# Metabolic Engineering and Systems-Level Analysis of *Escherichia coli* Central Metabolism through Mass Spectrometry

**Dissertation**

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Vanessa Pahl
aus Heidelberg

Tübingen
2023

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

| | |
|---|---|
| Tag der mündlichen Qualifikation: | 22.02.2024 |
| Dekan: | Prof. Dr. Thilo Stehle |
| 1. Berichterstatter/-in: | Prof. Dr. Hannes Link |
| 2. Berichterstatter/-in: | Prof. Dr. Karl Forchhammer |

# Table of Contents

*für meine Familie*

# Abbreviations

| | |
|---|---|
| **aTc** | anhydrotetracycline |
| **ALE** | adaptive laboratory evolution |
| ***argA*** | N-acetylglutamate synthase |
| ***argO*** | arginine exporter |
| ***argR*** | arginine transcriptional regulator |
| **bp** | base pairs |
| **C1** | one carbon |
| **CID** | collision induced dissociation |
| **$CO_2$** | carbon dioxide |
| **CRISPRi** | CRISPR interference |
| ***crp*** | transcriptional dual regulator |
| **dCas9** | deactivated Cas9 |
| ***dnaX*** | DNA polymerase III subunit γ |
| ***E. coli*** | *Escherichia coli* |
| **ESI** | electrospray ionization |
| **f6-p** | fructose 6-phosphate |
| **FACS** | fluorescence-activated cell sorting |
| **FI-MS** | flow injection mass spectrometry |
| **FI-TOF** | flow injection time of flight mass spectrometry |
| **g6-p** | glucose 6-phosphate |
| **HILIC** | hydrophilic interaction liquid chromatography |
| **HPLC** | high-pressure liquid chromatography |
| **kDa** | kilo Dalton |
| **LC-MS/MS** | liquid chromatography coupled to tandem mass spectrometry |
| **MEP pathway** | methylerythritol phosphate pathways |
| **MRM** | multiple reaction monitoring |
| **MS** | mass spectrometry |
| **msec** | milli seconds |
| **NAD** | nicotinamide adenine dinucleotide |
| **NADH** | nicotinamide adenine dinucleotide reduced |
| **NGS** | next-generation sequencing |
| ***pgi*** | phosphoglucoisomerase |
| **PPP** | pentose phosphate pathway |
| **QqQ** | triple quadrupole instrument |
| **r5-p** | ribose 5-phosphate |
| ***rpoB*** | RNA polymerase subunit β |
| **rPP cycle** | reductive pentose phosphate cycle |

| | |
|---|---|
| **ru5-p** | ribulose 5-phosphate |
| **sgRNA** | single guide RNA |
| **TS** | temperature-sensitive |
| **ubi** | ubiquitin |

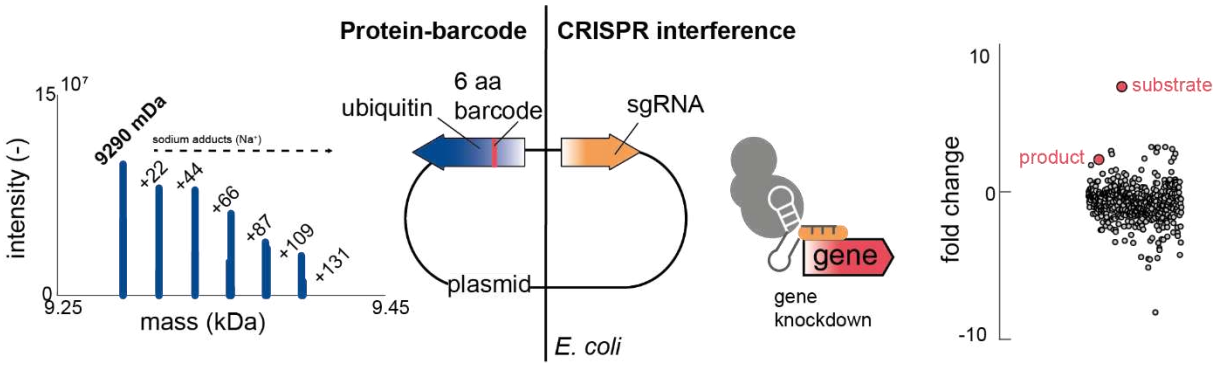
# Summary

Metabolism is the core of what we consider to be a living cell. It covers all chemical reactions that are necessary to break down nutrients and convert them into energy and cellular building blocks for growth. *Escherichia coli* is arguably the most studied model organism. Traits that make *E. coli* especially suitable for genetic engineering and convenient to study bacterial metabolism include: rapid growth, high yields, cost effectiveness, genetic accessibility and robust handling. To study metabolism, metabolomics serves as a valuable tool within the field of systems biology, allowing researchers to explore concentration changes of metabolites and metabolic fluxes. Here, I show in four structured chapters recent advances in mass spectrometry-based metabolomics to engineer and investigate *E. coli* central metabolism on a systems-level.

In **Chapter 1**, I introduce a novel protein-based barcoding system designed for tracking and phenotyping CRISPR interference (CRISPRi) strains in real-time. Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems have become the gold standard for RNA-guided gene editing, facilitating powerful CRISPR screens for unbiased gene function investigations. The CRISPR interference (CRISPRi) system, a versatile tool for gene knockdown, relies on a catalytically inactive Cas9 protein (dCas9) that forms a complex with a single guide RNA (sgRNA). This complex can bind to specific target regions, blocking transcription initiation, transcriptional elongation, or transcription factor binding. Importantly, gene knockdown with CRISPRi is both inducible and reversible, enabling the perturbation and analysis of essential genes. To target complete genomes, CRISPRi libraries have been designed, comprising thousands of strains. However, simultaneous tracking and phenotyping of these strains in high-throughput screens has been limited, primarily due to the need to sequence the sgRNA. Using the sgRNA as barcode has several drawbacks, including sequencing cost, sample preparation, and incompatibility with other screening methods like proteomics and metabolomics. In this chapter, I present an alternative method for barcoding CRISPRi strains using a small protein, ubiquitin. Ubiquitin is expressed with a six-amino-acid *N*-terminal barcode from the same plasmid as the sgRNA. By combining top-down proteomics and metabolomics

techniques, I achieved the simultaneous tracking and phenotyping of seven ubiquitin-barcoded CRISPRi strains, which targeted metabolic enzymes in central metabolism. The use of flow injection mass spectrometry (FI-MS) allowed for the detection of native barcode proteoforms and the annotation of 157 metabolites in just a 30-second run. Analysis of metabolites revealed metabolic signature of CRISPRi strains with local perturbations affecting single pathways. Additionally, I demonstrated real-time detection of protein barcodes by injecting living cells into the mass spectrometer. This innovation opens up new possibilities for rapid characterization of the metabolome within CRISPRi libraries, promising significant advancements in multi-phenotyping CRISPRi screens with potential applications across various organisms.

