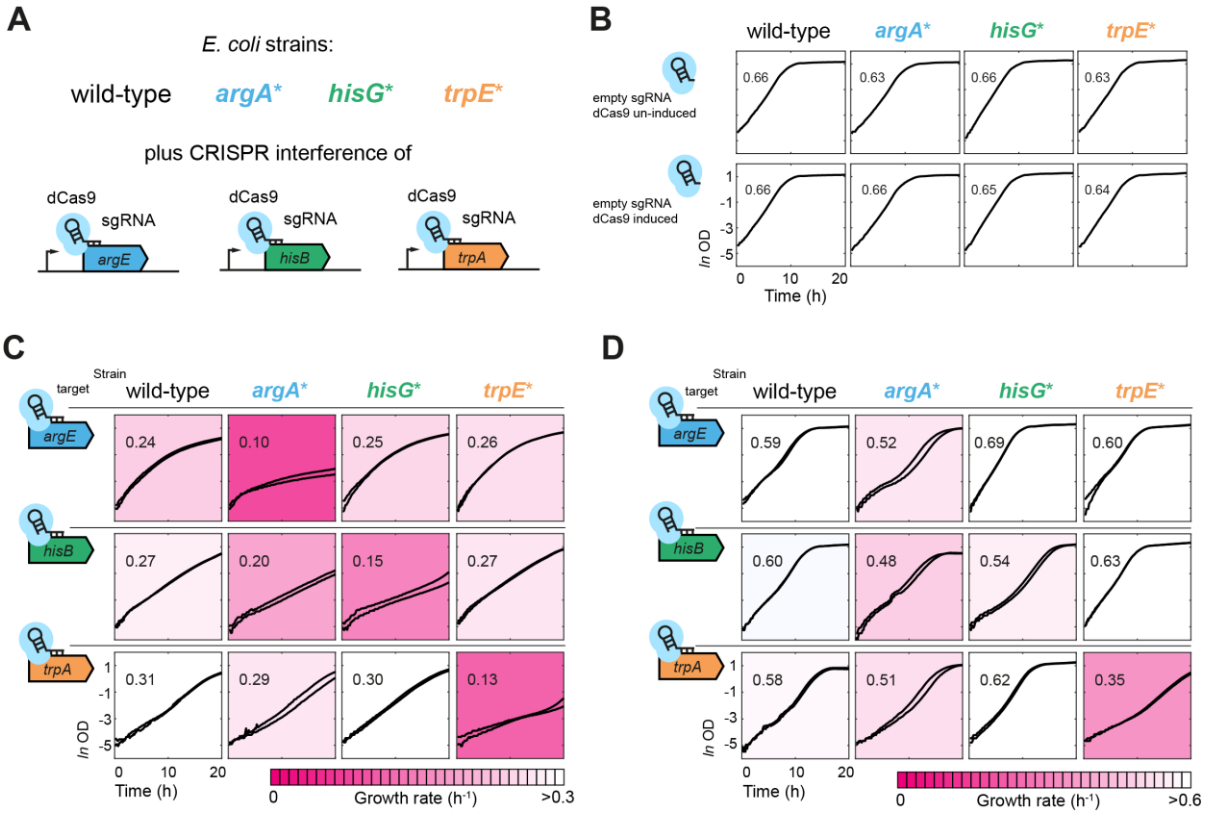


Figure 5. Enzyme Overabundance Achieves Robustness Against Perturbations of Gene Expression by CRISPR Interference. (A) CRISPR interference in wild-type cells and the allosteric feedback mutants *argA**, *hisG** and *trpE**. Strains were transformed with single guide RNAs targeting genes of the arginine (*argE*), histidine (*hisB*) and tryptophan (*trpA*) pathway, as well as an empty sgRNA without target. (B) Growth of wild-type, *argA**, *hisG** and *trpE** with the empty control sgRNA. Upper panels show un-induced cultures and lower panel induced cultures (100 µM IPTG). Growth curves show means from n=3 cultures cultivated in minimal glucose medium in a plate reader. Numbers are specific growth rates (in h^{-1}), and were estimated by linear regression between OD 0.2 and 0.6. (c) Growth of wild-type, *argA**, *hisG** and *trpE** with sgRNAs targeting *argE*, *hisB* and *trpA.* dCas9 expression was induced with 100 µM IPTG. Growth curves are means of n=3 cultures; two curves per graph show experiments that were performed at different days. Numbers and colors indicate specific growth rates (in h^{-1}), which were estimated by linear regression between 5 and 15 hours. All axes have ranges shown in the lower left graph. (D) Same as C) but without induction of dCas9. Growth rates were estimated by linear regression between OD 0.2 and 0.6. All axes have ranges shown in the lower left graph.

Discussion

In this study we explored the consequences of missing allosteric feedback inhibition in seven *E. coli* mutants with dysregulated amino acid biosynthesis pathways: arginine (*argA**), histidine, (*hisG**), tryptophan (*trpE**), leucine (*leuA**), threonine (*thrA**), isoleucine (*ilvA**), and proline (*proB**). In all mutants, the amino acid product of the feedback-dysregulated pathway increased, showing that allosteric feedback inhibition is relevant to maintain end-products at a desired level. In five mutants (*argA**, *trpE**, *hisG**, *thrA**, and *leuA**), we observed a downregulation of enzymes in the dysregulated pathways, presumably because high end-products caused stronger inhibition of enzyme expression. However, these low enzyme levels did not limit biosynthetic flux, thus indicating that wild-type cells maintain higher enzyme levels than would be necessary to ensure sufficient biosynthetic flux (enzyme

overabundance). These results are consistent with enzyme overabundance in other pathways (Davidi and Milo, 2017; O'Brien et al., 2016), and the observation that enzymes are rarely operating at maximal capacity (Fendt et al., 2010; Hackett et al., 2016).

Both model analysis and dysregulated mutants indicate that enzyme overabundance is enforced by allosteric feedback inhibition, which maintains low end-product levels and thereby increases production of enzymes. In case of amino acid biosynthesis, it is likely that low end-products de-repress transcription, because amino acid levels are known signals for transcription factors and transcriptional attenuation (Cho et al., 2012). Additionally, GFP-promoter fusions indicated regulation at the transcriptional layer in the *argA**, *trpE** and *thrA**mutant. It will be important to clarify if enzyme overabundance emerges also from other inhibitory interactions, which are abundant in metabolic networks (Alam et al., 2017). Besides inhibition of enzymes by metabolites, other sources for enzyme overabundance might be post-translational modifications. For example, it was recently shown that deleting kinases in yeast has a strong effect on enzyme levels (Zelezniak et al., 2018), pointing towards a similar interplay between post-translational modifications of enzymes and enzyme level regulation.

The strongest and most localized decrease of enzyme levels occurred when we removed allosteric feedback inhibition in the arginine, tryptophan and histidine pathway. Removing transcriptional regulation in the same pathways caused higher expression of enzymes, which is in agreement with previous reports of a role for transcriptional regulation in minimizing protein costs in metabolic pathways (Chubukov et al., 2012; You et al., 2013). This antagonistic regulation by allosteric feedback inhibition and transcriptional regulation enables an optimal balance of enzyme levels in amino acid metabolism of wild-type cells. Optimization of enzyme levels has been shown for the global *E. coli* proteome (Scott et al., 2010; You et al., 2013), for the lac system (Dekel and Alon, 2005), and for a single enzyme in the methionine pathway (Li et al., 2014). Here we provided first indication that enzyme abundance is optimized in the arginine, histidine and tryptophan pathway, to meet multiple, conflicting objectives – robustness and efficiency. Using a simplified model of amino acid metabolism, we show that cells can solve this tradeoff between protein costs and robustness through the interplay of allosteric feedback inhibition and enzyme level regulation. CRISPRi of metabolic enzymes in the dysregulated arginine, tryptophan and histidine pathways showed that allosteric feedback inhibition provides a substantial robustness benefit against perturbations of gene expression. While such robustness effects were attributed to allosteric feedback by previous modelling approaches (Chandra et al., 2011; Grimbs et al., 2007), we quantified it *in vivo* by studying mutants lacking allosteric control. During the lifetime of a cell, perturbations of gene expression could result from stochastic effects at the level of transcription or in response to fluctuating environments.

In conclusion, our case study of *E. coli* amino acid metabolism demonstrated that regulation of enzyme activity and enzyme abundance are not isolated from each other, but interact to control metabolism. Allosteric feedback inhibition sets amino acid concentrations, which are signals for enzyme level regulation. Considering the high precision of metabolite concentrations (Fuhrer et al., 2017; Mülleder et al., 2016), it seems possible that the proposed regulatory principle goes beyond *E. coli* amino acid metabolism.

Material & Methods

Experimental Model and Subject Details

Strains and Culture

E. coli MG1655 (DSMZ No. 18039) was the wild-type strain. Chemically competent *E. coli* TOP10 (One Shot™ TOP10, Invitrogen) were used for cloning. All mutants created in this study derive from the MG1655 strain and are listed in Key Resources Table. Genomic point mutations were created by scarless Cas9 Assisted recombineering (Reisch and Prather, 2015). Therefore, we constructed 7 specific sgRNA-plasmids, derived from the backbone plasmid pKDsgRNA-ack (Addgene #62654). The sgRNAs consist of a gene specific 20 base pair region (*argA*: ggtcgagggattccgccatt; trpE: acacaactggtgaaaaagcg; hisG: tggaaaaactgaaagcgctg; thrA: tggtgctgattacgcaatca; leuA: cggtaaagatgcgctgggtc; ilvA: caacacgctgggtacgtact; proB: cgacaccctgcgagcgttgc), which pairs adjacent to a NGG PAM site. Each sgRNA-plasmid was transformed together with pCas9-CR4 (Addgene #62655) into MG1655 wild-type cells. The resulting strains were grown at 30°C (pKDsgRNA-ack is temperature sensitive at 37°C) and supplemented with arabinose (final concentration 1.2 %) to induce the λ-Red recombinase genes which are located on the sgRNA-plasmid. The induced strains were then transformed with the 70-80 bp homologous oligonucleotides (**Table S2**), which contain the desired base pair exchanges of PAM site and the point mutation disrupting allosteric feedback (*argA*^{H15Y}, *trpE*^{S40F}, *hisG*^{E271K}, *thrA*^{S345F}, l*euA*^{G462D}, *ilvA*^{L447F}, *proB*^{D107N}). Cells were plated on LB agar containing 100 ng ml^{-1} anhydrotetracycline (aTc) to induce Cas9 expression, which recognizes the sgRNA adjacent to the PAM sequence and cleaves the chromosomal DNA. Only cells that successfully integrated the homologous oligonucleotides will survive due to the modified PAM sequence which prevents Cas9 recognition. Thereby we counter selected for clones harboring the desired amino acid exchanges, which were verified by sequencing. The transcriptional knockout mutants Δ*argR* and Δ*trpR* were constructed with the same cloning procedure according to the noSCAR protocol, while Δ*hisL* was constructed by P1 Phage transduction with the donor strain JW2000-1 (*ΔhisL*) from the Keio collection (Baba et al., 2006).

All cultivations were performed using M9 minimal medium with 5 g L^{-1} glucose (or the respective carbon source in **Figure S4**). The M9 medium consisted of the following components (per liter): 7.52 g Na_2HPO_4 2 H_2O, 5 g KH_2PO_4, 1.5 g $(NH_4)_2SO_4$, 0.5 g NaCl. The following components were sterilized separately and then added (per liter of final medium): 1 ml 0.1 M $CaCl_2$, 1 ml 1 M $MgSO_4$, 0.6 ml 0.1 M $FeCl_3$, 2 ml 1.4 mM thiamine-HCL and 10 ml trace salts solution. The trace salts solution contained (per liter): 180 mg $ZnSO_4$ 7 H_2O, 120 mg $CuCl_2$ 2 H_2O, 120 mg $MnSO_4$ H_2O, 180 mg $CoCl_2$ 6 H_2O. Where appropriate, 50 µg mL^{-1} kanamycin, 34 µg mL^{-1} chloramphenicol, 15 µg mL^{-1} gentamycin, 50 µg mL^{-1} spectinomycin or 100 µg mL^{-1} ampicillin was added. For cultivations in microtiter plates, LB pre-culture in 96-deep-well format plates were inoculated from glycerol stocks and grown to an exponential stage. From this first pre-culture a second M9 pre-culture in 96-deep-well plates was inoculated 1:100 and incubated overnight at 37 °C under shaking. Finally, 96-well flat transparent plates (Greiner Bio-One International) containing 150 µl M9 minimal medium were inoculated 1:150 from the overnight culture. Online measurements of optical density at 600 nm (OD_{600}) were performed at 37°C with shaking in a plate reader (Epoch, BioTek Instruments Inc, USA; Spark 10M, Tecan Trading AG, Switzerland). For induction of CRISPRi, IPTG was added to the main culture to a final concentration of 100 µM. Growth rates were calculated as dln(OD)/dt by linear regression over the indicated time windows. For cultivations in shake flask, 5 ml LB pre-culture in cultivation tubes were inoculated from glycerol stocks and grown to an exponential stage. From this first pre-culture, 5 ml of a second M9 glucose pre-culture in cultivation tubes was inoculated 1:100 and incubated overnight at 37°C in a rotary shaker. For the main culture, a 500 ml shake flask containing 35 ml M9 minimal medium (5 g L^{-1} glucose) was inoculated 1:150 from the overnight culture, and incubated at 37 °C under shaking at 220 rpm.

