Understanding metabolic robustness of *Escherichia coli* using genetic and environmental perturbations

Dissertation

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer.nat.)

des Fachbereichs Biologie der Philipps-Universität Marburg

Vorgelegt von

Stefano Donati

Aus Mailand, Italien

Marburg (Lahn), 2020

Die vorliegende Arbeit wurde in der Zeit von Oktober 2016 bis Juli 2020 unter der Betreuung von Dr. Hannes Link am Max-Planck-Institut für terrestrische Mikrobiologie in Marburg angefertigt.

Erstgutachter:	Dr. Hannes Link	
Zweitgutachter:	Prof. Lennart Randau	
Weitere Mitglieder der Prüfungskommission:		Prof. Victor Sourjik
		Prof. Tobias Erb

Vom Fachbereich für Biologie der Philipps-Universität Marburg als Dissertation angenommen am

17/07/2020

Mündliche Prüfung am

24/08/2020

Die während der Promotien erzielten Erbebnisse wurden zum Teil in folgenden Originalpublikationen veröffentlicht:

Stefano Donati, Timur Sander, Hannes Link. Crosstalk between transcription and metabolism: how much enzyme is enough for a cell? WIREs Systems Biology and Medicine 2018, 10:e1396. DOI: 10.1002/wsbm.1396.

Nicht veröffentlichte Arbeiten:

Stefano Donati, Michelle Kuntz, Vanessa Pahl, Dominik Beuter, Timo Glatter, José Vicente Gomes Filho, Lennart Randau, Hannes Link. The metabolome buffers CRISPRi-knockdowns of enzymes in *E. coli* metabolism. Under review, Cell Systems.

Dušica Radoš, Stefano Donati, Martin Lempp, Hannes Link. Homeostasis of the *Escherichia coli* biosynthetic metabolome across different environments.

Dedicato alla mia famiglia e ai miei amici

Table of Contents

List of Figures

List of Supplementary Figures

List of Supplementary Tables

Abbreviations

Δ	gene deletion
% (v/v)	percent per volume
% (w/v)	percent per volume
Amp	Ampicillin
aTc	Anhydrotetracycline
ATP	Adenosine triphosphate
bp	basepairs
Cas protein	CRISPR associted protein
Cmp	Chloramphenicol
CRISPR	Clustered Regulary Interspaced Short Palindromic Repeats
CRISPRi	CRISPR interference
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
dsDNA	Doublestranded DNA
et al.	*et alii*, and others
FBA	Flux balance analysis
g	gramm
GFP	Green gluorescent protein
h	Hours
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Km	Kanamycin
Kb	Kilobases
L	Liter
LB medium	Lysogeny broth medium
LC	Liquid chromatography
Ln	Natural logarithm
Log2	Binary logarithm
M	Molar (mol/L)
min	Minutes
mRNA	Messenger RNA
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
µ	Micro (10^{-6})
n	Nano (10^{-9})
N-terminal	Amino-terminal
nt	Nucleotides
OD600nm	Optical density at 600 nm
PAM	Protospacer adjacent motive
PCR	Polymerase chain reaction
pH	Potential of hydrogen
RNA	Ribonucleic acid
rpm	Revolutions per minute
RSD	Relative standard deviation
s	Seconds
sgRNA	Single guide RNA
SVD	Singular value decomposition
T	Tau, response time
Ypet	Yellow fluorescent protein

Summary

Metabolism provides the essential biochemical intermediates and energy that enable life and its growth. In this thesis we studied robustness of *Escherichia coli* metabolism, by perturbing it with different methods and measuring the response at a molecular level.

In **Chapter 1**, we introduce the latest insight into metabolic regulation and optimality in microbial model organisms. Overall, we identified and described two major gaps in knowledge: the limited amount of known metabolite-protein interactions and the unknown objectives towards which cells optimize their enzyme levels. Moreover, we provide a short introduction to the relevant methods utilized in this thesis.

In **Chapter 2**, we describe a series of experiments which confirmed that CRISPRi is a reliable tool to specifically perturb metabolism in *E. coli*. We showcase the advantage of using a CRISPRi system integrated in the genome, which is suitable to apply inducible knockdowns of essential genes. We demonstrate this by characterizing growth for a library of over 100 strains and verifying inducibility and specificity with proteomics data.

In **Chapter 3** we applied the validated CRISPRi setup to perturb and study metabolism systematically. First, we used a pooled CRISPRi library to knock down all metabolic genes in *E. coli*. By following the appearance of growth defects with next generation sequencing, we show that metabolic enzymes are expressed at higher levels than strictly necessary. We then focused on a panel of 30 CRISPRi strains and characterize their response to lower enzyme levels with metabolomics and proteomics. We show that the metabolome can buffer perturbations of enzyme levels in two different stages: first, metabolites increase enzyme activity to maintain optimal growth and only later they activate gene regulatory feedbacks to specifically upregulate perturbed pathways.

In **Chapter 4** we employed a different approach to perturb bacterial metabolism, by growing *E. coli* in different environmental conditions and measuring the response at the metabolome level. We could show that in exponentially growing cells key biosynthetic products as amino acids and nucleotides are kept at relatively stable levels across different environments. We compared our dataset to a matching published proteomics dataset, showing that unlike the proteome, metabolite levels are independent from growth effects.

Zusammenfassung

Der Stoffwechsel, oder auch Metabolismus, stellt die essentiellen Bausteine und die Energie bereit, die Leben und zelluläres Wachstum voraussetzen. In dieser Doktorarbeit wurde die Robustheit des Metabolismus von *Escherichia coli* untersucht, indem er mit verschiedenen Methoden perturbiert und die zelluläre Antwort auf molekularer Ebene verfolgt wurde.

In **Kapitel 1** werden die neuesten Erkenntnisse über die Regulation und Optimalität des Metabolismus in mikrobiellen Modellorganismen betrachtet. Zusammenfassend ließen sich zwei große Probleme feststellen: Zum einen die niedrige Zahl an nachgewiesenen Metabolit-Protein Interaktionen und zum anderen die unbekannten Ziele, auf deren Grundlage Bakterien ihre Enzymlevel regulieren und einstellen. Darüber hinaus werden in diesem Kapitel die für diese Arbeit relevanten und verwendeten Methoden besprochen.

In **Kapitel 2** werden eine Reihe von Experimenten beschrieben, die bestätigen, dass CRISPRi eine zuverlässige Methode ist, um den Metabolismus in *E. coli* spezifisch zu perturbieren. Außerdem werden die Vorteile von einem genomisch integriertem CRISPRi-System gezeigt, das dazu verwendet werden kann die Expression von essentiellen Genen induzierbar zu reprimieren. Die Induzierbarkeit und Spezifität konnten durch ein Wachstumsscreening von 100 Stämmen und Proteom-Analysen belegt werden.

In **Kapitel 3** wird dargestellt, wie das im vorherigen Kapitel beschriebene CRISPRi- System verwendet wurde, um den Metabolismus systematisch zu perturbieren und zu untersuchen. Zunächst wurde das Wachstum von Stämmen in einer gepoolten CRISPRi-Library, welche alle Gene im zentralen Stoffwechsel von *E. coli* beinhaltete, mittels Next-Generation Sequencing verfolgt. Hierbei konnte gezeigt werden, dass Enzyme im zentralen Metabolismus in höheren Mengen von der Zelle hergestellt werden, als es für die Aufrechterhaltung des Wachstums nötig wäre. Es wurden 30 CRISPRi Stämme mit Hilfe von Metabolomics und Proteomics genauer untersucht, um die zelluläre Antwort auf niedrigere Enzymlevel zu studieren. Hierbei konnte festgestellt werden, dass das Metabolom die Störung von Enzymleveln auf zwei unterschiedliche Wege puffern kann. Zunächst erhöhen Metabolite die Aktivität von Enzymen, um optimales Wachstum zu gewährleisten, und erst später aktivieren sie genregulatorische Feedback-Mechanismen, um perturbierte Stoffwechselwege spezifisch hochzuregulieren.

In **Kapitel 4** wird geschildert, wie eine alternative Methode, nämlich das Wachstum unter verschiedenen Bedingungen, genutzt wurde, um den Metabolismus zu perturbieren und anschließend die metabolische Antwort zu bestimmen. Hierbei konnte gezeigt werden, dass in exponentiell wachsenden Zellen unter verschiedenen Wachstumsbedingungen die Konzentrationen von Schlüsselbausteinen, wie Aminosäuren und Nukleotiden, stabil gehalten werden. Diese Daten wurden zudem mit einem passenden, bereits publizierten Proteomics Datensatz verglichen und es konnte gezeigt werden, dass Metabolitkonzentrationen, im Gegensatz zu Proteinkonzentrationen, unabhängig von Wachstumseffekten sind.

1 - Introduction

1.1 - Crosstalk between transcription and metabolism: how much enzyme is enough for a cell?

The function and structure of metabolic and transcriptional networks are well characterized. Transcription is the first step in the control of gene expression. Metabolism governs the supply of energy and cellular building blocks. Besides regulatory interactions within each of the two networks, mutual feedback is abundant between them. Already in the 1950s the discovery of the *lac* operon showed that transcription impacts metabolic operation (***metabolic gene expression,*** **Figure 1**)[1]. A few years later, the discovery of allosteric metabolite-protein interactions provided a mechanism for metabolite driven transcription (***metabolic feedback on transcription,*** **Figure 1**)[2]. In our view, the crosstalk between metabolism and transcription results from two interdependent processes: information from transcriptional networks to metabolism is transmitted by ***metabolic gene expression***, while metabolic information is conveyed via ***metabolic feedback on transcription***.

In the past decade systems biology has mostly been focused on genomes, transcriptomes and proteomes due to the availability of advanced and sensitive technologies. Recent improvements in metabolomics methods have now enabled metabolites to become the focus of many studies [3–6]. The fundamental challenge for understanding how metabolites regulate transcriptional programs lies in identifying metabolites that are key signals for transcriptional regulators. This is illustrated by the fact that the master regulatory metabolite of catabolic genes in *E. coli* – alpha-ketoglutarate - was identified only recently [7], despite the fact that the regulatory mechanism has been known since the early 1950s (*carbon catabolite repression*). Recent findings suggest that such metabolic feedback on transcription could govern global gene regulation and metabolism. For example, a recent study in yeast discovered that

seemingly pathway specific amino acid auxotrophies change up to 80% of transcripts and affect almost all metabolism [8].

The potentially widespread presence of cross-talk raises the question of what cellular function emerges from this interdependence between transcription and metabolism. Here, we reviewed the past five years of literature that addressed these questions in microbial model organisms (mainly *Escherichia coli* and yeast). We will discuss the putative function of crosstalk in optimizing enzyme levels, and focus on system-level studies that either used experimental transcript and metabolite data or took more theoretical approaches based on large metabolic and transcriptional networks.

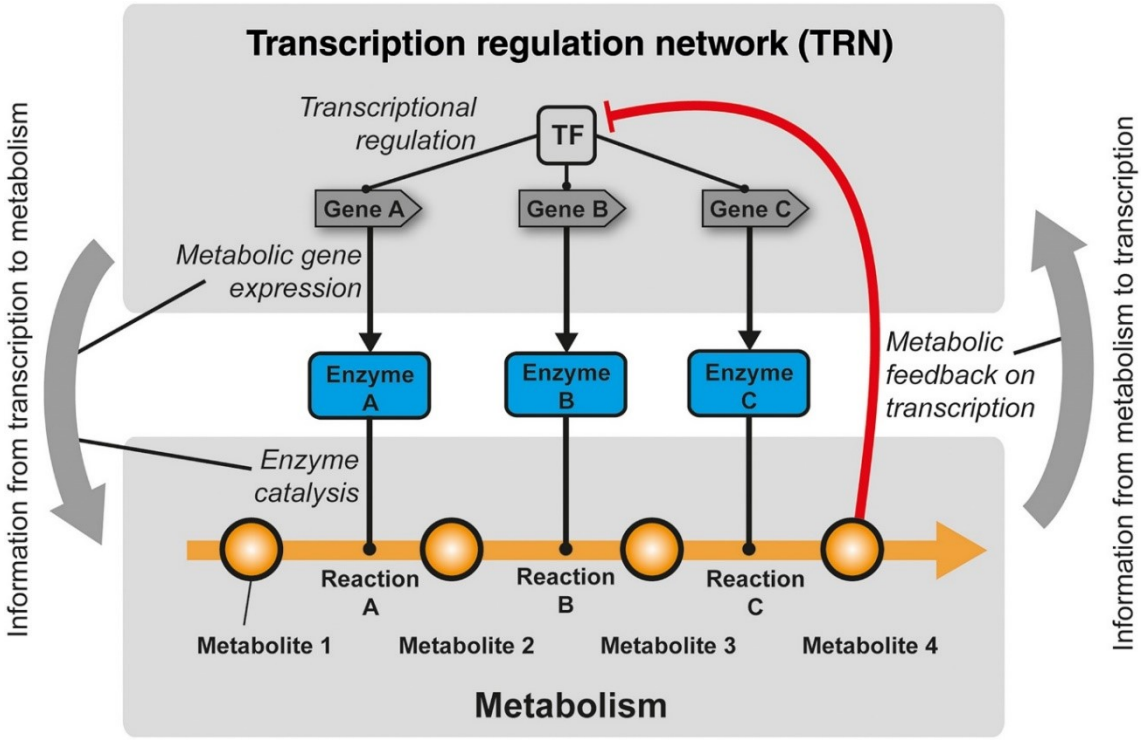


Figure 1: Schematic of the cellular processes that are involved in crosstalk between transcription and metabolism. Metabolic gene expression represents transcription and translation of a gene into a metabolic enzyme. Each enzyme catalyzes a certain reaction within the metabolic pathway (enzyme catalysis). Metabolic homeostasis implies that reaction rates A, B, and C are equal, that is, metabolite concentrations are constant and metabolic flux through the pathway is constant. Metabolites can interact with transcriptional regulatory proteins and modulate their activity (metabolic feedback on transcription). The example here shows an inhibition of a transcription factor (TF) by metabolite 4. The transcription rates of genes A, B, and C are then regulated by the activity of the transcription factor (transcription regulation).

1.1.1 - Metabolomics and transcriptomics studies indicate extensive crosstalk

Intracellular metabolite concentrations are conserved between organisms. Absolute concentrations of primary metabolites in *E. coli*, yeast and mammalian cells are remarkably similar, with amino acids constituting the most abundant fraction of a cell's metabolome [9]. Despite such similarities, recent studies suggest that metabolite concentrations contain a high amount of information. Ralser and colleagues measured absolute amino acid concentrations in 4913 yeast strains, each bearing a different single gene deletion [10]. For one third of these strains, the gene deletions caused significant changes in the levels of (individual) amino acids that were unexpectedly precise and specific. Each amino acid responded individually to gene perturbations, such that the signature of all 20 amino acids allowed functional annotation of genes as genes with similar amino acid signatures tended to be functionally related. The precision was so high that ribosomal genes, for instance, mapped to the ribosome structure. Similar to the case in yeast, a genome-wide study measuring 1,432 metabolites in *E. coli* showed that metabolite profiles were specific for gene deletions and enabled the inference of novel metabolism-related functions of many orphan genes [11]. Given that the metabolome holds such precise information on a cell's status, to which extent is this information used by cells for gene regulation? The first evidence that metabolic information is extensively used in gene regulation networks comes from a comprehensive transcriptome analysis of yeast [12]. In this study transcript profiles of 1484 yeast strains with single gene deletions were measured using RNA sequencing. The gene regulation network inferred from these data indicated a striking number of metabolic feedback circuits, with many metabolic genes assigned as nodes in incoherent network motifs. The incoherent or "counter-intuitive" motifs were then interpreted as crosstalk between metabolism and transcription. A drawback of all three studies is the use of relatively extreme genetic perturbations. It would be important to address whether more subtle changes in gene expression yield similarly informative metabolite profiles. Combining CRISPR-based transcriptional perturbations of gene expression with high-throughput data has the potential to address such questions [13], but has not been applied to metabolite data so far.

1.1.2 - Information flow from transcription to metabolism

Of the two aspects of cross-talk between transcription and metabolites, the expression of metabolic genes is the better characterized one as it follows the canonical flow of genetic information and has been studied extensively over the past decades. The comprehensive set of a cell's metabolic reactions can be identified by combining whole genome sequencing, omics data and the abundant biochemical knowledge [14]. The resulting set of metabolic reactions is then used to reconstruct genome-scale metabolic models (GSMs), where each reaction is associated to one or more proteins and the respective gene. These so-called gene-protein-reaction relationships are a holistic but static representation of metabolic gene expression. They list components, but do not allow the prediction of abundance or activity changes upon perturbations. The latest GSM of *E. coli* comprises 1366 genes and associates them to 2251 biochemical reactions and 1136 metabolites [15]. The latest GSM of the yeast *S. cerevisiae* includes 904 genes, 1412 reactions and 1228 metabolites [16]. The static gene-protein-reaction relationships can be advanced with transcriptional regulation by assigning Boolean rules that define on/off states, as shown by Palsson and colleagues [17]. In this framework, simple regulatory rules depending on environmental stimuli or internal metabolic flux distributions decide if a gene is expressed (on) or not (off), and therefore if a metabolic reaction can occur. Boolean rule–based methods have been extended to larger models [18,1920] and were included in the first whole-cell model of a living organism [21]. As these coarse binary rules can only define two states (on/off) they were later refined using probabilities to describe expression states [22]. In this method, probabilistic weights were inferred from large datasets of gene expression, and then used to constrain metabolic fluxes based on the probability that a certain gene is transcribed.

However, a caveat of the underlying metabolic models is that they are a purely stoichiometric representation of metabolism and therefore work with reaction rates (metabolic flux) instead of the potentially more informative metabolite concentrations. As a consequence, GSMs cannot evaluate how up- or down regulating an enzyme affects metabolite concentrations, and how these concentration changes propagate and alter metabolic fluxes. To address these questions, smaller dynamic metabolic models

including reaction-substrate relationships such as Michaelis-Menten kinetics, and allosteric regulation of enzymes have been developed [23]. The currently most advanced kinetic model of *E. coli* metabolism comprises 457 reactions and 295 allosteric interactions [24], and was able to reproduce more than half of the measured metabolite concentrations. In conclusion, large genome-size metabolic models can - to some extent - evaluate the effect of gene expression on **metabolic fluxes**. However, evaluating the effect of gene expression on **metabolite concentrations** remains a fundamental challenge due to missing kinetic information.

1.1.3 - Information flow from metabolism to transcription

The transmission of metabolic feedback on transcription constitutes the reverse direction of information flow in cross-talk. Metabolic feedback is mediated by transcription regulation networks (TRNs), which describe the relationship between genes and their transcriptional regulators. Typical transcriptional regulators in prokaryotes are transcription factors (TFs), sigma factors and nucleoid proteins [25]. In eukaryotic cells epigenetic DNA modifications, histone modifications and chromatin remodeling proteins constitute an additional layer of transcriptional regulation [26]. TRNs have a hierarchical structure as transcriptional regulators can control other transcriptional regulators. The *E. coli* TRN was reconstructed using manually curated experimental data [27] and is the currently most complete TRN available. It covers 210 out of the ~300 predicted transcription factors and 3261 interactions between transcription factors and genes. In comparison, the *S. cerevisiae* TRN contains only 147 experimentally validated transcription factors out of a total of 250 predicted ones [28]. Results from an *E. coli* study mapping physical DNA interactions for 116 transcription factors with SELEX indicate that transcription factors might have a wider DNA-binding spectrum than previously assumed [29]. More than 80% of transcription factors were associated with more than 10 genes, with binding occurring in both noncoding and coding regions. However, even for the well-studied microbe *E. coli* the structure of the TRN is ambiguous, as topologies of TRNs depend on the method used to assign interactions of transcription factors and target genes. The total number of functional interactions can

be overestimated in pure binding-based methods, as false positive interactions arise from non-functional binding. Likewise, false negative interactions arise if regulation is condition-dependent, leading to an underestimation. Studies that assess TRNs with and without effectors, such as iron [30] or amino acids [31] enable mapping the condition dependent logic of TRNs.

Once the structure of a TRN is defined the key challenge lies in identifying the most relevant input signals of transcriptional regulators. Previous findings in yeast show that many transcription factors are activated through post-transcriptional mechanisms [32]. This observation was recently confirmed at the proteome-scale by absolute protein abundance data of *E. coli* in various environmental conditions [33]. Indeed, transcriptional regulators were the most constant class of protein across all conditions. Post-translational regulation mechanisms can be broadly divided into activity changes resulting from an upstream (externally stimulated) signaling cascade or from internal metabolic signals (e.g. reviewed in [34]). While external signaling cascades and two component systems are well characterized, our knowledge on transcriptional regulators that sense internal metabolites is scarce. Even in the case of the arguably best-studied model microbe *E. coli*, direct interactions with metabolites have only been shown for 47 out of the 210 transcription factors [35]. This directly follows from the relative lack of scalable methods for identifying metabolite-protein interactions systematically. The gold standard for testing the effects of metabolites on transcriptional regulators are still low-throughput *in vitro* assays. Although new methods based on microarrays or affinity purification enable large-scale discovery of physical interactions between proteins and metabolites, they are limited to very stable interactions, which mostly occur with lipids and other hydrophobic metabolites [36,37]. More recent methods to detect conformational changes by NMR [38] and proteomics [39] have so far focused on enzyme-protein interactions only. In summary, even though transcriptional circuits are well defined the lack of scalable methods for identifying metabolite-transcription factor interactions hinders the discovery of regulatory metabolites. An alternative to mapping physical interactions is the inference of cross-talk from multi-omics data, as discussed in the next section.

1.1.4 - Inference of cross-talk from multi-omics data

Multi-omics datasets quantifying the abundance of transcripts, proteins and metabolites and the fluxes through metabolic networks across environments can theoretically be used to infer cross-talk. However, this requires systematic and often difficult integration of the different datatypes [40]. Furthermore, experimental perturbations rapidly propagate within and between networks, which complicates inference of causal interactions. For this purpose, theoretical frameworks like Metabolic Control Analysis provide relationships between local and global responses in metabolic and transcriptional networks [41]. In practice, however, these frameworks require very specific perturbations of single network components, such as up-and downregulation of single enzymes. This contrasts with the mostly very unspecific environmental perturbations, such as growth on different nutrients or stresses, used for generating the large majority of multi-omics datasets. However, in these cases other data-driven modeling approaches using either transcription rates or metabolic fluxes as starting points have been used, which will be discussed in the following two sections.

Regulation of transcription rates

To address whether metabolite levels are predictors of transcription rates, Kochanowski *et al.* measured metabolite concentrations and transcription rates of central metabolism promoters in 26 environmental conditions in *E. coli* [42]. The study revealed that across environments about 30% of transcriptional changes were largely caused by activity changes of two transcription factors, Crp and Cra. Importantly, the authors were able to recover the known effector metabolites of both transcription factors, cyclic AMP and fructose-1-phosphate/fructose-1,6-phosphate, respectively, without prior knowledge. The remaining 70% of transcriptional changes in this study were caused by global growth-dependent regulation. Such large-scale transcriptional changes after environmental perturbations seem to be a general growth rate dependent effect in yeast and *E. coli* [43]. So far only few other studies integrated metabolites with transcript data and focused mainly on coordinated responses in dynamic conditions [44,45].

Regulation of metabolic fluxes

The majority of multi-omics studies use metabolic fluxes as starting points and search for correlation with other data types. The predominant conclusion from such studies is that transcripts are poor predictors of metabolic fluxes both in central carbon metabolism [46,4746] and on a genome-scale [4849]. This is exemplified by flux and transcript changes between carbon and nitrogen limited yeast matching in just 53 out of 2194 flux-transcript pairs [48]. A systematic survey of computational methods arrived at the same conclusion [50], by showing that computational predictions of metabolic models were generally not improved by fitting measured transcript data. A recent multi-omics study in yeast by Rabinowitz and coworkers provided an explanation for this lack of correlation by showing that fluxes are mainly regulated at the post-transcriptional level [51]. By systematically fitting enzymes, fluxes and metabolites to Michaelis-Menten type kinetics, the authors found that it was mostly substrates that controlled flux through their associated reaction. Substrate-dependent flux implies that most enzymes are not operating at their maximal possible velocity (v_{max}), but rather at sub-saturating conditions. This finding agrees with the tendency for overabundant enzymes in microbes as discussed in more detail in the next section and illustrated in **Figure 2**.

In summary, recent studies show that is in principle possible to infer metabolite-transcription factor interactions from multi-omics data [42]. The scalability of this approach is an important aspect for enabling integration of large-scale transcriptional (e.g. Ref [52]) and high-throughput metabolomics data [53]. Moreover, the results from studies using metabolic flux as starting points for data integration support the prevailing opinion that fluxes are rarely controlled by abundance changes of enzymes [54]. This raises the question why, if not for controlling metabolic flux, cells regulate enzyme levels.

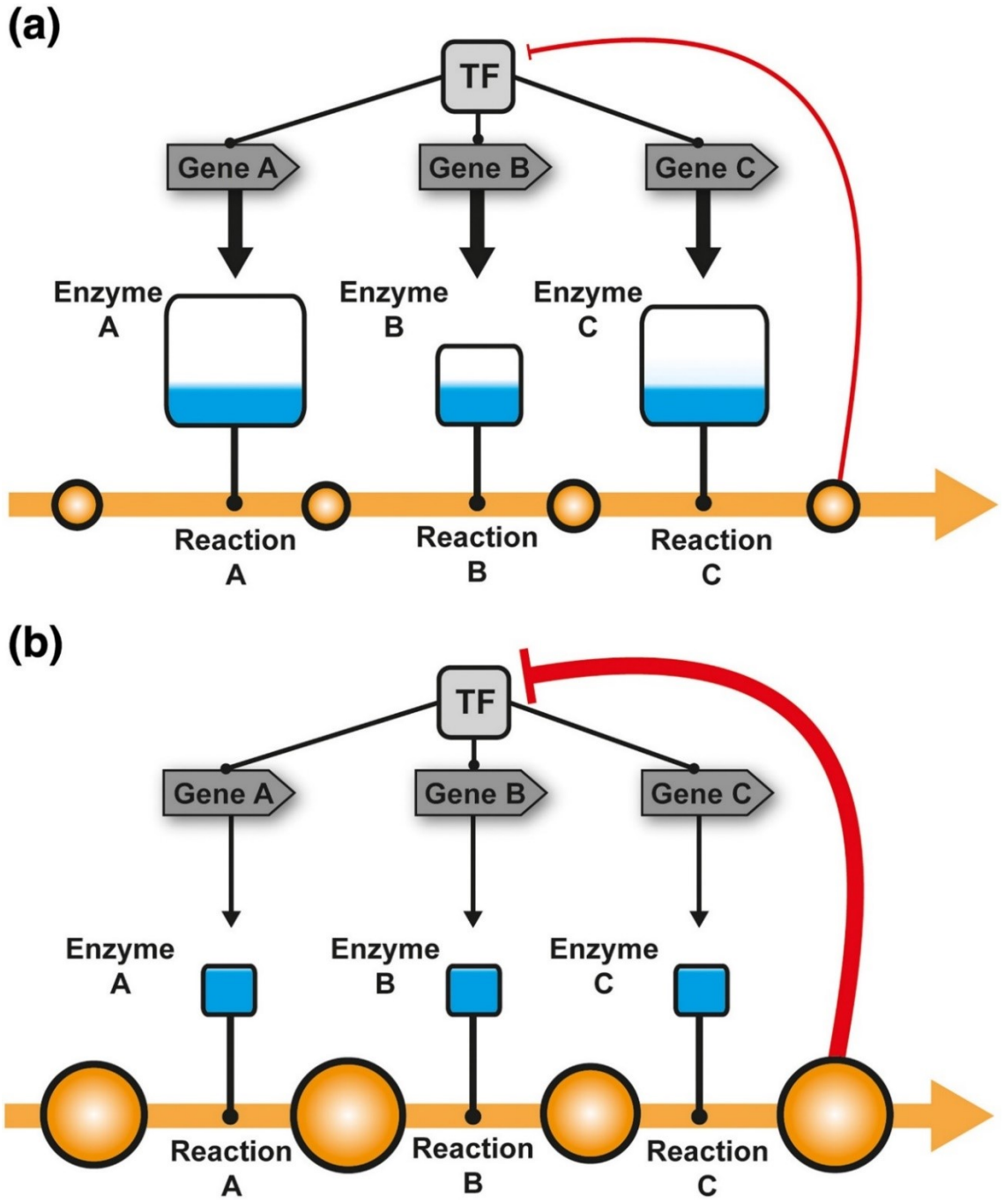


Figure 2: Robust versus efficient enzyme levels. (a) Enzymes in the metabolic pathway are overabundant and not operating at their full catalytic potential. The size of the enzyme indicates abundance and the blue fill activity. For example, enzyme B operates at half-maximal velocity (v max). Perturbations can be compensated without regulating enzyme abundance by transcription (e.g., changing substrate concentrations). (b) Enzymes in the metabolic pathway are expressed at the minimum level that allows the same metabolic flux as in (a). All enzymes operate at their full catalytic potential (v max) and are saturated with substrates (the size of metabolites indicates the concentration). Decreasing a single enzyme level results in a metabolic bottleneck and in flux limitations. Transcriptional feedback regulation can compensate such perturbations, but slower than in (a).

1.1.5 - Crosstalk regulates enzyme levels

Proteome data suggests enzyme overabundance

The studies in the previous section investigated regulation of transcription rates and metabolic flux. A series of recent studies has taken enzyme levels into consideration. A comparative study of predicted metabolic flux based on an *E. coli* GSM and quantitative proteomics data by Palsson and colleagues revealed that measured enzyme levels are for the most part higher than predicted [55]. Noor *et al.* demonstrated that for central metabolism such enzyme overabundance can be explained by enzyme saturation and thermodynamic effects [56]. Other approaches that compare enzyme abundance and enzyme kinetics were recently reviewed by Davidi and Milo [57].

Several hypotheses have been proposed for explaining the existence of pervasive enzyme overabundance [54]: On the one hand enzyme overabundance could simply be a result of imperfect regulatory mechanisms. However, there could also be an important functional role for this phenomenon, for example by providing a buffering mechanism against internal and external fluctuations or by enabling flux control via fast-acting allosteric interactions. In the latter case the unused enzyme fraction can be activated instantaneously, whereas expressing new enzymes would take time and result in a potential fitness cost. Such an fast acting allosteric mechanism can, for example, be observed in *E. coli* glycolysis: it allows unused glycolysis enzyme to be allosterically activated within 5 seconds of a shift from gluconeogenetic to glycolytic nutrients [58]. Having established that cellular enzyme levels tend to be higher than absolutely required to explain flux we will in the following sections focus on the cellular objectives that define enzyme levels and on the role of cross-talk in regulating enzyme levels.

Which cellular objectives define enzymes levels?

From an evolutionary perspective we would expect that cells tune enzyme levels to optimize fitness parameters and overall physiology. However, to which degree enzyme levels have been optimized through the course of evolution is a fundamental and longstanding question [59]. The most direct way to test optimality of enzyme levels for

fitness is to vary the levels of single enzymes *in vivo* and measure growth or other fitness parameters. In yeast, such a titration of the levels of enzymes and other proteins has recently been achieved using barcoded promoters with a 500-fold expression range [60]. The study demonstrated that changing single enzyme levels below and above wild-type levels rarely improves growth, implying that enzyme levels of the wild type strain are already optimized for maximum growth in glucose. However, the observation that wild-type expression was only optimal for growth on glucose, but not on galactose, indicates that it is difficult for cells to optimize enzymes levels in all possible conditions, as already suggested before [54]. Similarly, Milo et al. show that in various conditions around 30% of the total enzyme pool in *E. coli* carries no metabolic flux, suggesting again imperfect regulation [57]. Furthermore, optimality criteria of enzyme levels may involve multiple, even conflicting objectives that are not reflected by growth rates or metabolic flux alone. This is illustrated by a study in *B. subtilis* , which used CRISPR-based repression of single essential genes, resulting on average in a three-fold down-regulation of protein levels [61] . While 80% of all strains with single gene knockdowns showed growth similar to wild-type, 95% had problems to resume growth after stationary phase (reflected by long lag-phases). This implies that expression of most genes may be optimized for dynamic conditions and to a lesser extent to achieve maximal growth rates. In a different approach, Alper and colleagues used a CRISPR-based system to gradually express enzymes in metabolically engineered yeast and test the sensitivity of metabolic flux [62]. In a third approach Panke et al used ribosomal binding site libraries to unravel design principles of optimal enzyme levels in a synthetic metabolic pathway [63]. In conclusion, novel genome engineering methods enable exploring the relationship between cellular fitness and enzyme levels at large-scale. These studies show that enzyme levels seem to be optimal for different fitness aspects in different species: for maximum growth on glucose in yeast [57] and for growth resumption after stationary phase in *B. subtilis* [61]. Next, we searched for studies demonstrating that cross-talk between metabolism and transcription controls enzyme levels dynamically.

Cross-talk enables self-optimization of enzymes levels

As discussed in the last section, there seems to be optimization of enzyme levels to some extent. But how do cells achieve these optimal levels? A series of studies provide

evidence that cross-talk might be an important mechanism to "self-optimize" enzyme levels, meaning that optimal enzyme levels are an emergent property of cross-talk between transcription and metabolism. To test this hypothesis, cross-talk between metabolism and transcription can be disrupted by externally added regulatory metabolites. Alon and colleagues recently used such an approach to examine cross-talk between central carbon metabolism and transcription of catabolic genes [64]. By supplying *E. coli* cells externally with the metabolite cyclic AMP (cAMP) the authors were able to gradually change transcription of catabolic genes. Their results demonstrate that expressing catabolic genes at wild-type levels is optimal for growth on some nutrients, but sub-optimal on others. Therefore, in some environments, crosstalk via carbon catabolite repression enables *E coli* to optimize levels of catabolic enzymes. The more detailed mechanism underlying cAMP regulation of carbon catabolite repression is described in Ref. [7].

Self-optimization not only seems to occur for catabolic enzyme levels but also for ribosomes [61,65]. Bruggeman and colleagues showed that optimal ribosome levels result from cross-talk between amino acids and transcription of ribosomal genes. To this end, they developed a small-scale model of amino acid and protein synthesis, which achieved self-optimization of ribosome levels [66]. Mechanistically the model represented the well-studied transcriptional program centered around the secondary messenger (p)ppGpp. A similar function of (p)ppGpp mediated cross-talk between amino acids and transcription of ribosomal genes was suggested by Scott *et al.* [67]. Cross-talk functions as a "supply driven activation" of ribosomal gene transcription, which is a simple yet robust mechanism to optimize ribosome levels. However, "supply driven activation" alone could be sub-optimal in dynamic conditions. A small-scale model suggests that feedback inhibition by transcriptional repressors plays an important role during nutritional up- and downshifts [68]. In this scenario strong transcriptional feedback would emulate an "on-off" control strategy, which enables additional optimization of the levels of amino acid biosynthesis enzymes in dynamic conditions. While the examples above show that cells optimize the larger fractions of their proteome by cross-talk (the sum of catabolic enzymes and the sum of ribosomal proteins), it is currently not clear if this happens for smaller fractions or individual enzymes as well. Some support for the presence of self-

optimization in smaller metabolic modules (between 150 and 250 enzymes per fraction) comes from clustering of quantitative proteome data [69]. Finally, several studies provide evidence that *E. coli* optimizes a single bottle-neck enzyme in the methionine biosynthesis pathway [35], and that cross-talk in glycolysis may be geared towards optimal glycolysis flux [70].

1.1.6 – Discussion

Several lines of evidence have shown that microbial cells express more enzymes than they absolutely need for maintaining physiological flux [51,54,55]. Given the potential burden resulting from overexpressing enzymes it seems likely that this overabundance has a functional role in microbial physiology. The prevailing opinion is that higher than needed enzyme levels prevent bottlenecks in metabolic pathways: by default, each enzyme operates at sub-maximal velocity (v_{max}), which allows metabolites to rapidly change the current reaction velocity (e.g. by substrate-saturation or allosteric feedback). However, an alternative hypothesis is that cells actively accumulate enzymes in nutrient rich conditions, to protect against potential stresses in future conditions.

In the previous paragraphs we have summarized studies that support a role of crosstalk between transcription and metabolism in adjusting enzyme levels. There is evidence for both small (individual enzymes) and large-scale (catabolite repression) optimization. But there are more than 1000 distinct enzymes in an *E. coli* cell and it is unknown if of each of them is regulated individually. If the level of a single enzyme accidently falls below a flux limiting threshold this can be counteracted by two distinct mechanisms: global upregulation of all enzymes, which is probably more costly, and upregulation of just the critical flux-limiting enzyme. The current literature provides evidence for both scenarios. On one hand, recent findings based on metabolomics data indicate that metabolites carry very specific information about cellular processes [10,11]. Such localized and specific changes in metabolite levels could enable highly precise crosstalk to control levels of each enzyme individually. On the other hand inference from multi-omics data has revealed only sparse crosstalk between metabolism and transcription [42,47]. In fact, global

growth-rate dependent transcriptional regulation seems a major driver of gene expression [43]. It will be important to clarify if these global effects result from the very unspecific environmental perturbations applied in the studies. Global regulation could be the main driver upon broad-ranging external perturbations, whereas more localized internal perturbations invoke more specific metabolite-driven gene regulation.