## Synopsis

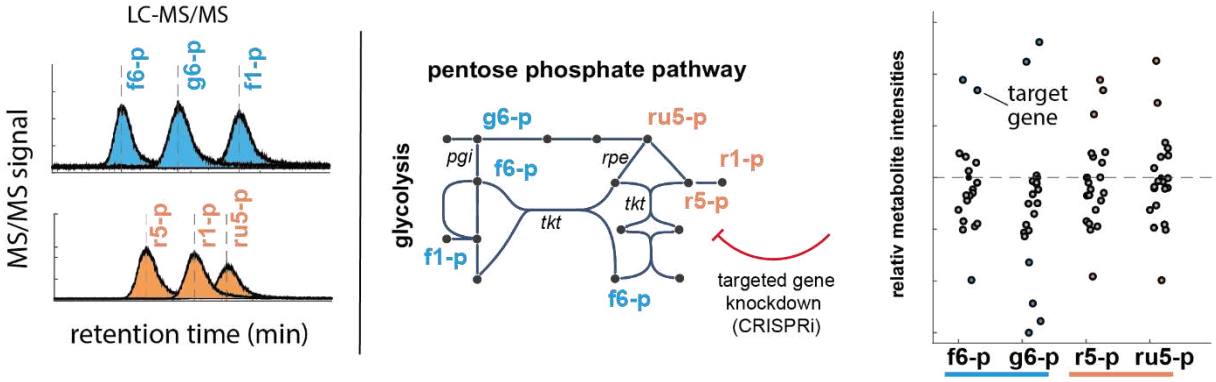



**Protein barcode for CRISPR interference strains.** Protein-barcode is co-expressed from the same plasmid as the CRISPR interference system (sgRNA). Flow injection mass spectrometry enabled simultaneous detection of protein-barcode and metabolic profile.

- Novel barcoding system for CRISPRi strains with small protein (ubiquitin)
- Flow injection mass spectrometry enable tandem-readout of barcode and metabolic fingerprint
- Time-efficient workflow for ubiquitin-barcode identification using top-down proteomics methods (spectral deconvolution)
- Real-time detection of protein barcodes

In **Chapter 2**, I present a method for the quantification of sugar phosphate isomers within the central metabolism of *E. coli*. Core pathways of *E. coli* central metabolism, the Pentose Phosphate Pathway (PPP) and glycolysis, have been conserved across various organisms throughout evolution, making them subjects of significant interest. One of the challenges in studying these pathways is the detection of hexose phosphates and pentose phosphates, which are crucial intermediates but often present as a pool due to their similar masses and chemical properties. Furthermore, quantifying metabolites in the PPP, poses difficulties because of the reversibility and interconnectedness of reactions, leading to low abundance due to high turnover. In this study, I use LC-MS/MS, employing reverse-phase chromatography with extended gradients to effectively separate and quantify three hexose phosphates (glucose 6-p, fructose 6-p, fructose 1-p) and three pentose phosphates (ribose 1-p, ribose 5-p, ribulose 5-p) based on retention time. To perturb metabolite levels *in vivo*, I applied CRISPR interference for targeted knockdown of 21 enzymes involved in PPP and glycolysis. The results of *in vivo* metabolite quantification revealed significant accumulations of sugar phosphates in particular strains, notably those affecting the genes *pgi*, *tktA*, and *rpe*. Knockdown of Phosphoglucoisomerase (Pgi), a key enzyme in glycolysis at the branchpoint to PPP, revealed an inversion of product and substrate, indicating a directional shift in reaction during growth with different carbon sources (xylose vs. glucose). The study also included an analysis of the two transketolases, TktA and TktB, which catalyse distinct reactions within the core of the oxidative PPP. It was observed that only the knockdown of TktA resulted in the accumulation of r5-p, its direct substrate, and ru5-p, an upstream metabolite, while the knockdown of its isoenzyme, Transketolase 2 (TktB), had no significant impact on metabolite levels. Additionally, the analysis of the *rpe* strain revealed the activation of a bypass reaction. Targeted metabolomics showed that knockdown of most targets in PPP and glycolysis did not perturb sugar-phosphate levels, highlighting the metabolic robustness of these pathways. This research provides valuable insights into the coordination of microbial metabolism through the characterization of three key enzymes. Understanding the central metabolism of *E. coli* is essential for various fields, and the methodology developed here can be applied to a wide range of organisms and pathways, such as cyanobacteria and the Calvin cycle, which overlap with the PPP.

## Synopsis




**Quantification and separation of sugar phosphates in *E. coli* using LC-MS/MS.**

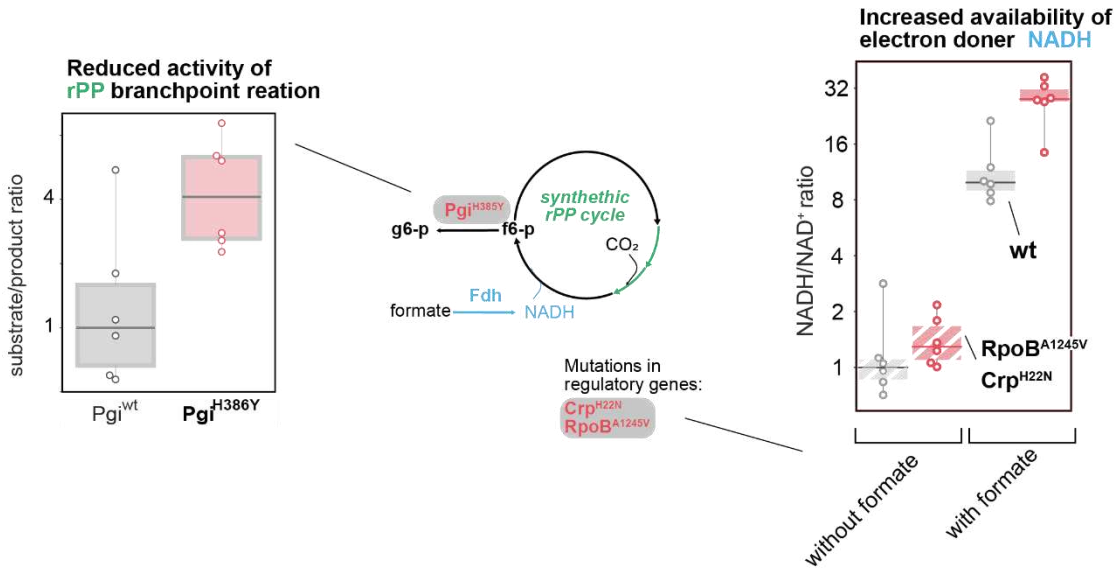
Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) achieved separation and quantification of sugar phosphates without the need for additional reagents (ion pairing). Perturbation of hexose and pentose phosphates in central metabolism using targeted gene knockdown. Metabolomics methods reveal insights into key enzymes (Pgi, Rpe, TktA) in PPP and glycolysis.