Method Details

CRISPR Interference

CRISPR interference experiments were performed with a single plasmid (pNUT1533) expressing the sgRNA from a constitutive and the dCas9 protein from an IPTG inducible Ptac promotor. For construction of this plasmid, the sgRNA and its constitutive promotor were amplified from the pgRNA plasmid (Addgene #44251) and the dCas9 gene was amplified from the pdCas9 plasmid (Addgene #44249). The promotor of dCas9 was replaced by an IPTG inducible Ptac promotor. To assure tight regulation of dCas9 expression, the gene coding for the lacIQ1 repressor (Glascock and Weickert, 1998) was added to the vector. The two single fragments were joined together by PCR and the resulting fragment was inserted into pNUT542 with PacI and NotI restriction enzymes (New England Biolabs, USA). This plasmid was used as a backbone for cloning of the specific plasmids targeting the arginine (pNUT1533-*argE*), histidine (pNUT1533-*hisB*) and tryptophan pathway (pNUT1533-*trpA*).

Therefore, sgRNAs guide sequences were customized by site-directed mutagenesis using the primer listed in **Table S6**.

Metabolite Measurements

Shake flask cultivations on M9 glucose were performed as described above. Cells were grown to an optical density (OD_{600}) of 0.5 and 2 mL culture aliquots were vacuum-filtered on a 0.45 µm pore size filter (HVLP02500, Merck Millipore). Filters were immediately transferred into 40:40:20 (v-%) acetonitrile/methanol/water at -20°C for extraction. Extracts were centrifuged for 15 minutes at 13,000 rpm at -9 °C. Centrifuged extracts were mixed with ^{13}C-labeled internal standard and analyzed by LC-MS/MS, with an Agilent 6495 triple quadrupole mass spectrometer (Agilent Technologies) as described previously (Guder et al., 2017). An Agilent 1290 Infinity II UHPLC system (Agilent Technologies) was used for liquid chromatography. Temperature of the column oven was 30 °C, and the injection volume was 3 µL. LC solvents A were water with 10 mM ammonium formate and 0.1% formic acid (v/v) (for acidic conditions); and water with 10 mM ammonium carbonate and 0.2% ammonium hydroxide (for basic conditions). LC solvents B were acetonitrile with 0.1% formic acid (v/v) for acidic conditions and acetonitrile without additive for basic conditions. LC columns were an Acquity BEH Amide (30 x 2.1 mm, 1.7 µm) for acidic conditions, and an iHILIC-Fusion(P) (50 x 2.1 mm, 5 µm) for basic conditions. The gradient for basic and acidic conditions was: 0 min 90% B; 1.3 min 40 % B; 1.5 min 40 % B; 1.7 min 90 % B; 2 min 90 % B. Absolute concentrations of amino acids in the ^{13}C-labeled internal standard were determined with authentic standards. Quantification of intracellular metabolite concentrations was based on the ratio of ^{12}C and ^{13}C peak heights, and a specific cell volume of 2 µL mg^{-1} was used to calculate the cell volume.

Proteomics

Shake flask cultivations on M9 glucose were performed as described above. Cells were grown to an optical density (OD_{600}) of 0.5 and 2 mL culture aliquots were transferred into 2 ml reaction tubes and washed two times with PBS buffer (0.14 mM NaCl, 2.7 mM KCL, 1.5 KH_2PO_4, 8.1 Na_2HPO_4). Cell pellets were resuspended in 300 µl lysis buffer containing 100 mM ammonium bicarbonate, 0.5 % sodium laroyl sarcosinate (SLS) and 5 mM Tris(2-carboxyethyl)phosphine (TCEP). Cells were lysed by 5 minutes incubation at 95 °C and ultra-sonication for 10 seconds (Vial Tweeter, Hielscher). Cells were again incubated for 30 minutes at 90 °C followed by alkylation with 10 mM iodoacetamide for 30 minutes at 25 °C. To clear the cell lysate, samples were centrifuged for 10 minutes at 15,000 rpm and the supernatant transferred into a new tube. Proteins in the cell lysates were digested with 1 µg trypsin (Promega) overnight at 30 °C. To remove the SLS by precipitation, trifluoroacetic acid (TFA) was added to a final concentration of 1.5 % and rested at room temperature for 10 minutes. Samples were

centrifuged for 10 minutes at 10,000 rpm and the supernatant used for C18 purification. The peptide purification was performed using the C18 microspin columns according to the manufactor's instructions (Harvard Apparatus). Eluted peptide solutions were dried and resuspended in 0.1 % TFA. The concentration of peptides in the samples was measured with a colorimetric peptide assay (Pierce™ Quantitative Colorimetric Peptide Assay, Thermo Fischer Scientific). Analysis of peptides was performed by liquid chromatography-mass spectrometry. Analysis of peptides was performed by liquid chromatography-mass spectrometry, carried out on a Q-Exactive Plus instrument connected to an Ultimate 3000 RSLC nano with a Prowflow upgrade and a nanospray flex ion source (Thermo Scientific). Peptide separation was performed on a reverse-phase HPLC column (75 µm x 42 cm) packed in-house with C18 resin (2.4 µm, Dr. Maisch GmbH, Germany). The following separating gradient was used: 98 % solvent A (0.15% formic acid) and 2 % solvent B (99.85 acetonitrile, 0.15 % formic acid) to 25 % solvent B over 105 minutes and to 35 % solvent B for additional 35 minutes at a flow rate of 300 nl/min. The data acquisition mode was set to obtain one high resolution MS scan at a resolution of 70,000 full width at half maximum (at m/z 200) followed by MS/MS scans of the 10 most intense ions. To increase the efficiency of MS/MS attempts, the charged state screening modus was enabled to exclude unassigned and singly charged ions. The dynamic exclusion duration was set to 30 seconds. The ion accumulation time was set to 50 ms for MS and 50 ms at 17,500 resolution for MS/MS. The automatic gain control was set to 3x106 for MS survey scans and 1x105 for MS/MS scans. Label-free quantification (LFQ) of the data was performed using Progenesis QIP (Waters), and for MS/MS searches of aligned peptide features MASCOT (v2.5, Matrix Science) was used. The following search parameters were used: full tryptic search with two missed cleavage sites, 10ppm MS1 and 0.02 Da fragment ion tolerance. Carbamidomethylation (C) as fixed, oxidation (M) and deamidation (N,Q) as variable modification. Progenesis outputs were further processed with SafeQuant.

Kinetic Flux Profiling

Incorporation of ^{15}N label into amino acids was measured with a filter cultivation method (Link et al., 2013). Briefly, cells were cultured on M9 glucose medium, which contains unlabeled ammonium sulfate as sole nitrogen source. At mid-exponential phase when cells reached ODs between 0.4 and 0.6, 2 mL of the culture was vacuum-filtered, and cell-loaded filters were continuously perfused with M9 glucose medium containing labeled ammonium-^{15}N sulfate. Filters were repeatedly loaded and perfused with ^{15}N-medium for different lengths of time: 0, 30, 60, 120 and 180 seconds. Subsequently, filters were immediately transferred into 40:40:20 (v-%) acetonitrile/methanol/water kept at -20 °C. Extracts were centrifuged for 15 minutes at 13,000 r.p.m. at -9 °C and the supernatant was directly used for LC-MS/MS. For LC separation of tryptophan, proline, threonine and (iso)leucine a ZIC-pHILIC column (150 x 2.1 mm, 5 µm, Merck) was used, and an Acquity BEH Amide (100 x 2.1 mm, 1.7 µm,

Waters) for LC separation of histidine and arginine. Buffers were as described for metabolite measurements and gradients were for Acquity BEH Amide: 0 min 90% B; 2.6 min 40 % B; 3 min 40 % B; 3.4 min 90 % B; 5 min 90 % B. For ZIC-pHILIC: 0 min 90% B; 4.5 min 40 % B; 5 min 40 % B; 6 min 90 % B; 8 min 90 % B. Transitions for all isotopologues per amino acid were measured by LC-MS/MS and the amount of each isotopologue was used to calculate the fraction of unlabeled amino acid F^U as:

$$F^U = \frac{M^0}{\sum_0^N M^{+i}} = \frac{Peak\ Area\ (unlabeled\ AA)}{Sum\ of\ Peak\ Area\ (all\ AA\ \text{isotopologues})}$$

Where M^0 is the amount of the unlabeled amino acid, M^{+1} is the amount of all isotopologues with one ^{15}N atom, etc. N is the number of ^{15}N atoms in the amino acid: arginine (N = 4 from 2x glutamate, 1x glutamine, 1x aspartate), tryptophan (N = 2 from 1x glutamine, 1x serine), histidine (N = 3 from ATP, 1x glutamate), threonine (N = 1 from glutamate), proline (*N* = 1 from glutamate), iso-/leucine (N = 1 from glutamate). Fluxes were estimated based on equations for kinetic flux profiling (Yuan et al., 2008), which considers the decay of the unlabeled fraction F^U:

$$F^U = \left[\frac{(1-a)(1-b)}{k_{pc}-k_{aa}}\right]\left[k_{pc}\, e^{-k_{aa}t} - k_{aa}\, e^{-k_{pc}t}\right] + [1-(1-a)(1-b)]$$

The rate constant k_{aa} is the flux into the amino acid (*flux*$_{aa}$) divided by their absolute concentration: k_{aa} = *flux*$_{aa}$ / c_{aa}. The rate constant k_{aa} was obtained by fitting the equation to the measured unlabeled fraction F^U. The rate constant k_{pc} describes labeling of upstream nitrogen precursor. Because amino acids like arginine receive ^{15}N label from several sources, the rate constant of precursor labeling k_{pc} was unknown. To account for this uncertainty the parameter k_{pc} was randomly sampled between boundaries of 0.8 min^{-1} and 14.2 min^{-1}, which are the highest and lowest first order rate constants measured for nitrogen assimilation in *E. coli* (Yuan et al., 2006). a and b consider amino acid production from degradation of protein and other macromolecules and they were estimable parameters within bounds of 0 and 0.2.

GFP-promoter Fusions

GFP reporter plasmids for detection of promotor activity of *argA*, *trpL*, *hisL* and *leuL* were obtained from a library of fluorescent transcriptional reporters for *E. coli* (Zaslaver et al., 2006). Since the original plasmids pUA66-P*hisL-gfp* and pUA66-P*leuL-gfp* lacked parts of the attenuator region, we

modified the respective promotor resulting in the plasmids pP*hisL-gfp* and pP*leuL-gfp*. Therefore, we amplified leader sequence including the rho-independent terminator of *hisL* and *leuL* from chromosomal DNA of *E. coli* MG1655 (P*hisL*: hisL_fwd_gfp ccgctcgaggctttcatcattgttgccg, hisL_rev_gfp ccgggatcccgcagaatatcaatcggc; P*leuL*: leuL_fwd_gfp ccgctcgagttgtccccttttttcctcg, leuL_rev_gfp ccgggatccgatggtttgcaccgattc). The resulting two single fragments were introduced into an empty pUA66 backbone with the restriction enzymes XhoI and BamHI. The threonine reporter plasmid which was not available in the library was constructed with the same strategy. The attenuator region of *thrL* was amplified with the primer pair thrA_fwd_gfp (ccgctcgagactgcaacgggcaatatg) and thrA_rev_gfp (ccgggatcctcggcatcgctgatattg) and the single fragment was introduced into pUA66 (XhoI and BamHI) resulting in pP*thrL-gfp*.