We have focused on studies in microbial model organism, which leveraged the system-level understanding about metabolism and transcriptional regulation in these cells. A direct benefit of such studies is that methods can be transferred and applied in medical and biotechnological research. For example, a novel proteomics-based method to detect metabolite-protein interactions in yeast was recently transferred to T-cells [71]. The method allowed identification of transcription factors that interact with the amino acid L-arginine and thereby promote anti-tumor activity. It remains to be seen if regulatory principles in simple microbial model systems apply to higher, multicellular organisms. Given the strong conservation of core cellular processes across evolutionary scales it seems likely that at least the more general principles are conserved. We have discussed the central role of alpha-ketoglutarate as a master regulatory metabolite of catabolic genes in bacteria. The same signal plays an important role in other organisms: Alpha-ketoglutarate concentrations control mTORC1 activity in mammalian cells [72], and increased alpha-ketoglutarate levels are associated with an extended life span in *Caenorhabditis elegans* [73]. However, currently it still unclear if the same signal has the same function in these cells, or if its function has been altered in higher organism.

Finally, biotechnological applications that utilize microbial model organism to produce chemicals will directly benefit from insights about regulatory cross-talk in these cells. Especially metabolic engineering applications require high enzyme levels to enforce high fluxes in synthetic metabolic pathways. However, strong overexpression of heterologous enzymes results in burden and instability in production strains [74]. Understanding principles that optimize enzyme levels in natural pathways could serve as blue-print to control enzyme levels in synthetic pathways dynamically. The great challenge lies in designing synthetic cross-talk, but linking transcription factors to new metabolites is already possible (the *lac* repressor in this case) [75].

Finally, the answer to the question how much enzyme is enough certainly depends on the organism, the context and the metabolic function of the enzyme. Even in simple and well-studied model organisms we are just starting to be able to address this question systematically. In our view, the control of enzyme levels is intimately linked with crosstalk between transcription and metabolism, and therefore central to understanding the role of this cross-talk. It is likely that cells have acquired a certain "knowledge" of how much enzyme is enough in various environments. The question is to what extent single metabolite concentrations encode this knowledge.

1.2 - Mass spectrometry-based methods for systems biology studies

Systems biology is a holistic approach to biological research as it strives to understand biological complexity as a whole, rather than as the sum of its single entities. The key driving force of this approach has been the development of various *omics* methods to measure molecules in a high throughput manner and their eventual combination in multi-omics studies [76]. In the context of metabolism, the two main molecular entities of interest are metabolites and enzymes. The comprehensive measurement of metabolites is defined as metabolomics. In order to study enzyme abundances, a direct measure of proteins through proteomics has been shown to be more suitable than indirect methods, as in example RNAseq. Comparative studies have shown that in various prokaryotic and eukaryotic organisms transcript levels do not correlate with levels of the respective protein [77,78]. This is due to various reasons, such as different translational rates, post-translational modifications, different degradation rates, etc. The direct measurement of proteins and metabolites in a high throughput manner can be achieved through mass-spectrometry (MS), an analytical technique which relies on measuring the mass-to-charge ratio (m/z) of electrically charged molecules.

MS-based proteomics enables to measure the composition of protein samples [79], and a typical method to measure such complex mixtures from biological samples is *shotgun* proteomics [80]. Proteomic workflows (**Figure 3**) start with the extraction of proteins from

a biological sample and their digestion into peptides, using a protease with a defined proteolytic specificity. The digested samples can then be purified and run through a liquid chromatography (LC) column which separates peptides based on their interaction with the column. The eluting peptides are then ionized and by entering the mass-spectrometer they are separated and detected based on their m/z. In tandem MS (MS/MS), the charged peptides exiting the first mass-spectrometer are further fragmented to undergo a second MS measurement, allowing the identifications of ions that had similar m/z ratios in the first mass spectrometer. The detected mass spectra can be used to identify peptide structures. The identified peptides, are then compared to a peptide list obtained by an *in silico* digestion of the theoretical proteome, inferred from the relative genome of the biological sample. Based on the mapped peptides, the relative protein abundance can be estimated [80]. This methodology can be applied to obtain relative quantifications of protein levels between different samples. In order to measure protein concentrations in absolute terms, samples can be spiked with isotopically labeled standards before being measured. Such methods have been used to estimate absolute concentrations of thousands of proteins, effectively quantifying the proteome compositions of organisms [81].

Due to the small mass of metabolites, MS-based metabolomics methods do not require fractioning of the collected samples, which is a lengthy procedure in MS-based proteomics. MS-based metabolomics methods can be performed in an untargeted manner, measuring known and unknown chemical species of a sample, or in a targeted manner, in which only a subset of well annotated metabolites is measured. Untargeted metabolomics is characterized by speed and throughput of measurements, rendering these methods powerful tools for large screenings of biological samples [82]. On the other hand, targeted metabolomics focuses on analyzing a defined set of biochemically characterized molecules, enabling quantitative measurements of metabolites and estimation of their concentrations. Advancements in manufacturing of shorter LC columns with decreased particle sizes have considerably increased the speed at which metabolomics samples can be eluted into the mass-spectrometer. Novel targeted LC-MS/MS methods can precisely measure hundreds of metabolites with run-times reduced from 30-60 minutes to few minutes [83,84]. In a typical metabolomics LC-MS/MS

workflow (**Figure 3**) to measure intracellular metabolites, samples are first filtered to dispose of their cultivation media and then lysed using a variety of solvents. The utilization of cold acidic acetonitrile solvents for sample preparation can preserve unstable compounds as nucleotide triphosphates [85]. Prepared samples are then eluted with LC to separate metabolites and subsequently measured by MS/MS in a similar manner as for proteomics. Measured m/z spectra can be then used to quantify in relative terms metabolite abundances. Higher precision of measurements can be obtained by spiking samples with isotopically labeled standards, improving peak identification throughout long series of measurements and enabling absolute quantification of metabolite concentrations [86,87].

In this work, we employed shotgun proteomics and/or targeted metabolomics to measure thousands of proteins and hundreds of metabolites (**Chapter 3-4**). Moreover, we studied the response of the *E. coli* metabolome under different environmental conditions (**Chapter 5**). Analyzing omics data from a range of different strains/conditions allowed us to infer basic principles of microbial metabolism.

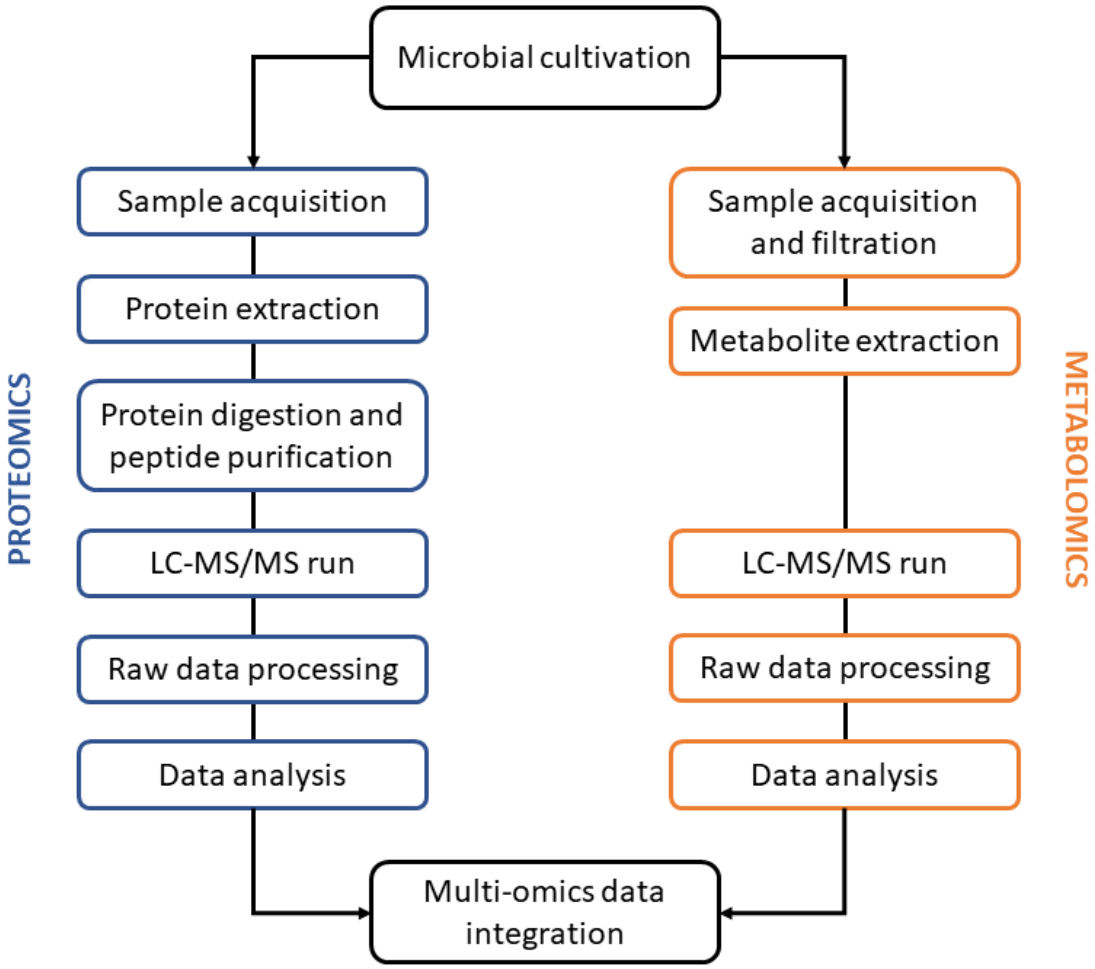


Figure 3: Scheme summarizing an MS-based multi-omics workflow. Shotgun proteomics is highlighted in blue and targeted metabolomics in orange.

1.3 - CRISPR interference, a tool for the control of gene expression

The discovery of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated proteins (Cas) has been a considerable breakthrough, with wide consequences for molecular biology and genetic engineering [88]. CRISPR/Cas systems are originally found in archaea and bacteria [89]. Most of them function as an immune adaptive defense mechanism against exogenous nucleic acids, as for example against phage infections [90]. CRISPR/Cas immunity relies on the formation of a complex between Cas proteins and short RNAs expressed from CRISPR arrays (crRNAs). crRNAs contain a

short sequence (spacer) which allows the complex to recognize and bind by base pairing a complementary target DNA or RNA (protospacer). In order to be recognized and bound by the complex, protospacer sequences need to be followed by a short protospacer adjacent motive (PAM), which serves cells to distinguish between self and non-self DNA. Upon recognition and binding of a correct protospacer, the endonuclease domains of Cas proteins can cleave the bound target. CRISPR/Cas effector complexes are characterized by a high binding specificity thanks to base-pairing, which has led to the repurposing of these systems for a number of applications.

In this context, the CRISPR/Cas effector complex from *Streptococcus pyogenes* is widely utilized, due to its stability in different organisms and its dependence on few elements: one single Cas protein (SpCas9) and two small RNAs. The crRNA contains the spacer sequence that guides the complex, while the trans-activating crispr RNA (tracrRNA) plays a role in the maturation of the crRNA [91]. The CRISPR/Cas9 system cleaving activity has been exploited most notably for genome engineering of bacteria and eukaryotes [92]. On the other hand, a successful application of the specific RNA-guided targeting has been the repurposing of SpCas9 to repress gene expression, technique known as CRISPR interference (CRISPRi) [93,94]. CRISPRi relies on SpCas9 proteins with mutated nuclease domains (dead Cas9 or dCas9) and a synthetic small guide RNA (sgRNA), comprising a spacer sequence of 20 nucleotides (**Figure 4A**). Upon their expression and formation, the dCas9/sgRNA complexes can recognize and strongly bind a target DNA sequence, without cleaving the target. Directing the complexes to genes of interest, using specifically designed sgRNAs, can prevent their transcription causing gene expression knockdowns. The strongest interference effects were demonstrated for spacer sequences designed to bind protospacers located in the non-coding strand of a gene at their 5' UTR or the coding region of a gene, in proximity to the translation initiation sequence (**Figure 4B**)[95].

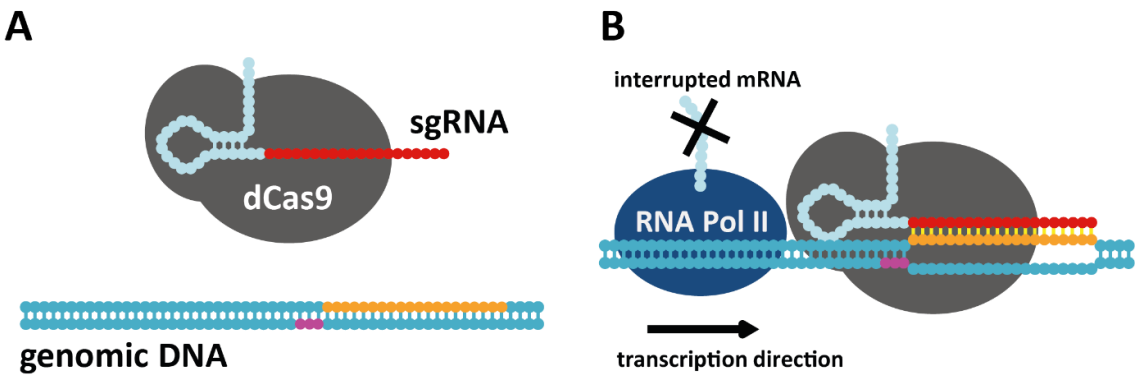


Figure 4: scheme of CRISPRi components and mechanism. (**A**) Components of a CRISPRi system. The scheme illustrates a dCas9/sgRNA complex. The spacer sequence is indicated in red. In orange, the protospacer sequence. In purple, the PAM. (**B**) Mechanism of a CRISPRi system. The dCas9/sgRNA complex scans the DNA to find a correct PAM and protospacer, and upon recognition binds the DNA. The scheme illustrates the strongest mode of interference, in which the complex is directed to stop RNA polymerases recruited to transcribe a gene into mRNA, and the spacer sequence binds the non-coding strand. Interrupting mRNA elongation effectively knocks down gene expression of the targeted gene.

CRISPRi can be easily reprogrammed by changing the 20-nucleotide sequence of the sgRNA. This enables the creation of large strain libraries which have been used extensively for genome-scale functional screenings, both in prokaryotes [96–98] and eukaryotes [96,99,100]. Moreover, CRISPRi has also been utilized in metabolic engineering studies to tune levels of enzymes and redirect flux towards production of particular compounds [101,102]. In this work, we exploited CRISPRi to study robustness of microbial metabolism. We designed and prepared CRISPRi strains in which we could artificially enforce lower enzyme levels. We then studied the response of cells, by measuring growth, proteins and/or metabolites. The initial design of the strains and testing of the CRISPRi experimental setup is described in **Chapter 2**. In **Chapter** we employed CRISPRi to perturb all metabolic genes in *E. coli* and to produce the biggest multi-omics dataset of CRISPRi strains to date.

2 – Characterization of CRISPRi-knockdowns of metabolic genes

2.1 - Results

2.1.1 - Comparison of different CRISPRi systems

A key factor for perturbing and studying metabolism is the utilization of expression systems orthogonal to the metabolic network of an organism. For this reason, we tested expression systems inducible by anhydrotetracycline (aTc) [103], a tetracycline which does not bloc bacterial protein synthesis. We compared two CRISPRi systems, inducible by aTc: a plasmid based CRISPRi system (pCRISPRi)[95] and a system relying on a dCas9 gene integrated in the genome of *E. coli*, with an optimized tetR promoter (YYdCas9)[104](**Figure 5A**). We transformed both strain with a plasmid expressing an empty control gRNA (control strain) or a plasmid containing a sgRNA with a spacer sequence targeting a gene (ilvC strain) encoding for the ketol-acid reductoisomerase. Upon growth on minimal medium with glucose, *ilvC* is known to be an essential gene [105]. We cultivated the four strains in minimal medium with glucose, supplementing or not 200nM of the inducer aTc. For the pCRISPRi system, we could observe a slight reduction in growth for the induced control strain, compared to the uninduced culture (**Figure 5B**). This might be caused by a growth burden, due to leaky expression of dCas9. This leaky expression was further confirmed by the fact that the strain containing a functional sgRNA, experienced an even greater growth reduction in absence of the inducer. When adding the inducer of dCas9, growth of the ilvC strain was severily impacted. In contrast, the YYdCas9 strains performed as expected (**Figure 5C**). Addition or not of the inducer did not cause differences in growth for the control strain. Moreover, the ilvC strain in absence of inducer grew exactly as the control strain, indicating low or negligible levels of dCas9 leaky expression. Finally, when cultivating the ilvC strain in presence of aTc, expression of dCas9 caused an expected growth reduction due to lower levels of the targeted *ilvC* gene and the relative essential enzyme. A YYdCas9 strain with gRNA targeting the essential cell division gene *ftsZ* confirmed the inducibility of the interference system

(**Figure S1**). Overall, these results suggest that the YYdCas9 system is more suitable for inducible CRISPRi experiments.

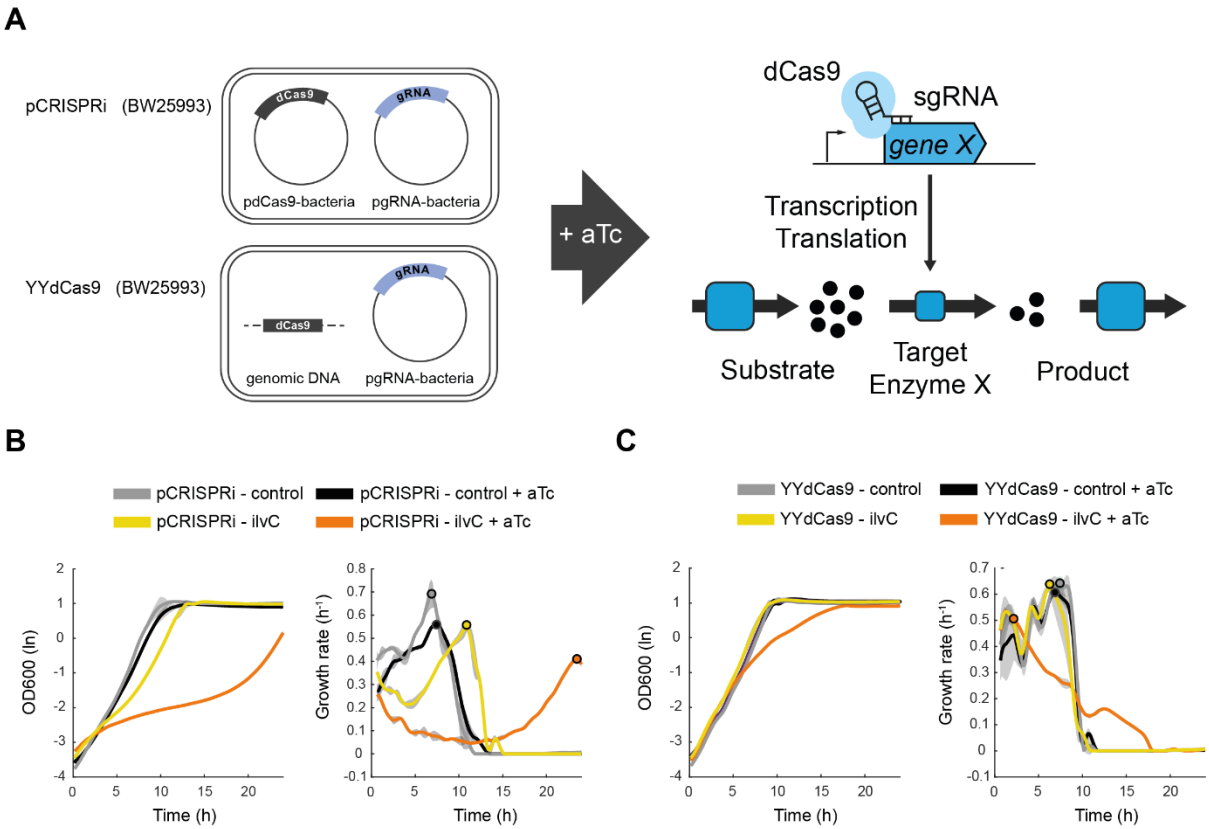


Figure 5: Comparison of different CRISPRi systems. (**A**) Scheme representing the two compared CRISPRi expression systems. Plasmids pdCas9-bacteria and pgRNA-bacteria were transformed in *E. coli* BW25993. pgRNA-bacteria plasmids were transformed into YYdCas9. Upon addition of aTc in cultures of strains with a functional sgRNA, we expected to observe a gene expression knockdown, consequent lower enzyme levels and possibly a growth defect. (**B**) Results of the growth screening for the pCRISPRi system. OD600 in ln scale and growth rate over time, lines represent average of replicates (n=3), grey shadowed areas represent standard deviation. aTc was added at T=0. (**C**) Results of the growth screening for the pCRISPRi system. OD600 in ln scale and growth rate over time, lines represent average of replicates (n=3), grey shadowed areas represent standard deviation. aTc was added at T=0.

2.1.2 - Screening of an arrayed library of CRISPRi strains

We then set to investigate a larger number of YYdCas9 strains. We selected 110 genes in primary metabolism of *E. coli* and designed pgRNA plasmids with sgRNAs targeting the strongest theoretical PAM site on the open reading frame of the gene [94]. The targets included genes encoding for enzymes in central carbon metabolism (25 enzymes),

biosynthesis of amino acids (34 enzymes), nucleotides (16 enzymes) and cofactors (31 enzymes). Four enzymes were involved in other pathways (fatty acid biosynthesis, phospholipid biosynthesis, lipopolysaccharide biosynthesis, Autoinducer-2 synthesis). 85 out of 110 chosen genes are essential for growth on glucose minimal medium (EcoCyc Database) [106]. When possible, we avoided targets that are in operons: 73 of the 110 target-enzymes are encoded by genes that are expressed singularly. The pgRNA plasmids were synthesised or cloned and transformed into YYdCas9 to obtain an arrayed library (**Table S1**). We measured growth 111 CRISPRi strains (110 + control strain) on glucose minimal medium, with and without induction of dCas9. All 110 CRISPRi strains grew similar to the control strain, when dCas9 was not induced (**Figure 6A**). To induce enzyme knockdowns, we added aTc at the start of cultivation and the induced cultures displayed a wide range of growth characteristics (**Figure 6A**). Knockdowns of 64 out the 110 enzymes caused a growth defect during the first 8 hours of exponential growth. However, the growth phenotypes appeared with a time delay of at least 3 hours after induction of the knockdown. Even strains with a very strong growth phenotypes grew initially like the control. We assumed that enzyme-levels would decrease exponentially after induction of dCas9 expression. Thus, growth phenotypes would appear early if the target-enzyme is expressed near a critical (flux limiting) level in the wild-type. In contrast, late or no growth phenotypes would indicate enzymes that are overabundant in the wild-type. To test this hypothesis, we calculated a response time (tau) as the time point when growth phenotypes appeared in the induced cultures (**Figure 6B**), and we calculated tau values for all strains (**Figure 6C**). We observed the shortest response time for knockdowns of MetE in methionine biosynthesis (3 h) and PyrE in nucleotide biosynthesis (3.3 h). This is consistent with our expectation that critical enzymes have the shortest response times, because MetE is a large and slow protein that limits overall protein synthesis [107]; and PyrE is sub-optimally expressed in K12-derived *E. coli* [108]. In central carbon metabolism, we observed the strongest response for knockdowns of the PTS system (PtsH) and the enolase (Eno): PTS is essential for carbon transport on glucose as sole carbon source [109], while Eno is a key limiting step for glycolysis in cells grown on glucose [110]. Finally, Idi is a known rate-limiting step for the synthesis of isoprenoids [111,112]. The median response time of all 110 target enzymes was 5.58 hours. Amino acid biosynthesis enzymes had the shortest response time (4.5 h). The higher sensitivity of

amino acid biosynthesis enzymes is expected, because their expression is tightly regulated by transcription factors and transcriptional attenuation [113]. Targets in cofactor metabolism and nucleotide biosynthesis had longer response times (6 h). Cofactor metabolism was the category with the highest fraction of target-enzymes that caused no growth phenotype at all (16 out of 31). The high robustness of cofactor metabolism matches previous reports showing that cofactors are stable and that their concentration is higher compared to the actual requirements for growth [114]. Overall, these results show that for a larger number of strains the chosen CRISPRi system is tight and inducible. Moreover, the delay between induction and appearance of growth phenotypes in 64 strains, as well as the absence of growth phenotypes in the other 46 strain, shows that *E. coli* is robust against knockdowns of enzymes. The response time to enzyme knockdowns might reflect whether enzymes are expressed near a critical level or if they are made in excess.

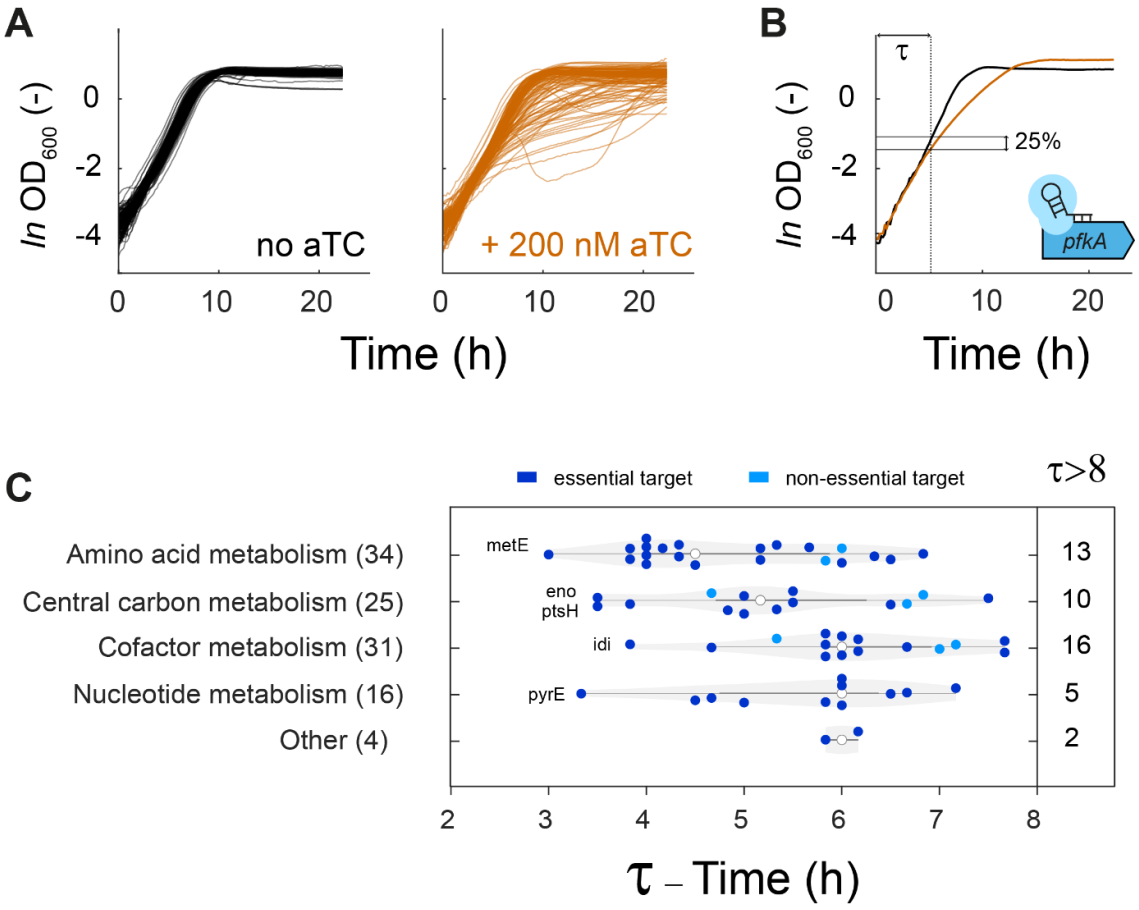


Figure 6: Growth screening of an arrayed CRISPRi library of 110 strains. (**A**) Growth curves of 110 CRISPRi strains carrying different sgRNA targeting expression of enzymes in the *E. coli* primary metabolism. In black are shown growth curves without addition of aTc, in orange with addition of aTc at Time = 0. Growth curves represent means of n=3 cultures. (**B**) Growth curves of the pfkA CRISPRi strain. Colors of curves as in Figure 1C. Growth curves represent means of n=3 cultures. Response time (t) was calculated as the time in which the induced/un-induced curves would diverge by more than 20% at OD > 0.12. (**C**) Response time (t) for the 110 strains of the CRISPRi library, divided by metabolic subsystems. t was calculated as described in Figure 2B using data from Figure 1A. Strains that experienced a t in the first 8 hours of exponential growth are shown in the plot, strains that did not display a t in the first 8 hours of exponential growth are counted in the panel at the right.

2.1.3 - Proteomics-based characterization of CRISPRi strains

In order to setup a sampling protocol for multi-omics data gathering, we first wondered whether the initial number of cells could affect the time in which growth phenotypes would arise. We inoculated different YYdCas9 strains (targeting *argA*, *argE*, *purM*) at different starting optical densities, and measured growth (**Table S2**). We could observe that regardless of the initial concentration of cells in the culture, the response time would be comparable for all the three strains. Therefore, a similar CRISPRi response happens in strains that are inoculated at lower or higher initial concentrations. This result implies that the initial inoculum can be adjusted to obtain enough biomass in exponential growth for metabolomics/proteomics sampling. We then moved to verify how proteins are affected by the chosen CRISPR interference setup, before and after a growth reduction takes place. Sampling cells earlier than the observed phenotype might prevent to detect significant changes in gene expression triggered by the metabolite perturbation. On the other hand, cells sampled after growth reduction might undergo global stress responses. We chose three strains from the library that displayed a growth reduction phenotype when subject to CRISPRi. The chosen strains included guides targeting essential genes in amino-acid biosynthesis (*metE*), glycolisis (*pfkA*) and isoprene biosynthesis (ispH). MetE and PfKA are the most active isoforms for their reaction (respectively 2.1.1.14 and 2.7.1.11), and are encoded as single ORFs. IspH is the only enzyme predicted to carry a key reaction in isoprene synthesis (EC 1.17.7.4) and is encoded as the last gene of its operon. After a 16h pre-culture in m9 minimal medium, we inoculated the strains at a starting OD of 0.05 in 35mL of medium in flasks, in presence or absence of aTc, in duplicates. For each flask, we took samples for proteomics at two timepoints (5mL of culture at OD=0.2, 2mL of culture at OD=0.5) and

measured OD600. Unlike the metE and pfkA strains, the ispH strain did not experience a reduction in growth when inducing dCas9 expression and grew as the control strain (**Figure 7A**).

Figure 7B displays the protein abundances of the interfered genes and of dCas9 for every strain, relative to the uninduced control. For all strains, when applying aTc, the dCas9 abundance increased by c.ca 5 times. Interestingly, dCas9 was detected also in cultures without aTc induction, meaning that the optimized tetR expression cassette of YYdCas9 permits a low basal level of expression. However, no strong reduction in the target proteins was detected in the strains cultivated without aTc. The low concentration of dCas9 might be insufficient to cause significant interference to the genes targeted by the dCas9-sgRNA complex. For all the tested strains cultivated with aTc expression of the targeted gene was reduced by more than 2-fold. We moved to verify how the rest of the proteome reacted to the CRISPRi induced bottlenecks.

From the 1760 detected proteins, we trimmed proteins which had either a lower peptide count than 2 or variability between replicates higher than 20%. Calculating fold change of the data over the protein dataset available for the control strain, led to high variability of differentially expressed genes (DEGs) (**Figure S2**). In example, for the metE strain many flagellar related genes resulted to be up-regulated for both conditions and both time-points compared to the Cntrl strain. Probably the metE strain underwent a common transposon related mutation [115]. Moreover, stress related proteins (as rpoS) appeared to be upregulated for strains sampled at T2. Therefore, we calculated the fold-change for every strain to the relative uninduced strain at T1 reducing the number of DEGs. Then we calculated fold-change for every dataset to the relative un-induced strain at the same time point, reducing the DEGs even further (**Figure S2**). Therefore, it appears that comparing data from induced cultivation to the not induced cultivation for the same strain contributes to reduce significantly noise in the data. **Figure 7C** depicts proteome changes for the strains at T2, normalized to the relative un-induced control. For the Control strain only dCas9 displays a significant upregulation, highlighting that expression of dCas9 does not cause significant perturbations at the proteome level. For the metE strain, several genes related to methionine synthesis, salvage and transport appear to be upregulated. Many of these genes are normally repressed by the TF metJ in its active

form, when binding S-adenosylmethionine (SAM). A reduction in methionine biosynthesis could lead to a reduction in SAM, reducing repression from metJ. For the pfkA strain, glycolytic genes did not appear to be upregulated significantly. Genes related to the glyoxylate cycle (aceA, aceB), maltose uptake and utilization (malM, malE, malF) and glutamate degradation (gadB) appeared to be up-regulated. A malate dehydrogenase (mqo) was strongly downregulated. For the ispH strain only the fkpB gene was downregulated, possibly because of its position in the *ispH* operon. However, a proteome-wide response upon 2 fold lower levels of IspH could not be observed.

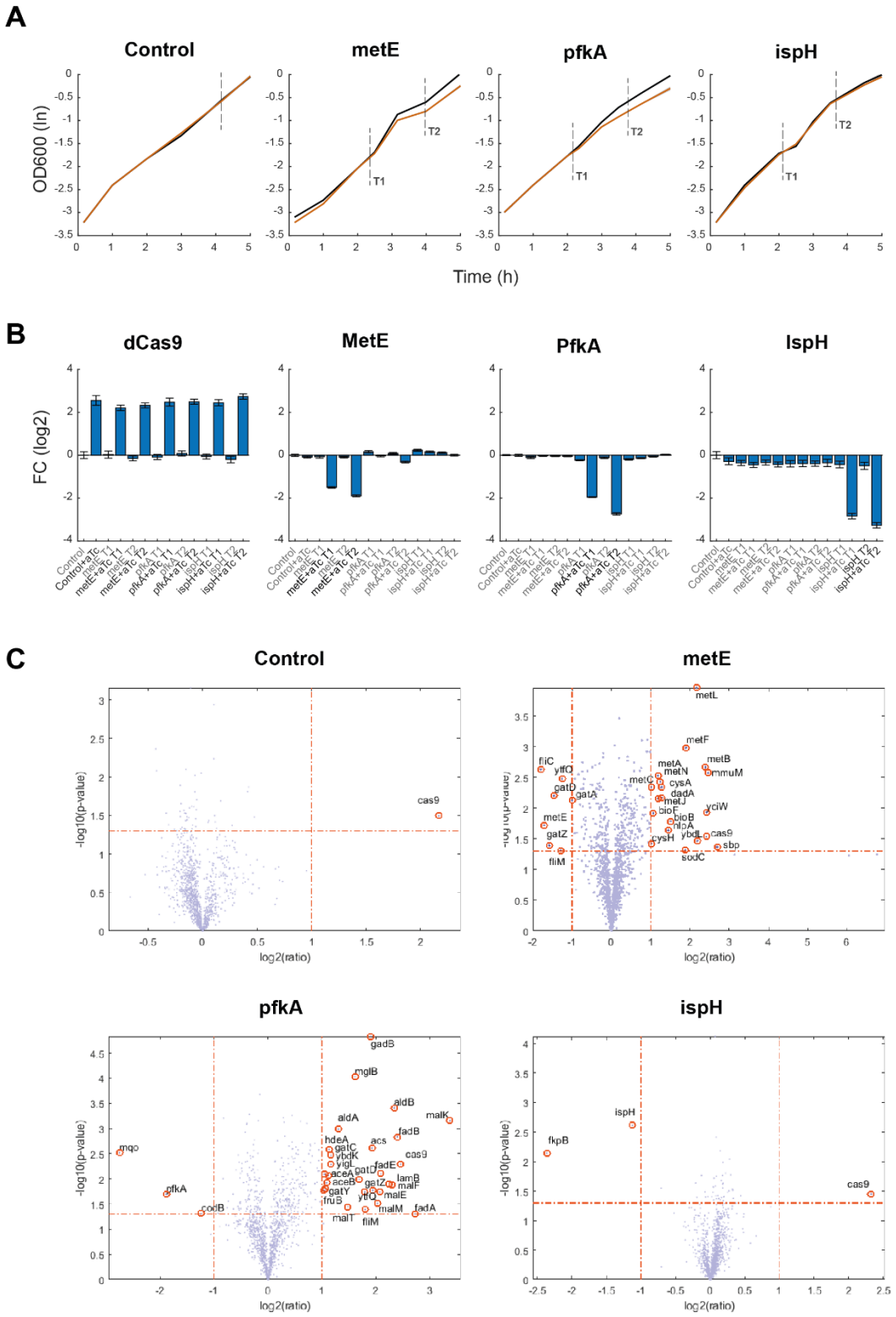


Figure 7: Characterization of 4 CRISPRi strains. (**A**) Growth curves for the 4 strains. Black lines represent growth of strains (n=2) without induction of dCas9 expression, orange lines represent strains growing in presence of aTc. Vertical dotted lines represent proteomics sampling timepoints (T1, T2). The control strain was sampled only at a single timepoint. (**B**) Bar plots representing abundance of selected proteins (n = 2), normalized to the proteome data of the uninduced control strain. Error bars represent the propagated error. (**C**) Volcano plots for proteomes of the CRISPRi strains, normalized to the proteome of the relative un induced strain at the same timepoint (T2). Red dotted lines represent cutoffs for fold change (FC>2, vertical lines) and significance (p-value>0.05, horizontal lines).