- Intracellular quantification of hexose and pentose phosphates using targeted metabolomics
- Hexose and pentose phosphates can be separated using chromatography without additional reagents
- Perturbation of hexose and pentose phosphates achieved through targeted gene knockdown (CRISPR interference) of enzymes in *E. coli* central metabolism
- CRISPRi-induced accumulations of hexose and pentose phosphates show, particularly when metabolites acted as substrates in reactions
- Potential for broad applicability across various organisms and pathways due to the conservation of central metabolism

In **Chapter 3**, I study the metabolic adaptations enabling *E. coli* to transition to chemoautotrophic growth, focusing on the role of autotrophic-enabling mutations and their impact on energy and carbon metabolism. Synthetic autotrophy holds great promise for microbial conversion of $CO_2$ into valuable compounds. *E. coli* has been successfully engineered to rely on $CO_2$ for biomass and formate for energy generation. To achieve this, the introduction of three key enzymes (Rubisco, Phosphoribulokinase,

Carbonic anhydrase) was sufficient to rebuild a reductive pentose phosphate pathway (rPP) for carbon fixation. The energy module converts formate into NADH and $CO_2$, catalysed by a heterologous formate dehydrogenase. The initial rational design of chemoautotrophy, solely inserting heterologous enzymes, proved unsuccessful. Instead, a process of iterative laboratory evolution led to the identification of three autotrophic-enabling mutations ($Pgi^{H385Y}$, $Crp^{H22N}$, $RpoB^{A1245V}$). Presumably the mutations tune metabolism, however their exact role is yet to show. In this study, I use metabolomics methods to investigate the impact of these mutations on energy and carbon metabolism in chemoautotrophic *E. coli*. My findings reveal insights, including the role of Phosphoglucoisomerase (Pgi), catalysing a branchpoint reaction to the rPP cycle. By measuring substrate and product levels, I observed a reduction in enzyme activity associated with the mutation. Consequently, intermediates within the carbon fixation cycle accumulated, indicating a constriction of efflux from the cycle. This suggests that the mutation functions as a stabilizing factor for the carbon-fixing cycle by limiting branching flux. Furthermore, I explore the roles of mutations in global regulators Crp and RpoB. Notably, the double mutant exhibited increased NADH/NAD+ ratios. NADH is important electron donors of the rPP cycle suggesting adaptations in energy metabolism during the trophic transition. The use of formate for generation of NADH generates $CO_2$ as by-product. In order to be fully sustainable, formate can be produced electrochemically. However, this strain already serves as a versatile platform for understanding and engineering the rPP cycle. Further, understanding these rewired metabolisms holds the potential to harness *E. coli* as a robust platform for sustainable carbon conversion and has implications for synthetic autotrophy, extending to natural autotrophs, cell-free systems, and the broader field of sustainable carbon conversion.

## Synopsis



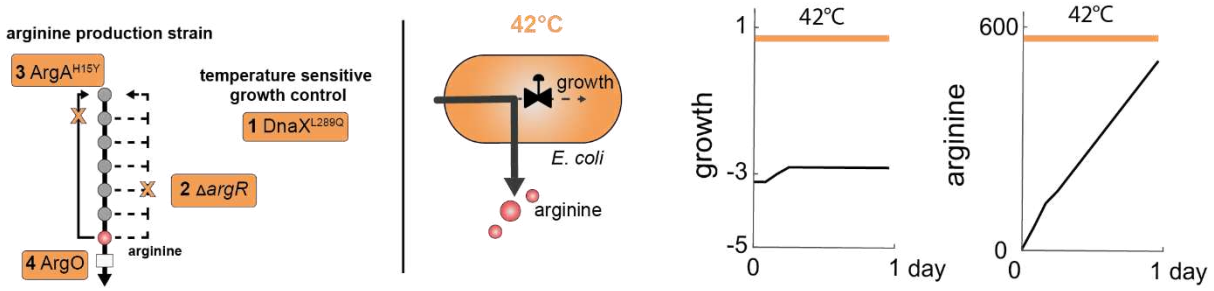

**Autotrophic enabling mutations affect energy and carbon metabolism in chemoautotrophic *E. coli*.** Relative quantification of intracellular metabolites to understand the role autotrophic-enabling mutations (PgiH$^{385Y}$, Crp$^{H22N}$, RpoB$^{A1245V}$) for metabolism.

- A mutation in branch point reaction (catalysed by Pgi) reduced efflux from carbon fixation cycle
- Mutations in regulatory genes (*crp*, *rpoB*) increased the abundance of the electron donor (NADH)
- Chemoautotrophic conversion of *E. coli* requires metabolic adaptations in both carbon and energy metabolism.

**Chapter 4** presents a temperature-controlled two-stage process designed to enhance arginine production in *E. coli*, by decoupling of growth and overproduction through temperature switches. Two-stage bioprocesses are widely employed to improve chemical production in microbes by separating growth (stage 1) and production (stage 2). Since many bioreactors have temperature control capabilities, temperature sensitive valves seem ideal to induce production. In this study, arginine production is enhanced in *E. coli* by utilizing a strain with specific mutations that disable transcriptional and allosteric regulation of the pathway (Δ*argR* and ArgA$^{H15Y}$). An additional version of an arginine exporter (ArgO) was introduced. To decouple growth from production, a temperature-sensitive DnaX$^{L289Q}$ mutation was inserted. The protein

remains intact at 30°C but loses enzyme activity at 42°C, leading to growth arrest. Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) was used for absolute quantification arginine levels throughout cultivations at 42°C. Insertion of the temperature sensitive mutation DnaX$^{L289Q}$ robustly stopped growth at 42°C, showcasing the effectiveness of the temperature-sensitive mutation for induced growth arrest. During the initial 6 h of the growth arrest, the arginine production rate was constant at 1.35 mmol gDW$^{-1}$ h$^{-1}$, showing its potential for large-scale arginine production. Furthermore, cultivation at lower temperatures (30°C) could rescue arginine production after 24 h of growth arrest, indicating reversibility of the temperature switch. This research provides a proof of concept for coupling temperature-sensitive mutations with two-stage processes to produce amino acids and other valuable compounds in *E. coli*. It sets the stage for further exploration and development to control microbial bioproduction.

## Synopsis




**Temperature controlled two-stage arginine production in *E. coli*.** The two-stage arginine production strain contained four modifications: (1) a point mutation DnaX$^{L289Q}$, (2) a gene deletion Δ*argR*, (3) a point mutation ArgA$^{H15Y}$ and (4) an additional copy of ArgO. Cultivating cells at 42°C achieved robust growth arrest while maintaining arginine production for 24 h. This research provides a proof of concept for decoupling growth from production using a temperature-sensitive mutation.

- Decoupling of growth from arginine overproduction with temperature switch
- Cultivation at high temperatures achieved robust growth arrest
- Arginine production maintained for 24 h
- Cultivation at lower temperatures rescued arginine production after long-term growth arrest, indicating reversibility

In this work, I explored innovative approaches in metabolic engineering and metabolomics to advance our understanding of bacterial metabolism, using the model organism *E. coli*. I introduced a novel protein-based barcode system for CRISPR interference strains, enabling real-time tracking and phenotyping while avoiding the need for sequencing sgRNAs. This method offers a cost-effective solution for simultaneous tracking and phenotyping of CRISPRi Libraries. Additionally, I developed a liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) approach to quantify sugar phosphate isomers in *E. coli*, providing insights into the central metabolism. By perturbing metabolic pathways through CRISPR interference, I revealed dynamic changes in hexose and pentose phosphates, underscoring the potential of this method for various organisms. Further, I investigated the metabolic adaptations of *E. coli* during its transition to chemoautotrophy, shedding light on the roles of autotrophic-enabling mutations in energy and carbon metabolism. This has implications for the conversion of $CO_2$ into valuable compounds in sustainable bioprocesses. Finally, I introduced a temperature-controlled two-stage process to produce arginine in *E. coli*, demonstrating the decoupling of growth and overproduction using a temperature switch. This strategy holds promise for microbial production of amino acids and valuable compounds.

Together, this work contributes to our understanding of cellular metabolism, offering innovative tools and methodologies for metabolic engineering, metabolomics, and the sustainable production of valuable compounds.