Flow Cytometry

Activity of the *argA*, *trpL* and *thrL* promoter was assayed using plasmid-based GFP reporters that were described in the previous section. Strains for flow cytometry were cultivated in three independent shake flasks (100 ml) containing 10 ml M9 minimal medium (5 g L^{-1} glucose; 50 µg mL^{-1} kanamycin) as described in Strains and Culture. After reaching an OD between 0.5 and 0.8 cells were diluted 1:2000 in tethering buffer (10 mM KH_2PO_4, 100 µM EDTA, 1 µM L-methionine and 10 mM lactic acid, pH=7.0) and fluorescence was measured with BD LSRFortessa SORP cell analyzer (BD Biosciences, Germany). 488-nm lasers, 600 long pass and a 520/30 band pass filters were used for detection of green fluorescence. Per sample, fluorescence of 10,000 single cells was measured. Before the measurements, cell aggregates were dispersed by vigorous mixing. BD FACSDiva software version 8.0 (BD Biosciences, NJ, USA) and FlowJo v10.4.1 (FlowJo LLC, Ashland, OR, USA) were used for analysis of the acquired data.

Purification and *In Vitro* Activity Assays of N-Acteylglutamate Synthase

E. coli BL21 cells harboring the overexpression vector pET28a(+)-*argA* respectively pET28a(+)-*argA*(H15Y) were cultivated at 37 °C (220 rpm) in 500 ml of LB medium (5 L shake flasks) containing 30 µg ml^{-1} kanamycine. When cells reached OD_{600} 0.6, the culture was shifted to 16 °C to cool down the cell broth. To induce protein expression, 10 µl of IPTG stock solution (final concentration is 10 µM) were added. The culture was incubated overnight at 16 °C (220 rpm). The cells were harvested by centrifugation at 6000 x *g* for 10 minutes at 4 °C. The supernatant was completely removed. The cell pellet was resuspended in Lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM Imidazol) (2-5 ml per gram wet weight). 50 µl protease inhibitor cocktail and 5 mg of DNAse I powder were added. Lysis of cells was performed by french press (1100 bar). The lysate was centrifuged at 4,000 x *g* for 45 minutes

at 4 °C to pellet the cellular debris. The supernatant was filtered using a 0.2-µm-pore-size syringe filter and transferred into a new collection tube. Purification was performed with columns purchased from GE Healthcare Life Science (His GraviTrap; 11-0033-99). 10 ml of equilibration buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM Imidazol) was added to the column. As soon as equilibration buffer flowed through, up to 35 ml of filtered supernatant were added to the column. The column was washed twice with 10 ml washing buffer (same as equilibration buffer). Elution of the protein was performed 3 times with 3 ml elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM Imidazol). Protein concentration of all fractions was determined (660 nm protein assay, life technologies PIERCE™ #22660). Activity of purified N-acetylglutamate-synthase (ArgA) as well as for the feedback-resistant version ArgA (H15Y) was assayed in 30 mM TRIS buffer (with 40 mM L-glutamate, 0.65 mM N-acetyl-CoA and 10 mM $MgCl_2$). To start the enzymatic reaction 10 µL of enzyme stock solution (0.15 mg/ml) was transferred to 90 µL assay buffer and mixed by pipetting up and down. To stop the reaction, 10 µL were transferred into 40 µl of 50:50 (v-%) acetonitrile/methanol at -20°C. Samples were taken every minute in a total time interval of 8 minutes. The reaction product N-acetylglutamate was measured by LC-MS and calibrated with authentic standards.

Kinetic Model

The stoichiometry of the model is shown in **Figure 4A**. Mass balancing results in the system of ordinary differential equations (ODEs), F, that is a temporal function of the state variables x and the kinetic parameters p:

$$F(x,p) = \frac{dx}{dt} = \begin{cases} \frac{dm_1}{dt} = r_1 - r_2 \\ \frac{dm_2}{dt} = r_2 - \alpha\ \mu \\ \frac{de_1}{dt} = \beta_1 - e_1\ \mu \\ \frac{de_2}{dt} = \beta_2 - e_2\ \mu \end{cases} \quad \text{(Equation 1)}$$

The five reactions (r_1, r_2, β_1, β_2, μ) are described by the following kinetic equations:

Reaction 1 is feedback inhibited by m_2 according to normal inhibition kinetics:

$$r_1 = k_{cat,1}\ e_1\ \frac{K_1}{K_1 + m_2} \quad \text{(Equation 2)}$$

In the model without allosteric regulation the equation reduces to:

$$r_1 = k_{cat,1}\, e_1 \qquad \text{(Equation 3)}$$

Reaction 2 follows Michaelis-Menten kinetics:

$$r_2 = k_{cat,2}\, e_2\, \frac{m_1}{m_1 + K_m} \qquad \text{(Equation 4)}$$

Expression rates of enzyme 1 and enzyme 2 follow inhibition kinetics

$$\beta_1 = \beta_{1,\max}\, \frac{K_2}{K_2 + m_2} \qquad \text{(Equation 5)}$$

$$\beta_2 = \beta_{2,\max}\, \frac{K_2}{K_2 + m_2} \qquad \text{(Equation 6)}$$

The growth rate depends on availability of the end-product m2:

$$\mu = \mu_{max}\, \frac{m_2}{m_2 + K_\mu} \qquad \text{(Equation 7)}$$

Dilution of metabolites by growth was not considered, due to large difference in time scales between growth dilution and metabolic flux. Dilution of enzymes by growth is included in equation 1, because the time scales of enzyme synthesis and growth dilution are closer.

Together, the kinetic equations include eight kinetic parameters k_{cat1}, k_{cat2}, $\beta_{1,max}$, $\beta_{2,max}$, K_1, K_2, K_m and α. The physiological ranges for these parameters were derived from literature values. The boundaries of enzyme turnover number ($k_{cat,1}$ and $k_{cat,2}$) are based on *in vitro* measured k_{cat} values of enzymes in amino acid biosynthesis (**Table S3**) and have values between 930 min^{-1} and 4140 min^{-1}. The maximal enzyme expression rates ($\beta_{1,max}$ and $\beta_{2,max}$) are defined by the translation rate of ribosomes according to equation 8. The equation considers the following parameters that were derived from the Bionumbers Database (Milo et al., 2010): average translation rate (r_T = 8.4 amino acids s^{-1}), the median and abundance weighted protein length (L = 209 amino acids), the fraction of active ribosomes (f_R = 0.8), the cellular volume ($V_{c,0.6}$ = 3 x 10^{-15} L) at a growth rate of μ = 0.6 h^{-1}, the Avogadro number (N_A =

6.02 x 10^{23} mol^{-1}), the amount of ribosomes per cell at that growth rate ($R_{0.6}$ = 8000 ribosomes $cell^{-1}$) and the fraction of ribosomes (p) that synthesize the enzyme:

$$\beta_{k,max} = \frac{r_t \cdot R_{0.6} \cdot f_R}{L \cdot N_A \cdot V_c} \cdot p \qquad \text{(Equation 8)}$$

The limits of $\beta_{k,max}$ are then derived by varying the fraction of ribosomes (p) that synthesize the enzymes in the pathway. According to the literature the maximal number for a single amino acid biosynthesis enzyme in *E. coli* is 7% (Li et al., 2014), therefore we set the boundaries to 1% and 10% (p = 0.01 - 0.1). The parameter limits for the Ki and Km values were set to 0.01 mM and 1 mM. The amino acid requirement (α = 86.6 mM) was a fixed parameter based on the average amino acid requirement of an *E. coli* cell (**Table S4**). We assumed that the amino acid limits the growth rate reaction only at very low concentrations. This reflects the low Km values of tRNA ligases. Therefor we fixed K_μ at a low value of 10^{-5} mM and set μ_{max} to the measured growth rate on glucose of 0.6 h^{-1}.

Steady State and Robustness Analysis

For steady state analysis a parameter set was randomly sampled from the intervals given above. With a specific parameter set the steady state concentrations of e_1, e_2, m_1 and m_2 were calculated numerically for each of the two models (complete model and single feedback model). Starting values of the numerical solver were 0.01 mM for m_1 and m_2, and 10^{-5} mM for e_1 and e_2. The convergence criterion was defined as $<10^{-8}$ change in all variables. To test stability of the steady state we calculated eigenvalues of the Jacobian matrix, and tested if all eigenvalues are negative ($\lambda < -10^{-5}$). This procedure was repeated until 5000 steady states (with different parameter sets) were achieved. Note that both models share the same parameter sets and reach the same steady state flux. In order to estimate robustness of the model against perturbations of the maximal enzyme expression rate $\beta_{2,max}$, we used a numerical parameter continuation method (Lee et al., 2014). The method is based on finding a connected path of steady state concentrations (x_{ss}: steady state concentration vector containing $e_{1,ss}$, $e_{2,ss}$, $m_{1,ss}$, $m_{2,ss}$), as a parameter, p, is varied. As the system is in steady state it follows that:

$$\frac{dx}{dt} = F(x_{SS}, p) = 0 \qquad \text{(Equation 9)}$$

The derivative of $F(x_{SS}, p)$ with respect to the parameters is also zero:

$$\frac{dF(x_{SS}, p)}{dp} = \frac{\delta F}{\delta x_{SS}} \cdot \frac{dx_{SS}}{dp} + \frac{\delta F}{\delta p} = 0 \qquad \text{(Equation 10)}$$

After rearranging Equation 10, Equation 11 is obtained:

$$\frac{dx_{SS}}{dp} = -\left(\frac{\delta F}{\delta x_{SS}}\right)^{-1} \cdot \frac{\delta F}{\delta p} \qquad \text{(Equation 11)}$$

which describes the changes in the steady-state concentrations as a kinetic parameter is varied iteratively. The iteration stops when one of the following three stability criteria is no longer fulfilled. 1st criterion: all real parts of the eigenvalues of the system's Jacobian need to be negative. This implies stability of a steady state. Furthermore, in equation 11 the inverse of the Jacobian Matrix ($\delta F/\delta x_{SS}$) is required. The inversion is only possible as long as the matrix is regular. Once an eigenvalue reaches zero, the Jacobian becomes singular and matrix inversion is no longer possible. This bifurcation point defines the boundary between the stable and unstable parameter space. In other words: after this point is passed, the system no longer returns to a stable steady state. By checking the eigenvalues of the Jacobian at each step, we make sure that the iteration is terminated when one eigenvalue becomes bigger than $\lambda = -10^{-5}$. 2nd criterion: all variables are required to be positive. 3rd criterion: a model is considered unstable when a certain time limit (t > 1 s) is exceeded, which can be the case when numerical errors occur during the numerical integration process. The maximum theoretical enzyme amount in the model was calculated as:

$$0 = \beta_{i,max} - e_{i,max}\,\mu \qquad \text{(Equation 12)}$$

After rearranging equation 12 and substituting the upper parameter bound of the maximum protein translation rate ($\beta_{i,max}^{ub}$), the maximum theoretical enzyme amount of each enzyme is:

$$e_{i,max} = \frac{\beta_{i,max}^{ub}}{\mu} = \frac{8.5 \cdot 10^{-4}\ mM\ min^{-1}}{0.01\ min^{-1}} = 0.085\ mM \qquad \text{(Equation 13)}$$

Considering that the model includes two enzymes, the maximum amount of total enzyme is 0.17 mM, which was defined as the maximal enzyme level (100%).

Quantification and Statistical Analysis

Statistical analysis was done with MATLAB. The statistical details of each experiment can be found in the respective figure caption. For proteomics and metabolomics n represents the number of independent shake flask cultures. In growth assays, n represents the number of independent microtiter plate cultures. For *in vitro* assays, n represents the number of independent reaction vessels.

Software

All codes for model analysis are available in the Github repository: https://github.com/nfarke/Sander_et_al.