2.2 - Discussion

Here, we set to characterize and optimize a setup to precisely perturb metabolic genes using CRISPRi. We show that a strain with a genomically integrated dCas9 *cassette* can obtain an inducible activation of CRISPRi, without affecting cell physiology due to burden or dCas9 toxicity, as it has been shown previously when overexpressing dCas9 [116]. Screening a large library of CRISPRi strains targeting metabolic genes confirmed inducibility of the strain. We could observe that interference of different genes enforces a wide range of response times. Known rate limiting enzymes as MetE and PyrE had the shortest response times, confirming the hypothesis that the degree of the response time might reflect whether enzymes are expressed near a critical level or if they are made in excess. Moreover, even for such rate-limiting enzymes the response occurred 3 hours after inducing CRISPRi, and for many other strains the response time was longer or absent. We then set a sampling experiment to verify the response of cells to CRISPRi at the molecular level. dCas9 levels increased in all induced cultures, while the targets of interference were correctly reduced in their concentrations in the relative conditions. We could observe that comparing data from induced and uninduced cultures sampled at the same time allows to avoid noise and identify specific responses caused by enzyme downregulation. When normalizing the data in such way, we could observe specific responses for 2 of the analyzed strains. In the case of the metE strain, we could observe a specific upregulation of related genes in methionine biosynthesis. When perturbing expression of *pfkA* we observed an upregulation of transporters and TCA cycle related genes. In both cases, the perturbation seemed to become stronger with time. In the case of the ispH strain, we could not observe a significant response in gene expression or in the metabolome. Interestingly, ispH is known to be an essential gene for *E. coli* in the tested cultivation conditions, as *metE* and *pfkA*. It could be speculated that an unknown isoenzyme or an enzyme with moonlighting activity [117] can compensate measured lower amounts of IspH. Alternatively, it could be possible that IspH does not function at full capacity and that therefore the enzyme was not diluted to a critical concentration. Overall, we show here that integrated dCas9 *cassette* integrated in the genome of *E. coli*

allows for precise and burdenless interference of gene expression, which can be exploited for large scale functional screenings and multi-omics studies.

3 - The metabolome buffers CRISPRi-knockdowns of enzymes in *E. coli* metabolism

3.1 - Results

3.1.1 - 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 [104], and a constitutively expressed single guide RNA (sgRNA) on a plasmid [94] (**Figure 8A**). To evaluate dynamics of gene interference with this CRISPRi system, we targeted a YPet reporter protein inserted in the *E. coli* genome [104]. These experiments showed an exponential decrease of the YPet content per cell, indicating a constant dilution of the YPet protein by growth (**Figure 8B**). 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 [118]. Moreover, YPet expression was only repressed in the presence of the dCas9 inducer aTc, showing tight control of the CRISPRi system (**Figure 8B**). 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 8C**). 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 8D**). 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).

In conclusion, CRISPRi allowed us to induce dynamic decreases of protein-levels (**Figure 8B**). 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 [108]. The comparably weaker growth defects of the other pyrimidine knockdowns indicated that these enzymes do not operate at a critical level. In other words, these enzymes are expressed at higher levels than absolutely necessary for UMP biosynthesis and growth (enzyme overabundance). However, an alternative explanation is that the sgRNAs targeting these genes are weaker or not functional. Therefore, we next targeted genes with several sgRNAs, and designed sgRNAs for all metabolism-related genes in *E. coli*.

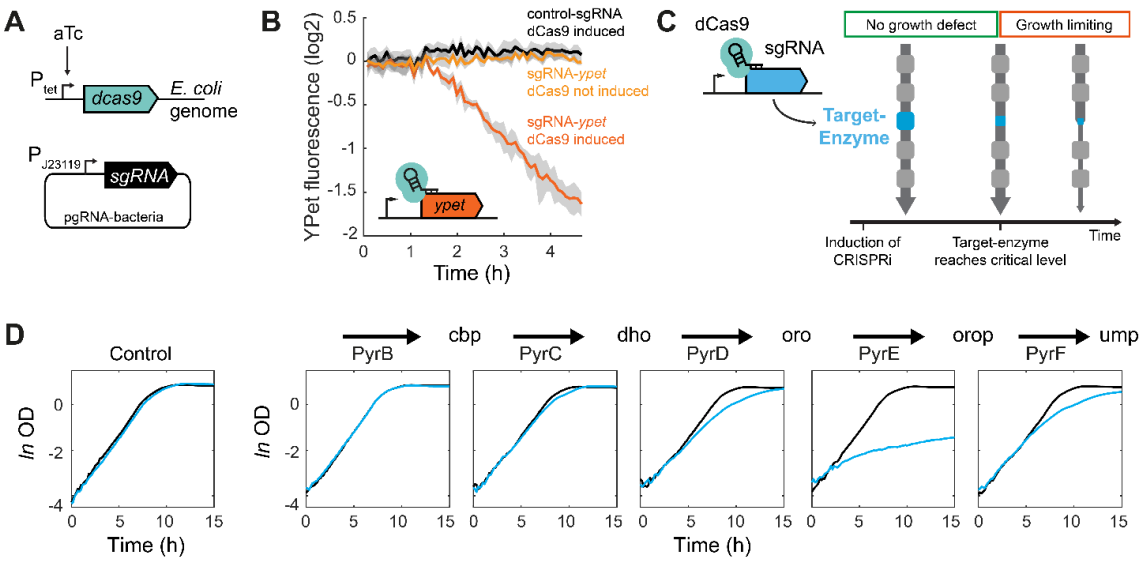


Figure 8. Dynamic knockdowns of enzymes with CRISPR interference

(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.

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

The latest genome-scale model of *E. coli* metabolism, *i*ML1515, includes 1515 genes [119] 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 9A**). 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 S3**). We cultured the library for 13 hours on glucose minimal medium without induction of dCas9, which hardly altered the composition of the library (**Figure S3**). 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 hours in intervals of 1 hour (**Figure 9A**). To assess reproducibility, we used two independent cultivations. Every two hours, the cultures were back-diluted into fresh medium, to avoid limitations of oxygen and nutrients. Growth of single CRISPRi strains was quantified as fold-change of sgRNA abundances, which was reproducible between the two experiments (**Figure S4**).

To explore dynamic patterns in sgRNA abundances, we performed k-means clustering with time profiles of fold-changes (**Figure 9B**). The abundances of 45% of the sgRNAs were constant for 14 hours (cluster A). Another 30% of the sgRNAs in cluster B showed a slight increase, suggesting that these guides produce a small fitness advantage. The remaining 25% of sgRNAs in cluster C and D caused fitness defects. Cluster C represents mild fitness defects (average fold-change 0.5 at 14h), and cluster D stronger fitness

defects (average fold-change 0.1 after 14h). Cluster C and D include 1789 sgRNAs, which target in total 748 genes. Out of these 748 genes, 387 genes have at least two sgRNAs in cluster C and D, and we considered these genes as potential metabolic-bottlenecks. According to simulations with the *i*ML1515 model, 277 of the 387 metabolic-bottleneck genes (71%) encode enzymes that carry metabolic flux with glucose as sole carbon source. 218 of the metabolic-bottleneck genes (56%) are essential for growth on glucose medium (**Figure 9C**). 89 genes (23%) are neither essential nor encode for enzymes with metabolic flux. Phenotypes of 9 out of these 89 genes can be explained by polar effects, as an essential or flux-carrying gene is encoded downstream of the targeted gene in the same operon. What caused fitness defects of the remaining 80 genes is unclear and it seems likely that these genes have previously unrecognized functions that affect fitness (e.g. transporters).

Next, we identified at which time point the knockdowns created a metabolic bottleneck and when they became growth limiting. Therefore, we estimated a "response-time" for the 387 metabolic-bottleneck genes. The response time was defined as the point when the fold-change of sgRNA abundance was 0.8. To obtain robust estimates of response times we fitted sigmoidal functions to the time-courses of fold-changes (**Figure 9D**). Genes that had the shortest response times and were therefore the most sensitive targets in our screen were: the *nrdA/nrdB* operon, *ppc*, the *ilvE/ilvD* operon and *fbaA* (**Figure 9D**). All enzymes encoded by these genes catalyze essential reactions in primary metabolism: biosynthesis of deoxyribonucleotides (NrdAB), branched chain amino acids (IlvE and IlvD), glycolysis (FbaA) and anapleurosis (Ppc). Notably, all sgRNAs that target these genes had similar dynamics, suggesting that the repression efficiency of sgRNAs has little influence on fitness defects of sensitive targets.

In summary, only 6 out of 1513 metabolism-related genes were very sensitive to knockdowns, as they had response times below 2 h. Another 32 targets had response times below 3 h (**Figure 9E**). The majority of knockdowns, however, responded late to induction of CRISPRi (on average 6.5 hours). This suggests that *E. coli* is robust against reducing the abundance of most metabolic enzymes and that only few enzymes (like NrdAB) are expressed at critical levels. Next, we wondered how strongly the abundance of target-enzymes decreased and which mechanisms buffered low enzyme-levels.

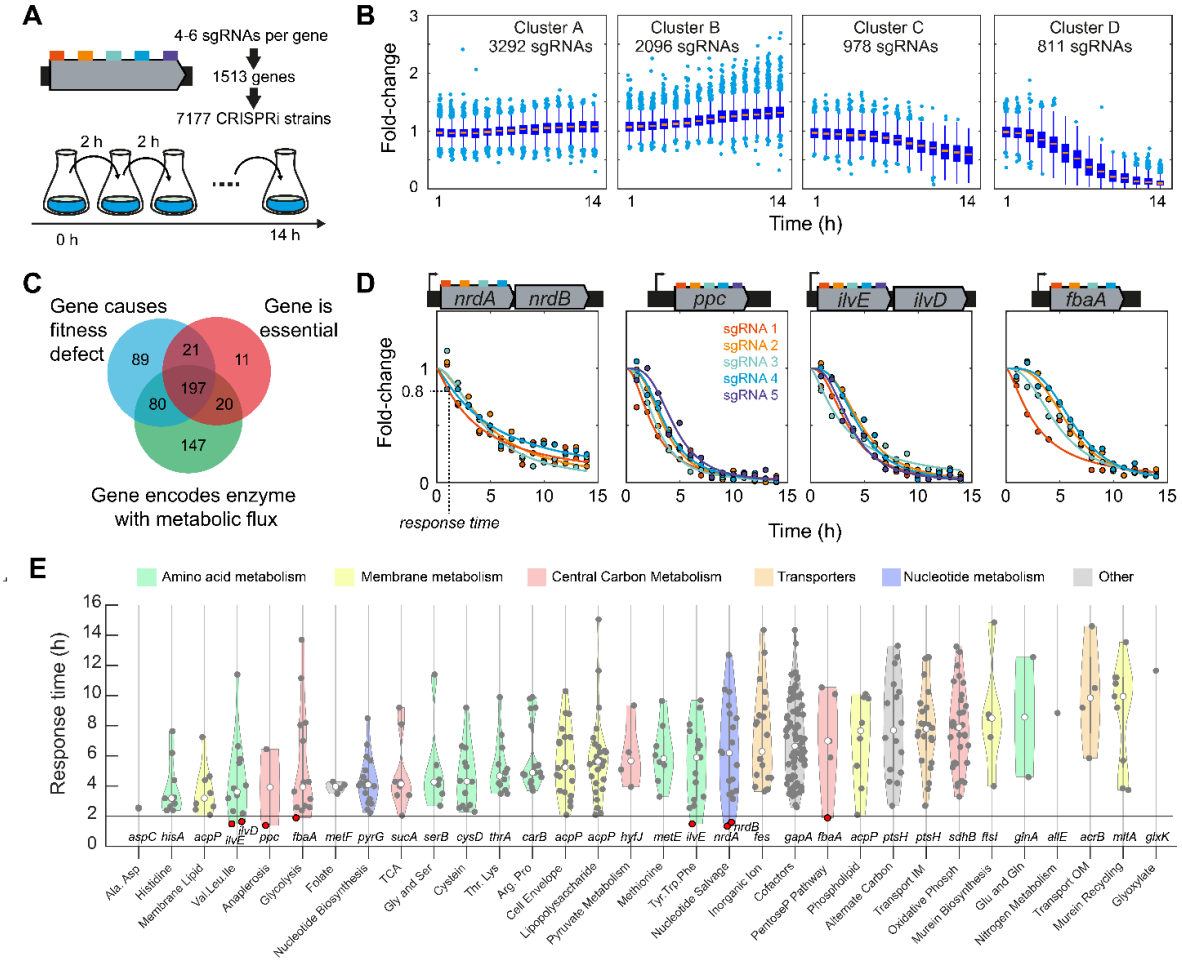


Figure 9. Dynamic knockdowns of 1513 genes in the metabolic network of *E. coli*

(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 8A). 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. (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. (C) Venn diagram showing the overlap between 387 genes that caused fitness defects in the CRISPRi screen (blue, genes with at least 2 sgRNAs in cluster C and D), genes that are essential on glucose minimal medium (red), and genes that encode enzymes with metabolic flux (green). (D) Fold-changes of all sgRNAs targeting the 4 most sensitive targets in the CRISPRi library. 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 0.8. (E) Response times of all 387 genes that showed fitness defects in the CRISPRi screen. Shown is the average response time of the 2 strongest sgRNAs of each target-gene. Target-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.

3.1.3 - CRISPRi achieves similar and specific decreases of enzyme-levels

To probe how strongly CRISPRi decreases the abundance of target-enzymes, we measured the proteomes of 30 CRISPRi strains (**Figure 10A, Figure S5**). The 30 target-enzymes included one of the most sensitive enzymes in our pooled 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 12-well plates and measured their proteomes 4.5 hours after dCas9 induction, which is the time when growth phenotypes appeared in the first strains (**Figure 10C**). Each strain was cultured in triplicates with and without induction of dCas9, resulting in a total of 180 proteome samples. CRISPRi downregulated specifically the target-enzymes, since target-enzymes decreased only in the respective knockdown and remained stable in the other strains (**Figure 10B**). The average decrease of target-enzymes was 5.1-fold, and decreases varied between 8-fold (PyrF) and 2.6-fold (MetE). Target-enzymes hardly decreased in un-induced cultures (**Figure 10B**), confirming that CRISPRi is tight and inducible. The degree of downregulation of the target-enzyme did not correlate with the reduction of the growth rate at the time point of sampling (**Figure 10C**, **Figure S6**), showing that different repression efficiencies of CRISPRi were not responsible for the different growth phenotypes. In summary, 4.5 h after induction of CRISPRi, target enzymes decreased on average 5-fold. Decreases of 19 target-enzymes caused no growth defect. For 10 target-enzymes growth rates declined just before the time point of sampling. Therefore, *E. coli* metabolism tolerates substantial decreases of enzyme-levels and we next wondered which mechanisms enable this robustness.

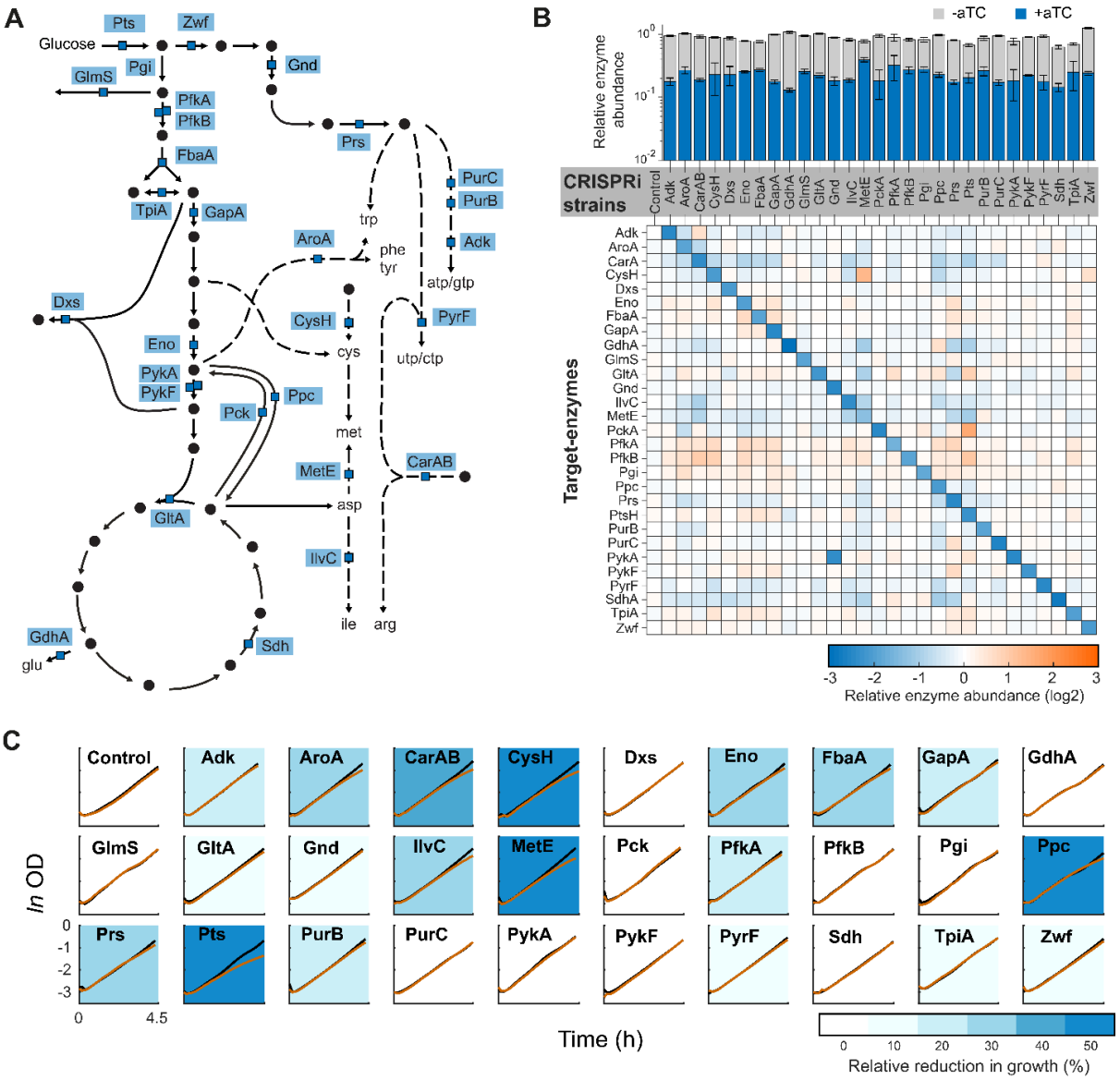


Figure 10. Growth defects and abundances of target-enzymes in 30 CRISPRi strains

(A) 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 Supplementary Figure S2. (B) 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. (C) Growth curves of the 30 CRISPRi strains. Uninduced 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.

3.1.4 - Substrates and allosteric effectors buffer decreases of enzyme-levels

To understand how *E. coli* metabolism compensated the ~5-fold decrease of enzyme-levels, 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). Metabolites that responded most strongly to knockdowns of enzymes were often substrates of the respective reactions (**Figure 11A**). In 18 out of 29 knockdowns, the substrate increased more than 2-fold and was one of the most abundant metabolites. Products, in contrast, were more stable than substrates (**Figure 11A**, **Figure S7**). In the strains with strong substrate responses, the respective products decreased less than 2-fold, except of the Gnd-product ribose-P and the MetE-product methionine. The low methionine levels in the MetE strain imply that the growth defect of this strain is caused by a bottleneck in the methionine pathway. We hypothesized that the high concentration of substrates increases the active site occupancy of enzymes, which in turn increases their activity. Thereby, substrates could buffer knockdowns of enzymes by increasing their activity.

To test the hypothesis that substrates buffer decreases of enzyme-levels, we measured metabolites in the PfkA strain dynamically after induction of CRISPRi (**Figure 11B**). PfkA catalyzes an essential rate-limiting step in glycolysis, which is the phosphorylation of fructose-6-phosphate into fructose-1,6-bisphosphate. If fructose-6-phosphate buffers the decrease of PfkA, then its concentration should increase slowly and early after induction of the PfkA knockdown (resembling dynamics of the target-enzyme). In contrast, a rapid and late response of fructose-6-phosphate would indicate that the substrate increases because glycolysis is blocked. Indeed, the pool of hexose-phosphates (which include fructose-6-phosphate) increased early and slowly after induction of dCas9 (**Figure 11B**), supporting our hypothesis that substrates buffer enzyme knockdowns. The delay of 60 minutes between induction and increase of hexose-phosphates matches the activation time of CRISPRi that we measured with YPet (**Figure 11B**). Thus, as soon as PfkA levels decrease, the concentration of hexose-

phosphate starts to increase, which maintains stable glycolysis and constant levels of the PfkA-product fructose-1,6-bisphosphate. Therefore, dynamic metabolite responses to knockdowns reveal the capacity of substrates to buffer changes in enzyme-levels.

In 4 strains allosteric effectors of the target enzyme responded most strongly to the knockdown (CarAB, GlmS, Ppc and Zwf, **Figure 11A**). Ornithine, for example, is one of the many allosteric effectors of CarAB and ornithine increased more than 500-fold in the CarAB knockdown. Thus, allosteric activation of CarAB by ornithine seems to buffer the knockdown of the enzyme. Similarly, knockdown of Ppc resulted in a decrease of aspartate (13-fold) and malate (16-fold), which are allosteric inhibitors of Ppc. The decreases of aspartate and malate suggested that low abundance of Ppc is doubly compensated by relieving two negative allosteric feedbacks. The increase of the Ppc substrate PEP could additionally increase the activity of the enzyme. In the GlmS and Zwf strain we observed a similar relieve from allosteric inhibition, because their respective reaction product glucoseamine-P and NADPH decreased. NADPH is the allosteric inhibitor of Zwf [120], and glucoseamine-P is a potent inhibitor of GlmS activity [121].

To confirm that allosteric regulators buffer knockdowns, we measured again the dynamic response of metabolites in the CarAB knockdown (**Figure 11C**). Similar to the PfkA knockdown, ornithine responded slowly and early after induction of the CarAB knockdown, showing that ornithine compensated decreasing CarAB levels for ~2 hours. After 2 hours, the concentration of arginine decreased, which is the end-product of the arginine biosynthesis pathway and consumes the CarAB-product carbamoyl-phosphate. The other pathway that uses carbamoyl-phosphate is pyrimidine nucleotide biosynthesis, but the pyrimidine end-products CTP and UTP did not change in the CarAB knockdown. The higher robustness of pyrimidine over arginine biosynthesis is probably due to the higher demand for amino acids than for nucleotides. Therefore, CarAB reached a critical level after 160 minutes, when ornithine cannot compensate the CarAB knockdown anymore and arginine starts limiting cell growth.

In summary, substrates and allosteric effectors buffer decreases of enzyme-levels by increasing their activity. This buffering effect of the metabolome explains why *E. coli* grows normally for 2-3 hours after induction of enzyme knockdowns. Once this buffer is

exhausted, the reaction-product of the target-enzyme (or end-products of the target-pathway) decrease and will eventually limit growth. We then wondered how cells respond to such metabolic bottlenecks at the level of enzyme-level regulation, and we took a closer look at the proteome data.

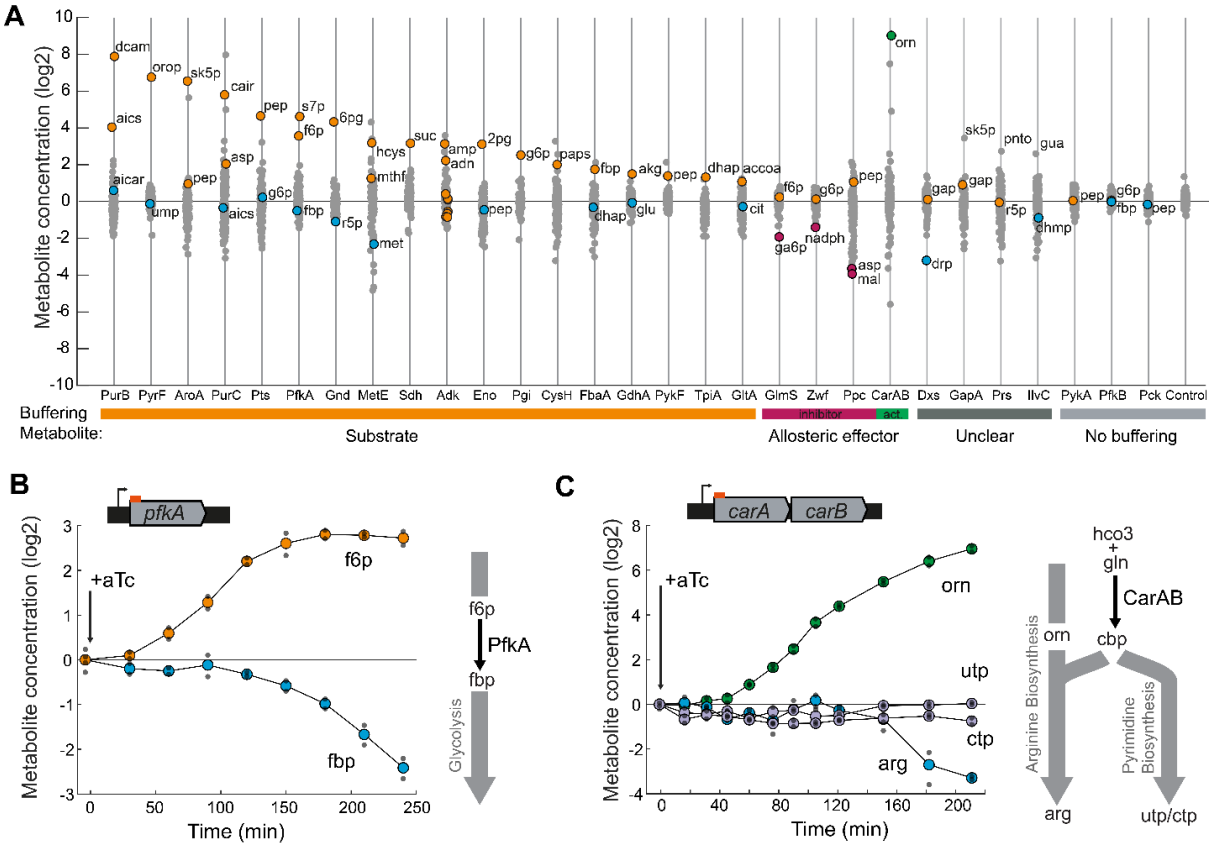


Figure 11. Metabolome of 30 CRISPRi strains and dynamic metabolite responses

(A) Intracellular concentration of 119 metabolites in the 30 CRISPRi strains. Metabolite levels are shown as log2 fold-change between induced and un-induced cultures. Samples were collected at the end of the 4.5 hours cultivations (see **Figure 10C**). Data are represented as mean (n = 2). Substrates of the target enzyme are shown in orange, products in blue, allosteric inhibitors in magenta and allosteric activators are green. (B) Time-course of hexose-phosphate (f6p) and fructose-1,6-bisphosphate (fbp) in the PfkA knockdown. Metabolite levels are normalized to the time point before induction. The culture was induced with aTc at t = 0h. Black dots are measurements in n = 2 cultures and colored dots are the mean. (C) Same as in (B) for ornithine (orn), arginine (arg), utp and ctp in the CarAB knockdown. Note that isomers were not separated: g6p and f6p is the total pool of hexose-p, r5p is the total pool of pentose-p, dhap and gap is the total pool dhap/gap.

3.1.5 - Metabolites cause a compensatory upregulation of enzymes in the target-pathway

Proteome data showed that CRISPRi achieves specific and in average 5-fold reduction of the abundance of target-enzymes (**Figure 10B**). In 20 of 29 knockdowns the target-enzyme was the most strongly downregulated protein among all 1506 measured proteins (**Figure 12A**). The number of significantly changed proteins (2-fold, p-test<0.05) had a strong correlation with the reduction in growth of the respective knockdown (**Figure 12B**). This means that strains with a growth defect had stronger proteome changes, whereas the proteome was stable in strains without a growth defect. The latter group consisted of 19 strains with less than 20 significantly changed proteins, despite low levels of the target-enzymes. The constant proteome in these 19 strains confirmed that knockdowns are buffered at the metabolome level and not at the proteome level.

We then analyzed if the 10 strains with stronger proteome changes showed a global growth-dependent proteome response [65] or if proteome changes were specific. Because the average similarity of proteome changes between pairs of CRISPRi strains was only 6% (**Figure S8**), we concluded that each knockdown caused specific proteome changes. As expected, knockdowns of enzymes that are close in the metabolic network caused more similar proteome responses. For example, the most similar proteome changes occurred in knockdowns of neighboring enzyme pairs: Pgi and Gnd (40% similarity), GapA and FbaA (39% similarity) (**Figure S6**). Therefore, decreasing target-enzymes to a critical growth-limiting level enforces specific proteome changes, which affected different metabolic subsystems (**Figure S9**).

In five metabolic pathways we noticed a compensatory upregulation, because enzymes within the metabolic pathway of the target-enzyme were upregulated (**Figure 12C**). For example, all enzymes in the arginine biosynthesis pathway increased in the CarAB strain, demonstrating that enzyme-level regulation responds to the arginine starvation in this strain (**Figure 11A** and **Figure 11C**). Similarly, the methionine starvation in the MetE strain (**Figure S10**) caused a compensatory upregulation of enzymes in the methionine pathway. In the IlvC strain, enzymes in valine and isoleucine biosynthesis were upregulated. Enzymes involved in sulfur assimilation increased in the CysH strain, and

enzymes in aromatic amino acid biosynthesis were upregulated in the AroA strain. All of these upregulated pathways include the target-enzyme (which is downregulated due to CRISPRi). However, we observed the same response for distal target-enzymes that are not in the upregulated pathway. For example, methionine enzymes increased also in the Ppc strain, thus indicating that limiting anapleurosis has the strongest effects on methionine biosynthesis. Aromatic amino acid biosynthesis was also upregulated in knockdowns of enzymes in lower glycolysis (Eno, FbaA, GapA and TpiA). The knockdowns in lower glycolysis consistently caused low levels of phosphoenolpyruvate (PEP)(**Figure S10**), which is a precursor for aromatic amino acids biosynthesis. Therefore, limited supply of PEP for aromatic amino acid biosynthesis pathway caused the same compensatory upregulation as a knockdown of AroA within the pathway.

The compensatory upregulation occurred probably at the level of transcription, because expression of four pathways is regulated by related transcription factors: the arginine pathway by ArgR, the methionine pathway by MetJ, sulfur assimilation by CysB and aromatic amino acid biosynthesis by TrpR. We expected that these transcription factors responded to the knockdown because the concentration of their allosteric regulators changed accordingly. Indeed, the CarAB strain had the lowest levels of arginine across all 30 CRISPRi strains (**Figure 12C**). This reduces the activity of ArgR and thereby de-repressed expression of arginine biosynthesis enzymes in the CarAB strain. Similarly, the MetE and Ppc strains had low levels of S-adenosylmethionine (SAM). Low SAM levels reduce MetJ activity and thereby caused the compensatory upregulation of the methionine pathway. Changes of acetyl-serine explained upregulation of sulfur assimilation, and low valine levels caused upregulation of the valine and isoleucine biosynthesis enzymes (in this case via transcriptional attenuation). Only the response of the TrpR regulator, tryptophan, did not match the upregulation of enzymes in aromatic amino acid biosynthesis: in most strains with a transcriptional response, tryptophan levels did not decrease more than 2-fold and they even increased for the GapA condition. This suggests that additional regulators might control expression of these enzymes.

To obtain additional evidence that the compensatory upregulation occurs at the level of transcription, we used transcriptional fluorescent reporters [122] (**Figure 12D**). GFP

expression from a MetJ regulated promoter confirmed the compensatory upregulation in the MetE strain. Similarly, an ArgR regulated promoter was upregulated in the CarAB strain. Promoter activity increased with a 2-hour delay after inducer addition. This is also the time when growth defects appeared in these strains, thus indicating that the compensatory upregulation started when the target-enzyme reached a critical level.

In conclusion, knockdowns without growth defect had a stable proteome despite a 5-fold decrease of target enzymes. Stronger proteome changes occurred only in strains that had a reduction in growth at the time point of sampling. These proteome changes were specific, because different proteins changed in different knockdowns. In 9 knockdowns the responses were remarkably precise: either enzymes within the target pathway were upregulated (MetE, CysH, IlvC, AroA strains) or enzymes that utilize the reaction product of the target were upregulated (CarAB, Ppc, Eno, FbaA, GapA and TpiA strains). Once more, metabolites were buffering the knockdowns, in these cases by allosteric interactions with transcription factors (MetJ, ArgR, CysB, TrpR) or transcriptional attenuation (valine and isoleucine). Thus, the metabolome buffers knockdowns both at the level of enzyme activity and enzyme abundance.

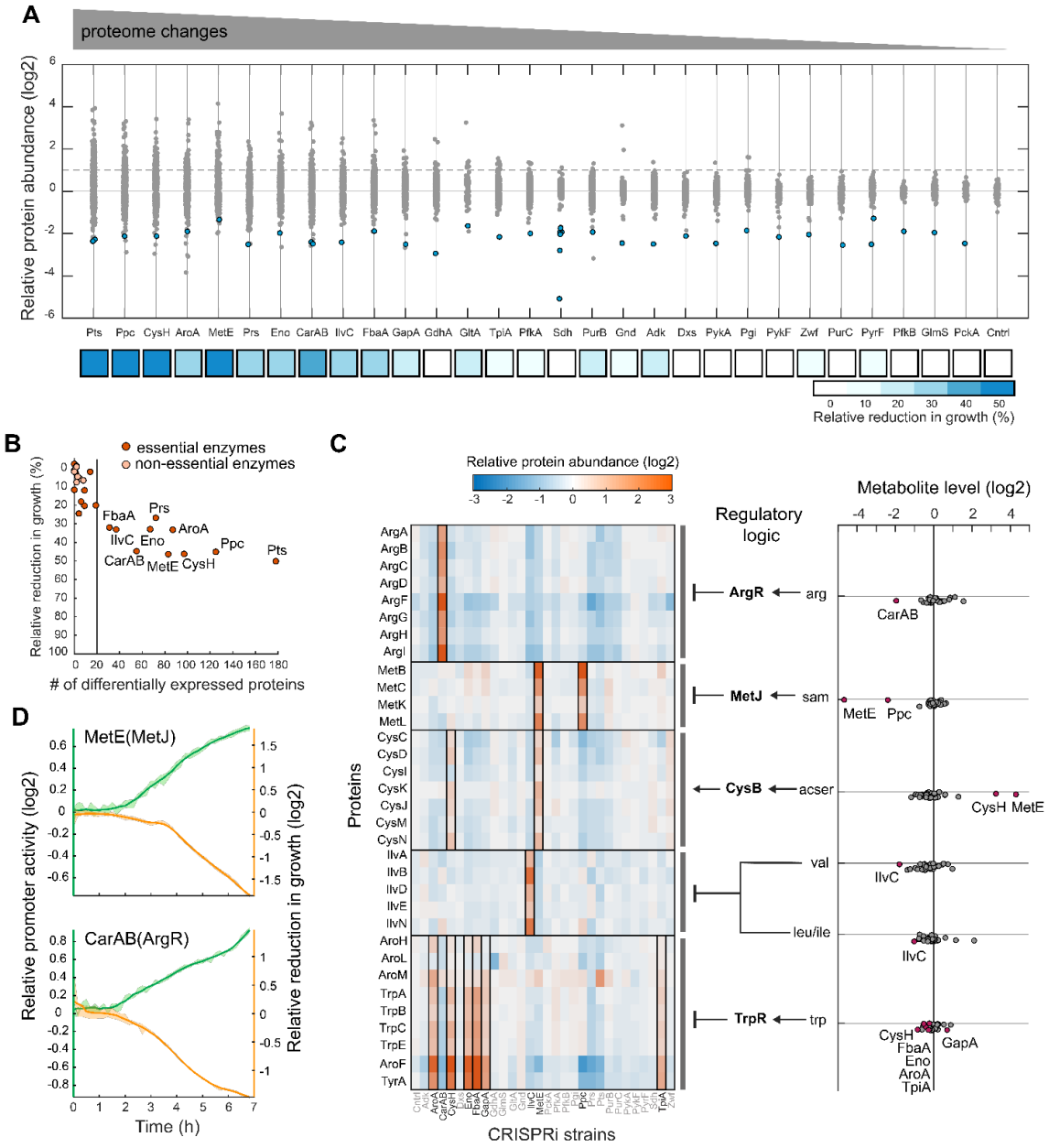


Figure 12. Localized proteome changes occur in CRISPRi strains with a growth defect

(A) Abundance of 1506 proteins in the 30 CRISPRi strains. Protein levels are shown as log2 fold-change between induced and un-induced cultures. Data are shown as mean (n = 3). Strains are ordered based on the number of differentially expressed proteins (FC>2, p-value<0.05). Blue dots indicate target-enzymes. Colored boxes show the reduction in growth rates at the time of sampling (see also **Figure 3**). (B) Correlation between the number of differentially expressed proteins (FC>2, p-value<0.05) and the reduction in growth rates for the 29 CRISPRi strains. (C) Fold-changes of enzymes in pathways of arginine biosynthesis, methionine biosynthesis, sulfur assimilation (Cys), valine/isoleucine biosynthesis and aromatic amino acids biosynthesis. 10 CRISPRi strains with a compensatory upregulation are shown in bold. (D) Concentration of metabolites that are allosteric effectors of transcription factors ArgR, MetJ, CysB and TrpR. 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). (E) The MetE and CarAB knockdowns were transformed with GFP reporter plasmids. CarAB expressed an ArgR regulated promoter (pUA66-*argE-gfp*). MetE expressed 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 between induced and un-induced cultures is shown in orange. Curves are means of n=3 cultures, shadows represent standard deviation. aTc was added at t = 0 h.