# Zusammenfassung der Arbeit

Der Stoffwechsel ist das Herz dessen, was wir als eine lebende Zelle betrachten. Er umfasst alle chemischen Reaktionen, die notwendig sind, um Nährstoffe abzubauen und sie in Energie und zelluläre Bausteine für das Wachstum umzuwandeln. *Escherichia coli* ist wohl das am meisten erforschte Modellorganismus. Merkmale, die *E. coli* besonders geeignet für die Genetik und die Untersuchung des bakteriellen Stoffwechsels machen, sind: schnelles Wachstum, hohe Ausbeute, Kosteneffizienz, genetische Zugänglichkeit und robuste Handhabung. Um den Stoffwechsel zu erforschen, dient Metabolomik als wertvolles Werkzeug im Bereich der Systembiologie und ermöglicht es Forschern, Konzentrationsänderungen von Metaboliten und Reaktionsraten zu untersuchen. In dieser Arbeit zeige ich in vier strukturierten Kapiteln aktuelle Fortschritte in der Massenspektrometrie basierenden Metabolomik zur Manipulation und Untersuchung des zentralen Stoffwechsels von *E. coli* auf systemischer Ebene.

Im **ersten Kapitel** stelle ich ein neuartiges proteinbasiertes Barcodesystem vor, das für das Nachverfolgen und die Phänotypisierung von CRISPR-Interference (CRISPRi)-Stämmen in Echtzeit entwickelt wurde. „Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated“ (Cas)-Systeme sind zum Goldstandard für RNA-geführte Genbearbeitung geworden und ermöglichen einen hohen Durchsatz durch CRISPR-Screenings für die unvoreingenommene Erforschung von Genfunktionen. Das CRISPR-Interference (CRISPRi)-System, ein vielseitiges Werkzeug zur Genunterdrückung, beruht auf einem katalytisch inaktiven Cas9-Protein (dCas9), das mit einer sogenannten „single guide RNA“ (sgRNA) einen Komplex bildet. Dieser Komplex kann an spezifische Zielregionen binden und Proteinsynthese oder die Bindung von Transkriptionsfaktoren blockieren. Die Genunterdrückung mit CRISPRi ist sowohl induzierbar als auch reversibel und ermöglicht damit die Störung und Analyse essentieller Gene. Zur Untersuchung ganzer Genome wurden CRISPRi-Bibliotheken entwickelt, die Tausende von Stämmen umfassen. Die gleichzeitige Verfolgung und Phänotypisierung dieser Stämme in Hochdurchsatz-Screenings war bis jetzt jedoch begrenzt, hauptsächlich aufgrund der Notwendigkeit, die sgRNA zu sequenzieren. Die Verwendung der sgRNA als Barcode hat mehrere Nachteile,

einschließlich Sequenzierungskosten, Probenpräparation und Inkompatibilität mit anderen Screening-Methoden wie Proteomik und Metabolomik. In diesem Kapitel stelle ich eine alternative Methode zur Markierung von CRISPRi-Stämmen vor, bei der ein kleines Protein, Ubiquitin, verwendet wird. Ubiquitin wird mit einem sechsaminosäure *N*-terminalen Barcode aus demselben Plasmid wie die sgRNA exprimiert. Durch die Kombination von Top-Down-Proteomics und Metabolomik-Techniken gelang es mir, die gleichzeitige Verfolgung und Phänotypisierung von sieben mit Ubiquitin markierten CRISPRi-Stämmen zu erreichen, die auf metabolische Enzyme im zentralen Stoffwechsel abzielten. Die Verwendung von Flow injection Massenspektrometrie (FI-MS) ermöglichte die Erkennung nativer Barcode-Proteoformen und die Annotation von 157 Metaboliten in nur einem 30-Sekunden-Lauf. Die Analyse der Metabolite ergab metabolische Signaturen von CRISPRi-Stämmen mit lokalen Störungen, die nur einzelne Soffwechselwege betrafen. Darüber hinaus zeige ich die Echtzeit-Erkennung von Proteinbarcodes durch die Injektion lebender Zellen in das Massenspektrometer. Diese Innovation eröffnet neue Möglichkeiten für die schnelle Charakterisierung des Metaboloms innerhalb von CRISPRi-Bibliotheken und verspricht bedeutende Fortschritte bei multiplen Phänotypisierungs-CRISPRi-Screenings mit Anwendungen in verschiedenen Organismen.

Im **zweiten Kapitel** präsentiere ich eine Methode zur Quantifizierung von Zuckerphosphat-Isomeren im zentralen Stoffwechsel. Die Kernwege des zentralen Stoffwechsels *von E. coli*, der Pentosephosphatweg (PPP) und die Glykolyse, wurden im Verlauf der Evolution in verschiedenen Organismen konserviert, was sie zu Gegenständen von erheblichem Interesse macht. Eine der Herausforderungen bei der Untersuchung dieser Wege besteht in der Erkennung von Hexosephosphaten und Pentosephosphaten, die entscheidenden Zwischenprodukte sind, aber aufgrund ihrer ähnlichen Masse und chemischen Eigenschaften oft zusammen detektiert werden. Die Quantifizierung von Metaboliten im PPP ist aufgrund der Reversibilität und der Vernetzung von Reaktionen, die zu einer niedrigen Abundanz aufgrund hoher Umsatzraten führen, zusätzlich schwierig. In dieser Studie verwende ich LC-MS/MS mit einer umgekehrten Phasenchromatographie und erweiterten Gradienten, um drei Hexosephosphate (Glucose 6-p, Fructose 6-p, Fructose 1-p) und drei

Pentosephosphate (Ribose 1-p, Ribose 5-p, Ribulose 5-p) basierend auf der Retentionszeit effektiv zu trennen und zu quantifizieren. Um Metabolit Konzentrationen *in vivo* zu verändern, wandte ich CRISPR-Interferenz zur gezielten Unterdrückung von 21 Enzymen an, die am PPP und der Glykolyse beteiligt sind. Die Ergebnisse der *in vivo*-Metabolit-Quantifizierung zeigten signifikante Anhäufungen von Zuckerphosphaten in bestimmten Stämmen, insbesondere solchen, die die Gene *pgi*, *tktA* und *rpe* beeinflussten. Die Unterdrückung der Phosphoglucoisomerase (Pgi), einem Schlüsselenzym in der Glykolyse an der Verzweigungsstelle zum PPP, zeigte eine Umkehrung von Produkt und Substrat, was auf eine Richtungsänderung in der Reaktion während des Wachstums mit unterschiedlichen Kohlenstoffquellen (Xylose vs. Glucose) hinweist. Die Studie umfasste auch eine Analyse der beiden Transketolasen, TktA und TktB, die gleich zwei unterschiedliche Reaktionen im Kern des oxidativen PPP katalysieren. Es wurde beobachtet, dass nur die Unterdrückung von TktA zu einer Anhäufung von r5-p, seinem direkten Substrat, und ru5-p, einem vorgelagerten Metaboliten, führte, während die Unterdrückung seines Isoenzyms, Transketolase 2 (TktB), keine signifikanten Auswirkungen auf die Metabolit-Level hatte. Darüber hinaus zeigte die Analyse des *rpe*-Stamms die Aktivierung einer Bypass-Reaktion. Die gezielte Metabolomik zeigte, dass die Unterdrückung der meisten Ziele im PPP und in der Glykolyse die Zuckerphosphat-Level nicht beeinträchtigte, was die metabolische Robustheit dieser Wege unterstreicht. Diese Forschung bietet wertvolle Einblicke in die Koordination des mikrobiellen Stoffwechsels Charakterisierung von drei Schlüsselenzymen. Das Verständnis des zentralen Stoffwechsels von *E. coli* ist für verschiedene Bereiche unerlässlich, und die hier entwickelte Methodik kann auf eine Vielzahl von Organismen und Wegen angewendet werden, wie zum Beispiel Cyanobakterien und den Calvin-Zyklus, die mit dem PPP überlappen.