Supplemental Information

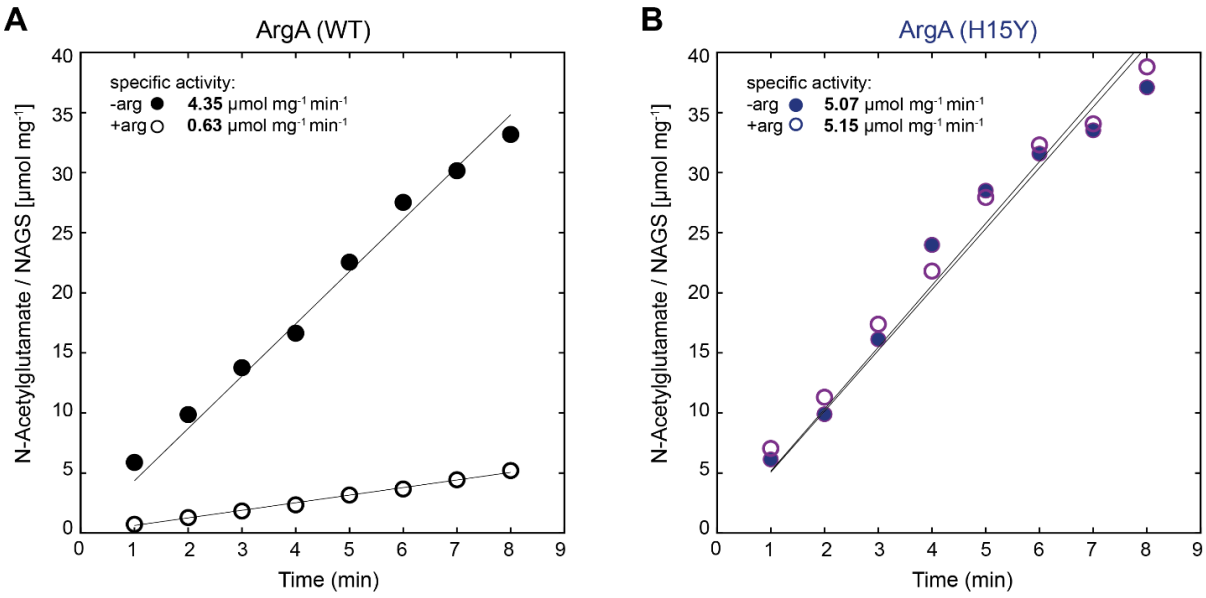


Figure S1. Related to Figure 1; *In vitro* kinetics of N-acetylglutamate-synthase (NAGS) from *E. coli* (ArgA) in the **A** native and **B** the allosteric feedback resistant version ArgA (H15Y). Dots represent means from n=2 independent assays (filled = no arginine; empty = 1 mM arginine). Activity of His-tagged purified enzymes was assayed in 30 mM TRIS buffer (40 mM L-glutamate, 0.65 mM Acetyl-CoA and 10 mM $MgCl_2$). For sampling 10 µL of reaction solution was transferred into 40 µL of 50:50 (v-%) acetonitrile/methanol at -20°C. The reaction product N-acteylglutamate was measured by LC-MS/MS. Specific activity in [µmol mg^{-1} min^{-1}] was calculated from linear regression through the 8 time points.

Figure S2. Related to Figure 1; Relative concentrations of 110 intracellular metabolites in wild-type *E. coli* and seven dysregulated mutants (n = 3).

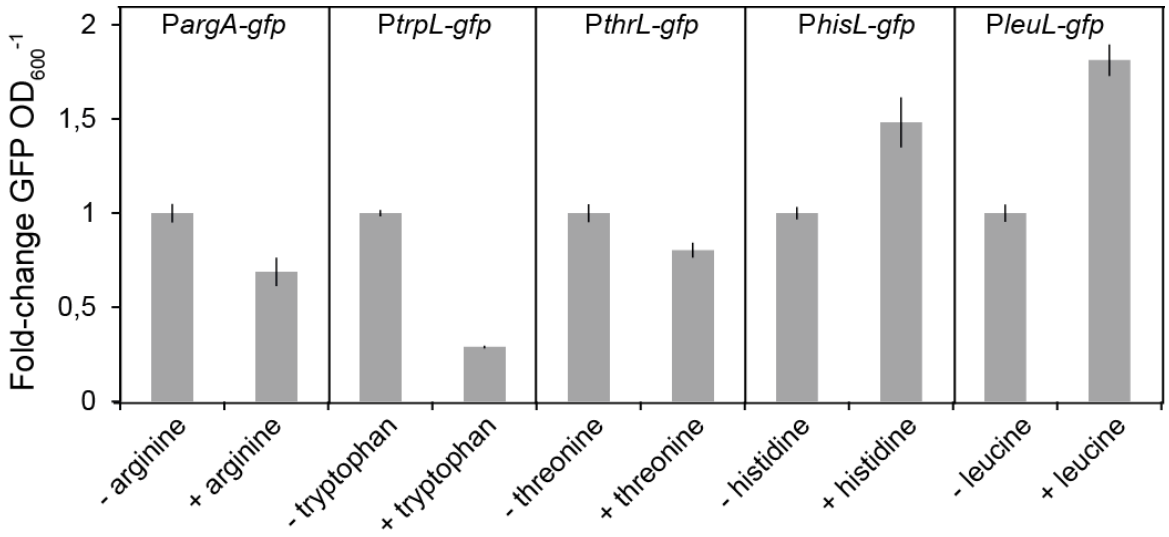


Figure S3. Related to Figure 2; GFP expression of promotor fusions P*argA-gfp*, P*trpL-gfp*, P*thrL-gfp*, P*hisL-gfp* and P*leuL-gfp* in *E. coli* wild-type with and without addition of external amino acids. Bar plots show fold-changes of GFP per OD_{600} relative to the condition without external amino acids (n=3). Cells were grown in M9 minimal medium (5 g L^{-1} glucose) and GFP expression was measured in mid-exponential phase at OD_{600} ~0.5 with a plate reader. Amino acids were supplemented to a final concentration of 2 mM.

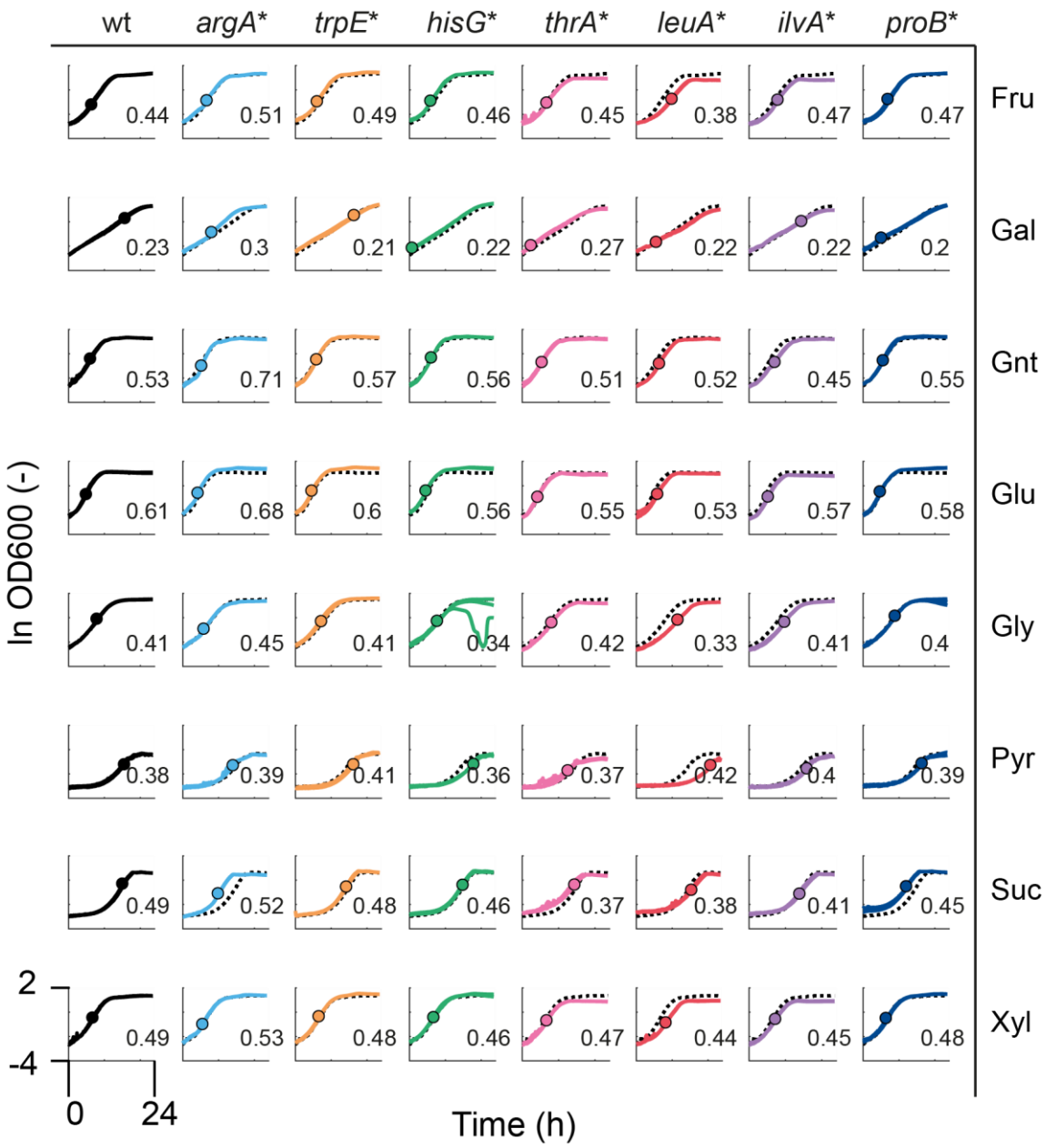


Figure S4. Related to Figure 3; Growth of wild-type *E. coli* and 7 mutants (see also Figure 1A) on fructose (Fru), galactose (Gal), gluconate (Gnt), glucose (Glu), glycerol (Gly), pyruvate (Pyr), succinate (Suc), and xylose (Xyl). Shown are three cultivations in microtiter plates. The dashed line is the mean of the wild-type in the particular condition (n = 3). Numbers are the maximal growth rates in h^{-1}, which is reached at the time indicated by dots. All x-axes range from 0 to 24 hours. All y-axes range from -4 to 2 ($lnOD_{600}$).

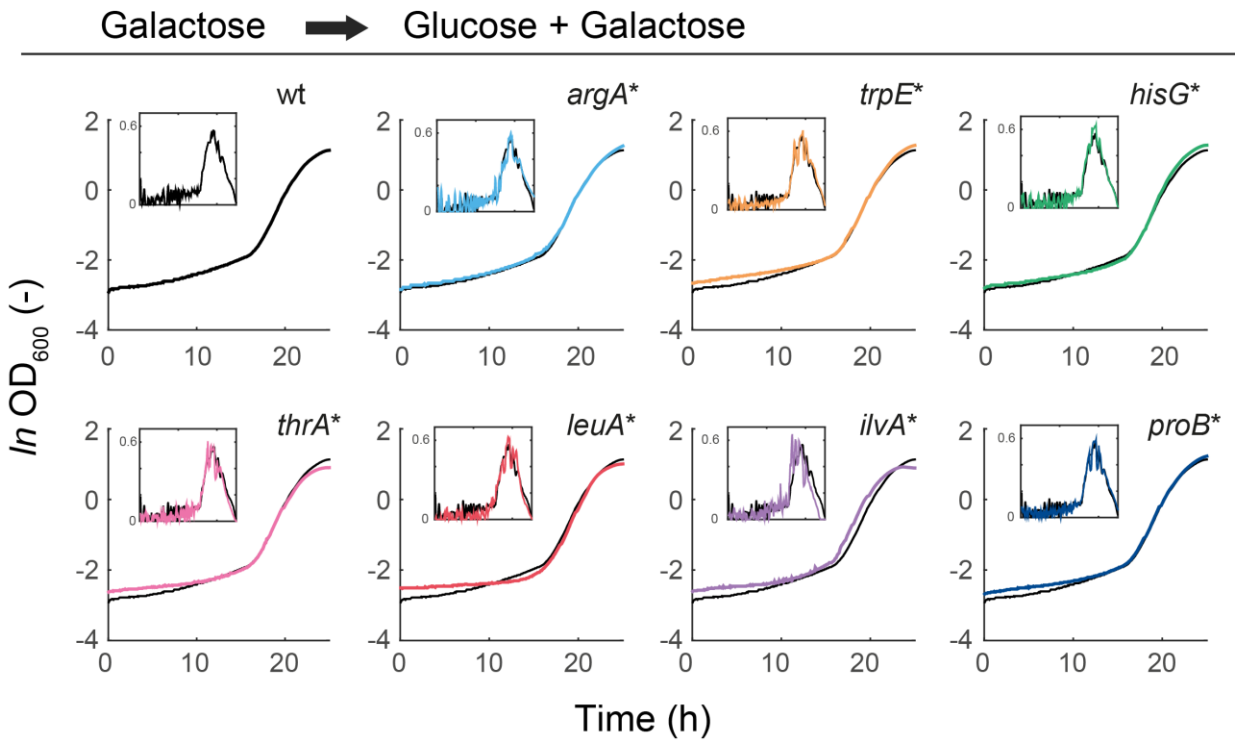


Figure S5. Related to Figure 3; Growth of wild-type *E. coli* and the seven dysregulated mutants in shifts from galactose to glucose. For up-shifts from galactose to glucose, cells were grown in M9 minimal medium with 5 g L^{-1} galactose and glucose was added to a final concentration of 5 g L^{-1} at an OD of 0.1. Shown are means of n = 3 cultures. Inserts show the growth rate during the same time period. Growth rates were estimated by linear regression over a moving 30-minute window. The same wild-type growth curve is shown in each graph in black as a referenc