3.1.6 - 6-phosphogluconate buffers knockdowns in the pentose-phosphate pathway

Apart from compensatory responses at the level of the target-enzyme itself, we noticed a third compensatory mechanism. Knockdown of Gnd in the pentose-phosphate pathway activated an alternative pathway, the Entner-Doudoroff (ED) pathway. This is consistent with previous reports about increased flux through the ED pathway in a Gnd knockout [123,124]. The ED pathway utilizes the Gnd substrate 6-phosphogluconate, which accumulated in the Gnd knockdown (**Figure 13A**). Thus, higher expression of the ED pathway in the Gnd knockdown might enable overflow of the excess of 6-phosphogluconate.

We wondered if upregulation of the ED pathway was also caused by a metabolite. Transcription of the ED pathway is regulated by the two transcriptional repressors KdgR and GntR. While KdgR controls only the two ED enzymes (Edd and Eda), GntR has additional targets in uptake of gluconate (**Figure 13A**). The increase of an additional GntR target (*gntT*) suggested that GntR responded to the knockdown of Gnd (**Figure 13B**). The activity of GntR is allosterically inhibited by gluconate [125]. Therefore, we assumed that accumulation of 6-phosphogluconate produced small amounts of gluconate, which inhibited GntR and de-repressed transcription of Edd and Eda (**Figure 13C**). In the un-induced Gnd strain the concentration of gluconate was 50 µM, which is comparable to previous measurements in *E. coli* [126]. Induction of the Gnd knockdown, led to an increase of gluconate to 184 µM. This increase in gluconate concentrations was probably sufficient to inhibit GntR and increase expression of the ED pathway. The presence of gluconate in the GND knockdown and its regulatory role in bypassing the oxidative PPP, reveals that gluconate acts as regulatory metabolite, which does not participate in metabolism but in regulation. Thus, gluconate could sense imbalances between the oxidative and the non-oxidative PPP, and adapt expression of the ED pathway accordingly. We expected that we can alter this regulation by disrupting the

interconversion of 6-phosphogluconate to gluconate. Indeed, deletion of gluconate kinase (*gntK*) prevented re-phosphorylation of gluconate, such that gluconate levels increased even further: 246 µM in the un-induced Gnd knockdown and 620 µM in the induced Gnd knockdown. The high gluconate levels in the Δ*gntK* strain upregulated again the ED pathway (with and without induction of the Gnd knockdown). ED enzymes were even stronger overexpressed in the Δ*gntK* strain than in the strain with only knockdown of Gnd. The higher expression of ED enzymes almost prevented the accumulation of 6-phosphogluconate in the Gnd knockdown, confirming that the ED pathway enables overflow of an excess of 6-phosphogluconate.

In summary, 6-phosphogluconate levels are sensitive to imbalances in the oxidative pentose phosphate pathway. The ED pathway responds to 6-phosphogluconate via an indirect interaction with gluconate-GntR, and enables overflow of an excess of 6-phosphogluconate.

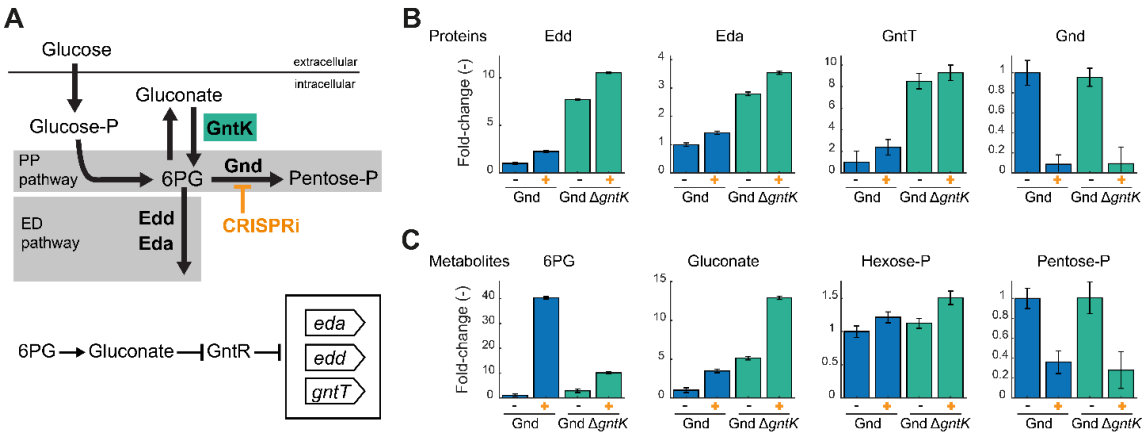


Figure 13. 6-phosphogluconate coordinates the Entner-Doudoroff pathway with the Pentose-Phosphate pathway.

(A) Metabolic map 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), as well as 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 YYsCas9 strain (blue) and the YYsCas9-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. Data are means of n = 3 cultures, error bars are propagated errors. (C) Same as in (B) for intracellular metabolites (6PG: 6-phosphogluconate).

3.2 - 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 limit metabolic flux and eventually cellular fitness. Theories like Metabolic Control Analysis predicted that metabolism is insensitive to the exact abundance of single enzymes [127–130], 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 [131] and *E. coli* [132]. But how changes of enzyme-levels affect metabolism is largely unexplored. In this study, we used CRISPRi to perturb the expression of enzymes and investigated the effects on metabolism and fitness systematically.

First, we used CRISPRi to knockdown 1513 genes that are included in the latest genome-scale model of *E. coli* metabolism, *i*ML1515 [119]. We leveraged an inducible CRISPRi system for dynamic knockdowns of enzymes. Knockdown of 387 out of the 1513 metabolism-related genes caused a fitness defect. However, fitness defects were delayed relative to the addition of inducer, on average 6.5 h. At this time point, enzyme-levels should be markedly reduced, since proteome data for 29 strains showed a ~5-fold reduction of target-enzymes after 4.5 h. Thus, decreasing enzymes below endogenous levels had no immediate effect on cellular fitness. Previous reports that support this observation showed that metabolic enzymes are expressed in excess [133–135] and that *E. coli* keeps reserve capacities of enzymes [120,131].

Yet, few enzymes affected fitness almost immediately after induction of CRISPRi (NrdAB, Ppc, IlvE and FbaA). A hypothesis is that these enzymes are rate-limiting control points in the metabolic network of *E. coli*. The most sensitive target was ribonucleoside-diphosphate reductase (NrdAB), which seems a reasonable control point because it supplies deoxyribonucleotide triphosphates (dNTPs) for DNA replication. Previous work confirms a rate-limiting function of NrdAB, because its concentration is directly proportional to the rate of DNA synthesis [136]. 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* [137]. The high sensitivity of the *ilvE*/*ilvD* operon is probably due to the frameshift mutation upstream of *ilvG*, which causes suboptimal expression of the operon [138].

In 30 CRISPRi strains, we measured the metabolome and proteome, in order to examine how decreases of enzyme-levels affected metabolism. In theory, metabolism could respond in two distinct ways to knockdown of an enzyme: either by global upregulation of the whole metabolic network, or by specific regulation of just the critical target-enzyme. Here, we observed the latter case, since metabolome and proteome responses were specific and centered around the target-enzyme. For example, substrates or allosteric effectors of the target-enzyme were among the top responding metabolites. Dynamics of these local metabolite concentration changes support the hypothesis that they increased the activity of the target-enzymes, and that this buffered knockdowns. This observation matches reports about the concentration of intracellular metabolites, which are often near or even below binding constants of substrates or allosteric effectors [126,139]. Moreover, the metabolome buffered knockdowns at the level of transcription. For example, allosteric regulators of transcription factors (arginine, SAM, acetyl-serine) and transcriptional attenuation by valine were responsible for a compensatory upregulation of enzymes in pathways that were most seriously affected by the knockdown. Thus, CRISPRi enforces strong responses of metabolites that interact with the target-enzyme directly or indirectly with regulators of the target-enzyme. From a methodological perspective, this highlights the large potential of CRISPR perturbations to infer functionally relevant interactions between metabolites and proteins, which are currently measured by spiking purified metabolites into cell extracts [140].

In conclusion, our study shows that the metabolome can buffer decreases of enzyme-levels, because enzyme-levels are higher than absolutely needed to maintain metabolic flux. This mechanism may ensure a high constancy of metabolic flux despite expression noise [141,142] or mutations that occur during the evolution of metabolic networks [143].

4 - Homeostasis of the *Escherichia coli* biosynthetic metabolome across different environments

4.1 - Results

4.1.1 - A comprehensive and systematic dataset of the *E. coli* metabolome

In this work, we measured >100 hydrophilic intracellular metabolites across 16 environmental conditions and in 3 *E. coli* reference strains grown on glucose, to explore the response of these compounds to different environments. For all conditions, we cultivated *E. coli* in three independent shake-flask batch-cultures. All cultures were sampled in mid-exponential phase when they reached an OD_{600} of 0.5, to ensure that oxygen and carbon sources were not limiting growth (**Figure 14A,** left panel). An exception were stationary cultures, which were sampled 1 and 2 days after entering stationary phase (sampling OD_{600} of 5.2 and 5.1, respectively). For all 19 conditions, fast filtration and extraction in cold acetonitrile (ACN) was used to obtain metabolite extracts, which we mixed with ^{13}C internal standard and then measured by LC-MS/MS [83,86]. To obtain absolute metabolite concentrations, we first calibrated the ^{13}C internal standard with authentic standards (**Figure 14A,** middle panel) and used the ratio of ^{12}C and ^{13}C signals in our samples for quantification. Additionally, we used the correlation of signals in the ^{12}C and ^{13}C channels to score the quality of LC-MS/MS signals. Out of the 147 metabolites targeted by our method, we retained for further analysis 124 metabolites with high quality peaks (**Figure 14A**). For 67 metabolites we could determine absolute intracellular concentrations across all conditions.

The median error between replicates was of 15 % (relative standard deviation, **Figure 14A,** right panel), showing that measurements were robust and reproducible. Moreover, the energy charge (given as ([ATP]+0.5[ADP])/([ATP]+[ADP]+[AMP]) was high and remarkably constant in all growing cultures, ranging between 0.93 to 0.97 (**Figure 14A,** right panel). This confirms that sampling was fast and efficient, because ineffective sampling and quenching would immediately affect ATP levels, which have a turn-over

time of few seconds [144]. The low energy charge in stationary phase cells was expected and decreased from day 1 to day 2. Changes of the metabolomes between conditions were sufficient to group cells grown on minimal media, complex medium or in stationary phase (**Figure S11**). Metabolomes of cells grown under stresses as high temperature, low pH and high osmotic stress clustered together with metabolomes of cells grown in minimal media, suggesting that these conditions do not impact strongly metabolite concentrations.

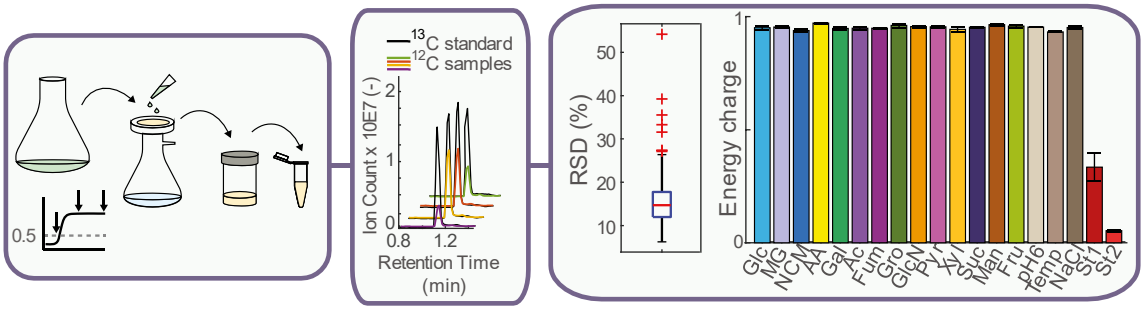


Figure 14. General overview of the experiments and data quality assessment. Cell cultures were grown aerobically, and sampled at OD_{600} of 0.5, or at 1 or 2 days of stationary phase. Quick filtration and quenching/extraction in ACN:methanol:H_2O were applied to get the intracellular metabolites. Samples were analyzed by LC-MS. Median error between experiments was 15%. High energy charge (EC) values were calculated for all experiments, suggesting an efficient nucleotide extraction procedure.

4.1.2 - The metabolome of growing cells is independent from growth effects

Next, we compared intracellular concentrations measured in this study, with those from previous studies [86,145]; intracellular levels of metabolites show a good match, especially with the more recent study, despite somewhat different cultivation conditions (**Figure S12**). After validating results from our measurments against published datasets, the next step was to compare it to a matching dataset of absolute protein levels of *E. coli* [146]. In this case, cells were cultivated under identical cultivation conditions, resulting in strongly comparable growth rates (**Figure S13**). The key question in the comparison of the two datasets was whether metabolite levels are also affected by growth effects, as

it has been shown for protein levels. We compared metabolite levels to the relative growth rates for conditions with strains in exponential growth on minimal media: 13 metabolites (out of 124, 10.5%) displayed a growth dependent tendency (R^2>0.4) (**Figure 15A**, **Table S3**), including the known regulator of catabolite repression, cAMP [145,147]. Metabolites with the strongest correlation with growth rates were IMP and 5-Amino-1-(5-phospho-D-ribosyl)imidazole (CAIR), which are both precursors for the biosynthesis of purines (**Figure S14**). Interestingly, data from Kochanowski et al. (2017), for cells grown on different carbon sources also suggests an overall independence between growth rates and metabolite levels (**Figure S15**).

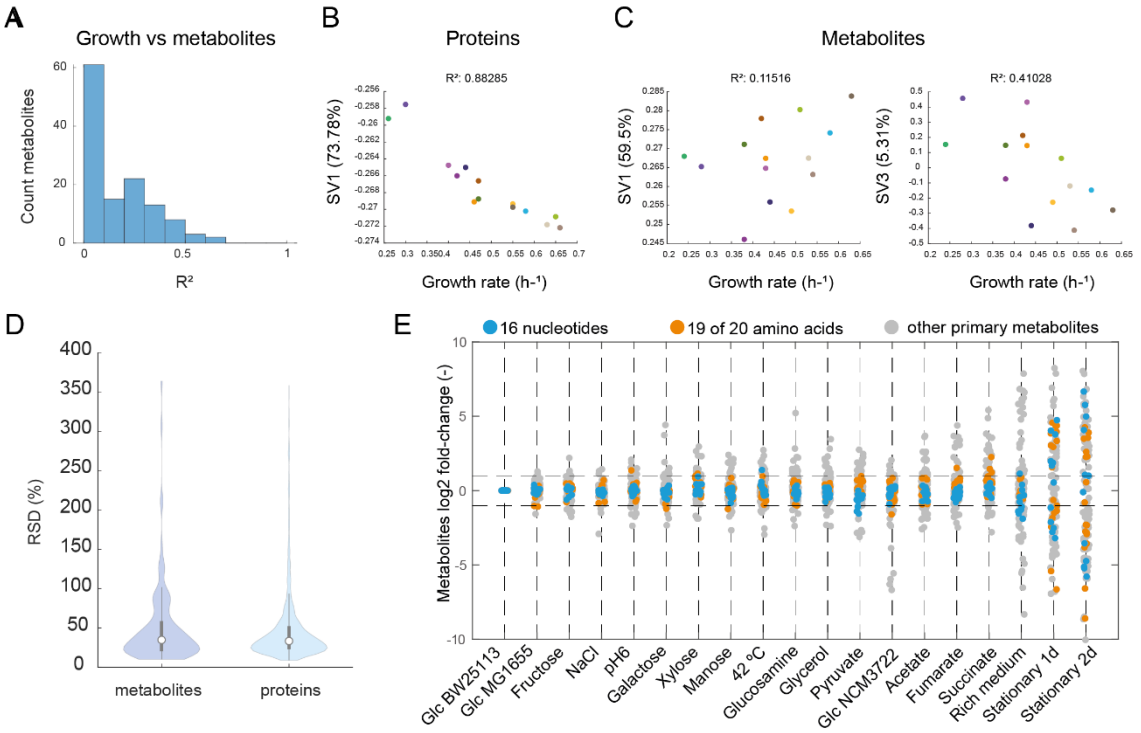


Figure 15: Analysis of the metabolome dataset and comparison with the matching proteome dataset from Schmidt et al. 2016. (**A**) histogram counting metabolites and their relative correlation coefficient between their levels and the relative growth rates. R^2 indicates the squared Person correlation coefficient. Most metabolites display a very low growth-dependency. (**B**) First singular vector (SV1, explaining 73.78% of the total variance) for the proteome dataset plotted against the respective growth rate, colors represent the different conditions based on Figure 1 (right panel). R^2 indicates the squared Person correlation coefficient between the first singular vector and the relative growth rates. (**C**) First singular vector (SV1, explaining 59.5% of the total variance) and third singular vector (SV3, explaining 5.31% of the total variance) for the metabolome dataset, plotted against the respective growth rate. Colors represent the different conditions based on Figure 1 (right panel). R^2 indicates the squared Person correlation coefficient between the relative singular vector and the relative growth rates. The first singular vector displays no correlation with the growth rate, while the third singular vector, albeit having a weak growth dependency, accounts for a small portion of the variability in the dataset. (**D**) Distributions of relative standard deviation (RSD) for each metabolite and each protein in the respective datasets. The median values in the distributions were respectively of 32.9% and 34.8%. (**E**) Detailed

overview of the relative metabolite levels for all conditions normalized (in log2) to the metabolite levels of *E. coli* BW25113 grown on M9 minimal medium with glucose as sole carbon source.

We next applied singular value decomposition [148] to deconvolute the matching proteomics or metabolomics datasets, for the same growth conditions mentioned above. In particular, we looked at which component correlates most strongly with growth rate values. In the case of the proteome data, the first component correlated strongly with growth and could explain together 73.8% of the variance in the dataset (**Figure 15B**, **Table S4**). In the case of the metabolome data, the first component showed no correlation with growth rates (**Figure 15C**). The third component showed the highest growth dependency. However, this component could explain only 5.3% of the variation in the data. Therefore, unlike for proteins, metabolite concentrations do not seem to be affected by global effects caused by changes in the growth rate. We compared then the variation for all metabolites or proteins: the median variation was comparable (relative standard deviation, 32.9% and 34.8% respectively, **Figure 15D**) indicating that metabolites and proteins have a similar degree of variation in different conditions. However, this variation is not driven by the same factor: protein variation is mostly affected by growth, unlike metabolites. Metabolite levels relative to the glucose condition are shown with a high resolution in **Figure 15E**. Metabolomes of conditions grown in minimal media display a limited variability, and in particular biosynthetic end-products like amino acids and nucleotides varied in almost all cases less than 2-fold. Thus, we took a closer look at amino acid and nucleotide metabolism.

4.1.3 - Homeostasis of amino acid metabolism

To visualize the variation and identify outliers, amino acid concentrations were normalized to the median value of all conditions, and log2 of these values were plotted as an expanded boxplot (**Figure 16A**). Values that fell within the whiskers of the boxplot were grey, and outliers were plotted in a color of the matching experimental condition. Among all conditions, the most stable amino acid was methionine (RSD of about 13%), and the most variable amino acid was aspartate (RSD of 65%). Variation of amino acids

between conditions did not correlate with the quality of MS signal, error between the replicates, or the absolute level of amino acids (R^2 values <0.1, **Figure S16**).

We noticed that, on particular carbon sources, the concentration of individual amino acids was especially high. This phenomenon could in many cases be explained by the proximity of the entry point of the carbon source. For example, pentose phosphate pathway (PPP) derived phenylalanine peaked in xylose-grown cells and several amino acids that come from TCA cycle precursors were particularly high on fumarate (lysine, asparagine, glutamine, aspartate, asparagine, arginine). Histidine was high on mannose, suggesting that this carbon source supports high flux through PPP. Interestingly, high level of phenylalanine was reached on fumarate, which is not entering the metabolism near PPP. This effect in fumarate grown cells can be explained by increased availability of aspartate for the last transamination step in phenylalanine biosynthesis [149]. Similar effect could be responsible for high tyrosine abundance on fumarate.

Amino acids regulate their biosynthetic pathways in two different ways: 1) feedback inhibition via allosteric regulation of enzymes in the pathway by its end-product, or 2) control of expression levels of their biosynthetic pathway via transcriptional attenuation or through interaction with the specific transcription factors. To visualize the free amino acid concentrations in the context of their regulatory role, we plotted the intracellular concentrations together with the known inhibitory half-saturation constants (Ki) for allosteric enzymes involved in their biosynthesis, and with the dissociation constants (Kd) of the transcription factors that control their biosynthesis (**Figure 16B**). Interestingly, the 5 most abundant free amino acids in *E. coli* do not directly control their biosynthetic enzymes/pathways neither allosterically or transcriptionally. For the remaining amino acids, regulatory feedback loops are described, and strikingly, their intracellular concentrations match well with Ki or Kd values of the proteins involved in the regulation. This is in agreement with previous studies [86,150,151] which found that at physiologically relevant concentrations of metabolites, the majority of enzyme inhibitor sites are occupied but far from being saturated, in contrast to substrate binding sites which are at or near saturation. We show that the same applies to the amino acid binding sites of their transcription regulators.

Another factor which may have an influence on the levels of intracellular concentrations of amino acids is their biosynthetic cost on the same substrate, in this case, glucose (**Fig 16C**). Remarkably, the abundance of free amino acids seems to weakly anticorrelate with their biosynthetic cost, similarly to what was found for amino acid concentrations in mammalian cells and serum by Zhang et al. (2018). Notably, all most abundant amino acids for which no allosteric or transcriptional feedback is known (Glu, Asp, Gly, Ala), are also among the amino acids with the lowest biosynthetic cost. Therefore, energetically cheaper amino acids might not require a tight control on their biosynthesis, which leads to higher concentrations and higher availability for proteins synthesis.

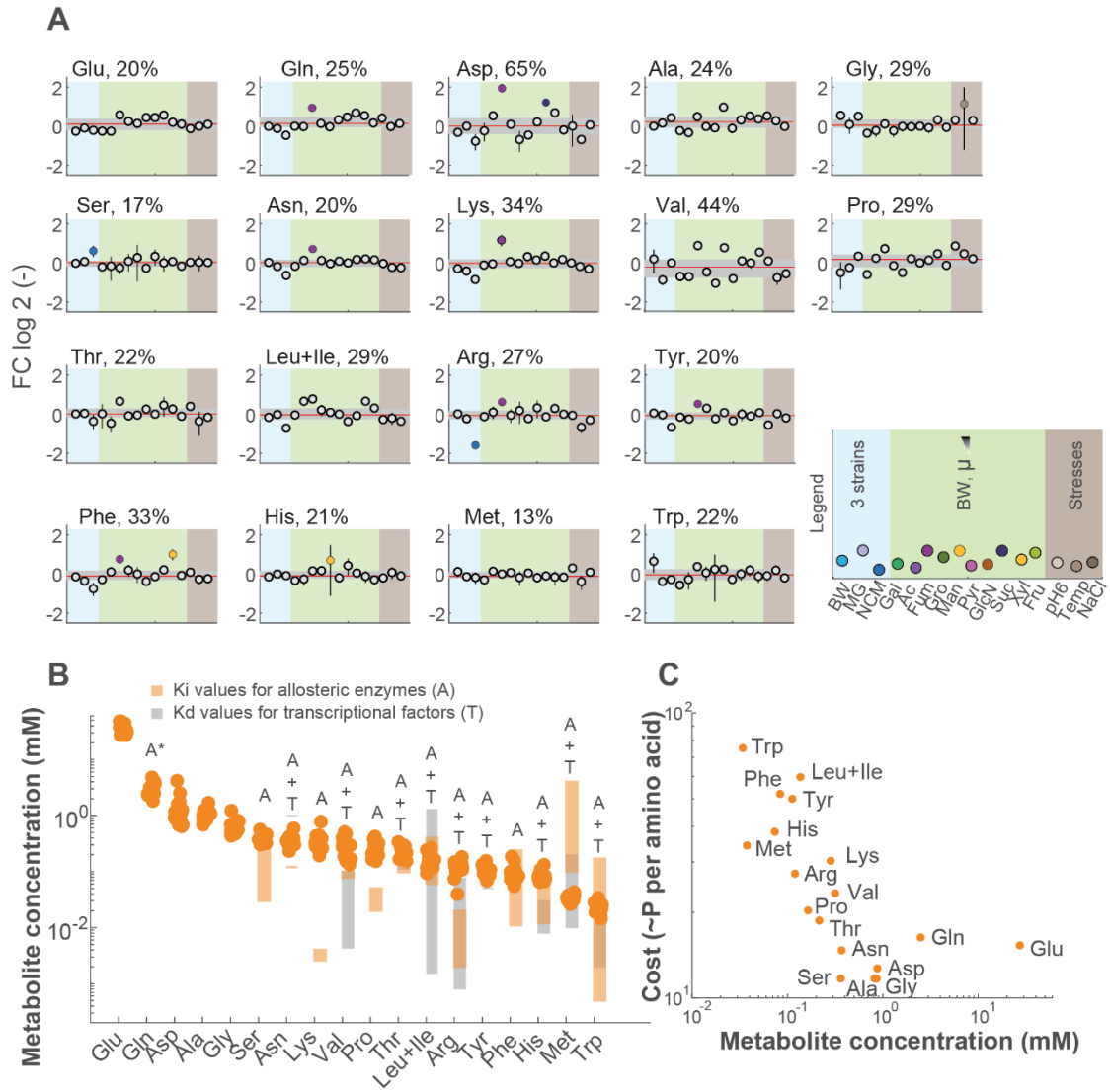


Figure 16. Amino acid pools in *E. coli*. **(A)** Variation of 19 amino acids in *E. coli* strains growing in M9 medium supplemented with 13 different carbon sources and under stress conditions. **(B)** Absolute intracellular concentrations of free amino acids plotted together with Ki values of the allosterically regulated enzymes in their biosynthetic pathway (orange), and Kd values for the transcriptional factors

involved in regulation of their biosynthesis (grey). **(C)** Absolute intracellular concentrations of free amino acids plotted against the energetic cost of their biosynthesis (calculated by Akashi and Gujobori 2002).

4.1.4 - Nucleotide levels remain stable despite environmental or genetic perturbations

Next, we examined the stability end products in the form of nucleotide mono-, di- and triphosphates, which were remarkably constant across all conditions (**Fig 17A**). In case of ATP there was a 14% variation among exponentially growing cells. Nucleotide concentrations remained stable with increase in growth rate [152,153]. Some of the most obvious outliers in nucleotide concentrations were cells grown in synthetic complex medium, which points to the difference between *de novo* synthesis and salvaging of adenine and uracil. For example, high UMP levels on synthetic rich medium may have an impact on the initial steps of *de novo* synthesis, as UMP is an allosteric inhibitor of carbamoyl phosphate synthetase [154]. Interestingly, UTP levels were constant and there was no evidence that transcriptional attenuation of the pyrBI operon was responsible for the regulation of protein abundance in *de novo* synthesis pathway. This means that changes in UMP levels from 0.03 mM to 0.07 mM, might decide between *de novo* synthesis and salvaging of purine nucleotides in *E. coli*. Such ultrasensitive regulation has been show for allosteric enzymes in glycolysis [86].

Despite end product stability, nucleotide precursor levels differed in three *E. coli* strains BW25113, MG1655 and NCM3722 grown on glucose minimal medium. While BW25113 and MG1655 had almost identical metabolite profiles, the NCM3722 strain showed some striking differences. In particular, intermediates in pyrimidine nucleotide biosynthesis (N-carbamoyl-L-aspartate, dihydroorotic acid and orotate) were more than 10-fold lower in NCM3722 than in the other two strains (**Fig 17B**). We assumed that the high concentration of these metabolites reflects the known frameshift mutation in MG1655-derived *E. coli* strains that causes low expression of *pyrE* encoded orotate phosphoribosyltransferase [108]. The proteome data confirms this pyrimidine bottleneck at *pyrE*, because BW25113 and MG1655 have less than 50 copies of the enzyme, compared to 500 copies in NCM3722 (**Fig 17B**, upper panel, data from Schmidt et al. 2016). While expression of *pyrE* was reduced in MG1655 and BW25113, other enzymes were upregulated, indicating a compensatory response to the *pyrE* bottleneck.

Compared to NCM3722, the enzymes catalyzing the first committed reaction (PyrI and B) were particularly upregulated. We assume that the chronic overproduction of pyrimidine intermediates in MG1655 and BW25113 is a combined effect from the bottleneck in the middle of the pathway (*pyrE*) and high enzyme levels at the entry point (*pyrBI*) (**Fig 17B**, upper panel). This hypothesis is supported by metabolites and protein levels in synthetic rich medium that contains uracil. In this condition, pyrimidines are synthesized via uracil salvaging, and *de novo* synthesis is repressed. Consequently, N-carbamoyl-L-aspartate, dihydroorotic acid and orotate in BW25113 on synthetic rich medium decreased to comparable levels as in the NCM3722 strain on glucose. We also noticed that the bottleneck in pyrimidine nucleotide biosynthesis of MG1655 and BW25113 spreads into biosynthesis of purines as well. For example, levels of purine intermediates xanthosine-5P, IMP, and adenylosuccinic acid were 3-4 times higher in NCM3722 than in the other two strains on glucose (**Fig 17B**, lower panel). These differences may reflect the close coupling of the purine and pyrimidine nucleotide biosynthetic pathways. Such coupling can occur simply through the common precursor L-aspartate, or by crosstalk at the layer of transcriptional and allosteric regulation. In conclusion, nucleotide biosynthetic pathways are capable of maintaining stable concentrations of their final products against both environmental or genetic perturbations.

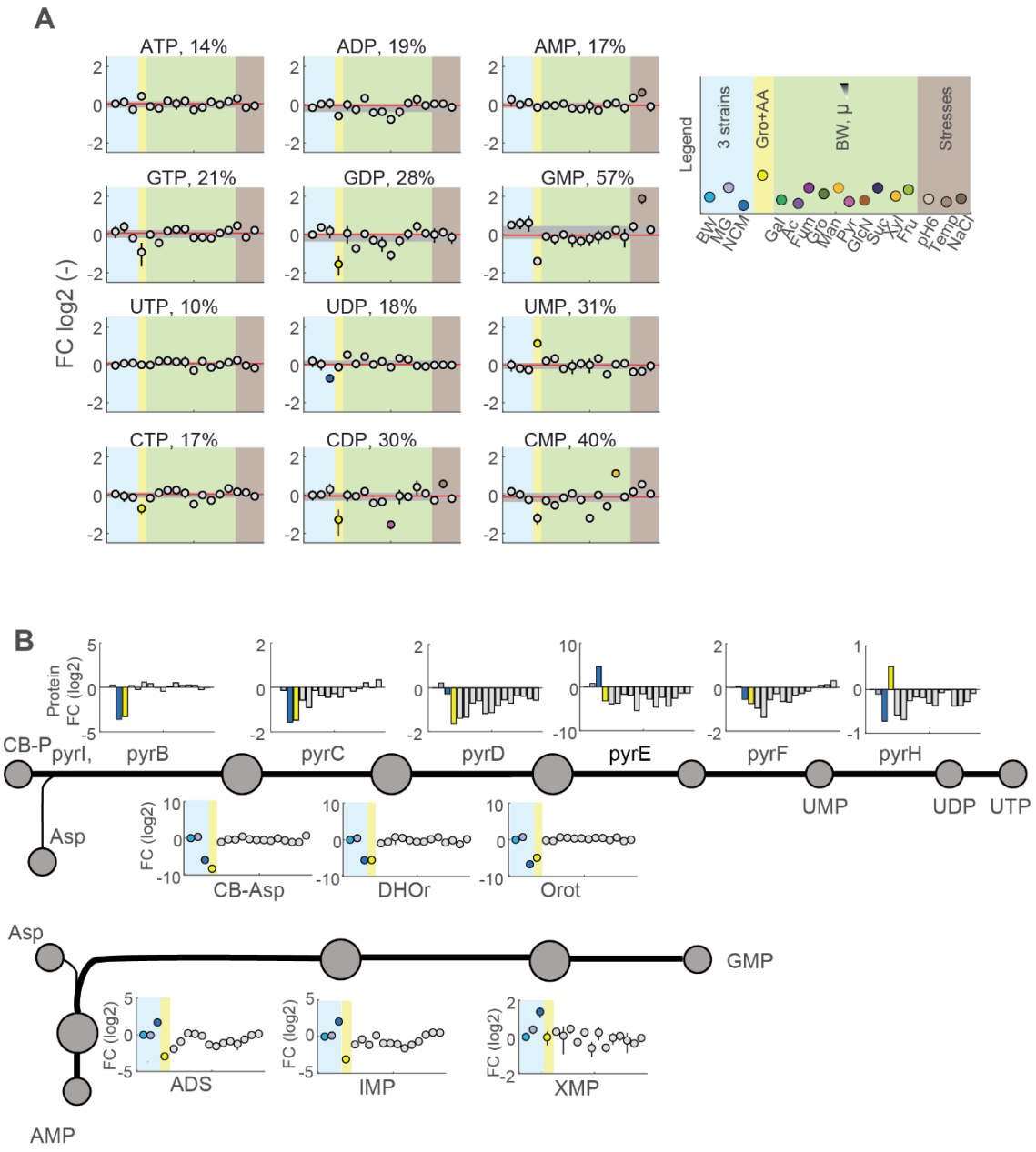


Figure 17. Nucleotide levels in *E. coli*. (**A**) Intracellular concentrations of nucletide mono-, di- and triphosphates remain stable across all conditions. Log2 was applied to median normalized values. Red line represents median of the 16 values, and grey box values that fall between 25th and 75th percentile. Outliers are depicted in color. Numbers associated to name of each metabolite represent RSD across presented conditions. (**B**) Purine and pyrimidine biosynthetic pathways. In the upper panel, protein data from Schmidt et al. 2016 are plotted as bars, and in the lower panel, relative intracellular concentrations are shown. BW25113 and MG1655 have a bottleneck caused by low levels of PyrE, which in turn results in upregulation of protein levels upstream and accumulation of upstream metabolites.

4.2 - Discussion

This work investigated the intracellular metabolome of *E. coli* BW25113 growing on different substrates, under stress conditions, and in different stages of stationary phase. Additionally, the dataset includes the intracellular metabolome of three *E. coli* reference strains (BW25113, NCM3722 and MG1655) during exponential growth on glucose minimal medium.

The type and extent of metabolome variability under steady-state growth conditions in substrate excess is revealed in this work. The vast majority of measured metabolites' levels were not affected by a growth-rate dependency, in contrast to the proteome under matching conditions. A similar independence from the growth rate was observed for amino acids in yeast mutant libraries which grew at various rates [155]. While the enzymatic machinery is adjusted in bulk during different growth conditions, metabolite levels do not change linearly with growth. The non-linear relationship between metabolite levels and growth/protein levels suggests that allosteric enzymes might be prevalent in the *E. coli* metabolism, as recently proposed [140]. Enzymes are usually thought to follow Michealis-Menten kinetics, implying that reaction rates are linearly related to metabolite and enzymes concentrations, or only to the latter in case of high metabolite concentrations. Kinetics of allosteric enzymes follow instead sigmoidal relationships, in which small changes in concentrations of metabolites can strongly impact reaction rates. Previous theoretical studies have hypothesized that metabolic networks are insensitive to changes of enzyme concentrations [156,157], which underlines the importance of metabolites in control of enzymes activity, and thus in the regulation of the metabolic fluxes/networks. The overall degree of variation between the metabolome and the proteome was similar, with metabolites having a slightly higher variation than the proteome. A similar observation was made in a multi-omics study of yeast [158]. In particular, key building blocks as amino acids and nucleotides showed a particularly low variation.

Most of the variation in amino acids levels could be explained by the proximity of the entry point of the carbon source into the metabolic network. By measuring amino acids

concentrations systematically, we could also suggest that i) cellular bioenergetics (i.e., the cost of biosynthesis of a particular amino acid) may play a role in adjusting the levels of amino acids and determining the extent of regulation of their biosynthesis, and ii) highlight the importance of allosteric and "local" transcriptional control for amino acid biosynthesis.

Bioenergetically cheaper amino acids, for which no allosteric or transcriptional control was shown, are the most abundant in the cell. It has been shown that the proteome of *E. coli* and *B. subtilis* has a bias towards less energetically costly amino acids [159]. This appears to be a widely spread concept, as rapidly proliferating cancer cells optimize their proteome expression patterns to utilize amino acids more economically [160]. In contrast, energetically expensive amino acids are present in low concentrations, and have tightly controlled *de novo* biosynthesis. Ki and Kd values of responsive enzymes and transcription factors that control amino acid biosynthesis show good match to levels of amino acids in the cells, pointing at the importance of combinatorial control of cellular metabolism [135].