Im **dritten Kapitel** untersuche ich die metabolischen Anpassungen, die es *E. coli* ermöglicht, auf chemolithoautotrophes Wachstum umzustellen, wobei ich mich auf die Rolle Autotrophie ermöglichender Mutationen und ihre Auswirkungen auf Energie- und Kohlenstoffstoffwechsel konzentriere. Die synthetische Autotrophie birgt großes Potenzial für die mikrobielle Umwandlung von $CO_2$ in wertvolle Verbindungen. Es wurde erfolgreich gezeigt, dass sich *E. coli* ausschließlich auf $CO_2$ für Biomasse und

Formiat für die Energieerzeugung verlässt. Hierfür reichten die Einführung von drei Schlüsselenzymen (Rubisco, Phosphoribulokinase, Carbonanhydrase) aus, um einen reduktiven Pentosephosphatweg (rPP) für die Kohlenstofffixierung aufzubauen. Das Energie-Modul wandelt Formiat in NADH und $CO_2$ um, katalysiert durch eine heterologe Formiatdehydrogenase. Die ursprüngliche rationale Gestaltung der Chemoautotrophie, die nur die Einführung heterologer Enzyme vorsah, war jedoch erfolglos. Stattdessen führte ein Prozess der wiederholten Labor-Evolution zur Identifizierung von drei Autotrophie ermöglichenden Mutationen ($Pgi^{H385Y}$, $Crp^{H22N}$, $RpoB^{A1245V}$). Vermutlich regulieren diese Mutationen den Stoffwechsel, jedoch ist ihre genaue Rolle noch zu klären. In dieser Studie verwende ich metabolomische Methoden, um die Auswirkungen dieser Mutationen auf den Energie- und Kohlenstoffstoffwechsel in einem chemoautotrophem *E. coli* zu untersuchen. Meine Ergebnisse geben Einblicke, in die Rolle der Phosphoglucoisomerase (Pgi), die eine Verzweigungsreaktion zum rPP-Zyklus katalysiert. Durch die Messung von Substrat- und Produktkonzentrationen habe ich eine Reduktion der Enzymaktivität in Verbindung mit der Mutation beobachtet. Folglich häuften sich Zwischenprodukte im Kohlenstofffixierungszyklus an, was auf eine Begrenzung des Ausflusses aus dem Zyklus hinweist. Dies deutet darauf hin, dass die Mutation eine stabilisierende Funktion für den kohlenstofffixierenden Zyklus hat, indem sie den Verzweigungsfluss begrenzt. Darüber hinaus erforsche ich die Rollen von Mutationen in den globalen Regulatoren Crp und RpoB. Bemerkenswerterweise zeigte eine Doppelmutante erhöhte $NADH/NAD^+$-Verhältnisse. NADH ist der wichtigste Elektronendonator des rPP-Zyklus, was auf Anpassungen im Energiestoffwechsel während des Trophieübergangs hinweist. Die Verwendung von Formiat zur Erzeugung von NADH erzeugt $CO_2$ als Nebenprodukt. Um vollständig nachhaltig zu sein, kann Formiat elektrochemisch hergestellt werden. Dennoch dient dieser Stamm bereits als vielseitige Plattform zur Erforschung und Entwicklung des rPP-Zyklus. Ein weiteres Verständnis dieser umgebauten Stoffwechselwege birgt das Potenzial, *E. coli* als robuste Plattform für die nachhaltige Kohlenstoffumwandlung zu nutzen und hat Auswirkungen auf die synthetische Autotrophie, die auf natürliche Autotrophen, zellfreie Systeme und das breitere Feld der nachhaltigen Kohlenstoffumwandlung ausgedehnt werden kann.

Das **vierte Kapitel** stellt einen temperaturgesteuerten zweistufigen Prozess vor, der darauf abzielt, die Argininproduktion in *E. coli* durch die Entkopplung von Wachstum und Überproduktion mittels Temperaturschaltern zu erhöhen. Zweistufige Bioprozesse werden weit verbreitet eingesetzt, um die chemische Produktion in Mikroorganismen durch die Trennung von Wachstum (Stufe 1) und Produktion (Stufe 2) zu verbessern. Da viele Bioreaktoren über Temperaturregelungsmöglichkeiten verfügen, scheinen temperatursensitive Ventile ideal zu sein, um die Produktion auszulösen. In dieser Studie wird die Argininproduktion in *E. coli* durch die Verwendung eines Stammes mit spezifischen Mutationen erhöht. Die transkriptionelle und allosterische Regulation des Stoffwechselweges (Δ*argR* und $ArgA^{H15Y}$) wurden ausschalten, um erhöhte Arginin Endproduktion zu ermöglichen. Eine zusätzliche Version eines Argininexporters (ArgO) wurde eingeführt. Um das Wachstum von der Produktion zu entkoppeln, wurde eine temperatursensitive $DnaX^{L289Q}$-Mutation eingefügt. Das Protein bleibt bei 30°C intakt, verliert jedoch bei 42°C seine Enzymaktivität, was zum Wachstumsstopp führt. Flüssigchromatographie gekoppelt mit Tandem-Massenspektrometrie (LC-MS/MS) wurde zur absoluten Quantifizierung von Arginin während der Kultivierung bei 42°C verwendet. Die Einführung der temperatursensitiven Mutation $DnaX^{L289Q}$ stoppte das Wachstum bei 42°C robust, was die Effektivität der temperatursensitiven Mutation für den induzierten Wachstumsstopps betont. Während der ersten 6 Stunden des Wachstumsstillstands war, die Argininproduktionsrate konstant bei 1,35 mmol $gDW^{-1}h^{-1}$, was ihr Potenzial für die groß angelegte Argininproduktion zeigt. Darüber hinaus konnte die Kultivierung bei niedrigeren Temperaturen (30°C) die Argininproduktion nach 24 Stunden des Wachstumsstillstands wiederherstellen, was auf die Umkehrbarkeit des Temperaturschalters hinweist. Diese Forschung liefert einen Nachweis für die Kopplung temperatursensitiver Mutationen mit zweistufigen Prozessen für die Produktion von Aminosäuren und anderen wertvollen Verbindungen in *E. coli*. Sie bildet die Grundlage für weitere Entwicklungen zur Kontrolle von Mikrobieller Produktion.