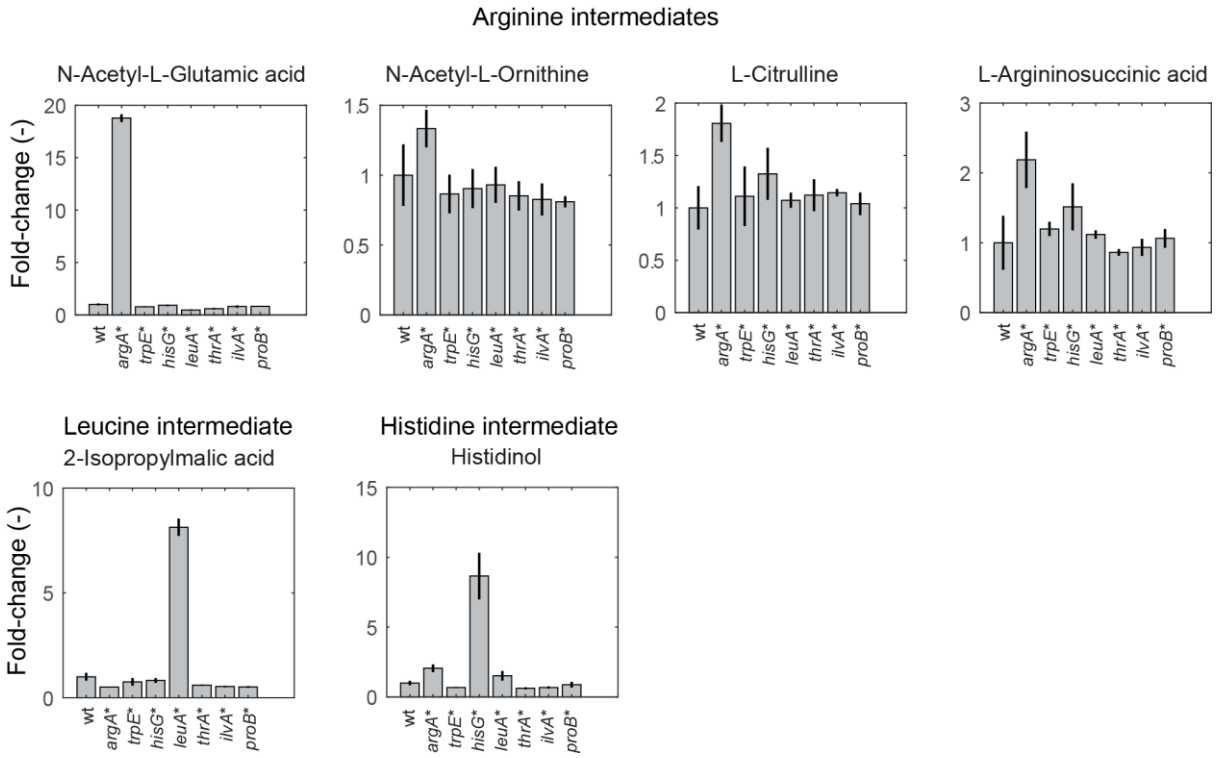


Figure S6. Related to Figure 4; Intermediates in dysregulated pathways measured by LC-MS in wild-type *E. coli* and seven dysregulated mutants (n = 3).

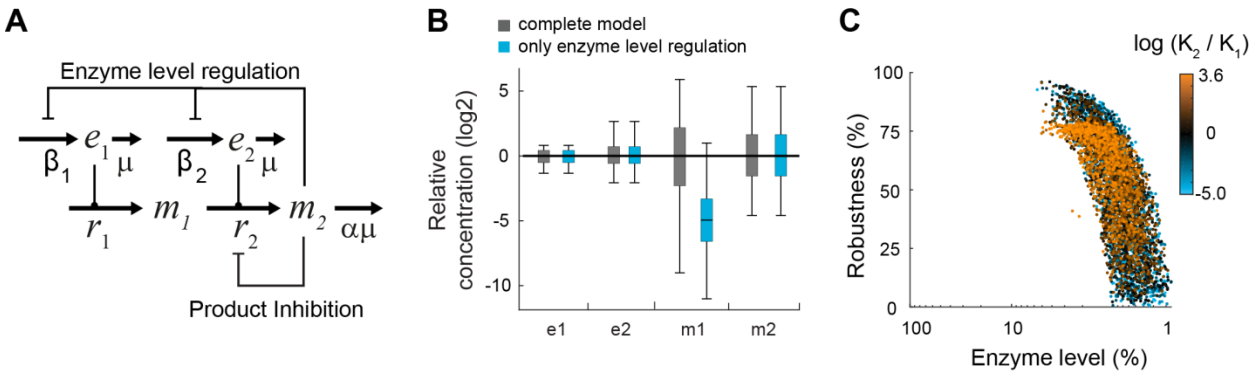


Figure S7. Related to Figure 4;

(A) Model with product inhibition, instead of allosteric feedback inhibition. Metabolite 2 inhibits reaction 2 by competitive product inhibition, which was modelled using the following equation:

$$r_2 = k_{cat,2} \cdot e_2 \cdot \frac{m1}{m1 + Km \cdot \left(1 + \frac{m2}{K1}\right)}$$

(B) Steady state concentrations of e_1, e_2 m_1 and m_2 calculated with 5000 simulations for the complete model (grey), and the model with only enzyme level regulation (blue). Boxes contain 50% and whiskers 99% of the simulated concentrations. All concentrations are normalized to the median concentrations of the complete model.

(C) Enzyme levels (sum of e_1 and e_2) and robustness against perturbations of $\beta_{2,max}$ for 5000 simulations of the complete model (dots). The color of each dot shows the ratio of inhibition constants for allosteric feedback inhibition (K1) and enzyme level regulation (K2) in the respective model. Robustness corresponds to the percentage downregulation of $\beta_{2,max}$ that was tolerated by each model. 100% enzyme abundance corresponds to the maximum theoretical enzyme concentration in the model.

Table S1. Related to Figure 1; Mutations in allosteric enzymes that were investigated in this study.

Pathway	Gene	Enzyme	Mutation	Reference
L-arginine biosynthesis	*argA*	N-acetylglutamate synthase	H15Y	Rajagopal et al., 1998
L-isoleucine biosynthesis	*ilvA*	Threonine deaminase	L447F	LaRossa et al., 1987
L-histidine biosynthesis	*hisG*	ATP phosphoribosyl transferase	E271K	Doroshenko et al., 2013
L-leucine biosynthesis	*leuA*	2-isopropylmalate synthase	G462D	Gusyatiner et al., 2002
L-proline biosynthesis	*proB*	Glutamate-5-kinase	D107N	Csonka et al., 1988
L-threonine biosynthesis	*thrA*	Aspartate kinase	S345F	Lee et al., 2003
L-tryptophan biosynthesis	*trpE*	Anthranilate synthase	S40F	Caligiuri and Bauerle, 1991

Table S2. Related to Figure 1; Oligonucleotides for recombineering

Gene	Oligonucleotides for recombineering (5'-3')	Protospacer sequence (5'-3')
argA	GTGGTAAAGGAACGTAAAACCGAGTTGGTCGAGGGAT TCCGCTATTCAGTTCCCTATATCAATACCCACCGGGGAA	GGTCGAGGGATT CCGCCATT
ilvA	GGAATCACCGGGCGCGTTCCTGCGCTTTCTCAACACG CTGGGTACGTACTGGAACATTTCTTTGTTCCACTATCG	CAACACGCTGG GTACGTACT
hisG	GTCAGCAGCAAAACCCTGTTCTGGGAAACTATGGAAA AACTGAAAGCGCTGGGGCCAGTTCAATTCTGGTCCTG	TGGAAAAACTGA AAGCGCTG
leuA	CTGGTGAAATACAGCCTGACCGCCAAAGGACACGGTA AAGATGCGCTGGATCAGGTGGATATCGTCGCTAACTAC	CGGTAAAGATGC GCTGGGTC
proB	ACCCGTGCTAATATGGAAGACCGTGAACGCTTCCTGAACGCTCGCGACAC CCTGCGAGCGTTGCTCGATAACAATATC	CGACACCCTGCG AGCGTTGC
thrA	GCGCGCGTCTTTGCAGCGATGTCACGCGCCCGTATTT TCGTGGTGCTGATTACGCAATCATCTTCCGAATACAGC	TGGTGCTGATTA CGCAATCA
trpE	CTTATCGCGACAATCCCACTGCGCTTTTTCACCAGTTGTGTGGGGATCGTC CGGCAACGCTGCTGCTGGAATTCGCAGAT	CGCTTTTTCACC AGTTGTGT

Table S3. Related to Figure 4; Literature k_{cat} values for enzymes in amino acid biosynthesis. The values were collected from the BRENDA database, and from Davidi and Milo, 2017. - indicates that no value could be found in both sources. The 25th and 75th quartiles of these k_{cat} values are 930 min^{-1} and 4140 min^{-1}, respectively.

Name	k_{cat}, s^{-1}	Name	k_{cat}, s^{-1}	Name	k_{cat}, s^{-1}
argA	654.00	cysK	378.50	ilvN	40.00
argB	-	cysM	24.00	leuA	-
argC	-	cysN	-	leuB	69.00
argD	-	cysQ	11.00	leuC	-
argE	1800.00	dadX	33.66	leuD	-
argF	-	dapA	104.00	lysA	33.00
argG	-	dapB	382.00	lysC	22.13
argH	-	dapD	36.00	metA	22.00
argI	-	dapE	-	metB	121.00
aroA	32.00	dapF	84.00	metC	34.10
aroB	14.00	gdhA	37.00	metE	3.50
aroC	39.00	glnA	33.00	metH	-
aroD	75.00	gltB	-	metL	-
aroE	237.00	gltD	-	pheA	32.00
aroF	-	glyA	10.00	proA	10.00
aroG	4.20	hisA	7.20	proB	53.00
aroH	-	hisB	-	proC	717.00
aroK	-	hisC	-	prs	-
aroL	-	hisD	12.00	serA	29.00
asd	-	hisF	-	serB	-
asnA	-	hisG	-	serC	1.80
asnB	4.50	hisH	-	thrA	-
aspC	-	hisI	-	thrB	17.00
avtA	-	ilvA	-	thrC	-
cysC	50.00	ilvB	38.50	trpA	-
cysD	-	ilvC	0.30	trpB	-
cysE	772.00	ilvD	69.00	trpC	18.77
cysH	-	ilvE	-	trpE	-
cysI	47.00	ilvH	-	tyrA	71.00
cysJ	-	ilvI	-	tyrB	-

Table S4. Related to Figure 4; Amino acid requirements of *E. coli* (Monk et al., 2017). The mean of 86.6 mM was used as parameter α in the model.