Similarly, levels of nucleotides, and in particular nucleotide triphosphates, were also remarkably conserved against environmental perturbations. We could show that despite perturbations of pathway intermediates due to mutations, end products were kept constant. In the case of the pyrimidine pathway, homeostasis could also be enforced by fluctuations in the levels of UMP, which strongly propagated at the level of gene expression. Interestingly, metabolites in the *de novo* biosynthesis of purine nucleotides (CAIR, IMP) showed the highest correlation with the growth rate. The purine *de novo* biosynthesis pathway plays a key role, as it is involved in the biosynthesis of adenylate nucleotides. ATP demand is coupled with glycolytic flux [161], and this has been recently confirmed by a study in which antibiotic-induced adenine starving increased purine biosynthesis and metabolic activity [162]. While the adenylate charge ratio is kept constant among different growth rates and is a good indicator of metabolic activity [163], the absolute levels of adenylate nucleotides vary independently from the growth rate. This raises the question on whether intermediates upstream of AMP might be involved in coupling purine biosynthesis with growth.

In conclusion, our results shows the non-linear relationship between metabolite concentrations and growth in exponentially growing cells, which is not valid for enzymes. As the stoichiometry of metabolic enzymes of same pathways has been shown to be relatively conserved across evolution [164], metabolite concentrations might retain information on how reactions rates are adjusted at the molecular level. Precise measurements of larger portions of the metabolome coupled with novel methods to integrate multi-omics data, as machine learning [165], might help in the quest to unravel the large unknown network of metabolite-enzyme interactions.

5 - Conclusion and Outlook

In this thesis, we focused on studying basic principles of metabolic robustness in the bacterial model organism *Escherichia coli*. We employed different methods to perturb metabolism and analyzed the response at the metabolite and protein level.

First, we set to study how cells would respond to artificial localized perturbations of single enzymes. In **Chapter 2** we explored whether CRISPRi could be a suitable method to cause specific perturbations of metabolism. Proteomics data supported our hypothesis, showing that relatively low levels of a heterologously expressed dCas9 do not affect cell physiology and can enforce specific and dynamic reductions in abundance of a targeted enzyme. In **Chapter 3** we applied this methodology at different scales, in different experimental setups and with different readouts. We perturbed the whole metabolic network of *E. coli*, observing that the bacterium is remarkably robust against perturbations of single enzymes. Cells with progressively reduced enzyme levels could grow optimally for hours, before encountering a growth defect (response time). This result suggests that most enzymes are expressed at higher levels than strictly necessary for optimal growth (enzyme *overabundance*). This robustness principle was previously postulated in theoretical and computational studies [55,166] and proven for amino-acids biosynthesis pathways [135]. Enzyme *overabundance* has probably evolved to allow cells to resist environmental perturbations. However, which optimization principles guide this phenomenon and how it differs between different enzymes remain as open questions.

A thorough characterization of enzyme *overabundance* could be exploited for the construction of *minimal cells* from the perspective of gene expression. The *E. coli* proteome covers 55% of the whole dry weight of the cell [167], and metabolic enzymes cover 47.1% of the whole proteome [146]. By artificially reducing these fractions to the minimum necessary, cells could consume less nutrients to maintain and grow their biomass under controlled conditions with limited environmental perturbations. This could have important applications for industrial biotechnology, as production yields of

such synthetic strains could be increased. In this context, investigating *overabundance* of other proteomic fractions could also be of interest.

Another interesting observation derived from the CRISPRi experiments was that intracellular concentrations of substrates and allosteric effectors of the targeted enzymes were consistently strongly perturbed. By observing the composition of the metabolome over time, we hypothesize that these particular variations in metabolite concentrations, upon reduction of levels of an enzyme, might increase its activity maintaining optimal metabolic flux and cell growth. We could also observe that regulation of enzyme abundance, through known metabolite-transcriptional regulation feedbacks, was triggered only when the targeted enzyme reached a critically low level causing a growth phenotype (response time). Normally, cells should be able to increase levels of the critical enzymes due to these regulatory feedbacks. However, in our experimental setup, transcriptional upregulation could not overcome the knockdown enforced by the dCas9 complexes.

These combined results suggest a further robustness principle of metabolism, in which the metabolome provides a *buffering effect* that can counter a certain range of perturbations of enzyme levels. When this range is crossed, or in other words, when the *metabolome buffering effect* is exhausted, then specific gene expression regulatory circuits are activated to restore optimal enzyme levels. As cells are already expressing higher enzyme levels than strictly needed, specific gene regulatory mechanisms probably serve the purpose of countering strong genetic or dynamic environmental perturbations. Interestingly, due to the *metabolome buffering effect*, CRISPRi mediated enzyme knockdowns could be exploited to identify, *in vivo*, metabolites that regulate enzyme activity as allosteric effectors. If such identification of allosteric effectors could be proven at a systems level, it could be argued that metabolite concentrations are a driving evolutionary force for the emergence of allosteric regulation.

In **Chapter 4** we perturbed microbial metabolism by growing cells in different conditions and measuring the metabolome at steady-state. We observed that, when cells grow exponentially, key metabolites as amino-acids and nucleotides are kept at constant levels. Robustness in levels of these key metabolites is achieved by combined activities

of allosteric and transcriptional regulatory feedbacks [135]. We then compared our data with a matching published dataset of protein abundances in *E. coli* [146]. Unlike for proteins, we could observe that metabolite levels are not affected by growth effects. This fact suggests that absolute concentrations of metabolites within cells follow different optimality principles than proteins. A hypothesis could be that the metabolome maintains its independence from growth to enforce its *buffering effect*. Enzyme levels are tuned in bulk at different growth rates, meaning that their stoichiometry is kept constant. In fact, enzyme stoichiometry is quantitatively conserved even across evolution [164]. However, enzymes are known to possess a wide range of kinetic parameters [168] and their activity *in vivo* might generally not increase linearly with substrate concentrations, as most enzymes are thought to be heavily regulated at the allosteric level [140]. Therefore, growth independent changes in metabolite concentrations could be necessary to enable the *buffering effect* for enzymes that have a different range of kinetic parameters, but are co-expressed in bulk.

Measuring metabolite concentrations in large numbers of steady-state cultures could contain valuable information regarding enzyme kinetics. As an example, in yeast it has been shown that the precise measurement of amino acids levels in thousands of gene knockouts could lead to infer the deleted gene identity based on the metabolite signature [155]. In this context, advancements in machine learning methods for multi-omics data analysis and metabolic kinetic modeling will play a fundamental role in deconvoluting complex datasets [165,169]. Being able to obtain kinetic information from steady-state omics data would be a considerable advantage, as the construction of large datasets of dynamic data is still experimentally challenging. To this end, further improvements in precision and coverage of MS-based metabolomics and automation of experimental procedures will play an important role.

Overall, in this thesis we have applied a systems biology approach to investigate microbial metabolic robustness. We could observe two main mechanisms which render microbes robust against perturbations of metabolism: the constitutive overexpression of metabolic enzymes (*overabundance*) and specific changes in metabolite levels that prevent and precede specific regulation of enzyme abundance (*metabolome buffering effect*).

Metabolism is widely conserved among living beings. Therefore, it is possible that such principles could be conserved in more complex, fast growing living systems, in example cancer cells. Moreover, bacteria are extensively used in industrial biotechnology, and industrial strains often require the engineering of metabolic pathways. Hence, understanding metabolic robustness in this context, and how it can be possibly manipulated, could lead to more efficient industrial strains. In general, increasing our understanding of biological principles from a systems perspective has a fundamental role in enabling robust and reproducible biological engineering of synthetic organisms, known also as synthetic biology (**Figure 18**). In turn, the utilization of synthetic organisms for basic studies, can lead to insights on how biological systems function, as showcased in this thesis. Therefore, the combination of systems and synthetic biology approaches is of great importance for scientific discovery and technological advancement. In the context of industrial biotechnology, these combined methodologies will have an important impact for the transition towards a sustainable bio-based economy.

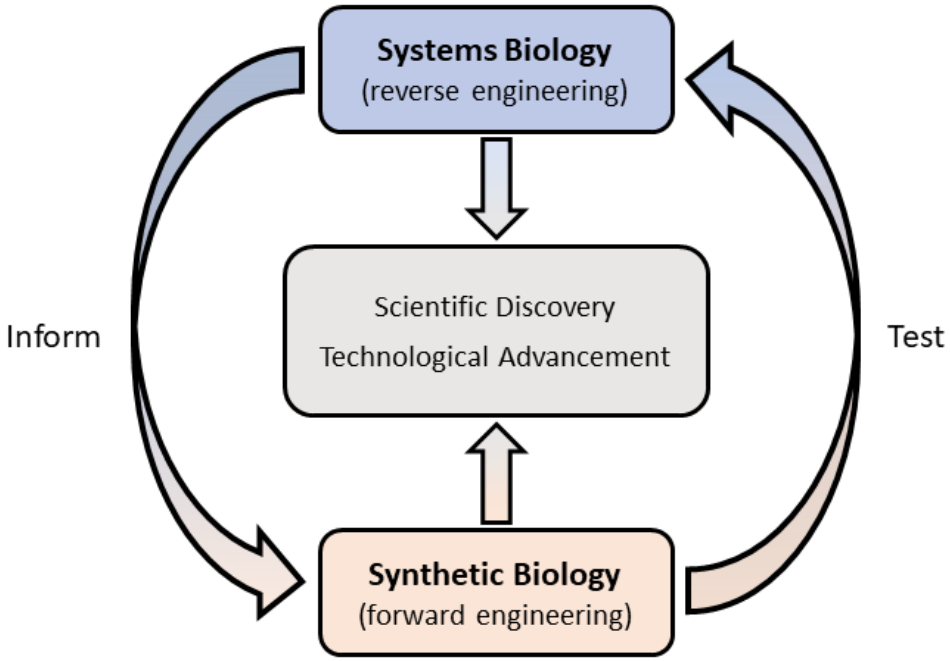


Figure 18: Scheme highlighting the mutual relationship between systems and synthetic biology. Systems biology tries to reverse engineer living beings by understanding how the ensemble of biological parts function as a system. The insight provided, is used by synthetic biology to forward engineer organisms that can be studied by systems biology approaches. This mutual relationship can lead, from both sides, to novel scientific discoveries and technological advancements.

6 - Materials and Methods

6.1 - Construction and cultivation of CRISPRi strains

The following paragraphs (6.1) refer to strains and experimental procedures showcased in Chapters 2 and 3.

6.1.1 - Construction of arrayed strains

E. coli DH5α was used for plasmid construction. *E. coli* BW25933 was transformed with pdCas9-bacteria to obtain the pCRISPRi strain. sgRNA guide sequences were designed with Matlab scripts by searching for the first NGG PAM site on the coding strand for each gene of interest. Adjacent to PAM sites, 20 nt regions were selected. Most plasmids were synthesised (Doulix srl). Alternatively, 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. All plasmids were verified by sequencing. pCRISPRi and *E. coli* YYdCas9 were then transformed with the relative pgRNA-bacteria plasmids.

For CRISPRi of YPet, the sgRNA targeted lacZ, the first gene of the operon that includes YPet [104]. The plasmid pUA66 was used to measure promoter activity [122]. The ΔgntK mutant was constructed by P1 Phage transduction of YYdCas9 using the donor strain JW3400 (ΔgntK) from the KEIO collection [170]. 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.

6.1.2 - 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 [119]. 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.

6.1.3 - 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. Additionally, pCRISPRi strains were cultivated in the presence of Chloramphenicol (35 µg mL^{-1}). To induce expression of the dCas9 protein in the YYdCas9 strain, aTc was added to a final concentration of 200 nM.

6.1.4 - General Cultivation conditions

Strains of interest were first recovered from glycerol stocks on fresh LB plates. Colonies were then inoculated into liquid LB cultures for 4-5 hours and then diluted 100x into 5mL of m9 minimal medium overnight. Precultures were then diluted into 96-well plates or flasks at different starting concentrations. For induction of dCas9 expression 200nM of aTc were added at the start of the the relative cultures. 96-well plates were then incubated in a plate reader (Biotek Synergy) for 24 hours measuring OD600. Flasks were incubated in a shaking incubator and samples for proteomics were obtained by temporarily moving the cultures in a thermostatic hood. All cultivations were carried out under shaking conditions at 37°C.

6.1.5 - 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 □L 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 at 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 490nm, emission 530nm) was measured in 10 min intervals using a plate reader (Tecan, Spark).

6.1.6 - Cultivation conditions for metabolome and proteome sampling

Single colonies of strains of interest 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 c.ca 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.

6.1.7 - 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).

6.2 - Cultivation of *E. coli* under different environmental conditions

The following paragraphs (6.2) refer to strains and experimental procedures showcased in Chapter 4.

6.2.1 - Strains and growth conditions

Escherichia coli strains BW25113, MG1655 and NCM3722 were used (Baba et al. 2006, Bachmann et al. 1996, Brown and Jun 2015). Experimental conditions were matching those from Schmidt et al. (2016). M9 minimal medium was prepared in the following way: to 800 ml of water, 200 ml of 5 × base salt solution (211 mM Na_2HPO_4, 110 mM KH_2PO_4, 42.8 mM NaCl, 56.7 mM $(NH_4)_2SO_4$, in H_2O, autoclaved), 10 ml of trace elements (0.63 mM $ZnSO_4$, 0.7 mM $CuCl_2$, 0.71 mM $MnSO_4$, 0.76 mM $CoCl_2$, in H_2O, autoclaved), 1 ml 0.1 M $CaCl_2$ solution, 1 ml 1 M $MgSO_4$ solution, 2 ml of 500 × thiamine solution (1.4 mM) and 0.6 ml 0.1 M $FeCl_3$ solution (all in in H_2O) were added. The resulting solution was filled up to 1 l with water and filter sterilized. 40 x stock solutions of individual carbon sources were prepared in H_2O and pH was set to 7 by 1 M NaOH or HCl. Final concentrations of individual carbon sources were: sodium acetate, 3.5 g/L, disodium fumarate, 2.8 g/L, galactose 2.3 g/L, glucosamine 2.1 g/L, glycerol 2.2 g/L, sodium pyruvate 3.3 g/L, disodium succinate hexahydrate, 5.7 g/L, glucose, fructose, mannose and xylose 5 g/L. For salt stress experiment, NaCl was added to M9 to a concentration of 50 mM; for pH stress, M9 pH was set to 6.0 by titrating with HCl. Amino acids were dissolved in M9 to following concentrations: alanine 1.0 mg/L (0.0 mM), adenine 10.2 mg/L (0.1 mM), arginine 51.1 mg/L (0.3 mM), asparagine 1.6 mg/L (0.01 mM), aspartic acid 81.8 mg/L (0.6 mM), cysteine 1.2 mg/L (0.01 mM), glutamate 15.2 mg/L (0.1 mM), glutamine 13.9 mg/L (0.1 mM), glycine 0.4 mg/L (0.01 mM), histidine 20.5 mg/L (0.1 mM), isoleucine 51.1 mg/L (0.4 mM), leucine 102.3 mg/L (0.8 mM), lysine 51.1 mg/L (0.4 mM), methionine 20.5 mg/L (0.14 mM), phenylalanine 51.1 mg/L (0.3 mM), proline 5.2 mg/L (0.05 mM), serine 9.2 mg/L (0.1 mM), threonine

102.3 mg/L (0.9 mM), tryptophan 51.1 mg/L (0.3 mM), tyrosine 51.1 mg/L (0.3 mM), valine 143.2 mg/L (1.2 mM) and uracil 20.5 mg/L (0.2 mM). To this synthetic complex medium, glycerol was added (final concentration 2.2 g/L). All chemicals used were from Sigma-Aldrich.

6.2.2 - Cultivation and growth rates

Cryostocks kept at -80°C were streaked out on LB plates (Luria Miller, Carl Roth) and incubated overnight at 37°C. Single colony was picked and inoculated into M9 preculture (7.5 ml in 100-ml flask), which was grown 6 - 10h, to be subsequently diluted in another equal preculture. From the second preculture growing exponentially, main culture was inoculated at an approximate OD 0.05 in 35 ml in 500-ml non-baffled wide-neck shake flask, covered by a 38-mm silicone sponge closure (BellCo glass). Cultivations were performed at 37 °C, 200 rpm and 5-cm shaking diameter (Infors HT Minitron), except for the heat-stressed cells which were grown at 42 °C. Growth was monitored by measuring the OD_{600}. Specific growth rates (μ) were calculated through linear regression of the plots of $\ln(OD_{600})$ *versus* time during the exponential growth phase. Further information on the growth rates and comparison with the results from the proteomics study can be found in Supplementary Table S1. Stationary cells were cultivated 24 and 48 h after reaching stationary phase.

6.2.3 - Sampling and sample preparation for metabolomics

All growing cultures were sampled at an OD_{600} of approximately 0.5. Further information on the OD values and sampling volumes can be found in Supplementary Table S1. For the intracellular metabolites, 2 ml of whole cell broth was vacuum-filtered through filter membrane (Durapore 0.45 µM Whatmann). Filter containing cells was quickly immersed into 1 ml of ACN:methanol:H_2O (40:40:20) at -20°C. Extraction was performed overnight at -20°C. Cell extracts were then centrifuged at –9°C, 13 000 rpm and 20 min

(HeraerusTM Pico 17TM ThermoScientific). An aliquot of the supernatant was immediately mixed with 13-C internal standard in equal proportion.

6.2.4 - 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 3). 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. 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 = 0h) to calculate fold-changes.

6.3 - 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 [83].

Additionally, for measurments of strains cultivated under different environmental conditions, we applied 2 separate LC methods for nucleotide and keto-acids determination. Nucleotides in cell extracts were measured by dedicated basic conditions method, in which LC parameters were kept. Keto acids were derivatized prior to measurement, using 50 µM freshly prepared phenylhydrazine (Sigma-Aldrich) (Zimmermann et al. 2014). Reaction was left to proceed for 1 h at -20°C in ACN:methanol:H2O solvent (40:40:20). Since the concentration of oxaloacetate and glyoxylate was too low for MS detection, derivatized mixture was concentrated approximately 10 times by drying in speedvac (RVC 2-25 Cdplus, Martin Christ) and separated using the acidic method with standardized LC run settings.

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.

Quantification of selected primary metabolites was performed by using $^{12}C/^{13}C$ ratio described by Bennett et al. (2008). The ^{13}C internal standard was quantified by using metabolite standards of known concentration at 10 different dilutions (**Figure S6**). Data analysis was performed using Matlab 2016b and 2017b. Obtained MS data was converted into a text file using MSConvert (Chambers et al. 2012). Further data analysis was performed by in-house Matlab-based software (Guder et al. 2017).

For normalization of intracellular metabolites, we followed an assumption that OD-specific cell volume is constant [171]. It is also worth to mention that in the case of metabolite being present in the medium, like in the case of amino acids in synthetic complex medium, the intracellular concentrations of these were not taken into account, because they are probably influenced by the extracellular amino acids.

6.4 - 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.

6.5 - Data analysis

6.5.1 - 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 datasets, proteins with an average variability between triplicates higher than 20% were removed. 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.

6.5.2 - Constraint-based modelling

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/ [172] and FBA simulations were applied using COBRApy [173] with parameters as described in Monk et al., 2017.

6.5.3 - Singular value decomposition

Singular value decomposition [148] was applied as previously described [145], transforming the data into *log* space and using the *svd* function in Matlab.

Supplementary Materials

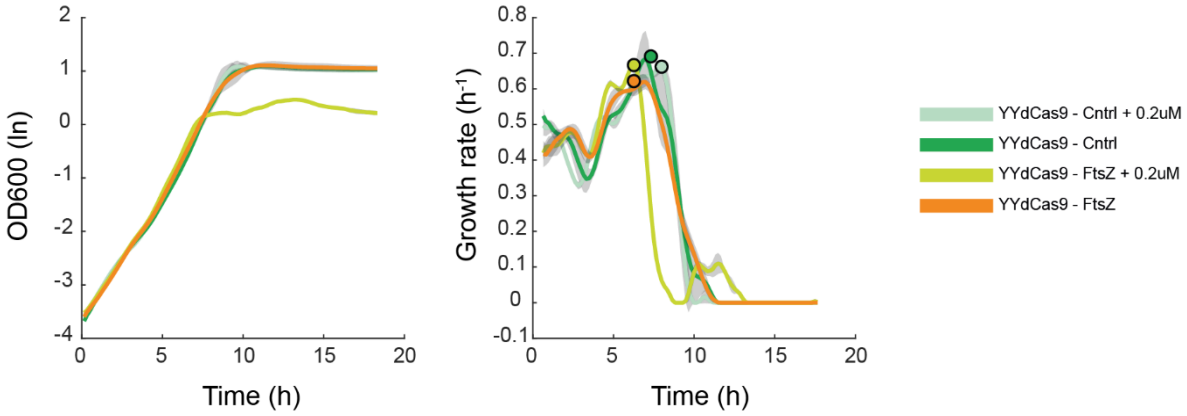


Figure S1: Behaviour of a YYdCas9 strain with interference of *ftsZ,* in comparison to the YYdCas9 control strain. OD600 in log scale and growth rate over time, lines represent average of replicates (n=3), grey shadowed areas represent standard deviation. Only when inducing interference, the ftsZ strain displays a growth phenotype, while in the uninduced state the strain grows comparably to the control.

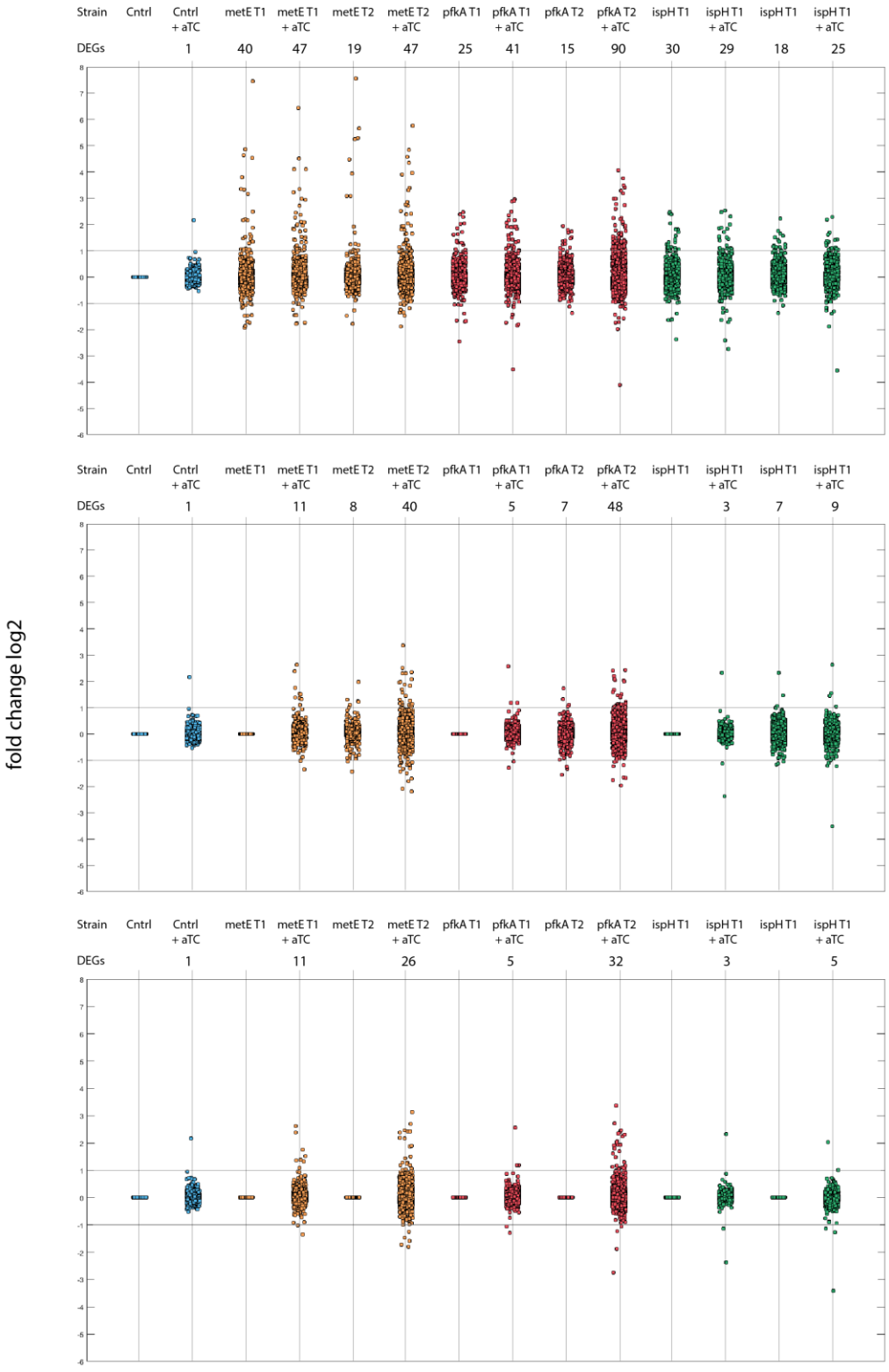


Figure S2: Differences in proteome data when normalizing for fold-change against different conditions. The top plot showcases proteomes of all conditions normalized to the un-induced control strain. The middle plot showcases proteomes normalized to the relative un-induced strain at T1. The bottom plot represents proteomes normalized against the relative un-induced strain at the relative time-point. The tables on top of the plots represent the number of differentially expressed genes (FC>2, grey horizontal lines).

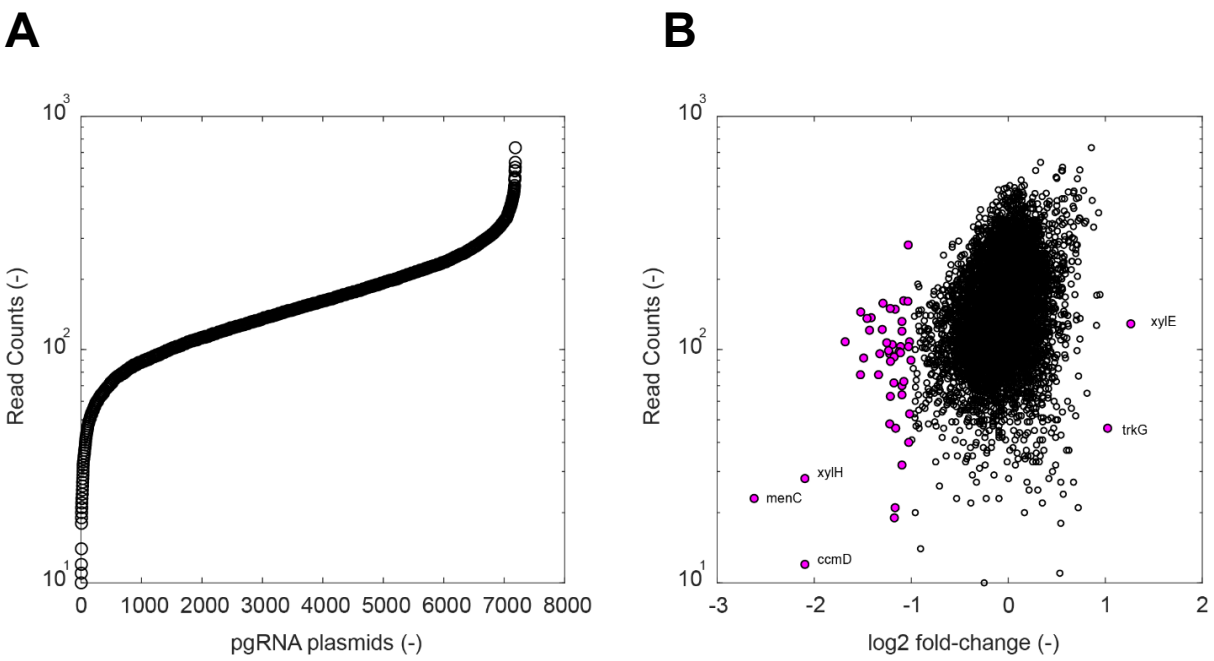


Figure S3. Details of the raw deep-sequencing data.

(A) Distribution 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 abundances 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. 47 sgRNAs showed a fold-change higher than 2 (magenta).

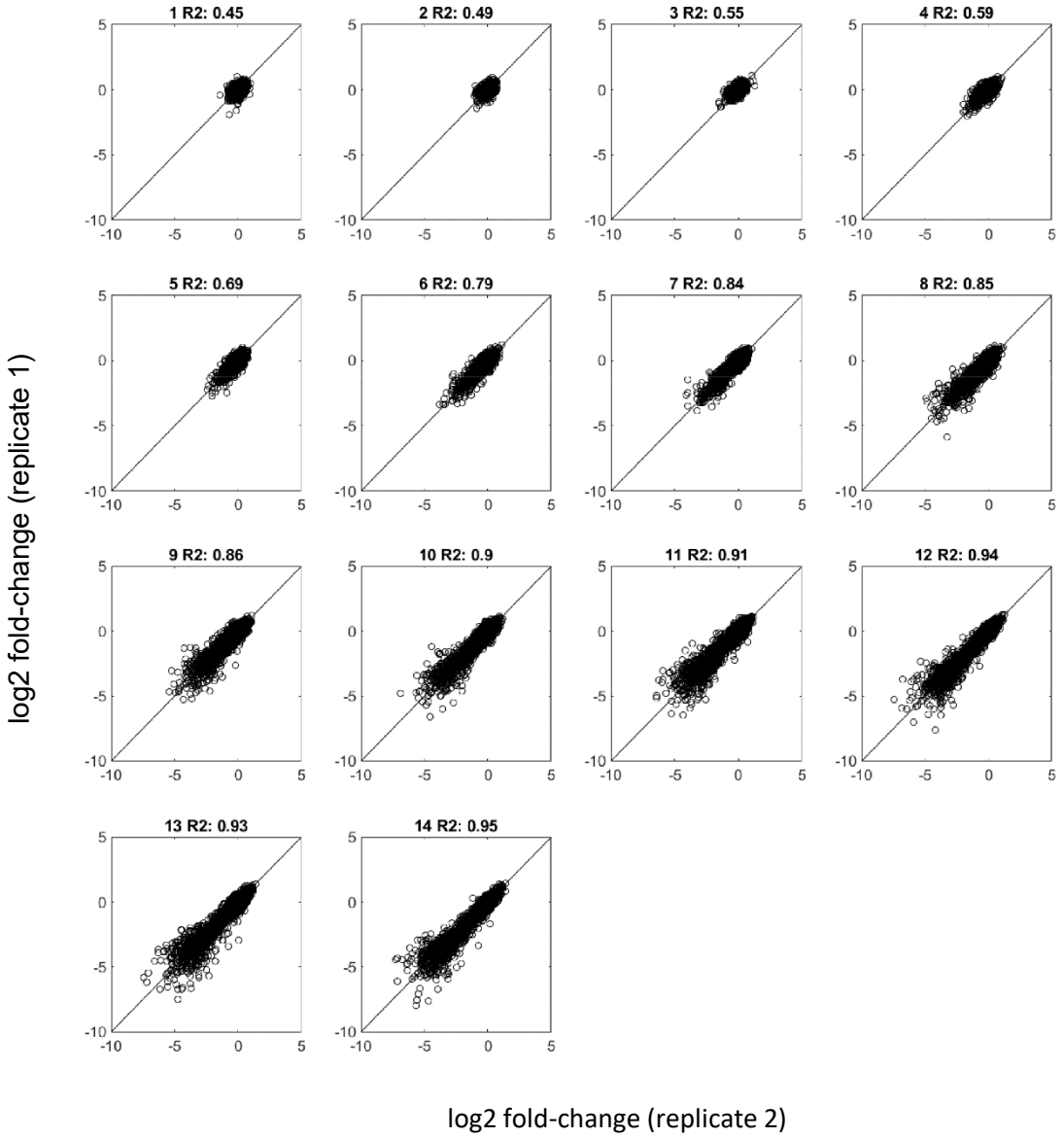


Figure S4. Fold-changes of sgRNA abundances in the two competition experiments. Each plot shows data for one of the 14 time points. R2 is the correlation coefficient.

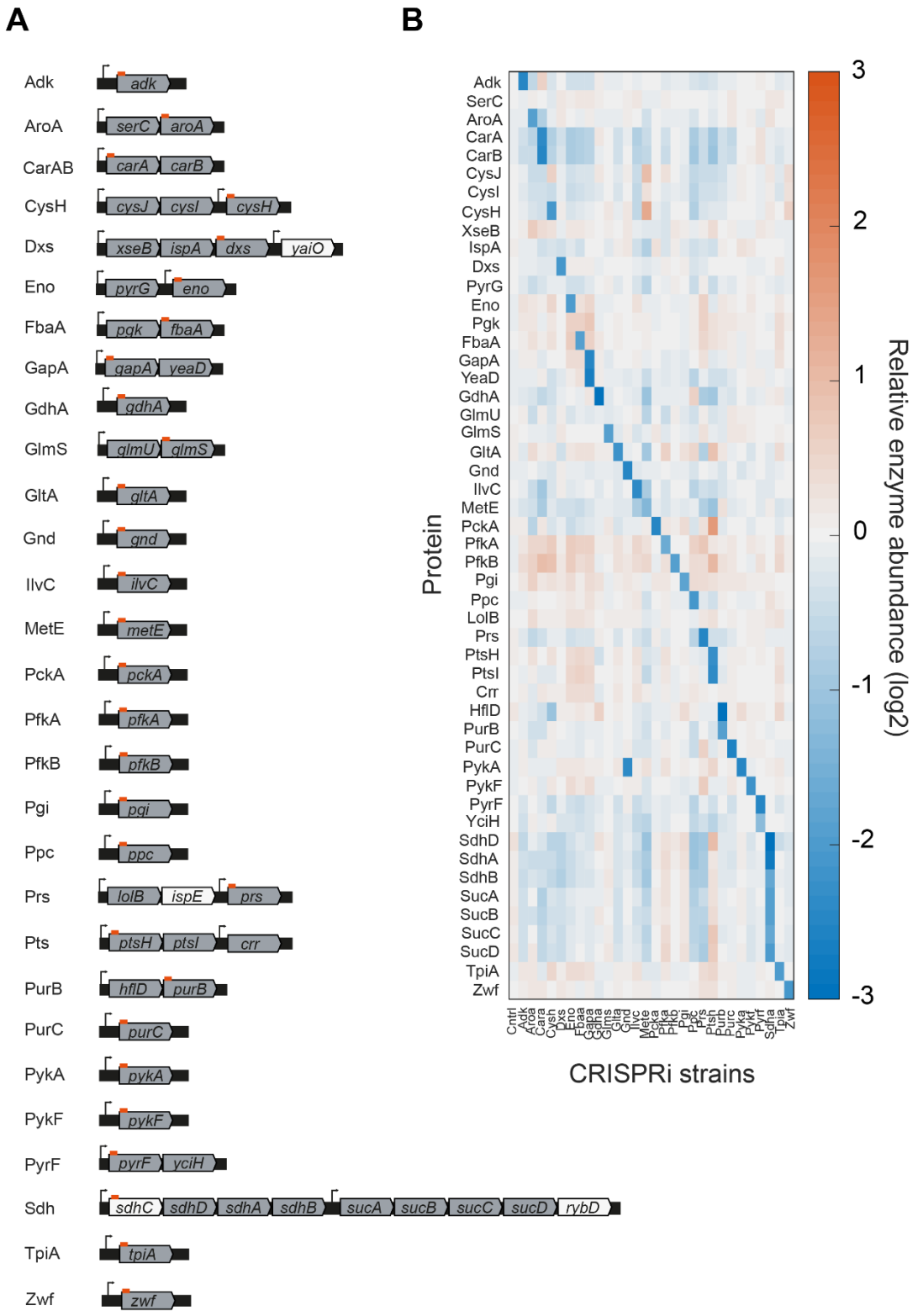


Figure S5. Related to Figure 10.

(A) Operon structure of the 29 target-genes, based on the Ecocyc database (ecocyc.org). The approximate locus that is targeted by sgRNAs is indicated in red (the exact targeted sequences are available in Supp Table 9). 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.

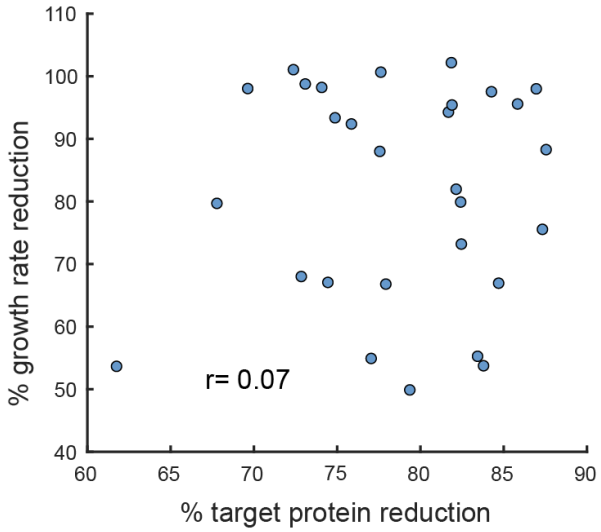


Figure S6. Related to Figure 10.