In dieser Arbeit habe ich innovative Ansätze in der Gentechnik und der Metabolomik erforscht, um unser Verständnis des bakteriellen Stoffwechsels mithilfe des Modellorganismus *E. coli* voranzutreiben. Ich habe ein neuartiges proteinbasiertes Barcodesystem für CRISPR-Interferenzstämme eingeführt, dass die

Echtzeitverfolgung und Phänotypisierung ermöglicht, ohne dass die Notwendigkeit besteht, sgRNAs zu sequenzieren. Diese Methode bietet eine kostengünstige Lösung für die gleichzeitige Verfolgung und Phänotypisierung von CRISPRi-Bibliotheken. Darüber hinaus habe ich eine Flüssigchromatographie gekoppelt mit Tandem-Massenspektrometrie (LC-MS/MS)-Methode entwickelt, um Zuckerphosphat-Isomere in *E. coli* zu quantifizieren, was Einblicke in den zentralen Stoffwechsel bietet. Durch die Störung von Stoffwechselwegen durch CRISPR-Interference habe ich dynamische Veränderungen in Hexose- und Pentosephosphaten aufgezeigt, was das Potenzial dieser Methode für verschiedene Organismen unterstreicht. Des Weiteren habe ich die metabolischen Anpassungen von *E. coli* während des Übergangs zur Chemoautotrophie untersucht und dabei die Rolle von Autotrophie ermöglichenden Mutationen im Energie- und Kohlenstoffstoffwechsel beleuchtet. Dies hat Auswirkungen auf die Umwandlung von $CO_2$ in wertvolle Verbindungen in nachhaltigen Bioprozessen. Schließlich habe ich einen temperaturgesteuerten zweistufigen Prozess zur Argininproduktion in *E. coli* eingeführt und damit die Entkopplung von Wachstum und Überproduktion mittels eines Temperaturschalters demonstriert. Diese Strategie birgt Potenzial für die mikrobielle Produktion von Aminosäuren und wertvollen Verbindungen.

Zusammenfassend trägt diese Arbeit zu unserem Verständnis des zellulären Stoffwechsels bei und bietet innovative Werkzeuge und Methoden für Systembiologie, Metabolomik und die nachhaltige Produktion wertvoller Verbindungen.

# Introduction

*Escherichia coli* is arguably the most studied model organism. Traits that make *E. coli* especially suitable for genetic engineering and convenient to study bacterial metabolism include: rapid growth, high yields, cost effectiveness, genetic accessibility and robust handling. Metabolomics serves as a valuable tool within the field of systems biology, allowing researchers to explore concentration changes of metabolites and metabolic fluxes.

## 1 Mass spectrometry enables metabolomics in *E. coli*

Bacterial metabolism is built on a complex network of pathways. The latest genome-scale model of *E. coli* accounts for 1,192 metabolites (Monk et al. 2017). This is one order of magnitude less compared to reported metabolites in humans (> $10^4$, Wishart et al. 2007) and even more metabolites were estimated for plants (~$10^4$ – $10^5$, Alseekh and Fernie 2018). Despite differences in the number of metabolites, humans, plants and bacteria share key metabolites that involve essential processes such as conversion of nutrients into energy, providing the building blocks for biomass. Metabolites can change within seconds and low quantities are inside each cell (~32 µM, Palsson and Abrams 2011). Mass spectrometry (MS) is the gold standard for the fast identification of a large number of different metabolites (Markley et al. 2017). Chemical diversity, rapid turnover and the broad abundance range currently prohibit measuring all metabolites in a single extraction or a single analysis. Consequently, many different extraction techniques and combinations of analytical methods have been developed in an attempt to achieve adequate metabolite coverage. Comparing mass spectrometry approaches, the main differences are: (i) chromatographic separation, (ii) ion source and (iii) the choice of mass analyser (Glish and Vachet 2003).

The chromatographic separation can occur prior to MS analysis, within the metabolomics workflow. Special separation of metabolites increases the selectivity and decreases negative effects like ion suppression and space charge effects (Alseekh et al. 2021). No single chromatographic method is suitable for all metabolites. Liquid chromatography (LC) provides the best combination of both versatility and robustness

compared to gas chromatography (GC), and capillary electrophoresis (CE) (Buescher et al. 2009). High pressure liquid chromatography is an advanced type of LC and has largely replaced most traditional LC methods. Grouping metabolites largely into polar and nonpolar (Lu et al. 2017), high-liquid chromatography (HPLC) is suitable for the detection of both and many different HPLC- columns and methods have been developed to cover a large variety of metabolites. Hydrophilic interaction liquid chromatography (HILIC) columns are often used to analyse highly polar metabolites (Tang et al. 2016). Unpolar metabolites and proteins are commonly detected using reverse-phase chromatography (C8 – C18 columns).

MS enables identification and quantification of molecules based on their mass to charge ratio (*m/z*). Ions are separated under vacuum by applying an electromagnetic field. Therefore, before entering the mass analyser metabolites need to be ionised. Most MS methods for metabolomics rely on electrospray ionization (ESI), which works effectively for charged metabolites. In ESI, a liquid sample is introduced into the mass spectrometer via a fine needle (Fig 1A, Fig 2B). The nozzle is typically charged with high voltages (> 2kV) to form fine aerosols of charged droplets. Depending on the polarity of the of the applied voltage (mode) ESI can result in positively or negatively charged ions. Protonation $[M+H]^+$ and deprotonation $[M-H]^-$ are thereby the main source of charging. Proteins, peptides, or nucleotides are often found with several sites of protonation. Multiple charging enables mass spectrometers to analyse higher-molecular-weight molecules within a lower *m/z* range. For example, cytochrome c, is a protein with a molecular mass of 12,360 Da. The addition of 10 protons shifts the *m/z* ratio of the proteoform below 2,000, which is in the working range of most mass analysers. Since analytes are kept intact during ionisation, ESI is considered a soft ionisation mode (Glish and Vachet 2003). ESI is commonly used in combination with liquid chromatography. Taken together, ESI is particularly suited for analysing polar and high-molecular-weight compounds, such as peptides, proteins, nucleic acids, and complex organic molecules as relevant in *E. coli* extracts.

**High throughput metabolomics using TOF-MS**

To perform metabolomics methods, I used ESI in two different instruments that rely on different mass analysers (Fig 1, 2). Time of flight mass spectrometry (TOF-MS), is

arguably the simplest mass analyser (Fig 1B). Essentially it measures the time of an ion to pass through a tube (typically 0.5-2 meters) to reach the detector. Based on the velocity the *m/z* of the ion can be determined. The instruments excel due to a high mass accuracy (<0.001 amu) even in the absence of chromatographic steps (Fuhrer et al. 2011). Omitting chromatographic steps can decrease the measurement time effectively. Flow injection time of flight mass spectrometry directly injects samples and thereby effectively decreases measurement time down to 15-30 sec per sample (Fuhrer et al. 2011, Farke et al. 2023). Using a TOF mass analyser enables to detect a broad range of masses (~50 -1700 *m/z*). This enables the detection of metabolites (smaller mass range) but also proteins (higher mass range). ESI-TOF-MS is commonly used in top-down proteomics for the detection of intact proteins and protein species (Chen et al. 2018). One shortcoming of TOF-MS is its susceptibility to ion suppression effects. When solutions contain high salt concentrations (that is > 1 mM), analyte ion formation is usually hindered, which means most biological samples should be desalted before analysis. In addition, when complex mixtures of compounds are present, the higher-concentration analytes can suppress ion formation by lower-concentration analytes, leading to less accurate quantification.

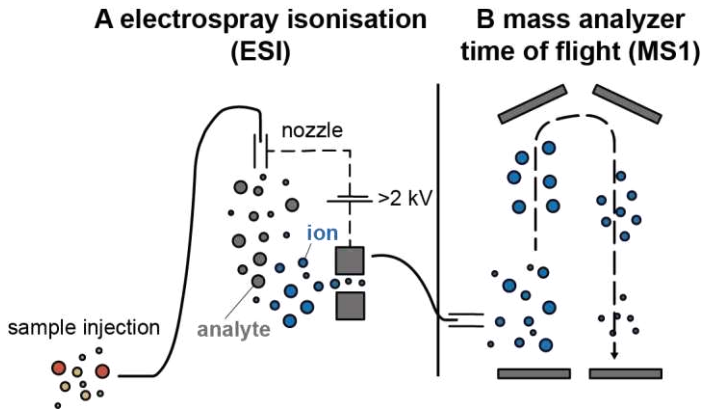



**Figure 1 Technical configurations of time-of-flight mass spectrometer (TOF-MS).**
**A** Electrospray ionisation (ESI). Sample is injected to mass spectrometer e.g. using a multisampler. In ESI the analyte is introduced via a fine needle. The nozzle is typically charged with high voltages (> 2kV) to form fine aerosols of charged droplets. Depending on the polarity of the applied voltage (mode) ESI can result in positively $[M+H]^+$ or negatively charged ions $[M-H]^-$ (Famiglini et al. 2021). **B** Time of flight mass analyser essentially measures the time of an ion to pass through the tube (ca. 0.5-2 m) to reach the detector. Based on the velocity the *m/z* of the ion can be determined.