Amino Acid	Coefficients, mmol g_{dw}^{-1}	alpha, mM
ala-L	0.499	166.4
arg-L	0.287	95.8
asn-L	0.234	78.1
asp-L	0.234	78.1
cys-L	0.089	29.7
gln-L	0.256	85.2
glu-L	0.256	85.2
gly	0.595	198.4
his-L	0.092	30.7
ile-L	0.282	94.1
leu-L	0.438	145.9
lys-L	0.333	111.1
met-L	0.149	49.8
phe-L	0.180	60.0
pro-L	0.215	71.6
ser-L	0.210	69.9
thr-L	0.247	82.2
trp-L	0.055	18.4
tyr-L	0.134	44.7
val-L	0.411	137.1
Mean	**0.260**	**86.6**

Table S5. Related to Figure 4; Inhibition constants of allosteric enzymes (K_i-value), transcriptional attenuation (tRNA-ligase K_m-value) and metabolite-transcription factor interactions (K_d-value). Values were obtained from EcoCyc (Keseler et al., 2017), Brenda (Schomburg et al., 2002) or RegulonDB (Gama-Castro et al., 2016). When more than one value was available, an upper and a lower bound are given. The grey background indicates the seven pathways that were investigated during this work. The Ki of ArgA was measured in this work with *in vitro* assays.

Biosynthesis pathway	Allosteric Feedback		K_i mM		Transcriptional Feedback			$K_{m/d}$ mM	
	Enzyme	Metabolite	LB	UB	Mechanism	Protein	Metabolite	LB	UB
Arginine	ArgA	arg	0.15		Repressor	ArgR	arg	0.28	
Asparagine	AsnA	asn	0.12		Repressor	AsnC	asn	1	
Cysteine	CysE	cys	0.001						
Histidine	HisG	his	0.012	0.1	Attenuation	his-tRNA ligase	his	0.008	0.03
Isoleucine	IlvA	ile	0.06		Attenuation	ile-tRNA ligase	ile	0.0036	1.3
Leucine	LeuA	leu	0.28		Attenuation	leu-tRNA ligase	leu	0.0015	0.05
Lysine	DapA	lys	0.21	3.9					
Methionine	MetA	met	0.1	4	Repressor	MetJ	sam	0.01	0.05
Phenylalanine	PheA	phe	0.1	0.6		TyrR	phe	>0.18	
Proline	ProB	pro	0.02						
Serine	SerA	ser	0.005	0.37					
Threonine	ThrA	thr	0.097	0.167	Attenuation	thr-tRNA ligase	thr	0.11	0.2
Tryptophan	TrpE	trp	0.17		Repressor	TrpR	trp	0.16	
Tryptophan	TrpE	trp	0.17		Attenuation	trp-tRNA ligase	trp	0.017	
Tyrosine	TyrA	tyr	0.1		Repressor	TyrR	tyr	0.18	
Valine	IlvB	val	0.078	0.1	Attenuation	val-tRNA ligase	val	0.0043	0.1

Table S6. Oligonucleotides used in this study.

Oligonucleotide	Sequence (5'-3')	Description
argA_Forward	GGTCGAGGGATTCCGCCATTG TTTTAGAGCTAGAAATAGCAAG	Forward primer used with **CPEC001** for amplification of fragment 1 for customzied pKDsgRNA targeted against argA
argA_Reverse	AATGGCGGAATCCCTCGACCG TGCTCAGTATCTCTATCACTGA	Reverse primer used with **CPEC002** for amplification of fragment 2 for customzied pKDsgRNA targeted against argA
ilvA_Forward	AGTACGTACCCAGCGTGTTGG TTTTAGAGCTAGAAATAGCAAG	Forward primer used with **CPEC001** for amplification of fragment 1 for customzied pKDsgRNA targeted against ilvA
ilvA_Reverse	CAACACGCTGGGTACGTACTG TGCTCAGTATCTCTATCACTGA	Reverse primer used with **CPEC002** for amplification of fragment 2 for customzied pKDsgRNA targeted against ilvA
hisG_Forward	CAGCGCTTTCAGTTTTTCCAGT TTTAGAGCTAGAAATAGCAAG	Forward primer used with **CPEC001** for amplification of fragment 1 for customzied pKDsgRNA targeted against hisG
hisG_Reverse	TGGAAAAACTGAAAGCGCTGG TGCTCAGTATCTCTATCACTGA	Reverse primer used with **CPEC002** for amplification of fragment 2 for customzied pKDsgRNA targeted against hisG
leuA_Forward	GACCCAGCGCATCTTTACCGG TTTTAGAGCTAGAAATAGCAAG	Forward primer used with **CPEC001** for amplification of fragment 1 for customzied pKDsgRNA targeted against leuA
leuA_Reverse	CGGTAAAGATGCGCTGGGTCG TGCTCAGTATCTCTATCACTGA	Reverse primer used with **CPEC002** for amplification of fragment 2 for customzied pKDsgRNA targeted against leuA
proB_Forward	GCAACGCTCGCAGGGTGTCGG TTTTAGAGCTAGAAATAGCAAG	Forward primer used with **CPEC001** for amplification of fragment 1 for customzied pKDsgRNA targeted against proB
proB_Reverse	CGACACCCTGCGAGCGTTGCG TGCTCAGTATCTCTATCACTGA	Reverse primer used with **CPEC002** for amplification of fragment 2 for customzied pKDsgRNA targeted against proB
thrA_Forward	TGATTGCGTAATCAGCACCAG TTTTAGAGCTAGAAATAGCAAG	Forward primer used with **CPEC001** for amplification of fragment 1 for customzied pKDsgRNA targeted against thrA
thrA_Reverse	TGGTGCTGATTACGCAATCAG TGCTCAGTATCTCTATCACTGA	Reverse primer used with **CPEC002** for amplification of fragment 2 for customzied pKDsgRNA targeted against thrA
trpE_Forward	ACACAACTGGTGAAAAAGCGG TTTTAGAGCTAGAAATAGCAAG	Forward primer used with **CPEC001** for amplification of fragment 1 for customzied pKDsgRNA targeted against trpE
trpE_Reverse	CGCTTTTTCACCAGTTGTGTG TGCTCAGTATCTCTATCACTGA	Reverse primer used with **CPEC002** for amplification of fragment 2 for customzied pKDsgRNA targeted against trpE
argR_Forward	ATTCTTCAATGGACTGGAGGG TTTTAGAGCTAGAAATAGCAAG	Forward primer used with **CPEC001** for amplification of fragment 1 for customzied pKDsgRNA targeted against argR
argR_Reverse	CCTCCAGTCCATTGAAGAATGT GCTCAGTATCTCTATCACTGA	Reverse primer used with **CPEC002** for amplification of fragment 2 for customzied pKDsgRNA targeted against argR
CPEC001	TTTATAACCTCCTTAGAGCTCGA	Reverse primer for amplification of fragment 1 for pKDsgRNA
CPEC002	CCAATTGTCCATATTGCATCA	Forward primer for amplification of fragment 2 for pKDsgRNA
Ec-F	GTTTTAGAGCTAGAAATAGCAAGTTAA AATAAGGC	Foward primer used with **guide_Rev** for ampflification of customized pNUT1533-ctrl
Ec-F-argE-mm5	TTTTTCATTGTTGACACCCCTCGTTTTAG AGCTAGAAATAGCAAGTTAAAATAAGG C	Foward primer used with **guide_Rev** for ampflification of customized pNUT1533-argE
Ec-F-trpA	TTCTTTGCGCTCCTTCAACTGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGC	Foward primer used with **guide_Rev** for ampflification of customized pNUT1533-trpA
Ec-F-hisB	TCACTCGGCGGTTCGCTAATCAGTTTTA GAGCTAGAAATAGCAAGTTAAAATAAG GC	Foward primer used with **guide_Rev** for ampflification of customized pNUT1533-hisB
Ec-R	ACTAGTATTATACCTAGGACTGAGCTA GC	Reverse primer for amplification of customized pNUT1533 plasmids
ArgA_fwd_NdeI	TGACCATATGATGGTAAAGGAACGTAA AAC	Amplification of genomic argA
ArgA_rev_BamHI	TGACGGATCCTTACCCTAAATCCGCCAT CA	Amplification of genomic argA
ArgA_H15Y_fwd	AGGGAACCGAATAGCGGAATCCCTC	Forward primer for amplification pET28a(+)-argA
ArgA_H15Y_rev	ATATCAATACCCACCGGG	Reverse primer for amplification pET28a(+)-argA
hisL_fwd_gfp	CCGCTCGAGGCTTTCATCATTGTTGCCG	Forward primer for amplification of hisL attenuator region
hisL_rev_gfp	CCGGGATCCCGCAGAATATCAATCGGC	Reverse primer for amplification of hisL attenuator region
leuL_fwd_gfp	CCGCTCGAGTTGTCCCCTTTTTCCTCG	Forward primer for amplification of leuL attenuator region
leuL_rev_gfp	CCGGGATCCGATGGTTTGCACCGATTC	Reverse primer for amplification of leuL attenuator region
thrA_fwd_gfp	CCGCTCGAGACTGCAACGGGCAATATG	Forward primer for amplification of thrL attenuator region
thrA_rev_gfp	CCGGGATCCTCGGCATCGCTGATATTG	Reverse primer for amplification of thrL attenuator region

References

Alam, M.T., Olin-Sandoval, V., Stincone, A., Keller, M.A., Zelezniak, A., Luisi, B.F., and Ralser, M. (2017). The self-inhibitory nature of metabolic networks and its alleviation through compartmentalization. Nat. Commun. *8*, 16018.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. a, Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. *2*, 2006.0008.

Caligiuri, M.G., and Bauerle, R. (1991). Subunit communication in the anthranilate synthase complex from *Salmonella typhimurium*. Science *252*, 1845–1848.

Chandra, F.A., Buzi, G., and Doyle, J.C. (2011). Glycolytic Oscillations and Limits on Robust Efficiency. Science *333*, 187.

Cho, B.-K., Barrett, C.L., Knight, E.M., Park, Y.S., and Palsson, B.Ø. (2008). Genome-scale reconstruction of the Lrp regulatory network in *Escherichia coli*. Proc. Natl. Acad. Sci. *105*, 19462–19467.

Cho, B.-K., Federowicz, S., Park, Y.-S., Zengler, K., and Palsson, B.Ø. (2012). Deciphering the transcriptional regulatory logic of amino acid metabolism. Nat. Chem. Biol. *8*, 65.

Christodoulou, D., Link, H., Fuhrer, T., Kochanowski, K., Gerosa, L., and Sauer, U. (2018). Reserve Flux Capacity in the Pentose Phosphate Pathway Enables *Escherichia coli*'s Rapid Response to Oxidative Stress. Cell Syst. *6*, 569-578.e7.

Chubukov, V., Zuleta, I.A., and Li, H. (2012). Regulatory architecture determines optimal regulation of gene expression in metabolic pathways. Proc. Natl. Acad. Sci. *109*, 5127–5132.

Chubukov, V., Uhr, M., Le Chat, L., Kleijn, R.J., Jules, M., Link, H., Aymerich, S., Stelling, J., and Sauer, U. (2013). Transcriptional regulation is insufficient to explain substrate-induced flux changes *in Bacillus subtilis.* Mol. Syst. Biol. *9*, 709.

Chubukov, V., Gerosa, L., Kochanowski, K., and Sauer, U. (2014). Coordination of microbial metabolism. Nat. Rev. Microbiol. *12*, 327–340.

Csonka, L.N., Gelvin, S.B., Goodner, B.W., Orser, C.S., Siemieniak, D., and Slightom, J.L. (1988). Nucleotide sequence of a mutation in the proB gene of *Escherichia coli* that confers proline overproduction and enhanced tolerance to osmotic stress. Gene *64*, 199–205.

Daran-Lapujade, P., Rossell, S., van Gulik, W.M., Luttik, M.A.H., de Groot, M.J.L., Slijper, M., Heck, A.J.R., Daran, J.-M., de Winde, J.H., Westerhoff, H.V., et al. (2007). The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. Proc. Natl. Acad. Sci. *104*, 15753–15758.

Davidi, D., and Milo, R. (2017). Lessons on enzyme kinetics from quantitative proteomics. Curr. Opin. Biotechnol. *46*, 81–89.

Dekel, E., and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. Nature *436*, 588–592.

Doroshenko, V.G., Lobanov, A.O., and Fedorina, E.A. (2013). The directed modification of *Escherichia coli* MG1655 to obtain histidine-producing mutants. Appl. Biochem. Microbiol. *49*, 130–135.