Reduction in growth rates of 29 CRISPRi strains is shown on the y-axis. The reduction of the target protein is shown on the x-axes.

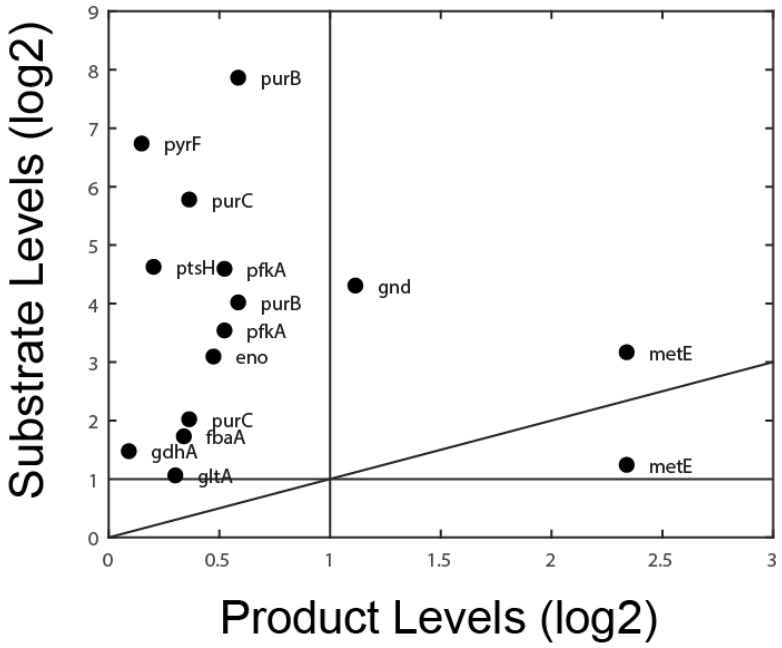


Figure S7. Related to Figure 11.

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).

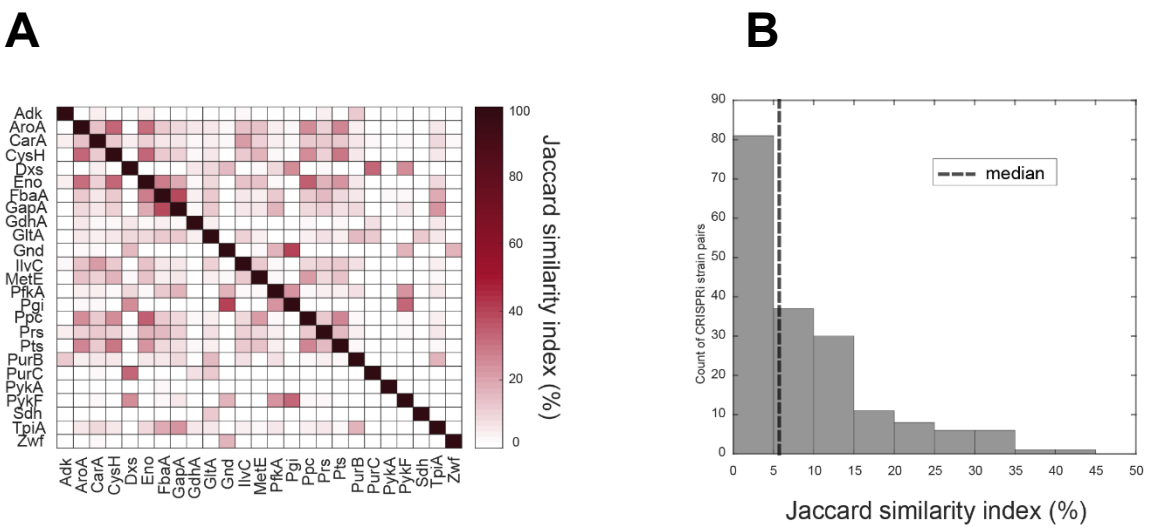
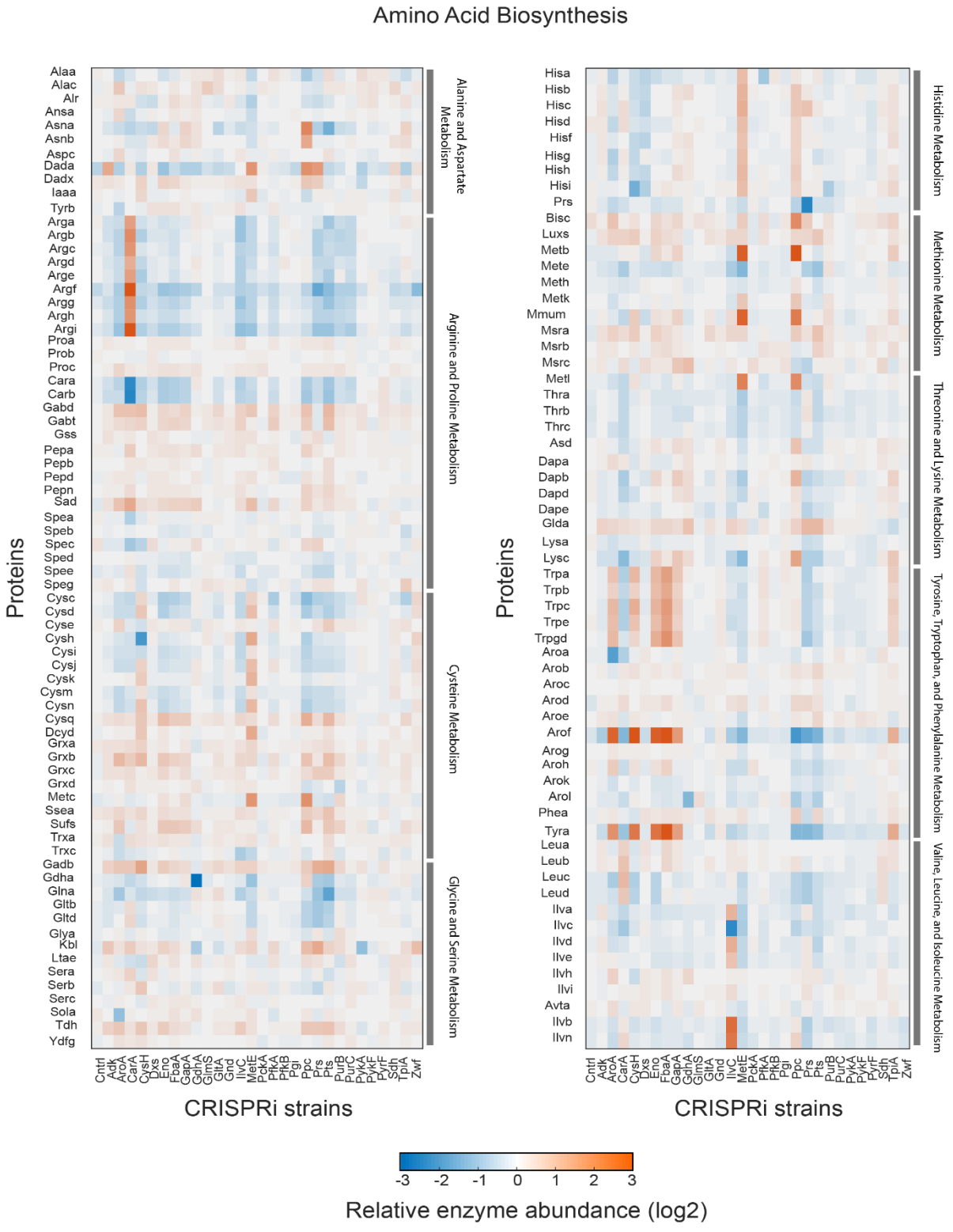


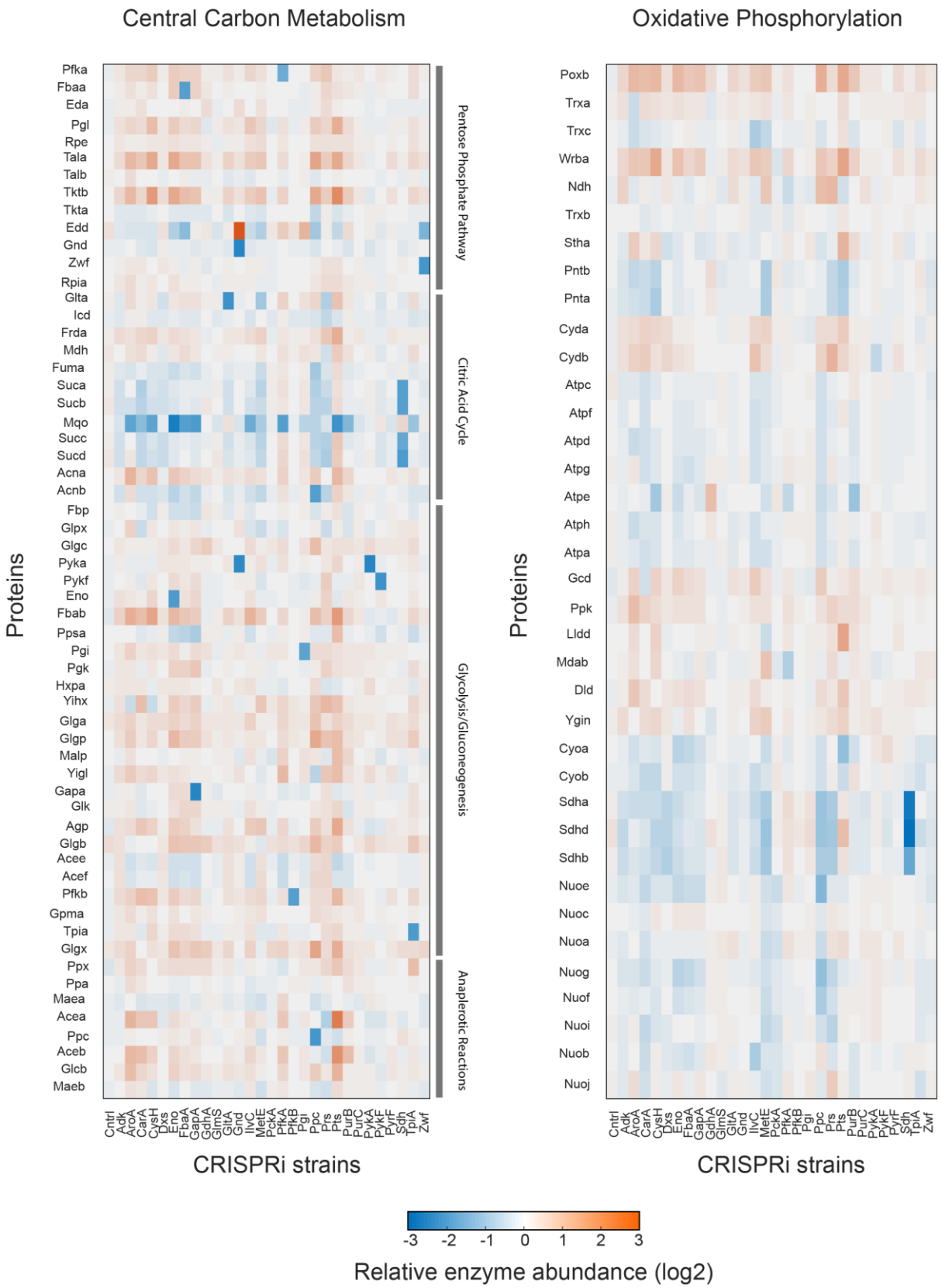
Figure S8. Similarity between the differentially expressed proteins of the 30 measured proteomes.

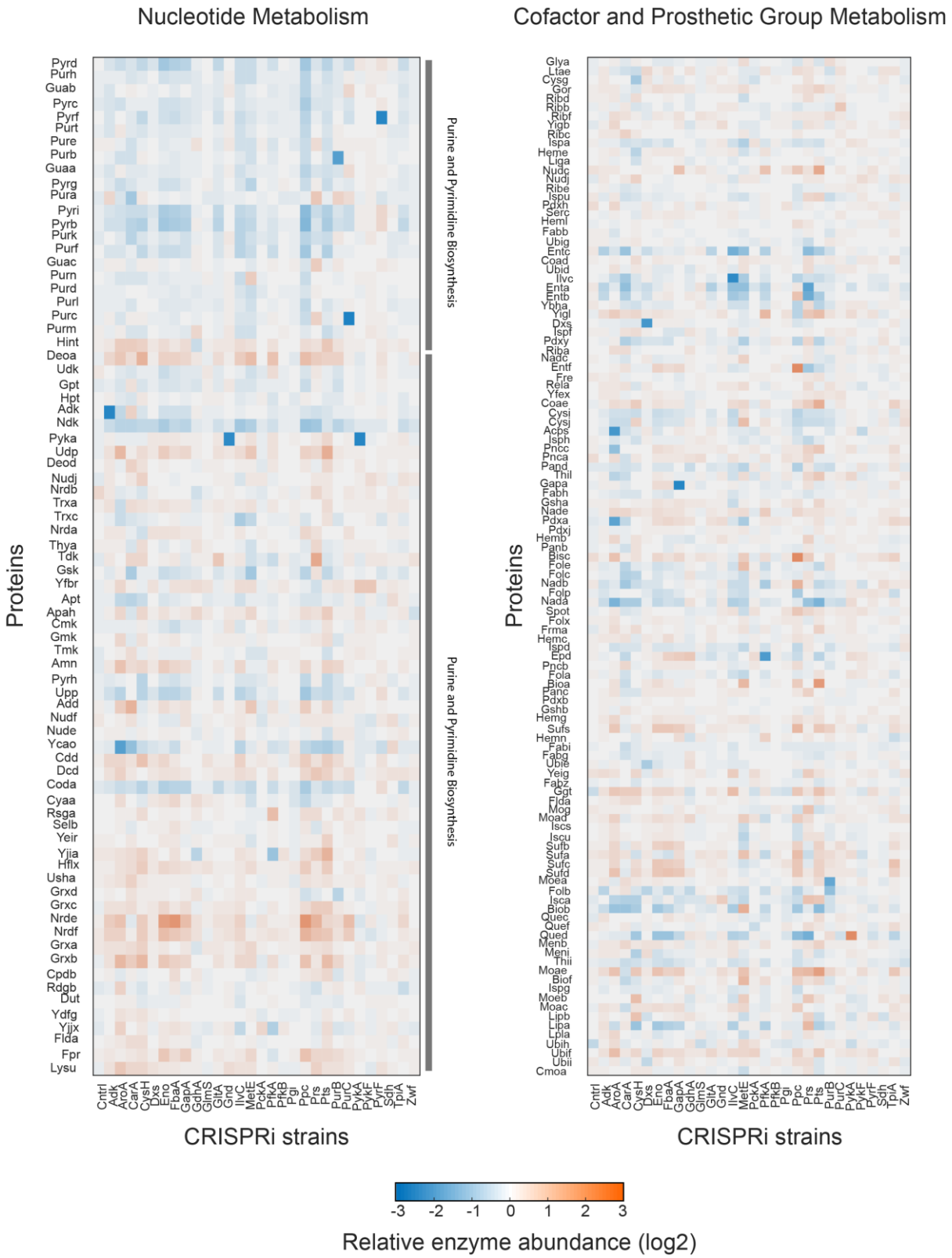
(A) Similarity matrix of differentially expressed proteins (FC=2, p-value<0.05) of the 30 measured proteomes. Similarity is defined as the Jaccard similarity index. Highest similarity was calculated for the pairs (in order) Gnd-Pgi (40%), FbaA-GapA (38.89%), Eno-Ppc (34.27%), AroA-CysH (33.33%), CysH-Eno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

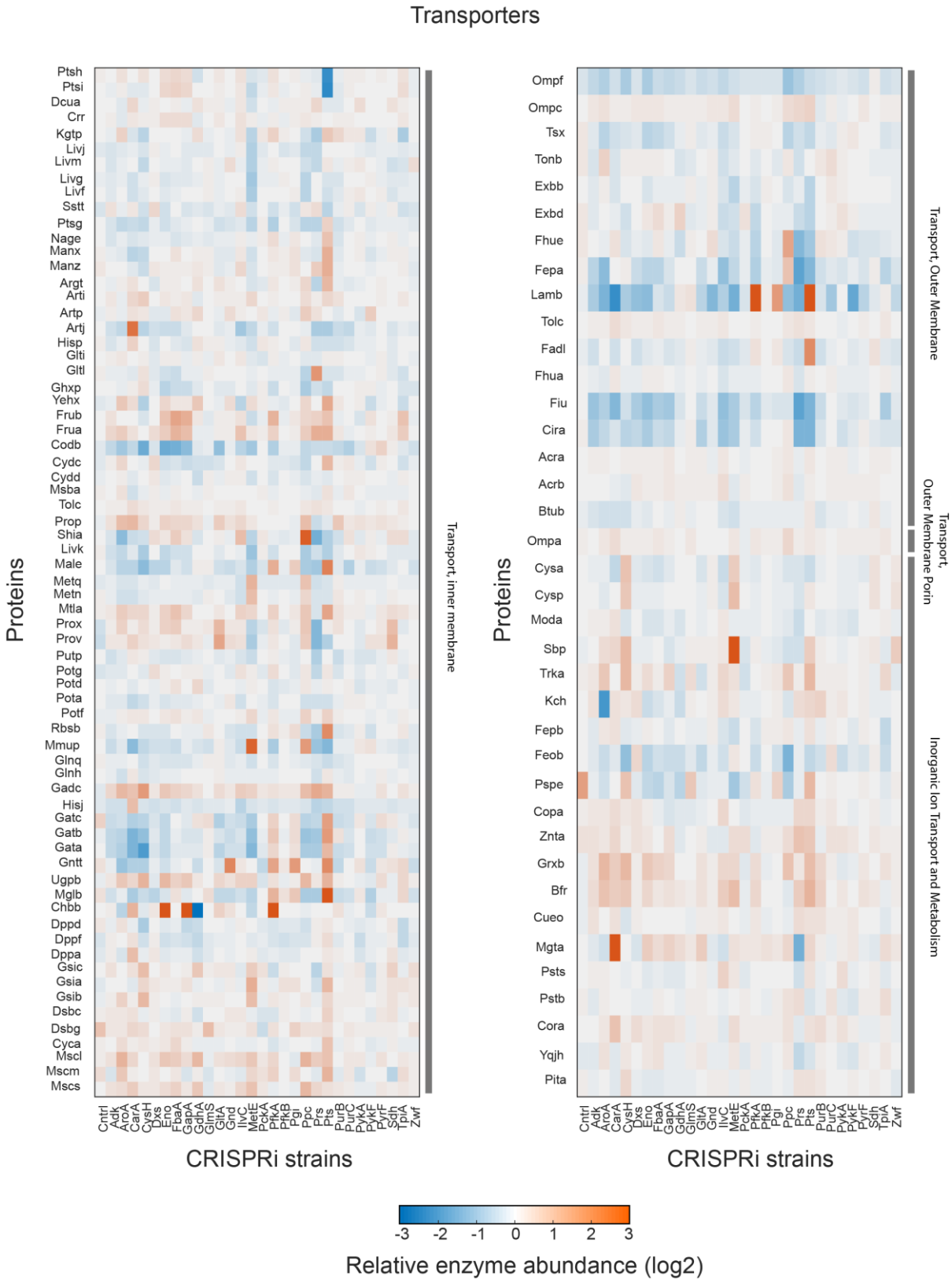
Figure S9. Related to Figure 12.

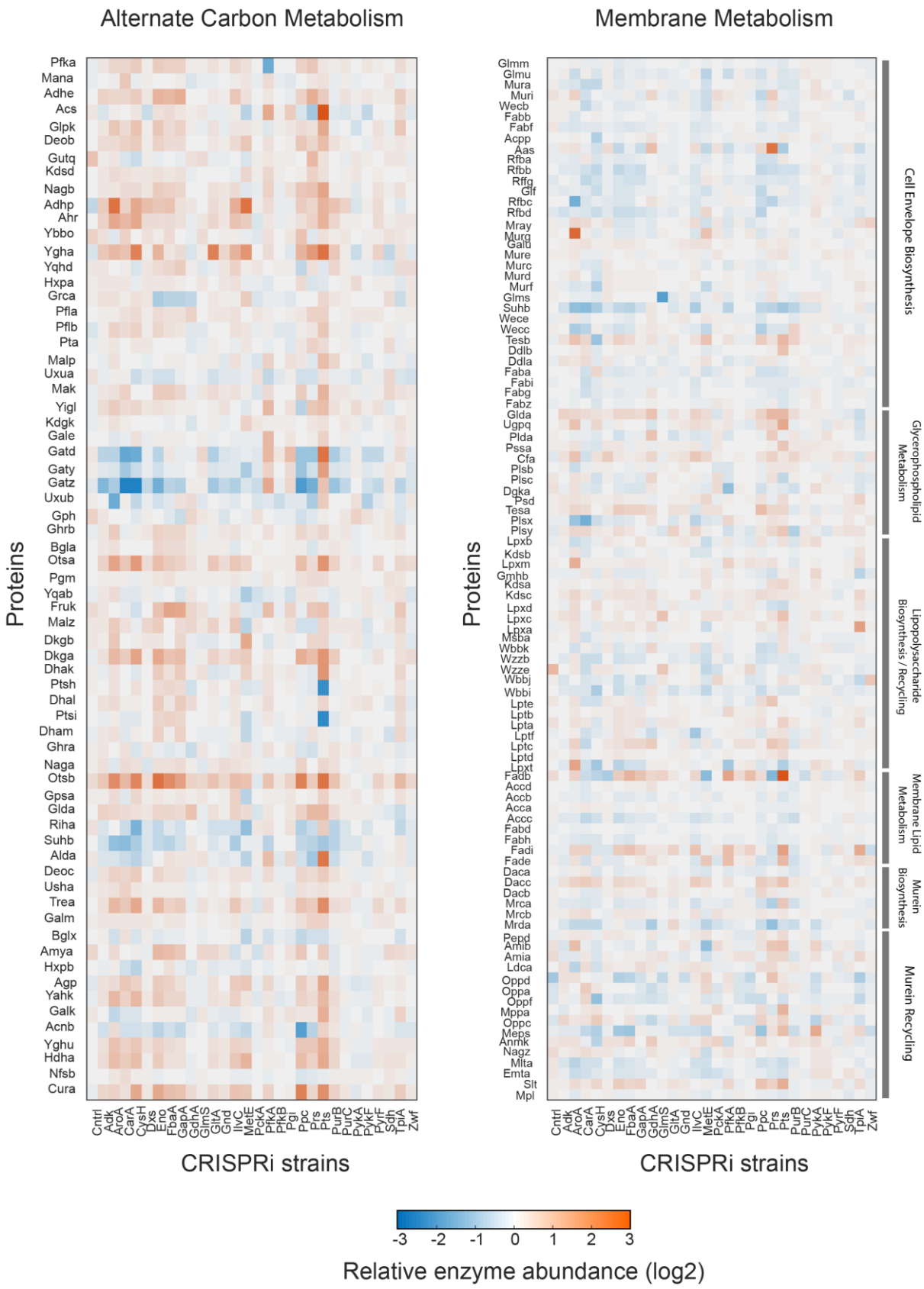
Heatmaps show log2 fold-changes of proteins between induced and un-induced cultures. Data was calculated using the means of n = 3 samples per strain. Data is organized based on metabolic subsystems in *i*ML1515.




Central Carbon Metabolism
Oxidative Phosphorylation
Proteins
Pfka
Fbaa
Eda
Pgl
Rpe
Tala
Talb
Tktb
Tkta
Edd
Gnd
Zwf
Rpia
Glta
Icd
Frda
Mdh
Fuma
Suca
Sucb
Mqo
Succ
Sucd
Acna
Acnb
Fbp
Glpx
Glgc
Pyka
Pykf
Eno
Fbab
Ppsa
Pgi
Pgk
Hxpa
Yihx
Glga
Glgp
Malp
Yigl
Gapa
Glk
Agp
Glgb
Acee
Acef
Pfkb
Gpma
Tpia
Glgx
Ppx
Ppa
Maea
Acea
Ppc
Aceb
Glcb
Maeb
Pentose Phosphate Pathway
Citric Acid Cycle
Glycolysis/Gluconeogenesis
Anaplerotic Reactions
Cntrl Adk AroA CarA CysH Dxs Eno FbaA GapA GdhA GlmS GltA Gnd IlvC MetE PckA PfkA PfkB Pgi Ppc Prs Pts PurB PurC PykA PykF PyrF Sdh TpiA Zwf
CRISPRi strains
Poxb
Trxa
Trxc
Wrba
Ndh
Trxb
Stha
Pntb
Pnta
Cyda
Cydb
Atpc
Atpf
Atpd
Atpg
Atpe
Atph
Atpa
Gcd
Ppk
Lldd
Mdab
Dld
Ygin
Cyoa
Cyob
Sdha
Sdhd
Sdhb
Nuoe
Nuoc
Nuoa
Nuog
Nuof
Nuoi
Nuob
Nuoj
-3 -2 -1 0 1 2 3
Relative enzyme abundance (log2)


Nucleotide Metabolism
Cofactor and Prosthetic Group Metabolism
Proteins
Purine and Pyrimidine Biosynthesis
Purine and Pyrimidine Biosynthesis
Pyrd
Purh
Guab
Pyrc
Pyrf
Purt
Pure
Purb
Guaa
Pyrg
Pura
Pyri
Pyrb
Purk
Purf
Guac
Purn
Purd
Purl
Purc
Purm
Hint
Deoa
Udk
Gpt
Hpt
Adk
Ndk
Pyka
Udp
Deod
Nudj
Nrdb
Trxa
Trxc
Nrda
Thya
Tdk
Gsk
Yfbr
Apt
Apah
Cmk
Gmk
Tmk
Amn
Pyrh
Upp
Add
Nudf
Nude
Ycao
Cdd
Dcd
Coda
Cyaa
Rsga
Selb
Yeir
Yjia
Hflx
Usha
Grxd
Grxc
Nrde
Nrdf
Grxa
Grxb
Cpdb
Rdgb
Dut
Ydfg
Yjjx
Flda
Fpr
Lysu
Glya
Ltae
Cysg
Gor
Ribd
Ribb
Ribf
Ygb
Ribc
Ispa
Heme
Liga
Nudc
Nudi
Ribe
Ispu
Pdxh
Serc
Heml
Fabb
Ubig
Entc
Coad
Ubid
Ilvc
Enta
Entb
Ybha
Ygl
Dxs
Ispf
Pdxy
Riba
Nade
Entf
Fre
Rela
Yfex
Coae
Cysi
Cysj
Acps
Isph
Pncc
Pnca
Pand
Thil
Gapa
Fabh
Gsha
Nade
Pdxa
Pdxj
Hemb
Panb
Bisc
Fole
Folc
Nadb
Folp
Nada
Spot
Folx
Frma
Hemc
Ispd
Epd
Pncb
Fola
Bioa
Panc
Pdxb
Gshb
Hemg
Sufs
Hemn
Fabi
Fabg
Ubie
Yeig
Fabz
Ggt
Flda
Mog
Moad
Iscs
Iscu
Sufb
Sufa
Sufc
Sufd
Moea
Folb
Isca
Biob
Quec
Quef
Qued
Menb
Meni
Thii
Moae
Biof
Ispg
Moeb
Moac
Lipb
Lipa
Lpla
Ubih
Ubif
Ubii
Cmoa
Cntrl
Adk
AroA
CarA
CysH
Dxs
Eno
FbaA
GapA
GdhA
GlmS
GlnA
Gnd
Ilvc
MetE
PckA
PfkA
PfkB
Pgi
Ppc
Pts
PurB
PurC
PykA
PykF
PyrF
Sdh
TpiA
Zwf
CRISPRi strains
CRISPRi strains
-3
-2
-1
0
1
2
3
Relative enzyme abundance (log2)


Transporters
Proteins
Ptsh
Ptsi
Dcua
Crr
Kgtp
Livj
Livm
Livg
Livf
Sstt
Ptsg
Nage
Manx
Manz
Argt
Arti
Artp
Artj
Hisp
Glti
Gltl
Ghxp
Yehx
Frub
Frua
Codb
Cydc
Cydd
Msba
Tolc
Prop
Shia
Livk
Male
Metq
Metn
Mtla
Prox
Prov
Putp
Potg
Potd
Pota
Potf
Rbsb
Mmup
Glnq
Glnh
Gadc
Hisj
Gatc
Gatb
Gata
Gntt
Ugpb
Mglb
Chbb
Dppd
Dppf
Dppa
Gsic
Gsia
Gsib
Dsbc
Dsbg
Cyca
Mscl
Mscm
Mscs
Transport, inner membrane
Cntrl
Adk
AroA
CarA
CysH
Dxs
Eno
FbaA
GapA
GdhA
GlmS
GltA
Gnd
IlvC
MetE
PckA
PfkA
PfkB
Pgi
Ppc
Prs
Pts
PurB
PurC
PykA
PykF
PyrF
Sdh
TpiA
Zwf
CRISPRi strains
Proteins
Ompf
Ompc
Tsx
Tonb
Exbb
Exbd
Fhue
Fepa
Lamb
Tolc
Fadl
Fhua
Fiu
Cira
Acra
Acrb
Btub
Ompa
Cysa
Cysp
Moda
Sbp
Trka
Kch
Fepb
Feob
Pspe
Copa
Znta
Grxb
Bfr
Cueo
Mgta
Psts
Pstb
Cora
Yqjh
Pita
Transport, Outer Membrane
Transport, Outer Membrane Porin
Inorganic Ion Transport and Metabolism
CRISPRi strains
-3
-2
-1
0
1
2
3
Relative enzyme abundance (log2)


Alternate Carbon Metabolism
Membrane Metabolism
Proteins
CRISPRi strains
Cell Envelope Biosynthesis
Glycerophospholipid Metabolism
Lipopolysaccharide Biosynthesis / Recycling
Membrane Lipid Metabolism
Murein Biosynthesis
Murein Recycling
-3 -2 -1 0 1 2 3
Relative enzyme abundance (log2)


Other
Genes
Rsga
Glya
Puru
Gcvh
Gcvt
Gcvp
Alda
Metf
Fold
Ghrb
Ghra
Gapa
Mgsa
Dkgb
Dkga
Glda
Hcha
Hmp
Alle
Adhe
Acs
Eda
Adhp
Grca
Pfla
Pflb
Pta
Purt
Acka
Ldha
Flda
Ydbk
Gltx
Can
Trxa
Trxc
Katg
Kate
Betb
Soda
Sodb
Sodc
Grxd
Grxc
Grxa
Grxb
Seld
Lnt
Lpp
Dsbc
Dsbg
Btue
Bcp
Dsba
Ahpf
Ahpc
Fpr
Tehb
Biomass and Mantainance Functions
Folate Metabolism
Glyoxylate Metabolism
Metabolite Repair
Methylglyoxal Metabolism
Nitrogen Metabolism
Pyruvate Metabolism
tRNA Charging
Unassigned
Cntrl
Adk
AroA
CarA
CysH
Dxs
Eno
FbaA
GapA
GdhA
GlmS
GltA
Gnd
IlvC
MetE
PckA
PfkA
PfkB
Pgi
Ppc
Prs
Pts
PurB
PurC
PykA
PykF
PyrF
Sdh
TpiA
Zwf
CRISPRi strains
-3
-2
-1
0
1
2
3
Relative enzyme abundance (log2)

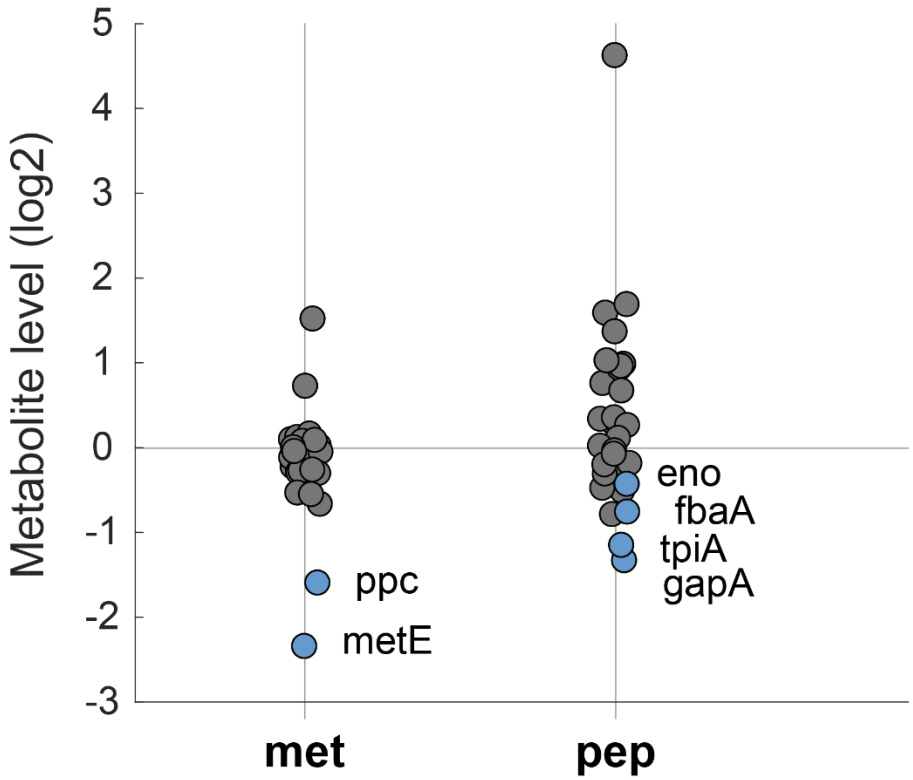


Figure S10. Related to Figure 12.

Fold-changes of methionine (met) and phosphoenolpyruvate (PEP) in the 30 different CRISPRi strains. Dots show means of n = 2 samples.