I employ the advantage of short measurement times for high-throughput detection of metabolites in **chapter 1**.

### Absolute quantification of metabolites using LC-MS/MS

In combination with TOF-MS (Fig 1), I used LC-MS/MS (Fig 2) with two quadrupoles as mass analysers (Fig 2C, 2E). In comparison to TOF, quadrupoles excel in quantification. Mass separation in a quadrupole is a result of ion motion in a dynamic (radio frequency or *rf*) electric field and is dependent directly on the *m/z* of the ion. Quadrupole mass analyser use four rods that have high electric currents alternating at specific frequencies to deflect all ions with undesired *m/z* values (Glish and Vachet 2003). Only ions within a selected *m/z* range have trajectories that can pass the quadrupole (Fig 2C). Triple quadrupole instruments (QqQ) are commonly used for MS/MS. The first and third quadrupoles are operated as mass spectrometers (Fig 2C, 2D), whereas the second (middle) quadrupole acts as the cell for collision-induced dissociation (CID) (Fig 2D).

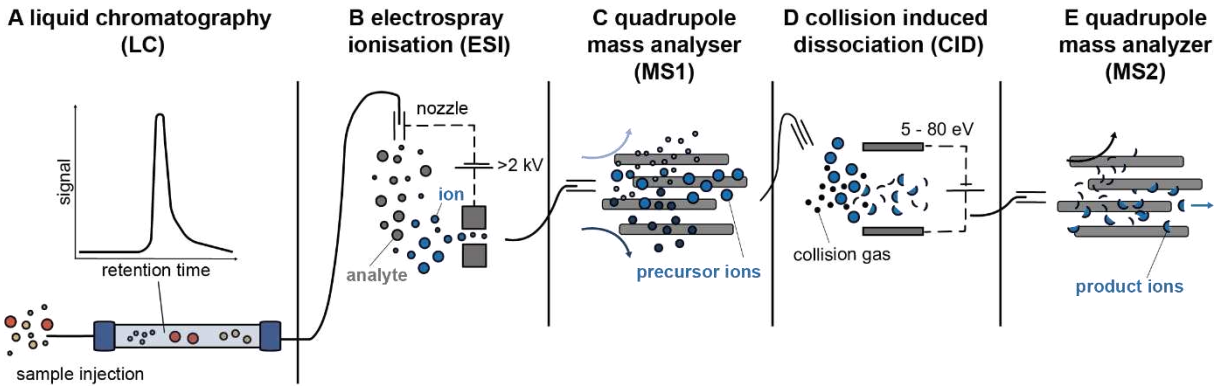



**Figure 2 Technical configurations of LC-MS/MS.**

**A** Mass spectrometry is coupled to liquid chromatography for separation and purification of metabolites (Bjerrum 2015) prior to analysis. **B** Electro spray ionisation (ESI). The analyte is introduced into the mass spectrometer via a fine needle. The nozzle is typically charged with high voltages (> 2kV) to form fine aerosols of charged droplets. Depending on the polarity of the of the applied voltage (mode), ESI can result in positively $[M+H]^+$ or negatively charged ions $[M-H]^-$ (Famiglini et al. 2021).**C** First Quadrupole to select precursor ion (MS1). Mass separation in a quadrupole is a result of ion motion in a dynamic (radio frequency or rf) electric field and is dependent directly on the *m/z* of the ion. Quadrupole mass analyser use four rods that have high electric currents alternating at specific frequencies to deflect all ions with undesired *m/z* values (Glish and Vachet 2003). Only ions within a selected *m/z* range have trajectories that can pass the quadrupole. **D** Collision-induced dissociation (CID). Precursor ions are

fragmented by applying an electric current (between 5-89 eV) into product ions. **E** Quadrupole mass analyser (as described in C). Only selected product ions can pass.

For CID precursor ions are fragmented by inert gas molecules (e.g. helium, nitrogen, or argon) into product ions. These fragmentation patterns are highly specific and provide a higher degree of confidence while increasing sensitivity.

Filtering metabolite-specific precursor and product ions is called tandem mass spectrometry (MS/MS). The characteristics of QqQ instruments include a fast scan speed (> 5 msec dwell time for a single transition) and fast positive-negative ion mode switching (Yuan et al. 2012). Multiple reaction monitoring (MRM) can achieve rapid scans over multiple (very narrow) mass windows to acquire traces of multiple product ions in parallel. To enable parallel measurements collision energy, precursor and product ions (*m/z*) can be preselected. This makes the QqQ instruments an ideal strategy for high throughput targeted metabolites analysis. The ability to measure many metabolites in hundreds of samples per day is essential for large-scale investigations of metabolic networks. Current high-throughput metabolomics methods detect many metabolites within minutes or seconds using direct or flow injection mass spectrometry (Beckmann et al. 2008, Madalinski et al. 2008, Fuhrer et al. 2011, Link et al. 2015). As a trade-off to the speed, these untargeted methods are not quantitative and the identification of metabolites can be ambiguous (Cajka and Fiehn 2016). Yet, absolute quantitation of selected metabolites plays an important role in hypothesis-driven studies, metabolic engineering approaches, and the development of metabolic models. I show relative and absolute quantification of metabolites with LC-MS/MS in **chapter 2**, **chapter 3** and **chapter 4**.

# 2 Systems-level analysis of *E. coli* central metabolism using metabolomics methods

### Metabolic profiling of *E. coli* to infer gene function

Targeted perturbations of cellular networks are key to understanding and engineering gene function. Previous research showed, the deletion of a gene causes a specific

metabolic fingerprint that allows tracking back the function of a gene (Fuhrer et al. 2017). Knockout mutant libraries enabled systematic analyses of gene-gene networks, gene-metabolite networks, and gene-regulatory networks (Kemmeren et al. 2014). However, gene deletions are static and irreversible perturbations and therefore permit the characterisation of genes, that are essential for cell growth. RNA interference (Na et al. 2013) or CRISPR interference (Qi et al. 2013) allow inducible knockdowns of essential genes, and these methods have been applied to construct synthetic regulatory circuits (Qi et al. 2013, Santos-Moreno et al. 2020) and dynamic growth switches (Li et al. 2016).

The CRISPR interference (CRISPRi) system is based on a catalytically dead Cas9 (dCas9) which binds to the DNA and sterically blocks DNA polymerase and transcription factors (Fig 3A). Similar to Cas9, dCas9 is guided by a single guide RNA (20-25 bp) allowing specific control of target gene expression. Repression efficiency depends on the position of the guide sequence relative to the promotor region. sgRNA sequences complementary to the starting region showed stronger effects than targeting regions inside or at the end of a gene (Qi et al. 2013, Lawson et al. 2017, Donati et al. 2021). Designing sgRNAs based on the target region can be employed to tune knockdown strength, providing a flexible tool for controlling gene expression. Different CRISPRi systems were developed for a variety of organisms and human cell lines e.g. cyanobacteria (Kirtania et al. 2019), *Bacillus subtilis* (Peters et al. 2016), mammalian cells (Du et al. 2017) including primary human T cells (Schmidt et al. 2022).