Fendt, S.-M., Buescher, J.M., Rudroff, F., Picotti, P., Zamboni, N., and Sauer, U. (2010). Tradeoff between enzyme and metabolite efficiency maintains metabolic homeostasis upon perturbations in enzyme capacity. Mol. Syst. Biol. *6*, 356.

Fuhrer, T., Zampieri, M., Sévin, D.C., Sauer, U., and Zamboni, N. (2017). Genomewide landscape of gene–metabolome associations in *Escherichia coli*. Mol. Syst. Biol. *13*, 907.

Gama-Castro, S., Salgado, H., Santos-Zavaleta, A., Ledezma-Tejeida, D., Muñiz-Rascado, L., García-Sotelo, J.S., Alquicira-Hernández, K., Martínez-Flores, I., Pannier, L., Castro-Mondragón, J.A., et al. (2016). RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. Nucleic Acids Res. *44*, D133-143.

Gerosa, L., Kochanowski, K., Heinemann, M., and Sauer, U. (2013). Dissecting specific and global transcriptional regulation of bacterial gene expression. Mol. Syst. Biol. *9*, 658.

Glascock, C.B., and Weickert, M.J. (1998). Using chromosomal lacIQ1 to control expression of genes on high-copy-number plasmids in *Escherichia coli*. Gene *223*, 221–231.

Goyal, S., Yuan, J., Chen, T., Rabinowitz, J.D., and Wingreen, N.S. (2010). Achieving Optimal Growth through Product Feedback Inhibition in Metabolism. PLoS Comput. Biol. *6*, e1000802.

Grimbs, S., Selbig, J., Bulik, S., Holzhütter, H.-G., and Steuer, R. (2007). The stability and robustness of metabolic states: identifying stabilizing sites in metabolic networks. Mol. Syst. Biol. *3*, 146.

Guder, J.C., Schramm, T., Sander, T., and Link, H. (2017). Time-Optimized Isotope Ratio LC-MS/MS for High-Throughput Quantification of Primary Metabolites. Anal. Chem. *89*, 1624–1631.

Gusyatiner, M.M., Ivanovskaya, L.V., Kozlov, Y.I., Lunts, M.G., and Voroshilova, E.B. (2005). DNA coding for mutant isopropylmalate synthase, l-leucine-producing microorganism and method for producing l-leucine.

Hackett, S.R., Zanotelli, V.R.T., Xu, W., Goya, J., Park, J.O., Perlman, D.H., Gibney, P.A., Botstein, D., Storey, J.D., and Rabinowitz, J.D. (2016). Systems-level analysis of mechanisms regulating yeast metabolic flux. Science *354*.

Hartl, J., Kiefer, P., Meyer, F., and Vorholt, J.A. (2017). Longevity of major coenzymes allows minimal de novo synthesis in microorganisms. Nat. Microbiol. *2*, 17073.

Hirasawa, T., and Shimizu, H. (2016). Recent advances in amino acid production by microbial cells. Curr. Opin. Biotechnol. *42*, 133–146.

Hofmeyr, J.-H.S., and Cornish-Bowden, A. (2000). Regulating the cellular economy of supply and demand. FEBS Lett. *476*, 47–51.

Kacser, H., and Burns, J.A. (1973). The control of flux. Symp. Soc. Exp. Biol. *27*, 65–104.

ter Kuile, B.H., and Westerhoff, H.V. (2001). Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. FEBS Lett. *500*, 169–171.

LaRossa, R.A., Dyk, T.K.V., and Smulski, D.R. (1987). Toxic accumulation of alpha-ketobutyrate caused by inhibition of the branched-chain amino acid biosynthetic enzyme acetolactate synthase in *Salmonella typhimurium*. J. Bacteriol. *169*, 1372–1378.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat. Protoc. *8*, 2180–2196.

Lee, J.-H., Lee, D.-E., Lee, B.-U., and Kim, H.-S. (2003). Global analyses of transcriptomes and proteomes of a parent strain and an L-threonine-overproducing mutant strain. J. Bacteriol. *185*, 5442–5451.

Lee, Y., Lafontaine Rivera, J.G., and Liao, J.C. (2014). Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways. Metab. Eng. *25*, 63–71.

Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell *157*, 624–635.

Link, H., Kochanowski, K., and Sauer, U. (2013). Systematic identification of allosteric protein-metabolite interactions that control enzyme activity *in vivo*. Nat. Biotechnol. *31*, 357–361.

Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010). BioNumbers—the database of key numbers in molecular and cell biology. Nucleic Acids Res. *38*, D750–D753.

Monk, J.M., Lloyd, C.J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., et al. (2017). iML1515, a knowledgebase that computes *Escherichia coli* traits. Nat. Biotechnol. *35*, 904–908.

Mori, M., Schink, S., Erickson, D.W., Gerland, U., and Hwa, T. (2017). Quantifying the benefit of a proteome reserve in fluctuating environments. Nat. Commun. *8*.

Mülleder, M., Calvani, E., Alam, M.T., Wang, R.K., Eckerstorfer, F., Zelezniak, A., and Ralser, M. (2016). Functional Metabolomics Describes the Yeast Biosynthetic Regulome. Cell *167*, 553-565.e12.

O'Brien, E.J., Utrilla, J., and Palsson, B.O. (2016). Quantification and Classification of *E. coli* Proteome Utilization and Unused Protein Costs across Environments. PLoS Comput. Biol. *12*, e1004998.

Rajagopal, B.S., DePonte, J., Tuchman, M., and Malamy, M.H. (1998). Use of inducible feedback-resistant N-acetylglutamate synthetase (argA) genes for enhanced arginine biosynthesis by genetically engineered *Escherichia coli* K-12 strains. Appl. Environ. Microbiol. *64*, 1805–1811.

Reaves, M.L., Young, B.D., Hosios, A.M., Xu, Y.-F., and Rabinowitz, J.D. (2013). Pyrimidine homeostasis is accomplished by directed overflow metabolism. Nature *500*, 237–241.

Reisch, C.R., and Prather, K.L.J. (2015). The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli*. Sci. Rep. *5*, 15096.

Reznik, E., Christodoulou, D., Goldford, J.E., Briars, E., Sauer, U., Segrè, D., and Noor, E. (2017). Genome-Scale Architecture of Small Molecule Regulatory Networks and the Fundamental Trade-Off between Regulation and Enzymatic Activity. Cell Rep. *20*, 2666–2677.

Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., and Heinemann, M. (2016). The quantitative and condition-dependent *Escherichia coli* proteome. Nat. Biotechnol. *34*, 104–110.

Schomburg, I., Chang, A., Placzek, S., Söhngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., et al. (2013). BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic Acids Res. *41*, D764-72.

Schuster, S., and Heinrich, R. (1987). Time hierarchy in enzymatic reaction chains resulting from optimality principles. J. Theor. Biol. *129*, 189–209.

Scott, M., Gunderson, C.W., Mateescu, E.M., Zhang, Z., and Hwa, T. (2010). Interdependence of cell growth and gene expression: origins and consequences. Science *330*, 1099–1102.

Umbarger, H.E. (1956). Evidence for a Negative-Feedback Mechanism in the Biosynthesis of Isoleucine. Science *123*, 848–848.

Yanofsky, C. (1981). Attenuation in the control of expression of bacterial operons. Nature *289*, 751–758.

You, C., Okano, H., Hui, S., Zhang, Z., Kim, M., Gunderson, C.W., Wang, Y.-P., Lenz, P., Yan, D., and Hwa, T. (2013). Coordination of bacterial proteome with metabolism by cyclic AMP signalling. Nature *500*, 301–306.

Yuan, J., Fowler, W.U., Kimball, E., Lu, W., and Rabinowitz, J.D. (2006). Kinetic flux profiling of nitrogen assimilation in *Escherichia coli*. Nat. Chem. Biol. *2*, 529–530.

Yuan, J., Bennett, B.D., and Rabinowitz, J.D. (2008). Kinetic flux profiling for quantitation of cellular metabolic fluxes. Nat. Protoc. *3*, 1328–1340.

Zaslaver, A., Mayo, A.E., Rosenberg, R., Bashkin, P., Sberro, H., Tsalyuk, M., Surette, M.G., and Alon, U. (2004). Just-in-time transcription program in metabolic pathways. Nat. Genet. *36*, 486–491.

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M.G., and Alon, U. (2006). A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. Nat. Methods *3*, 623–628.

Zelezniak, A., Vowinckel, J., Capuano, F., Messner, C.B., Demichev, V., Polowsky, N., Mülleder, M., Kamrad, S., Klaus, B., Keller, M.A., et al. (2018). Machine Learning Predicts the Yeast Metabolome from the Quantitative Proteome of Kinase Knockouts. Cell Syst. *7*, 269-283.e6.

Conclusion and Outlook

Key Findings

Chapter 1

A pooled CRISPRi screen, covering all metabolic genes of *E. coli*, enabled us to systematically search for bottleneck genes in metabolism and it revealed the consequences of lowering enzyme levels below wild-type levels. Surprisingly, the cell was able to buffer most knockdowns for hours after the induction of CRISPRi. Only 253 genes showed a growth defect below 14 hours for at least two guides, while the different targeting positions of the guides hardly affected the repression strength. Only 11 genes which showed a metabolic bottleneck did not carry flux on M9 medium with glucose or are known to be essential, however 3 of them can be explained by polar effects because an essential or flux-carrying gene is encoded downstream of the target gene. On average, the response time for the 253 bottleneck genes was 7.8 hours, while the 7 most sensitive targets (*ilvE/ilvD, ppc, sucA, lpxC, cysD, pyrG* and *nrdA/nrdB*) showed a response time below 4 hours. Hypothetically, these genes encode enzymes that catalyze rate-limiting steps, which is why they are expressed near-critical levels.

To understand how *E. coli* buffers the decrease of enzymes for such a long time, we measured the metabolome and proteome of 30 arrayed CRISPRi strains. We found three gene-specific buffering mechanisms: i) the knockdown of the carbamoyl phosphate synthetase (CarAB) is buffered by ornithine, which increases CarAB activity, ii) S-adenosylmethionine de-repressed the gene expression of the methionine pathway and counteracts the knockdown of homocysteine transmethylase (MetE) and iii) the knockdown of 6-phosphogluconate dehydrogenase (Gnd) is buffered by 6-phosphogluconate which activates a bypass.

Hence, the application of CRISPRi in a pooled library screening revealed bottlenecks in metabolism and the investigation of single strains identified local regulatory mechanisms.

Chapter 2

The metabolite measurements of chapter 1 showed that in CRISPRi knockout strains the substrate of the targeted reaction accumulates while the product level decreases. Hence, we used this inducible method to change the metabolite levels inside the cell and combined it with a reporter for transcription factor activity to systematically search for functional interactions *in vivo*.

We first combined the pooled CRISPRi library with a reporter plasmid with binding sites of ArgR. We were able to sort cells from the library which showed an increase of green fluorescence upon the induction of CRISPRi. After arraying these strains, we could show that 78% had a target in arginine biosynthesis, and 8% had off-targets for this pathway. The data indicated that the decreasing arginine levels in these strains led to a release of the arginine repressor from the promoter and therefore resulted in higher GFP levels, as expected.

We used the same approach to screen for CRISPRi targets whose changing metabolite levels had an effect on a reporter plasmid for the activity of Cra, which is a global transcriptional regulator of metabolism. For a long time, it was heavily discussed which effector metabolite binds and regulates Cra (Bley Folly et al., 2018; Chavarría and de Lorenzo, 2018). Here, we show the effect of a *fruK* knockdown on Cra in an *in vivo* context.

Thus, we can use CRISPRi as a tool to increase and decrease metabolite levels *in vivo*, which allows us to measure their effect on transcription factor activity and enables us to identify known and new interactions.

Chapter 3

This chapter is a resource and provides a detailed protocol for designing and constructing pooled CRISPRi libraries and gives a detailed protocol for designing and constructing libraries, performing fitness assays combined with Illumina sequencing and sorting single cells by FACS. New data or findings were not in the scope or focus of this chapter.