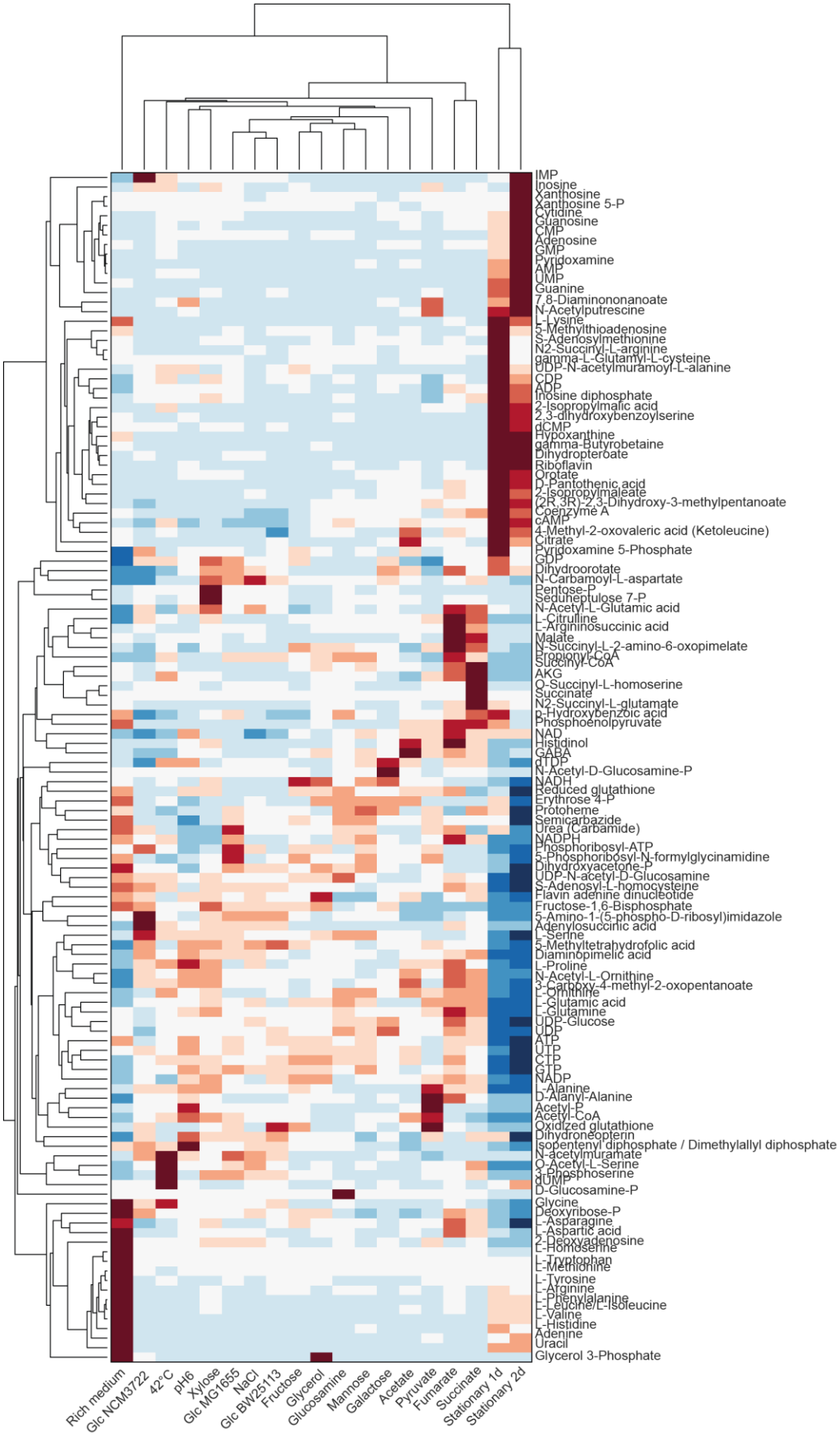

IMP
Inosine
Xanthosine
Xanthosine 5-P
Cytidine
Guanosine
CMP
Adenosine
GMP
Pyridoxamine
AMP
UMP
Guanine
7,8-Diaminononanoate
N-Acetylputrescine
L-Lysine
5-Methylthioadenosine
S-Adenosylmethionine
N2-Succinyl-L-arginine
gamma-L-Glutamyl-L-cysteine
UDP-N-acetylmuramoyl-L-alanine
CDP
ADP
Inosine diphosphate
2-Isopropylmalic acid
2,3-dihydroxybenzoylserine
dCMP
Hypoxanthine
gamma-Butyrobetaine
Dihydropteroate
Riboflavin
Orotate
D-Pantothenic acid
2-Isopropylmaleate
(2R,3R)-2,3-Dihydroxy-3-methylpentanoate
Coenzyme A
cAMP
4-Methyl-2-oxovaleric acid (Ketoleucine)
Citrate
Pyridoxamine 5-Phosphate
GDP
Dihydroorotate
N-Carbamoyl-L-aspartate
Pentose-P
Sedoheptulose 7-P
N-Acetyl-L-Glutamic acid
L-Citrulline
L-Argininosuccinic acid
Malate
N-Succinyl-L-2-amino-6-oxopimelate
Propionyl-CoA
Succinyl-CoA
AKG
O-Succinyl-L-homoserine
Succinate
N2-Succinyl-L-glutamate
p-Hydroxybenzoic acid
Phosphoenolpyruvate
NAD
Histidinol
GABA
dTDP
N-Acetyl-D-Glucosamine-P
NADH
Reduced glutathione
Erythrose 4-P
Protoheme
Semicarbazide
Urea (Carbamide)
NADPH
Phosphoribosyl-ATP
5-Phosphoribosyl-N-formylglycinamidine
Dihydroxyacetone-P
UDP-N-acetyl-D-Glucosamine
S-Adenosyl-L-homocysteine
Flavin adenine dinucleotide
Fructose-1,6-Bisphosphate
5-Amino-1-(5-phospho-D-ribosyl)imidazole
Adenylosuccinic acid
L-Serine
5-Methyltetrahydrofolic acid
Diaminopimelic acid
L-Proline
N-Acetyl-L-Ornithine
3-Carboxy-4-methyl-2-oxopentanoate
L-Ornithine
L-Glutamic acid
L-Glutamine
UDP-Glucose
UDP
ATP
UTP
CTP
GTP
NADP
L-Alanine
D-Alanyl-Alanine
Acetyl-P
Acetyl-CoA
Oxidized glutathione
Dihydroneopterin
Isopentenyl diphosphate / Dimethylallyl diphosphate
N-acetylmuramate
O-Acetyl-L-Serine
3-Phosphoserine
dUMP
D-Glucosamine-P
Glycine
Deoxyribose-P
L-Asparagine
L-Aspartic acid
2-Deoxyadenosine
L-Homoserine
L-Tryptophan
L-Methionine
L-Tyrosine
L-Arginine
L-Phenylalanine
L-Leucine/L-Isoleucine
L-Valine
L-Histidine
Adenine
Uracil
Glycerol 3-Phosphate
3
2
1
0
-1
-2
-3
Rich medium
Glc NCM3722
42°C
pH6
Xylose
Glc MG1655
NaCl
Glc BW25113
Fructose
Glycerol
Glucosamine
Mannose
Galactose
Acetate
Pyruvate
Fumarate
Succinate
Stationary 1d
Stationary 2d

Figure S11: agglomerative hierarchical clustering of the condition-dependent metabolome dataset. Data matrix is normalized along rows (metabolites) and data clustered using average linkage, dendrograms are determined with pairwise Euclidean distance. Upper dendrogram highlights the separation of metabolomes of cells growing exponentially on minimal media, compared to the rich medium condition and non-growing cells (stationary 1d/2d). In the rich medium condition, the additional nutrients are, as expected, are strongly enriched and cluster together.

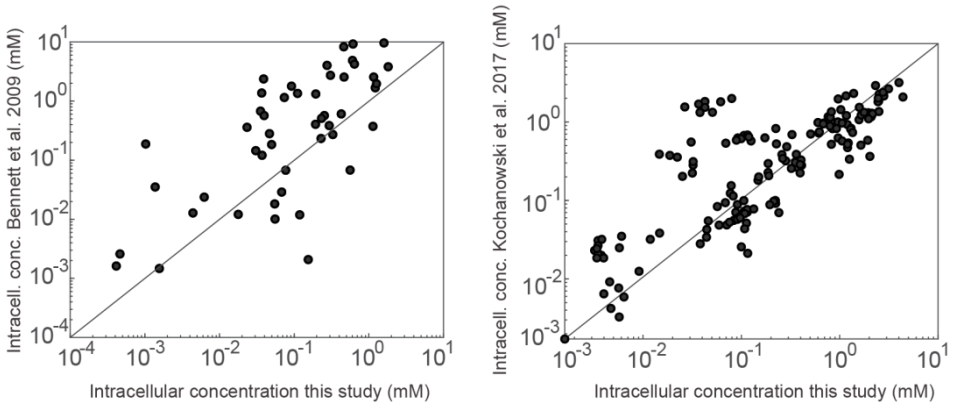


Figure S12: Comparison of absolute intracellular concentrations of metabolites with literature data. Only for conditions analyzed both in our study and in Bennett et al. 2009 (left) or and Kochanowski et al. 2017 (right).

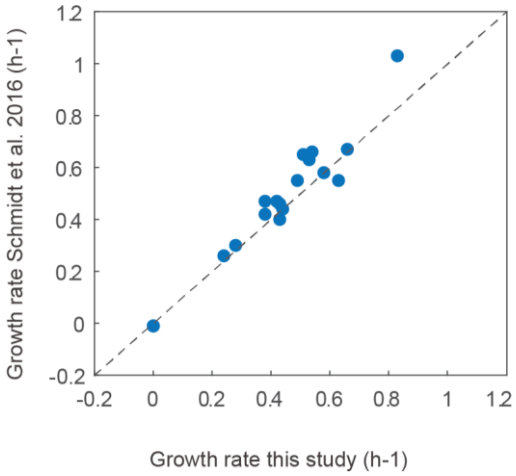


Figure S13: Comparison of growth rates for conditions analyzed both in our study and in Schmidt ed al. 2016.

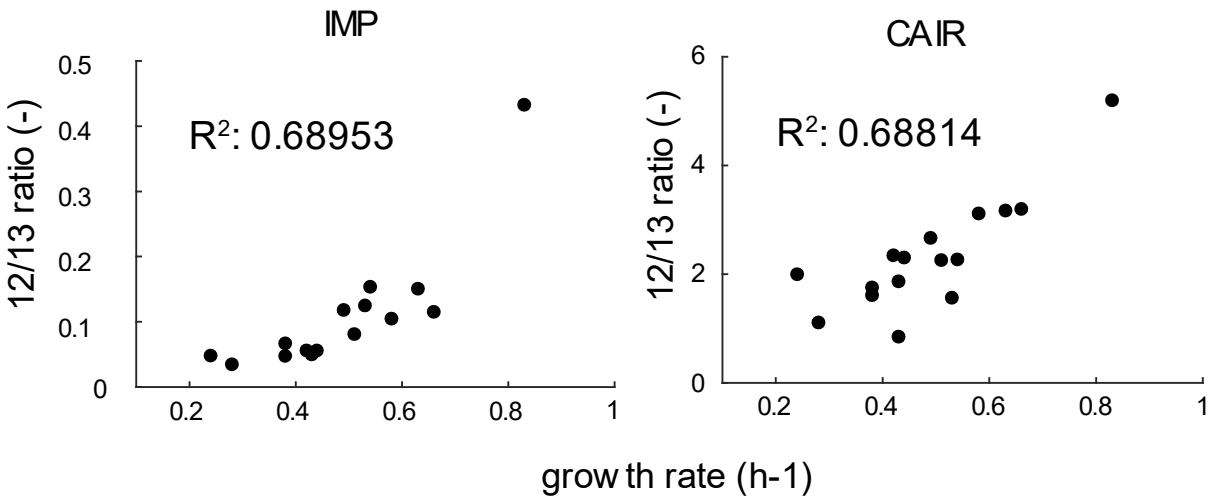


Figure S14: Metabolites with the highest correlation between their concentrations and growth rates. Shown all conditions with exponential growth in minimal medium. IMP (inosine monophosphate) and CAIR (5-Amino-1-(5-phospho-D-ribosyl)imidazole) are both precursors in the biosynthesis of purines. R^2 indicates the squared Person correlation coefficient.

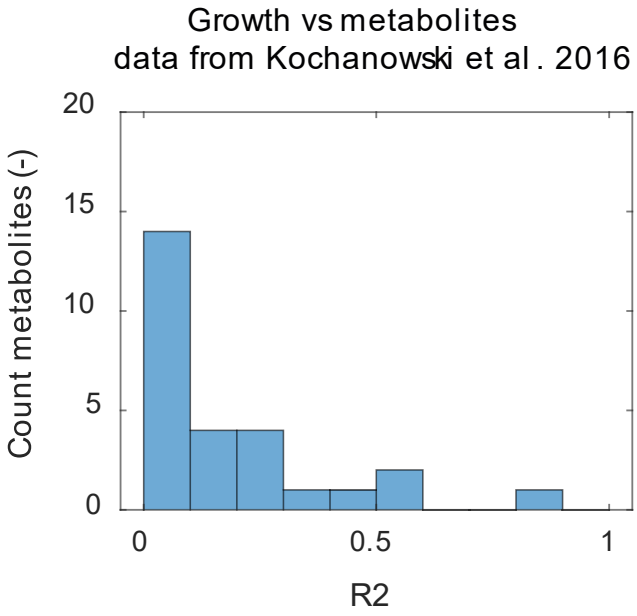


Figure S15: histogram representing correlation values between metabolite levels against the relative growth rate, data from Kochanowski et al. 2016. Correlations were plotted for all metabolites (27) which had a measured concentration among all 23 conditions. R^2 indicates the squared Person correlation coefficient. The highest scoring metabolite is cAMP, which was found also in our data to be among metabolites with a tendency to correlate with growth.

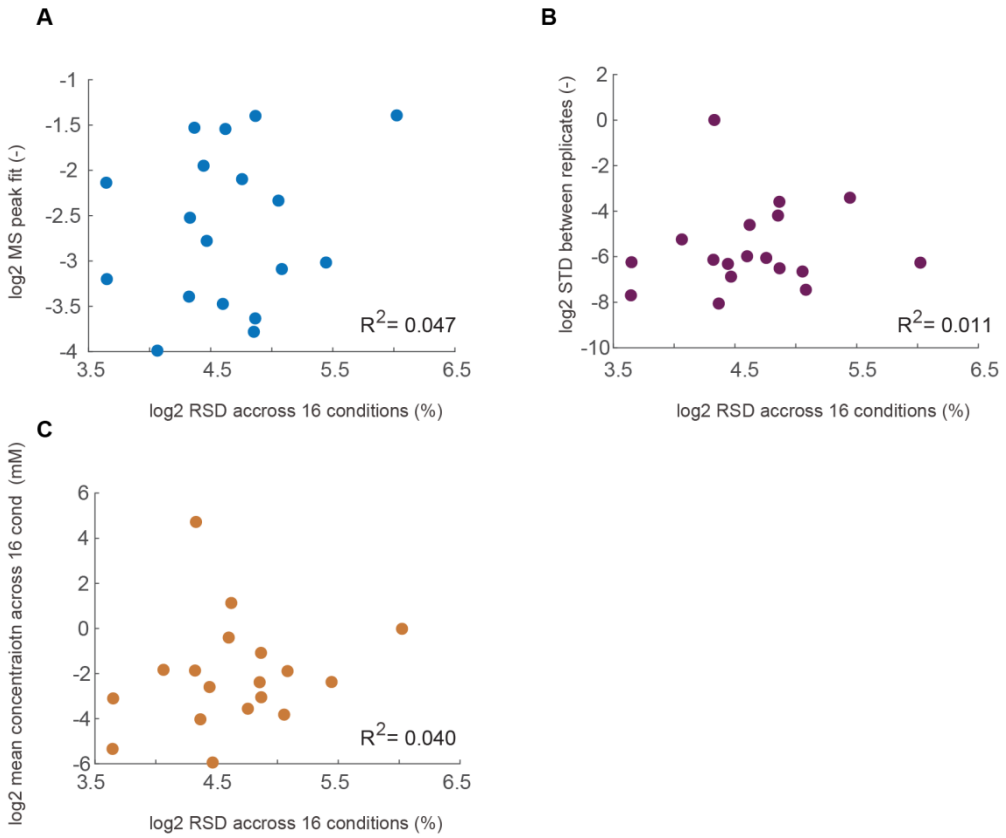


Figure S16: correlations between relative standard deviation (RSD) of amino-acids against the relative (A) fit of the raw signal peaks, (B) standard deviation between replicates and (C) the mean concentration between replicates across the 16 conditions.

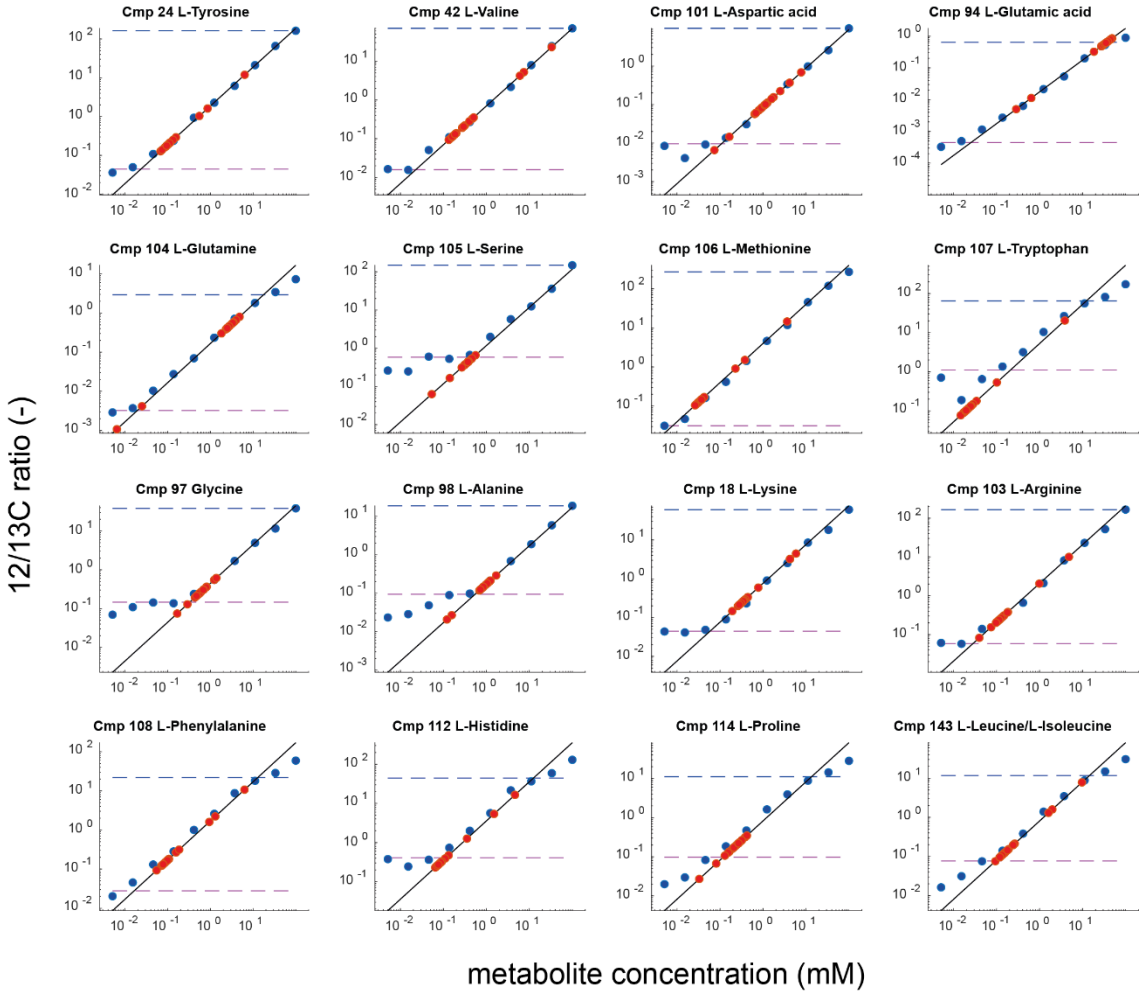
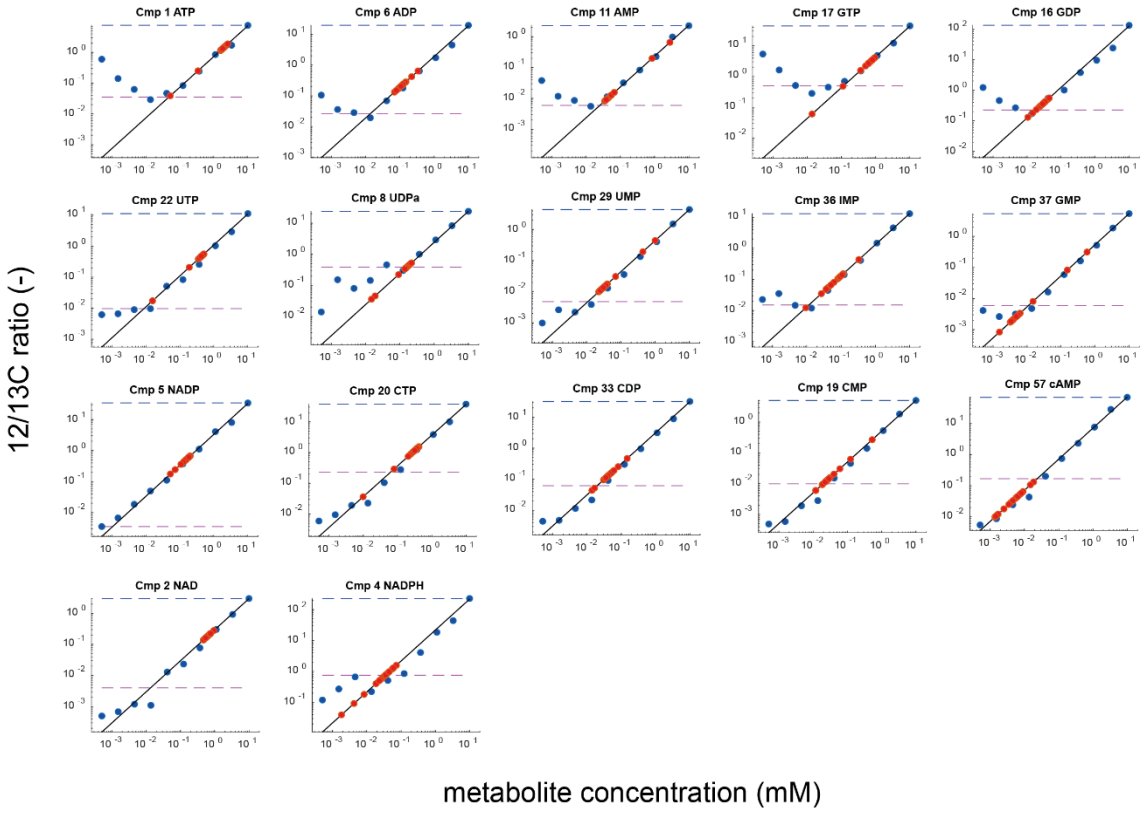


Figure S17: Calibration curves for absolute metabolite concentration calculations. 10 different concentrations of standards were measured (blue dots), and sample metabolites (red dots) were calibrating by fitting the measured standards by linear regression. Regressions were performed using a range for which the measured standards would not diverge from the linear fit by 25%.


Cmp 1 ATP
Cmp 6 ADP
Cmp 11 AMP
Cmp 17 GTP
Cmp 16 GDP
Cmp 22 UTP
Cmp 8 UDPa
Cmp 29 UMP
Cmp 36 IMP
Cmp 37 GMP
Cmp 5 NADP
Cmp 20 CTP
Cmp 33 CDP
Cmp 19 CMP
Cmp 57 cAMP
Cmp 2 NAD
Cmp 4 NADPH
12/13C ratio (-)
metabolite concentration (mM)

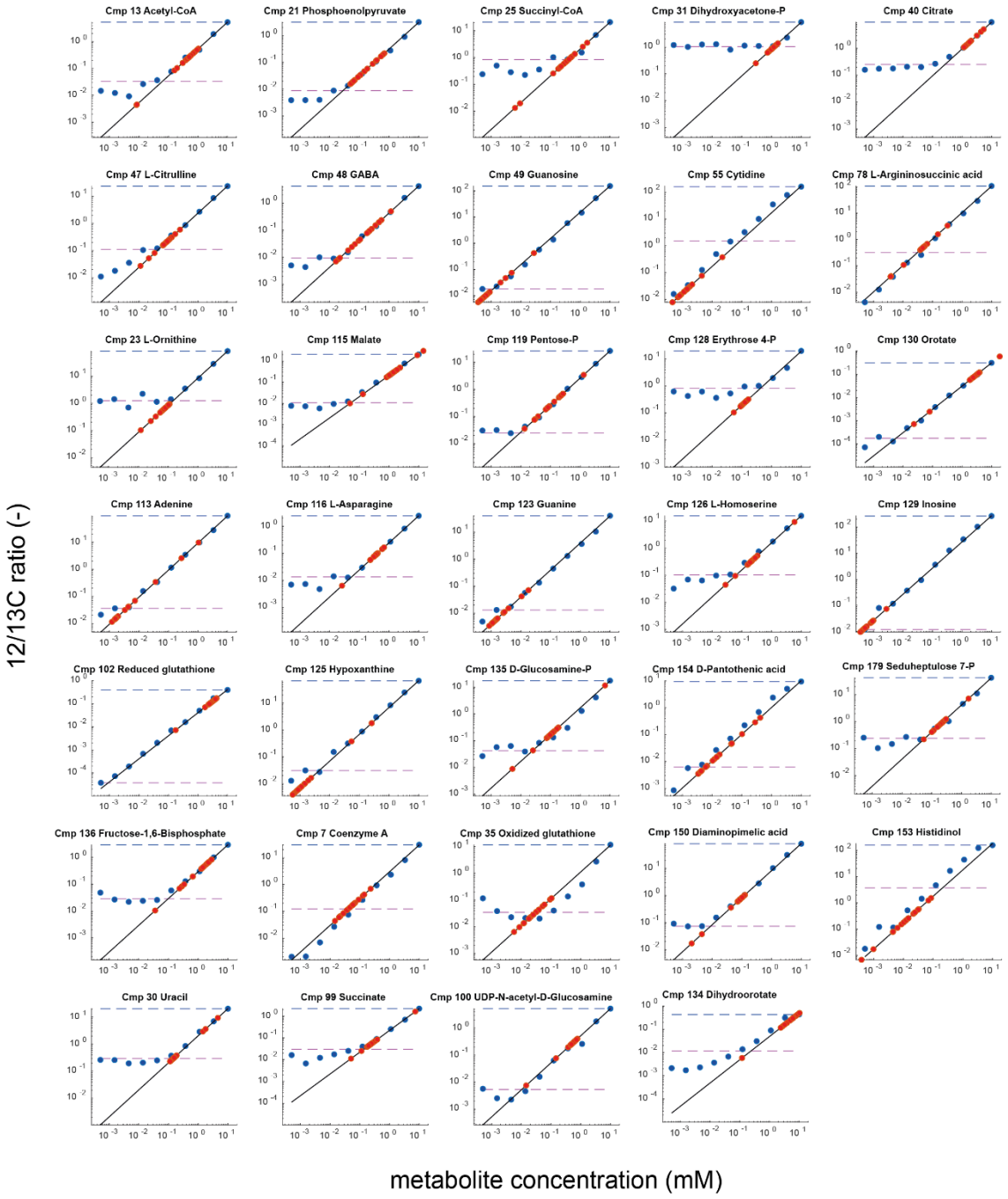

Cmp 13 Acetyl-CoA
Cmp 21 Phosphoenolpyruvate
Cmp 25 Succinyl-CoA
Cmp 31 Dihydroxyacetone-P
Cmp 40 Citrate
Cmp 47 L-Citrulline
Cmp 48 GABA
Cmp 49 Guanosine
Cmp 55 Cytidine
Cmp 78 L-Argininosuccinic acid
Cmp 23 L-Ornithine
Cmp 115 Malate
Cmp 119 Pentose-P
Cmp 128 Erythrose 4-P
Cmp 130 Orotate
Cmp 113 Adenine
Cmp 116 L-Asparagine
Cmp 123 Guanine
Cmp 126 L-Homoserine
Cmp 129 Inosine
Cmp 102 Reduced glutathione
Cmp 125 Hypoxanthine
Cmp 135 D-Glucosamine-P
Cmp 154 D-Pantothenic acid
Cmp 179 Seduheptulose 7-P
Cmp 136 Fructose-1,6-Bisphosphate
Cmp 7 Coenzyme A
Cmp 35 Oxidized glutathione
Cmp 150 Diaminopimelic acid
Cmp 153 Histidinol
Cmp 30 Uracil
Cmp 99 Succinate
Cmp 100 UDP-N-acetyl-D-Glucosamine
Cmp 134 Dihydroorotate
12/13C ratio (-)
metabolite concentration (mM)

Table S1 Bacterial strains and identification sequence for the arrayed CRISPRi library.

Number	*Background strain*	*Plasmid*	*Target*	*Spacer Sequence*
1	E. coli YYdCas9	pgRNA	accA	CAATCGGCTGTTCAAAATCA
2	E. coli YYdCas9	pgRNA	acnB	CACGCTCAGCTACGTGCTTA
3	E. coli YYdCas9	pgRNA	adk	TGAGTCCCTTTCCCCGCGCC
4	E. coli YYdCas9	pgRNA	alaC	TGCGCGTAAAGCGACGTTCA
5	E. coli YYdCas9	pgRNA	argA	GTATTGATATAGGGAACCGAATGG
6	E. coli YYdCas9	pgRNA	argE	TTTTTCATTGTTGACACACCTC
7	E. coli YYdCas9	pgRNA	aroA	AGCGATGGGTTGTAACGTCA
8	E. coli YYdCas9	pgRNA	aroL	CCCGAGGCCCGATCAGAAAAAG
9	E. coli YYdCas9	pgRNA	asd	AACCATGCGTTGCATGAGAA
10	E. coli YYdCas9	pgRNA	asnA	ACGTTGTTTGGCAATGTAAG
11	E. coli YYdCas9	pgRNA	aspC	AATCGGGTCGGCAGGAGCGG
12	E. coli YYdCas9	pgRNA	bioA	CCAGATATGGCGTTGGTCAA
13	E. coli YYdCas9	pgRNA	bioH	ATGAACATTCCCCTGACCTT
14	E. coli YYdCas9	pgRNA	carA	TATGGCCCGACCGTGAAACT
15	E. coli YYdCas9	pgRNA	coaA	TTGCGGTCAAACTGTAGGTA
16	E. coli YYdCas9	pgRNA	coaD	GTAATGGGATCGAAAGTACC
17	E. coli YYdCas9	pgRNA	cyaA	GTTTCAGAGTCTCAATATAG
18	E. coli YYdCas9	pgRNA	cysE	CAGTCCGCCAGCGTTCTGGCTT
19	E. coli YYdCas9	pgRNA	cysH	CAGGGCGTTTAGATCGAGTT
20	E. coli YYdCas9	pgRNA	dapB	CGGCTCCCGCGATGGCAACG
21	E. coli YYdCas9	pgRNA	dapD	ATCTCGGCACGGCGTTCAAAAG
22	E. coli YYdCas9	pgRNA	dfp	GAACGATTTTTTTACCGGCC
23	E. coli YYdCas9	pgRNA	dxr	CGAGCCGGTCGAGCCCAGAA
24	E. coli YYdCas9	pgRNA	dxs	CAGTGCCAGGGTCGGGTATT
25	E. coli YYdCas9	pgRNA	eno	ACCGATGATTTTTACGATTT
26	E. coli YYdCas9	pgRNA	fbaA	TCATCACCAGTGATTACGCC
27	E. coli YYdCas9	pgRNA	folA	TCGGCAGGCAGGTTCCACGGCA

28	E. coli YYdCas9	pgRNA	gadA	ACGTGAATCGAGTAGTTCTGAG
29	E. coli YYdCas9	pgRNA	gadB	ACGTTTTGATTCTGCGATAG
30	E. coli YYdCas9	pgRNA	gapA	GGAAAACAATGCGACCGATA
31	E. coli YYdCas9	pgRNA	gdhA	CGCGCTTTTGGACATGGTTG
32	E. coli YYdCas9	pgRNA	glmS	CAGACGACGTAAACCTTCAAGA
33	E. coli YYdCas9	pgRNA	glnA	CATCGTCAGTACGTGTTCAG
34	E. coli YYdCas9	pgRNA	gltA	AACAGCTGTATCCCCGTTGA
35	E. coli YYdCas9	pgRNA	glyA	CCACAGTTCGGCATCATAAT
36	E. coli YYdCas9	pgRNA	gmk	GGATTTACCCGCGCCACTGG
37	E. coli YYdCas9	pgRNA	gnd	GACTACGCCGATCTGTTGCT
38	E. coli YYdCas9	pgRNA	gpsA	CAGTCATTGAAGCATTACGT
39	E. coli YYdCas9	pgRNA	gshB	GATGTTGATGTTTGCGATG
40	E. coli YYdCas9	pgRNA	guaB	CGGTAGAGTGAGCAGGAACG
41	E. coli YYdCas9	pgRNA	hemB	TGCGCAGGCGACGAGGGCGT
42	E. coli YYdCas9	pgRNA	hemG	CAGTTCCGAAGCCAGGTAGG
43	E. coli YYdCas9	pgRNA	hemH	CAGGTTTGCCAGCAGGATAC
44	E. coli YYdCas9	pgRNA	hisB	TCACTCGGCGGTTCGCTAATCA
45	E. coli YYdCas9	pgRNA	hisG	TGAGTCATCACTTAAACGGC
46	E. coli YYdCas9	pgRNA	icd	GTGATCTTCTTGCCTTGTGC
47	E. coli YYdCas9	pgRNA	idi	TACTTTTCCAGCGTACCCGT
48	E. coli YYdCas9	pgRNA	ilvA	CCTTCCGGAGCACCGGACAG
49	E. coli YYdCas9	pgRNA	ilvC	CGCGCCATCGGCGAATTCATCG
50	E. coli YYdCas9	pgRNA	ispB	AACACCCGCCATATCTTGCG
51	E. coli YYdCas9	pgRNA	ispG	TTCTACGTTGAATTGGAGCC
52	E. coli YYdCas9	pgRNA	ispH	AACCACGCGGGTTGGCCAAC
53	E. coli YYdCas9	pgRNA	ispU	GCTGGCAATTTTTCGCTAAG
54	E. coli YYdCas9	pgRNA	kdsA	CGCCAAACAGTACGAACGGC
55	E. coli YYdCas9	pgRNA	leuA	GCCTGTTCACCGTCGCGCAATG
56	E. coli YYdCas9	pgRNA	LuxS	CGACTGTGAAGCTATCTAACAA
57	E. coli YYdCas9	pgRNA	LacZ	GGCCAGTGAATCCGTAATCA

58	E. coli YYdCas9	pgRNA	lysA	TCGGTGCTGAACAGTGAATG
59	E. coli YYdCas9	pgRNA	menD	GAATGACCGCCGCCCAGCGT
60	E. coli YYdCas9	pgRNA	metA	AGCTCGTCCGGCACACGAAT
61	E. coli YYdCas9	pgRNA	metC	ATTACGTGTCGCGTGTTTTT
62	E. coli YYdCas9	pgRNA	metE	CAGGCCAACGCGAGGGAAACCG
63	E. coli YYdCas9	pgRNA	metK	GGATGCCCTTCAGAGACGGACT
64	E. coli YYdCas9	pgRNA	nadA	GGAAAGGATAAATCGCCGTGTC
65	E. coli YYdCas9	pgRNA	nadB	ACACGTCACATGAATGTTCA
66	E. coli YYdCas9	pgRNA	nadC	TCAGGGTTATAGCGGCGAGG
67	E. coli YYdCas9	pgRNA	nadE	TCCTCTTCAGCATTAATCTG
68	E. coli YYdCas9	pgRNA	nadK	TCAGTGCAGTGGGGTGCCGT
69	E. coli YYdCas9	pgRNA	nrdA	TGCCGCCCAATCCAGAACGCGA
70	E. coli YYdCas9	pgRNA	pabA	AATCGTAGTTATCTATAAGC
71	E. coli YYdCas9	pgRNA	panC	TTGCTGACGCAGCAGCGGCA
72	E. coli YYdCas9	pgRNA	panD	CATGAGTCACTTTCACGCGG
73	E. coli YYdCas9	pgRNA	pck	ATAAGCCTCGAGTTCTTGCG
74	E. coli YYdCas9	pgRNA	pfkA	CGAATTGCGGCGTTCATGCC
75	E. coli YYdCas9	pgRNA	pfkB	ATTGTTGCGCTATCGAGAGA
76	E. coli YYdCas9	pgRNA	pgi	TGCCAGGCAGCGGTCTGCGT
77	E. coli YYdCas9	pgRNA	pgk	TTTCCCAGCAAGATCCAGAT
78	E. coli YYdCas9	pgRNA	pheA	TTTCTCTCGCAGCGCCAGTAAC
79	E. coli YYdCas9	pgRNA	ppc	ACTGACATTACTACGCAATG
80	E. coli YYdCas9	pgRNA	proB	GCCGAGTTTTACCACCAGCGTC
81	E. coli YYdCas9	pgRNA	proC	GGCAATCAGACCGCCGAGAA
82	E. coli YYdCas9	pgRNA	prs	CCAGCAAAAAGCTTCATATC
83	E. coli YYdCas9	pgRNA	ptsH	CGGTAATGGTAACTTCTTGC
84	E. coli YYdCas9	pgRNA	ptsI	TTTACCGAAAGCGATACCCG
85	E. coli YYdCas9	pgRNA	purA	TTTACCTTCGTCACCCCATT
86	E. coli YYdCas9	pgRNA	purB	ACAGGGGAAACGGCGGTCAGTG
87	E. coli YYdCas9	pgRNA	purC	CGGGTTTTCCGTGCTGTATA

88	*E. coli YYdCas9*	*pgRNA*	*purE*	*GACACGCGCCGGATTATTGC*
89	*E. coli YYdCas9*	*pgRNA*	*purL*	*ATTCGGAATGCCGACAGTGC*
90	*E. coli YYdCas9*	*pgRNA*	*pykA*	*TTTTGTTCTGCGAAGCCTTC*
91	*E. coli YYdCas9*	*pgRNA*	*pykF*	*TCCGATGGTGCAAACAATTT*
92	*E. coli YYdCas9*	*pgRNA*	*pyrB*	*AAATGATATGTTTCTGATATAG*
93	*E. coli YYdCas9*	*pgRNA*	*pyrC*	*CGGATCTTTAATACCTGGGA*
94	*E. coli YYdCas9*	*pgRNA*	*pyrD*	*AAAGGGCTTTACGAACGAAG*
95	*E. coli YYdCas9*	*pgRNA*	*pyrE*	*GCTAAGCGCAAATTCAATAAAC*
96	*E. coli YYdCas9*	*pgRNA*	*pyrF*	*AGGAGAATTCGTAACAGCGC*
97	*E. coli YYdCas9*	*pgRNA*	*pyrG*	*CAGAGAGGATACGACCCCGC*
98	*E. coli YYdCas9*	*pgRNA*	*ribA*	*GCCCCATGGGGTTGGCAGTT*
99	*E. coli YYdCas9*	*pgRNA*	*ribB*	*TTCGAAAGGCGTACCAAAAG*
100	*E. coli YYdCas9*	*pgRNA*	*ribC*	*GTCAATCGACACCAGTTTTG*
101	*E. coli YYdCas9*	*pgRNA*	*sdhC*	*GTCTGTAGGTCCAGATTAAC*
102	*E. coli YYdCas9*	*pgRNA*	*serA*	*GAAGGCTTTCCAGCGCCTTT*
103	*E. coli YYdCas9*	*pgRNA*	*sucA*	*GAGGTAAGAAGAGTCCAACC*
104	*E. coli YYdCas9*	*pgRNA*	*sucC*	*CGGTGCTGGTAAGCCATAGC*
105	*E. coli YYdCas9*	*pgRNA*	*thyA*	*AAAGCGTTCCGGTTCCGGTA*
106	*E. coli YYdCas9*	*pgRNA*	*tktA*	*ATTGGCAAGCTCTTTACGTG*
107	*E. coli YYdCas9*	*pgRNA*	*tpiA*	*TTCCAGTTACCCATCACTAA*
108	*E. coli YYdCas9*	*pgRNA*	*trpA*	*TTCTTTGCGCTCCTTCAACT*
109	*E. coli YYdCas9*	*pgRNA*	*tyrA*	*ATCAATTTGATCGCGTAATG*
110	*E. coli YYdCas9*	*pgRNA*	*ubiD*	*GCGTAAATCGTTATATTTCA*
111	*E. coli YYdCas9*	*pgRNA*	*zwf*	*AATGACCAGGTCACAGGCCT*

pgRNA-bacteria plasmids for strains 4, 11, 29, 47, 72, 82, 105, 106 were cloned by Gibson assembly. All remaining strains were prepared with plasmids synthetized by Doulix.