Chapter 4

The inference of dynamic metabolite and gene expression data enabled us to find new metabolite-transcription factor interactions. By switching *E. coli* from growth to glucose limitation, we were able to measure dynamically changing metabolites and transcript levels. In total, we measured 123 metabolites by LC-MS/MS and 4242 transcripts for 29 (35) different time points. The application of network component analysis and the good knowledge about the transcriptional regulatory network in *E. coli* allowed us to calculate the activity of 209 transcription factors and to correlate this data with metabolite levels. Using Hill-type kinetics, we were able to correlate the levels of cAMP with the transcription factor CRP, which is a known interaction and held as a validation of our approach. Performing the same correlation for all transcription factors and metabolites enabled us to predict putative regulatory metabolites of 71 transcriptional regulators, of which we validated 5 by *in vitro* binding affinity assays: dihydroxyacetone phosphate-DhaR, tyrosine-TrpR, glutamate-SgrR, tryptophan-SoxR and lysine-ArgR.

Hence, the external perturbation of the cell can also be applied to identify new metabolite-transcription factor interactions.

Chapter 5

Microbes have to face internal and external perturbations, and developed several mechanisms to buffer disruptions of their metabolism. As shown in chapter 1, this robustness can be mediated by the interaction of metabolites with proteins. We removed the allosteric feedback regulation of seven amino acid pathways in *E. coli* metabolism and measured decreasing enzyme levels of five of the seven dysregulated pathways. A decrease in metabolic flux was not observed. This indicated that enzyme levels in cells must be higher than needed to maintain metabolic homeostasis. Using a metabolic model and CRISPRi for the pathways of arginine, histidine and tryptophan biosynthesis, we showed that an overabundance of enzymes can improve the robustness of a pathway.

Thus, we could demonstrate that the fine-tuned interplay of enzyme-level regulation and allosteric feedback inhibition ensures a robust but also efficient synthesis of enzymes involved in arginine, histidine and tryptophan pathways in *E. coli*. Hence, CRISPRi can also be used to identify the function of regulatory mechanisms.

Outlook

In this thesis, we highlight the potential to investigate the regulation of metabolism via CRISPRi. Its application to map interactions between metabolites and transcriptional regulators (chapter 2) and to identify their function (chapter 5) opens new possibilities for the investigation of the regulatory network in *E. coli* and other bacteria. It enables us to find and examine interactions *in vivo*, which is why this method minimizes the identification of non-functional interactions. Undoubtedly, the conditions that are used to perform the assays play an important role for the outcome of the screenings and a combination of internal (CRISPRi, chapter 1,2 and 5) and external perturbations (chapter 4) could even raise the number of found interactions.

In chapter 2, we show that, for the transcription factors ArgR and Cra, the combination of a pooled CRISPRi library with a fluorescent transcriptional reporter allows us to identify already known and new metabolite- transcription factor interactions. The screening condition was based on the growth of *E. coli* on M9 minimal medium with glucose as the sole carbon source. As some pathways are only activated under certain conditions (Shimizu, 2013; Zampieri et al., 2019), we expect that by performing this assay with a different carbon source we would find many new PMIs, which could have been missed before. Moreover, some TFs like ArgR do not only act as repressors but also activators. Hence, changing

conditions could reveal new interactions with so far unknown metabolites that only activate or repress the TF under these circumstances.

As shown in chapter 4, another interesting approach could be to switch the CRISPRi library with a fluorescent reporter plasmid for TF activity between conditions and to take samples dynamically. Thus, it could be possible to identify interactions that for example only apply under the change of a carbon source. This method is not only limited to transcriptional regulators, but also a known binding site of any other protein could be fused with a fluorescent reporter to measure the influence of a metabolite on the binding activity. Furthermore, cells could be treated with cold-shocks, antibiotics or oxidative stress to test for interactions that are only active under stress conditions. As shown in chapter 1, these environments cannot only be applied to identify metabolite- transcription factor interactions, but also to find bottleneck genes for certain conditions. Hence, the before mentioned settings can also be used in fitness assays with CRISPRi libraries without the reporter plasmid, which is especially interesting for identifying target genes for new antibiotics (Jiang et al., 2020). Moreover, CRISPRi libraries may well also be used to investigate microbial communities. Microbial interaction is often achieved by the exchange of metabolites between species (Braga et al., 2016; D'Souza et al., 2018). However, the exact interaction network is often not completely determined (Blasche et al., 2021). Thus, interfering metabolism of one species via CRISPRi could help to identify which metabolite leads to a specific phenotype of another species.

In conclusion, the implementation of new tools like CRISPRi enables us to not only identify the function of genes, but also to reconstruct the way bacterial metabolism is regulated and to find the specific function and key signals of these regulations. Moreover, systematic techniques as proteomics, metabolomics and NGS are constantly developing, so a combination of all these approaches might help us to gain a full picture of bacterial metabolism and to determine the function of every single component.

References

Blasche, S., Kim, Y., Mars, R.A.T., Machado, D., Maansson, M., Kafkia, E., Milanese, A., Zeller, G., Teusink, B., Nielsen, J., et al. (2021). Metabolic cooperation and spatiotemporal niche partitioning in a kefir microbial community. Nature Microbiology *6*, 196–208.

Bley Folly, B., Ortega, A.D., Hubmann, G., Bonsing-Vedelaar, S., Wijma, H.J., van der Meulen, P., Milias-Argeitis, A., and Heinemann, M. (2018). Assessment of the interaction between the flux-signaling metabolite fructose-1,6-bisphosphate and the bacterial transcription factors CggR and Cra: interaction between the flux-signaling metabolite fructose-1,6-bisphosphate and the bacterial transcriptional factors. Molecular Microbiology *109*, 278–290.

Braga, R.M., Dourado, M.N., and Araújo, W.L. (2016). Microbial interactions: ecology in a molecular perspective. Brazilian Journal of Microbiology *47*, 86–98.

Chavarría, M., and de Lorenzo, V. (2018). The imbroglio of the physiological Cra effector clarified at last. Molecular Microbiology *109*, 273–277.

D'Souza, G., Shitut, S., Preussger, D., Yousif, G., Waschina, S., and Kost, C. (2018). Ecology and evolution of metabolic cross-feeding interactions in bacteria. Natural Product Reports *35*, 455–488.

Jiang, W., Oikonomou, P., and Tavazoie, S. (2020). Comprehensive Genome-wide Perturbations via CRISPR Adaptation Reveal Complex Genetics of Antibiotic Sensitivity. Cell *180*, 1002-1017.e31.

Shimizu, K. (2013). Metabolic Regulation of a Bacterial Cell System with Emphasis on *Escherichia coli* Metabolism. ISRN Biochemistry *2013*, 1–47.

Zampieri, M., Hörl, M., Hotz, F., Müller, N.F., and Sauer, U. (2019). Regulatory mechanisms underlying coordination of amino acid and glucose catabolism in Escherichia coli. Nature Communications *10*.

Acknowledgements

The adventure of my PhD is a chapter in the book of my life, which would not have been half as exciting and pleasant without all the main characters that accompanied me along the way.

First, I would like to thank **Hannes Link** for the opportunity to join his lab and to work with him. When I first visited your lab, only friendly, happy and satisfied PhD students have welcomed me. Now, after 3 years of working with you, I feel the same. You always know when to guide us and when to give us the freedom to find our way. You always had an open door, listened to me and after talking to you I had new energy to continue my way. I wish you all the best for the new start in Tübingen and hope you can continue like this.

Furthermore, I want to thank my thesis advisory committee members **Tobias Erb** and **Andreas Diepold** for the productive discussions, and **Lennart Randau** and **Roland Lill** for agreeing to join my thesis committee.

I am also grateful for all the former and current group members of the AG Link. Thank you, **Vanessa, Martin, Timur and Stefano,** for the pleasant time inside and outside the lab. Thank you, **Thorben, Dominik** and **Niklas,** for the great scientific discussions. Thank you, **Paul**, for the french spirit in the lab and your vibrant performances and thank you, **Ying,** for sharing all your travel experiences. Thank you, **Ale,** for the nice time via remote, **Andi,** for being such a great student and company in Marburg when we were left on our own, and **Chris,** for spending breaks together at the snack vending machine.

I also want to thank all the people making all the paperwork and organizational matters way easier: **Melissa, Sarah, Ali, Claudi, Inka, Manuel, David, Christina, Manuela, Jens, everyone from the Einkauf, Zrinka and Dušica**. Moreover, I would like to thank all the **collaborators and friends** from other labs and the people of the **IMPRS community**. Also, many thanks to the **IMPRS graduate school** for funding parts of my PhD.

Finally, I also want to thank all of the people who accompanied me outside the lab…

…**Tolga and Fran**: 3 people from 3 different countries finding each other in Marburg. Thank you for two amazing trips and all the evenings we spent together. You showed me a world that I did not know before and you were always there in moments when I needed someone to cheer me up, to celebrate or just to do nothing. I cannot wait for the next years with you.

…**all my friends and family** who supported me during this time. I am especially grateful for my parents who have made my studies possible and always believe in me.

…to all the amazing **women** who raised me, are part of my family or my friends. You show me every day how marvelous you are and that we can achieve anything we want to.

Publications of this thesis

Parts of this thesis that were published in peer-reviewed journals during the preparation of this thesis:

Chapter 1:

Kuntz M*, Donati S*, Pahl V, Farke N, Beuter D, Glatter T, Gomes-Filho JV, Randau L, Wang CY, Link H. Multi-omics Analysis of CRISPRi-Knockdowns Identifies Mechanisms that Buffer Decreases of Enzymes in E. coli Metabolism. Cell Syst. 2021 Jan 20;12(1):56-67.e6. doi: 10.1016/j.cels.2020.10.011. Epub 2020 Nov 24. PMID: 33238135.

*authors contributed equally

Chapter 4:

Lempp, M., Farke, N., **Kuntz, M.** et al. Systematic identification of metabolites controlling gene expression in E. coli. Nat Commun 10, 4463 (2019). https://doi.org/10.1038/s41467-019-12474-1

Chapter 5:

Sander T, Farke N, Diehl C, **Kuntz M**, Glatter T, Link H. Allosteric Feedback Inhibition Enables Robust Amino Acid Biosynthesis in E. coli by Enforcing Enzyme Overabundance. Cell Syst. 2019;8(1):66-75.e8. doi: 10.1016/j.cels.2018.12.005

Abgrenzung der Eigenleistung

Die in dieser Arbeit präsentierten Ergebnisse wurden von mir eigenständig und nur mit Hilfe der in den Kapiteln aufgeführten Personen angefertigt.

Kapitel 1: Mein Anteil an der Studie umfasst die Ausführung der Experimente mit der gepoolten CRISPRi library, als auch das Testen einzelner CRISPRi Stämme in Wachstumsexperimenten. Des Weiteren habe ich die Illumina Sequenzierung durchgeführt und die Proben entsprechend vorbereitet, als auch die Daten ausgewertet. Außerdem habe ich die entsprechenden Experimente geplant, als auch das Manuskript mitverfasst.

Kapitel 2: Mein Anteil and diesem Kapitel beinhaltet das Aufsetzen der Studie, die Ausführung der Wachstumsexperimente mit der CRISPRi library, die Durchführung des Cell-Sortings, als auch die Probenentnahme für Proteommessungen. Des Weiteren habe ich die Analyse der Daten durchgeführt und das Manuskript mitverfasst.

Kapitel 3: Mein Anteil an diesem Protokoll umfasst die Durchführung der Wachstumsexperimente mit der CRISPRi library, als auch das Cell-Sorting. Des Weiteren habe ich alle Anleitungen zur Durchführung der Experimente verfasst bzw. gesammelt und anschließend als Manuskript zusammengefasst.

Kapitel 4: Mein Anteil an der Studie umfasst die Implementierung zur Aufreinigung von Proteinen, als auch die experimentelle Aufreinigung von Proteinen zur Ausführung von Bindeassays.

Kapitel 5: Mein Anteil an diesem Kapitel umfasst die Klonierung von Plasmiden mit sgRNAs zur Anwendung von CRISPRi in allosterischen disregulierten Mutanten und die Ausführung von Wachstumsexperimenten mit diesen Stämmen.

Eigenständigkeitserklärung

Hiermit erkläre ich, dass die vorgelegte Dissertation von mir selbst und ausschließlich mit den angegebenen Hilfen verfasst, keine anderen als die angegebenen Quellen benutzt und alle übernommenen Zitate als solche gekennzeichnet wurden.

Diese Dissertation wurde in der vorliegenden oder einer ähnlichen Form noch bei keiner anderen in- oder ausländischen Hochschule anlässlich eines Promotionsgesuchs oder zu anderen Prüfungszwecken eingereicht.

Ort, Datum	Michelle Kuntz