Table S2: Occurrence of growth phenotypes at different starting optical densities (ODs). Data shown for three different YYdCas9 strains with gRNAs targeting *purM* (purine biosynthesis), *argE* and *argA* (arginine biosynthesis). Reported are the average (n=3) initial cell concentrations, the final OD at the time of occurrence of a growth defect, the time of occurrence of the defect and the number of OD duplications from the start of the culture.

PurM			
Initial OD	Final OD	Response Time	Duplications
0.08	0.1984	2.321	2.48
0.038	0.0935	2.488	2.46
0.021	0.0628	2.821	2.99
ArgE			
Initial OD	Final OD	Response Time	Duplications
0.082	0.216	2.321	2.63
0.038	0.107	2.321	2.81
0.0198	0.055	2.321	2.77
ArgA			
Initial OD	Final OD	Response Time	Duplications
0.0857	0.2616	2.321	3.05
0.0405	0.1312	2.488	3.24
0.0224	0.0726	2.654	3.24

Table S3: Correlation between metabolite concentrations and growth rates in different environmental conditions. R2 indicates the squared Pearson correlation coefficient.

Metabolite	**R2**
IMP	0.689534
5-Amino-1-(5-phospho-D-ribosyl)imidazole	0.688144
Dihydropteroate	0.570407
(2R,3R)-2,3-Dihydroxy-3-methylpentanoate	0.568763
Adenylosuccinic acid	0.551903
UDP	0.493183
cAMP	0.466123
Pyridoxamine	0.462529
N-acetylmuramate	0.458371
GABA	0.457916
Riboflavin	0.427516
NAD	0.418199
S-Adenosylmethionine	0.404423

Table S4: SVD of the metabolomics and proteomics datasets. The table displays, for each of the 14 singular vectors computed and the relative dataset, the correlation of each component to the growth rate values and how much each component contributes to explain the variance in the data. In bold, the 2 components for each dataset that had the highest correlation with the growth rate.

	proteome dataset		**metabolome dataset**	
component	**R2**	**% variance**	**R2**	**% variance**
SV1	**0.88**	**73.78**	0.12	59.50
SV2	0.40	4.44	0.21	7.53
SV3	0.07	3.10	**0.41**	**5.31**
SV4	0.36	2.54	0.12	4.20
SV5	0.07	2.05	0.02	3.88
SV6	0.01	2.03	0.10	3.18
SV7	0.00	1.86	0.02	2.95
SV8	0.00	1.74	0.01	2.72
SV9	0.01	1.64	0.01	2.48
SV10	0.03	1.63	0.01	2.07
SV11	0.00	1.54	0.05	1.95
SV12	0.01	1.34	0.00	1.66
SV13	0.00	1.26	0.01	1.34
SV14	0.02	1.05	0.03	1.24

Table S5: strains and reagents

REAGENT or RESOURCE	Source	IDENTIFIER
Bacterial and Virus Strains		
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
YYdCas9: BW25993 CRISPRi-pgRNA_cntrl: intC::tetR-dcas9-aadA lacY::ypet-cat	This study	N/A
YYdCas9: BW25993 CRISPRi-pgRNA_carAB: intC::tetR-dcas9-aadA lacY::ypet-cat pUA66-PargE-gfp	This study	N/A
YYdCas9: BW25993 CRISPRi-pgRNA_metE: intC::tetR-dcas9-aadA lacY::ypet-cat pUA66-PmetB-gfp	This study	N/A
BW25113: F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, ΔgntK768::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514	Baba et al. 2006	JW3400-1
Genotypes and spacer sequences of arrayed CRISPRi strains are listed in Table S7.		
Chemicals,Peptides, and Recombinant Proteins		
Acetonitrile	Honeywell Riedel-de Haën	Cat#14261-2L
Methanol	VWR	Cat#83638.320
Anhydrotetracycline	Sigma-Aldrich	Cat#1035708-25MG
IPTG	Roth	Cat#CN08.2
Ampicillin	Roth	Cat#K029.2
Kanamycin	Roth	Cat#T832.3
Critical Commercial Assays		

PierceTM Quantitative Colometric Peptide Assay	Thermo Fisher Scientific	Cat#23275
PierceTM BCA Protein Assay Kit	Thermo Fisher Scientific	Cat#23225
Recombinant DNA		
pgRNA-bacteria	Qi et al. 2013	Addgene plasmid #44251
pUA66-PargE-gfp: pPargE-gfp	Zaslaver et al. 2006	N/A
pUA66-PmetB-gfp: pPmetB-gfp	Zaslaver et al. 2006	N/A
Software and Algorithms		
Matlab R2018b (9.5.0.944444) for analysis of experimental data	mathworks.com	N/A
Python 3.7.4	python.org	N/A
COBRApy	opencobra.github.io/cobrapy	N/A
Progenesis QIP (Waters)	waters.com	N/A
MASCOT (v2.5, Matrix Science)	matrixscience.com	N/A
SafeQuant	https://cran.r-project.org/web/packages/SafeQuant/index.html	N/A

Table S6: oligonucleotides.

Oligonucleotide	Sequence (5'-3')	Description	Origin
psgRNAamp-F	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Amplification of pgRNA for Gibson Assembly with amplified spacer oligonucleotides	This study
psgRNAamp-R	ACTAGTATTATACCTAGGACTGAGCTAGC	Amplification of pgRNA for Gibson Assembly with amplified spacer oligonucleotides	This study
protoamp-F	TTGACAGCTAGCTCAGTCCTAGGTATAATACTAGT	Amplification of spacer oligonucleotide	This study
protoamp-R	GCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC	Amplification of spacer oligonucleotide	This study
FWD_cassette_seq1	CCGAGTTGCTCTTGCC	Sequencing of cloned CRISPRi plasmids	This study
Rev_pkD-pgRNA_seq2	GACTCGAGTAAGGATCCAGTTC	Sequencing of cloned CRISPRi plasmids	This study
OH_amp_fwd	TAAGGATGATTTCTGGAATTCTAAAG	Amplification of pooled oligonucleotides	This study
OH_amp_rev	GTGCCACTTTTTCAAGTTGATAAC	Amplification of pooled oligonucleotides	This study
EcF_forward	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC	Amplification of the pgRNA backbone for Gibson Assembly with amplified pooled oligonucleotides	This study
EcF_reverse	ACTAGTATTATACCTAGGACTGAGCTAGC	Amplification of the pgRNA backbone for Gibson Assembly with amplified pooled oligonucleotides	This study
NGS_F2_adapter	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGCAATAGGCGTATCACGAGG	Amplification of a 300 bp fragment of pgRNA including the sgRNA	This study
NGS_R2_adapter	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGACGGCGCTATTCAGATCC	Amplification of a 300 bp fragment of pgRNA including the sgRNA	This study
Custom_N705	CAAGCAGAAGACGGCATACGAGATGGACTCCTGTCTCGTGGGCTCGG	I7 oligo	This study
Custom_N706	CAAGCAGAAGACGGCATACGAGATTAGGCATGGTCTCGTGGGCTCGG	I7 oligo	This study
Custom_N721	CAAGCAGAAGACGGCATACGAGATTACGCTGCGTCTCGTGGGCTCGG	I7 oligo	This study
Custom_N503	AATGATACGGCGACCACCGAGATCTACACAGAGGATATCGTCGGCAGCGTC	I5 oligo	This study
Custom_N504	AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTC	I5 oligo	This study
Custom_N511	AATGATACGGCGACCACCGAGATCTACACCGGAGAGATCGTCGGCAGCGTC	I5 oligo	This study
Custom_N513	AATGATACGGCGACCACCGAGATCTACACCTAGTCGATCGTCGGCAGCGTC	I5 oligo	This study

References

1. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* **3**, 318–356 (1961).
2. Monod, J., Changeux, J.-P. & Jacob, F. Allosteric proteins and cellular control systems. *Journal of Molecular Biology* **6**, 306–329 (1963).
3. Chubukov, V., Gerosa, L., Kochanowski, K. & Sauer, U. Coordination of microbial metabolism. *Nature Publishing Group* **12**, 327–340 (2014).
4. Li, X. & Snyder, M. 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 screening studies. *BioEssays : news and reviews in molecular, cellular and developmental biology* **33**, 485–9 (2011).
5. de Lorenzo, V. From the *selfish gene* to *selfish metabolism* : Revisiting the central dogma: Insights & Perspectives. *BioEssays* **36**, 226–235 (2014).
6. Yugi, K. & Kuroda, S. Metabolism-Centric Trans-Omics. *Cell Systems* **4**, 19–20 (2017).
7. You, C. *et al.* Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* **500**, 301–6 (2013).
8. Alam, M. T. *et al.* The metabolic background is a global player in. 1–10 (2016) doi:10.1038/NMICROBIOL.2015.30.
9. Park, J. O. *et al.* Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nature Chemical Biology* **advance on**, 482–489 (2016).
10. Mülleder, M. *et al.* Functional Metabolomics Describes the Yeast Biosynthetic Regulome. *Cell* 553–565 (2016) doi:10.1016/j.cell.2016.09.007.
11. Fuhrer, T., Zampieri, M., Sévin, D. C., Sauer, U. & Zamboni, N. Genomewide landscape of gene–metabolome associations in *Escherichia coli*. *Molecular Systems Biology* **13**, 907 (2017).
12. Kemmeren, P. *et al.* Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* **157**, 740–752 (2014).
13. Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nature reviews. Genetics* **16**, 299–311 (2015).
14. Bordbar, A., Monk, J. M., King, Z. a & Palsson, B. O. Constraint-based models predict metabolic and associated cellular functions. *Nature reviews. Genetics* **15**, 107–20 (2014).
15. Orth, J. D. *et al.* A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Molecular systems biology* **7**, 535 (2011).
16. Mo, M. L., Palsson, B. Ø. & Herrgård, M. J. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. **17**, 1–17 (2009).
17. Covert, M. W. & Palsson, B. Ø. Transcriptional Regulation in Constraints-based Metabolic Models. **277**, 28058–28064 (2002).
18. Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J. & Palsson, B. O. Integrating high-throughput and computational data elucidates bacterial networks. **429**, 2–6 (2004).
19. Shlomi, T., Eisenberg, Y., Sharan, R. & Ruppin, E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Molecular Systems Biology* **3**, 101 (2007).
20. Shlomi, T., Eisenberg, Y., Sharan, R. & Ruppin, E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Molecular Systems Biology* **3**, 101 (2007).
21. Karr, J. R. R. R. *et al.* A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell* **150**, 389–401 (2012).

22. Chandrasekaran, S. & Price, N. D. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 17845–17850 (2010).
23. Link, H., Christodoulou, D. & Sauer, U. ScienceDirect Advancing metabolic models with kinetic information. *Current Opinion in Biotechnology* **29**, 8–14.
24. Khodayari, A. & Maranas, C. D. A genome-scale Escherichia coli kinetic metabolic model satisfying flux data for multiple mutant strains. *Nature Communications* **7**, 1–12 (2016).
25. Browning, D. F. & Busby, S. J. W. Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology* **14**, 638–650 (2016).
26. Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews. Molecular Cell Biology* **16**, 178–189 (2015).
27. Gama-castro, S. *et al.* RegulonDB version 9 . 0 : high-level integration of gene regulation , coexpression , motif clustering and beyond Liliana Porr on-Sotelo. **44**, 133–143 (2016).
28. Costanzo, M. C. *et al.* Saccharomyces genome database provides new regulation data. **42**, 717–725 (2014).
29. Ishihama, A., Shimada, T. & Yamazaki, Y. Transcription profile of Escherichia coli: Genomic SELEX search for regulatory targets of transcription factors. *Nucleic Acids Research* **44**, 2058–2074 (2016).
30. Seo, S. W. *et al.* Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in Escherichia coli. *Nature communications* **5**, 4910 (2014).
31. 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).
32. Conrad, M. *et al.* Nutrient sensing and signaling in the yeast Saccharomyces cerevisiae. *FEMS Microbiology Reviews* **38**, 254–299 (2014).
33. Schmidt, A. *et al.* The quantitative and condition-dependent Escherichia coli proteome. *Nature biotechnology* **34**, 104–110 (2015).
34. Ulrich, L. E., Koonin, E. V. & Zhulin, I. B. One-component systems dominate signal transduction in prokaryotes. *Trends in Microbiology* **13**, 52–56 (2005).
35. Li, G., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *CELL* **157**, 624–635 (2014).
36. Gallego, O. *et al.* A systematic screen for protein-lipid interactions in Saccharomyces cerevisiae. *Molecular systems biology* **6**, 430 (2010).
37. Li, X., Gianoulis, T. a, Yip, K. Y., Gerstein, M. & Snyder, M. Extensive in vivo metabolite-protein interactions revealed by large-scale systematic analyses. *Cell* **143**, 639–50 (2010).
38. Nikolaev, Y. V, Kochanowski, K., Link, H. & Sauer, U. Systematic identification of Protein − Metabolite Interactions in Complex Metabolite Mixtures by Ligand-Detected Nuclear Magnetic Resonance Spectroscopy. (2016) doi:10.1021/acs.biochem.5b01291.
39. Feng, Y. *et al.* Global analysis of protein structural changes in complex proteomes. *Nature Biotechnology* **32**, 1036–1044 (2014).
40. Kholodenko, B., Yaffe, M. B. & Kolch, W. Computational Approaches for Analyzing Information Flow in Biological Networks. *Science Signaling* **5**, re1–re1 (2012).
41. He, F., Fromion, V. & Westerhoff, H. V. (Im)Perfect robustness and adaptation of metabolic networks subject to metabolic and gene-expression regulation: marrying control engineering with metabolic control analysis. *BMC systems biology* **7**, 131 (2013).
42. Kochanowski, K. *et al.* Few regulatory metabolites coordinate expression of central metabolic genes in Escherichia coli. *Molecular Systems Biology* **13**, 903 (2017).
43. Keren, L. *et al.* Promoters maintain their relative activity levels under different growth conditions. *Molecular Systems Biology* **9**, 701 (2013).

44. 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).
45. Redestig, H. & Costa, I. G. Detection and interpretation of metabolite-transcript coresponses using combined profiling data. *Bioinformatics* **27**, 357–365 (2011).
46. Chubukov, V. *et al.* Transcriptional regulation is insufficient to explain substrate-induced flux changes in Bacillus subtilis. *Molecular Systems Biology* **9**, 1–13 (2013).
47. Gerosa, L. *et al.* Pseudo-transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data. *Cell Systems* **1**, 270–282 (2015).
48. Kerkhoven, E. J. *et al.* Regulation of amino-acid metabolism controls flux to lipid accumulation in Yarrowia lipolytica. *npj Systems Biology and Applications* **2**, 16005 (2016).
49. Monk, J. M. *et al.* Multi-omics Quantification of Species Variation of Escherichia coli Links Molecular Features with Strain Phenotypes. *Cell Systems* 238–251 (2016) doi:10.1016/j.cels.2016.08.013.
50. Machado, D. & Herrgård, M. Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism. *PLoS Computational Biology* **10**, (2014).
51. Hackett, S. R. *et al.* Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* **354**, (2016).
52. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866.e17 (2016).
53. 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* acs.analchem.6b03731 (2017) doi:10.1021/acs.analchem.6b03731.
54. Kochanowski, K., Sauer, U. & Chubukov, V. Somewhat in control-the role of transcription in regulating microbial metabolic fluxes. *Current Opinion in Biotechnology* **24**, 987–993 (2013).
55. O'Brien, E. J., Utrilla, J. & Palsson, B. O. Quantification and Classification of E. coli Proteome Utilization and Unused Protein Costs across Environments. *PLoS Comput Biol* **12**, e1004998 (2016).
56. Noor, E. *et al.* The Protein Cost of Metabolic Fluxes: Prediction from Enzymatic Rate Laws and Cost Minimization. *PLoS Computational Biology* **12**, 1–29 (2016).
57. Davidi, D. & Milo, R. Lessons on enzyme kinetics from quantitative proteomics. *Current Opinion in Biotechnology* **46**, 81–89 (2017).
58. Link, H., Kochanowski, K. & Sauer, U. Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. *Nature Biotechnology* 1–6 (2013) doi:10.1038/nbt.2489.
59. Dekel, E. & Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588–92 (2005).
60. Keren, L. *et al.* Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. *Cell* **166**, 1282-1294.e18 (2016).
61. Gross, C. *et al.* A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell* **165**, 1493–1506 (2016).
62. Deaner, M. & Alper, H. S. Systematic Testing of Enzyme Perturbation Sensitivities via Graded dCas9 Modulation in Saccharomyces cerevisiae. *Metabolic Engineering* **40**, 14–22 (2017).
63. Jeschek, M., Gerngross, D. & Panke, S. Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort. *Nature communications* **7**, 11163 (2016).

64. Towbin, B. D. *et al.* Optimality and sub-optimality in a bacterial growth law. *Nature Communications* **8**, 14123 (2017).
65. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. *Science (New York, N.Y.)* **330**, 1099–102 (2010).
66. Bosdriesz, E., Molenaar, D., Teusink, B. & Bruggeman, F. J. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *FEBS Journal* **282**, 2029–2044 (2015).
67. Scott, M., Klumpp, S., Mateescu, E. M. & Hwa, T. Emergence of robust growth laws from optimal regulation of ribosome synthesis. 1–14 (2014).
68. Pavlov, M. Y. & Ehrenberg, M. Optimal control of gene expression for fast proteome adaptation to environmental change. *Proceedings of the National Academy of Sciences* **110**, 20527–20532 (2013).
69. Hui, S. *et al.* Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. 1–15 (2015).
70. Kochanowski, K. *et al.* Functioning of a metabolic flux sensor in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1130–5 (2013).
71. Geiger, R. *et al.* L-Arginine Modulates T Cell Metabolism and Enhances Survival and Anti-tumor Activity. *Cell* **167**, 829-842.e13 (2016).
72. Lorendeau, D., Christen, S., Rinaldi, G. & Fendt, S. M. Metabolic control of signalling pathways and metabolic auto-regulation. *Biology of the Cell* **107**, 251–272 (2015).
73. Chin, R. M. *et al.* The metabolite α-ketoglutarate extends lifespan by inhibiting ATP synthase and TOR. *Nature* (2014) doi:10.1038/nature13264.
74. Borkowski, O., Ceroni, F., Stan, G. & Ellis, T. Overloaded and stressed: wholecell considerations for bacterial synthetic biology. *Current Opinion in Microbiology* **33**, 123130 (2016).
75. Taylor, N. D. *et al.* Engineering an allosteric transcription factor to respond to new ligands. *Nature methods* **In press**, 1–11 (2016).
76. Joyce, A. R. & Palsson, B. Ø. The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol* **7**, 198–210 (2006).
77. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–550 (2016).
78. Maier, T., Güell, M. & Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Letters* **583**, 3966–3973 (2009).
79. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
80. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C. & Yates, J. R. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **113**, 2343–2394 (2013).
81. Malmström, J. *et al.* Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **460**, 762–765 (2009).
82. Zampieri, M., Sekar, K., Zamboni, N. & Sauer, U. Frontiers of high-throughput metabolomics. *Current Opinion in Chemical Biology* **36**, 15–23 (2017).
83. 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).
84. McCloskey, D., Xu, J., Schrübbers, L., Christensen, H. B. & Herrgård, M. J. RapidRIP quantifies the intracellular metabolome of 7 industrial strains of E. coli. *Metabolic Engineering* **47**, 383–392 (2018).
85. Rabinowitz, J. D. & Kimball, E. Acidic Acetonitrile for Cellular Metabolome Extraction from *Escherichia coli*. *Anal. Chem.* **79**, 6167–6173 (2007).

86. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature Chemical Biology* **5**, 593–599 (2009).
87. Wu, L. *et al.* Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly 13C-labeled cell extracts as internal standards. *Analytical Biochemistry* **336**, 164–171 (2005).
88. Barrangou, R. & Horvath, P. A decade of discovery: CRISPR functions and applications. *Nat Microbiol* **2**, 17092 (2017).
89. Sontheimer, E. J. & Barrangou, R. The Bacterial Origins of the CRISPR Genome-Editing Revolution. *Human Gene Therapy* **26**, 413–424 (2015).
90. Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).
91. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).
92. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
93. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research* **41**, 7429–7437 (2013).
94. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).
95. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* **8**, 2180–2196 (2013).
96. Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867-1882.e21 (2016).
97. Peters, J. M. *et al.* A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell* **165**, 1493–1506 (2016).
98. Rousset, F. *et al.* Genome-wide CRISPR-dCas9 screens in E. coli identify essential genes and phage host factors. *PLOS Genetics* **14**, e1007749 (2018).
99. Liu, S. J. *et al.* CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* **355**, eaah7111 (2017).
100. Mandegar, M. A. *et al.* CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. *Cell Stem Cell* **18**, 541–553 (2016).
101. Kim, S. K., Seong, W., Han, G. H., Lee, D.-H. & Lee, S.-G. CRISPR interference-guided multiplex repression of endogenous competing pathway genes for redirecting metabolic flux in Escherichia coli. *Microb Cell Fact* **16**, 188 (2017).
102. Wu, J., Du, G., Chen, J. & Zhou, J. Enhancing flavonoid production by systematically tuning the central metabolic pathways based on a CRISPR interference system in Escherichia coli. *Sci Rep* **5**, 13477 (2015).
103. Skerra, A. Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in Escherichia coli. *Gene* **151**, 131–135 (1994).
104. Lawson, M. J. *et al. In situ* genotyping of a pooled strain library after characterizing complex phenotypes. *Molecular Systems Biology* **13**, 947 (2017).
105. Patrick, W. M., Quandt, E. M., Swartzlander, D. B. & Matsumura, I. Multicopy Suppression Underpins Metabolic Evolvability. *Molecular Biology and Evolution* **24**, 2716–2722 (2007).
106. Keseler, I. M. *et al.* The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res* **45**, D543–D550 (2017).
107. Li, G.-W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell* **157**, 624–635 (2014).
108. Jensen, K. F. 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. *Journal of Bacteriology* **175**, 3401–3407 (1993).

109. Joyce, A. R. *et al.* Experimental and Computational Assessment of Conditionally Essential Genes in Escherichia coli. *Journal of Bacteriology* **188**, 8259–8271 (2006).
110. Labhsetwar, P., Cole, J. A., Roberts, E., Price, N. D. & Luthey-Schulten, Z. A. Heterogeneity in protein expression induces metabolic variability in a modeled *Escherichia coli* population. *Proceedings of the National Academy of Sciences* **110**, 14006–14011 (2013).
111. Berthelot, K., Estevez, Y., Deffieux, A. & Peruch, F. Isopentenyl diphosphate isomerase: A checkpoint to isoprenoid biosynthesis. *Biochimie* **94**, 1621–1634 (2012).
112. Volke, D. C., Rohwer, J., Fischer, R. & Jennewein, S. Investigation of the methylerythritol 4-phosphate pathway for microbial terpenoid production through metabolic control analysis. *Microb Cell Fact* **18**, 192 (2019).
113. Chubukov, V., Gerosa, L., Kochanowski, K. & Sauer, U. Coordination of microbial metabolism. *Nature Reviews Microbiology* **12**, 327–340 (2014).
114. Hartl, J., Kiefer, P., Meyer, F. & Vorholt, J. A. Longevity of major coenzymes allows minimal de novo synthesis in microorganisms. *Nature Microbiology* **2**, (2017).
115. Barker, C. S., Prüß, B. M. & Matsumura, P. Increased Motility of Escherichia coli by Insertion Sequence Element Integration into the Regulatory Region of the flhD Operon. *JB* **186**, 7529–7537 (2004).
116. Cho, S. *et al.* High-Level dCas9 Expression Induces Abnormal Cell Morphology in *Escherichia coli*. *ACS Synth. Biol.* **7**, 1085–1094 (2018).
117. Jeffery, C. J. Moonlighting proteins: old proteins learning new tricks. *Trends in Genetics* **19**, 415–417 (2003).
118. Jones, D. L. *et al.* Kinetics of dCas9 target search in Escherichia coli. *Science* **357**, 1420–1424 (2017).
119. Monk, J. M. *et al.* iML1515, a knowledgebase that computes Escherichia coli traits. *Nature Biotechnology* **35**, 904–908 (2017).
120. Christodoulou, D. *et al.* Reserve Flux Capacity in the Pentose Phosphate Pathway Enables Escherichia coli's Rapid Response to Oxidative Stress. *Cell Systems* **6**, 569-578.e7 (2018).
121. Deng, M.-D. *et al.* Directed evolution and characterization of Escherichia coli glucosamine synthase. *Biochimie* **88**, 419–429 (2006).
122. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nature Methods* **3**, 623–628 (2006).
123. Jiao, Z., Baba, T., Mori, H. & Shimizu, K. Analysis of metabolic and physiological responses to *gnd* knockout in *Escherichia coli* by using C-13 tracer experiment and enzyme activity measurement. *FEMS Microbiology Letters* **220**, 295–301 (2003).
124. McCloskey, D. *et al.* Growth Adaptation of gnd and sdhCB Escherichia coli Deletion Strains Diverges From a Similar Initial Perturbation of the Transcriptome. *Frontiers in Microbiology* **9**, (2018).
125. Izu, H., Adachi, O. & Yamada, M. Gene organization and transcriptional regulation of the gntRKU operon involved in gluconate uptake and catabolism of Escherichia coli. *Journal of Molecular Biology* **267**, 778–793 (1997).
126. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature chemical biology* **5**, 593–9 (2009).
127. Chandra, F. A., Buzi, G. & Doyle, J. C. Glycolytic Oscillations and Limits on Robust Efficiency. *Science* **333**, 187 (2011).
128. Grimbs, S., Selbig, J., Bulik, S., Holzhütter, H.-G. & Steuer, R. The stability and robustness of metabolic states: identifying stabilizing sites in metabolic networks. *Mol. Syst. Biol.* **3**, 146 (2007).
129. Kacser, H. & Burns, J. A. The control of flux. *Symp. Soc. Exp. Biol.* **27**, 65–104 (1973).
130. Levine, E. & Hwa, T. Stochastic fluctuations in metabolic pathways. *PNAS* **104**, 9224–9229 (2007).

131. Hackett, S. R. *et al.* Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* **354**, aaf2786–aaf2786 (2016).
132. Gerosa, L. *et al.* Pseudo-transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data. *Cell Systems* **1**, 270–282 (2015).
133. Davidi, D. & Milo, R. Lessons on enzyme kinetics from quantitative proteomics. *Curr. Opin. Biotechnol.* **46**, 81–89 (2017).
134. O'Brien, E. J., Utrilla, J. & Palsson, B. O. Quantification and Classification of E. coli Proteome Utilization and Unused Protein Costs across Environments. *PLoS Comput Biol* **12**, (2016).
135. 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).
136. Gon, S. *et al.* A novel regulatory mechanism couples deoxyribonucleotide synthesis and DNA replication in Escherichia coli. *The EMBO Journal* **25**, 1137–1147 (2006).
137. Chao, Y. P. & Liao, J. C. Alteration of growth yield by overexpression of phosphoenolpyruvate carboxylase and phosphoenolpyruvate carboxykinase in Escherichia coli. *Appl. Environ. Microbiol.* **59**, 4261–4265 (1993).
138. Parekh, B. S. & Hatfield, G. W. 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. *Journal of bacteriology* **179**, 2086–2088 (1997).
139. Park, J. O. *et al.* Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nature Chemical Biology* **12**, 482–489 (2016).
140. Piazza, I. *et al.* A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* **172**, 358-372.e23 (2018).
141. Newman, J. R. S. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).
142. Taniguchi, Y. *et al.* Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
143. McCloskey, D. *et al.* Evolution of gene knockout strains of E. coli reveal regulatory architectures governed by metabolism. *Nature Communications* **9**, 3796 (2018).
144. Schneider, D. A. & Gourse, R. L. Relationship between Growth Rate and ATP Concentration in *Escherichia coli*: A BIOASSAY FOR AVAILABLE CELLULAR ATP. *J. Biol. Chem.* **279**, 8262–8268 (2004).
145. Kochanowski, K. *et al.* Few regulatory metabolites coordinate expression of central metabolic genes in *Escherichia coli*. *Mol Syst Biol* **13**, 903 (2017).
146. Schmidt, A. *et al.* The quantitative and condition-dependent Escherichia coli proteome. *Nature Biotechnology* **34**, 104–110 (2016).
147. You, C. *et al.* Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* **500**, 301–306 (2013).
148. Alter, O., Brown, P. O. & Botstein, D. Singular value decomposition for genome-wide expression data processing and modeling. *Proceedings of the National Academy of Sciences* **97**, 10101–10106 (2000).
149. Chao, Y.-P., Lai, Z. J., Chen, P. & Chern, J.-T. Enhanced Conversion Rate of L-Phenylalanine by Coupling Reactions of Aminotransferases and Phosphoenolpyruvate Carboxykinase in Escherichia coli K-12. *Biotechnol. Prog.* **15**, 453–458 (1999).
150. Park, J. O. *et al.* Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat Chem Biol* **12**, 482–489 (2016).
151. 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).
152. Buckstein, M. H., He, J. & Rubin, H. Characterization of Nucleotide Pools as a Function of Physiological State in Escherichia coli. *JB* **190**, 718–726 (2008).

153. Petersen, C. & Møller, L. B. Invariance of the Nucleoside Triphosphate Pools of *Escherichia coli* with Growth Rate. *J. Biol. Chem.* **275**, 3931–3935 (2000).
154. Anderson, P. M. Binding of allosteric effectors to carbamyl-phosphate synthetase from Escherichia coli. *Biochemistry* **16**, 587–593 (1977).
155. Mülleder, M. *et al.* Functional Metabolomics Describes the Yeast Biosynthetic Regulome. *Cell* **167**, 553-565.e12 (2016).
156. Kacser, H., Burns, J. A., Kacser, H. & Fell, D. A. The control of flux. *Biochemical Society Transactions* **23**, 341–366 (1995).
157. Levine, E. & Hwa, T. Stochastic fluctuations in metabolic pathways. *Proceedings of the National Academy of Sciences* **104**, 9224–9229 (2007).
158. Alam, M. T. *et al.* The metabolic background is a global player in Saccharomyces gene expression epistasis. *Nature Microbiology* **1**, (2016).
159. Akashi, H. & Gojobori, T. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proceedings of the National Academy of Sciences* **99**, 3695–3700 (2002).
160. Zhang, H. *et al.* Biosynthetic energy cost for amino acids decreases in cancer evolution. *Nat Commun* **9**, 4124 (2018).
161. Koebmann, B. J., Westerhoff, H. V., Snoep, J. L., Nilsson, D. & Jensen, P. R. The Glycolytic Flux in Escherichia coli Is Controlled by the Demand for ATP. *JB* **184**, 3909–3916 (2002).
162. Yang, J. H. *et al.* A White-Box Machine Learning Approach for Revealing Antibiotic Mechanisms of Action. *Cell* **177**, 1649-1661.e9 (2019).
163. Andersen, K. B. & von Meyenburg, K. Charges of nicotinamide adenine nucleotides and adenylate energy charge as regulatory parameters of the metabolism in Escherichia coli. *J. Biol. Chem.* **252**, 4151–4156 (1977).
164. Lalanne, J.-B. *et al.* Evolutionary Convergence of Pathway-Specific Enzyme Expression Stoichiometry. *Cell* **173**, 749-761.e38 (2018).
165. Noor, E., Cherkaoui, S. & Sauer, U. Biological insights through omics data integration. *Current Opinion in Systems Biology* **15**, 39–47 (2019).
166. Davidi, D. & Milo, R. Lessons on enzyme kinetics from quantitative proteomics. *Current Opinion in Biotechnology* **46**, 81–89 (2017).
167. Murrell, J. C. Physiology of the bacterial cell — A molecular approach. *Trends in Genetics* **7**, 341 (1991).
168. Bar-Even, A. *et al.* The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* **50**, 4402–4410 (2011).
169. Link, H., Christodoulou, D. & Sauer, U. Advancing metabolic models with kinetic information. *Current Opinion in Biotechnology* **29**, 8–14 (2014).
170. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* **2**, (2006).
171. Volkmer, B. & Heinemann, M. Condition-Dependent Cell Volume and Concentration of Escherichia coli to Facilitate Data Conversion for Systems Biology Modeling. *PLoS ONE* **6**, e23126 (2011).
172. King, Z. A. *et al.* BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research* **44**, D515–D522 (2016).
173. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology* **7**, 74 (2013).

Acknowledgments

This PhD has been the culmination of an adventure that started years ago, when I left my beloved hometown. The desire to pursue my scientific interests has brought me to live and work in several different environments. However, these last four years in Marburg have been hands down the most challenging, educational and fruitful time of my life so far. I would like to thank all the people that accompanied me, in one role or another, in this journey.

First and foremost, I would like to thank Hannes Link for his supervision and guidance during my time in Marburg. I have learnt so much from you, both scientifically and not. I am still convinced (and glad) that four years ago I decided to join the lab, stay and do my best with the project. I wish you all the best for your new upcoming beginning.

I would like to thank my thesis advisory committee members, Tobias Erb and Victor Sourjik. Thank you for taking the time to give me feedback during our yearly meetings and outside of them. I would also like to thank Lennart Randau for being part of my doctoral evaluation committee.

I would like to thank the International Max Planck Research School (IMPRS) for initially funding my PhD project, and the IMPRS Coordinators with whom I worked (with pleasure) as student representative, Zrinka Gattin and Dušica Radoš.

The people of AG Link played a fundamental role in providing a stimulating and fun environment during these years. I would like to thank my two buddies, Timur and Martin, aka Timurone and Martino. I could not have wished for better PhD companions, and I am so lucky to have found such great colleagues in the lab and friends outside of work. A special mention for Vanessa, my first student. I hope you have learnt as much as I did, it has been a pleasure to work with you and see you grow as a scientist. Many thanks also to all the other present and past members of the group: Dominik, Ying, Dusica, Thorben, Michelle, Paul, Niklas, Chris, Janhavi, Peter.

A big thank you to Timo Glatter. I won't lie, preparing hundreds of proteomics samples is a quite tedious task, but our conversations made it less boring and I am thankful to

have worked with such a great person and scientist. I would also like to thank all the people working in the institute and especially the staff, which has provided me with an excellent environment in which to pursue science: David, Melissa, Sarah, Manuel, Christina, et al.

When I started my PhD, I would not have been capable of tackling such a challenging project without all my previous teachers and mentors, in particular Mattia, Ruben and Pablo.

Work is important, but having a good time outside of work is as much important. I would like to thank all the people that made my time in Marburg more enjoyable: Manuel, Francisco, Marc, Tarryn, Marcel, Gabo, Simon, Basti, Vikus, Iria, Alberto, Bartek, Seba, Andre, Ilijana, Alex, Jannik, Carolina, Tiago, Vicente et al. I would also like to thank all my friends in Trento and around the world, especially Tiz, Genno, Nick, Ennio, Corra, Pepper, Elly, Babi, Ross, Enrico, et al. Your support has been important despite the distance.

Thank you, Hanna. I am truly blessed that our paths crossed in Marburg. Your support has been fundamental, especially through the hardest times of this PhD.

Ultimo, ma sicuramente non per importanza, il ringraziamento più grande va alla mia famiglia, Liliana, Paolo, Daniele, Jacek et al. Zawsze we mnie wierzyliście i zawsze mnie wspieraliście, pomimo odległości od domu. Vi voglio bene.

Per aspera ad astra!

Declaration of Contributions

The results presented in this work were carried out by me independently without help other than what listed in this declaration.

Characterization of CRISPRi-knockdowns of metabolic genes

Stefano Donati performed strain design, construction, experimental design, experiments, data analysis, wrote the manuscript. Vanessa Pahl performed growth experiments of the arrayed library. Timo Glatter performed proteomics measurements. Hannes Link conceived the study, discussed results and provided funding.

The metabolome buffers CRISPRi-knockdowns of enzymes in E. coli metabolism

Stefano Donati performed strain design, experimental design, experiments with arrayed CRISPRi strains, analysed data, performed FBA analysis and co-wrote the manuscript. Michelle Kuntz performed experiments with the pooled CRISPRi library and co-wrote the manuscript. Vanessa Pahl performed experiments with arrayed CRISPRi strains. Dominik Beuter constructed the pooled CRISPRi library. Michelle Kuntz, Jose Vicente Gomes Filho and Lennart Randau performed Illumina sequencing. Timo Glatter performed proteomics measurements. Hannes Link co-wrote the manuscript, conceived the study, discussed results and provided funding.

Homeostasis of the *Escherichia coli* biosynthetic metabolome across different environments

Stefano Donati performed data analysis and co-wrote the manuscript. Dušica Radoš performed experiments, metabolomics measurements, data analysis and co-wrote the manuscript. Martin Lempp performed data analysis. Hannes Link conceived the study, discussed results and provided funding.

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.

Marburg, 17/07/2020 *Stefano Donati*

_____ _____

Ort, Datum Stefano Donati