The original CRISPRi system design was a two-plasmid system with separated expression of dCas9 and sgRNA (Qi et al. 2013) (Fig 3B). Further applications are shown to condense the two-plasmid system to a one-plasmid system in order to be able to combine methods, involving otherwise three plasmids (Beuter et al. 2018, Sander et al. 2019b). One disadvantage of plasmid-based versions is leakiness of promoters, leading to plasmid instability and/or loss of the plasmid. Further, too high expression of dCas9 results in additional protein burden for the cell (Huang et al. 2015). Genomic versions are more laborious, however, have been shown to be tight and minimally invasive (Lawson et al. 2017, Yao et al. 2020, Donati et al. 2021).

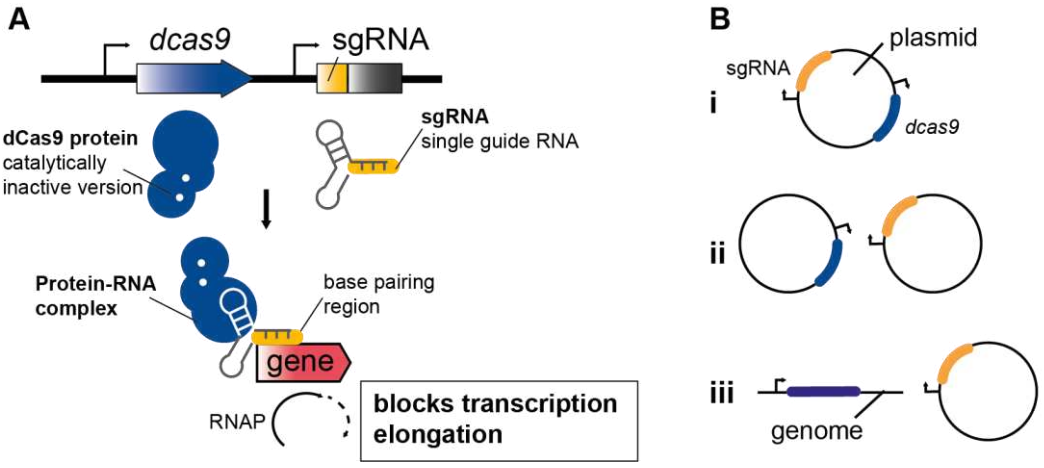



**Figure 3 The CRISPR interference system an overview**

**A** The CRISPR interference system consists of a protein - RNA complex. A catalytically inactive Cas9 protein (dCas9, blue), forms a complex with a single guide RNA (sgRNA, red). The sgRNA encodes 20-25 nt that are complementary to the target region. Since the dCas9 protein is defective in nuclease activity. the complex can bind to DNA and block RNA polymerase (RNAP, red) and transcript elongation. Depending on the location of the target region (closer to the transcription start site), transcription initiation and binding of transcription factors can be prohibited. Figure adapted from (Qi et al. 2013). **B C** CRISPRi expression systems. The simplest version consists of one plasmid encoding sgRNA and dCas9 (Beuter et al. 2018, Sander et al. 2019b). Different approaches include two-plasmid versions (Qi et al. 2013, Lawson et al. 2017) or genomic integration of dCas9 (Lawson et al. 2017, Yao et al. 2020, Donati et al. 2021).

Another advantage of this knockdown strategy is, that it allows control of multiple target genes in parallel (Qi et al. 2013). CRISPR libraries have been made possible by using the sgRNA as barcodes (Shalem et al. 2015). Oligonucleotides are synthesized and can be ordered as an oligo pool in a tube. In one reaction the pooled sgRNA sequences can be cloned into the vector plasmid with cloning methods such as Golden Gate and Gibson Assembly (Gibson et al. 2009, Engler and Marillonnet 2014). After pooled transformation, colonies can be either isolated for multiplexed (arrayed) screens or combined for pooled screens. A pooled approach is faster and unbiased but might require the cloning of individual strains for hypothesis-driven studies. Tracking of individual strains in pooled screens is performed by Next-Generation sequencing (NGS) or Fluorescence-activated Cell Sorting (FACS). Pooled screens are less laborious and thereby less expensive. However, arrayed screens enable to directly measure single phenotypes using metabolomics, proteomics or image analysis. In addition to tracking of the strains combined tracking and RNA sequencing was

performed (Adamson et al. 2016, Dixit et al. 2016, Datlinger et al. 2017). However, this tandem read-out is mainly limited by the cost of RNA sequencing (Bock et al. 2022). Moreover, phenotypes such as increasing or decreasing concentrations of specific metabolites, cannot be easily assessed in screens using an RNA barcodes. Recently developments showed protein barcodes for CRISPR libraries enabling the tandem-readout of 364 barcodes and multiple phenotypic markers such as phospho-signalling (Wroblewska et al. 2018). Genome-wide CRISPR screens however also showed great relevance to perform functional analyses of genes (Peters et al. 2016, Rousset et al. 2018, Donati et al. 2021), to enrich slow-growing cells (Beuter et al. 2018) or for the identification of targets for metabolic engineering (Li et al. 2020).

In **chapter 1** I present a novel protein barcoding system for CRISPRi strains. In **chapter 2** I use CRISPRi to study gene-function in *E. coli* central metabolism.

### Pentose phosphate pathway in *E. coli* - a metabolomics view

The pentose phosphate pathway (PPP) plays a pivotal role in maintaining carbon homeostasis, providing precursor molecules for nucleotide and amino acid biosynthesis, generating reducing agents for anabolism, and to counter oxidative stress. The PPP is divided into an oxidative branch and a non-oxidative branch (Fig 4). The oxidative branch converts glucose 6-phosphate (g6-p) into $CO_2$, ribulose 5-phosphate (ru5-p) and NADPH. The non-oxidative branch instead metabolizes the glycolytic intermediates fructose 6-phosphate (f6-p) and glyceraldehyde 3-phosphate (g3-p) as well as seduheptulose-7 phosphate (s7-p), yielding ribose 5-phosphate (r5-p) for the synthesis of nucleic acids and sugar phosphate precursors for the synthesis of amino acids. Whereas the oxidative PPP is unidirectional, the non-oxidative branch can supply glycolysis with intermediates derived from r5-p and vice versa. Depending on the chemical demand, reaction rates require constant adaption. High turnover of metabolites is one challenge for the quantification of metabolites within the PPP. Further, some intermediates of the PPP namely hexose phosphates and pentose phosphates are isomers or isobars with similar masses and similar chemical properties, making quantification of single metabolites using solely MS impossible. Current methods rely on additional sample treatment such as ion-pairing reagents (e.g. 3-tributylamine, Buescher et al. 2010). Ion-pairing reagents can improve the retention

and peak shape of polar metabolites such as hexose and pentose phosphates. But ion-pairing agents cause ion-suppression, lower sensitivity, greater retention time variability and poor analyte recovery (Gustavsson et al. 2001). **Chapter 2** presents an LC-MS/MS method developed to quantify metabolites in PPP and glycolysis without the addition of ion-pairing reagents.

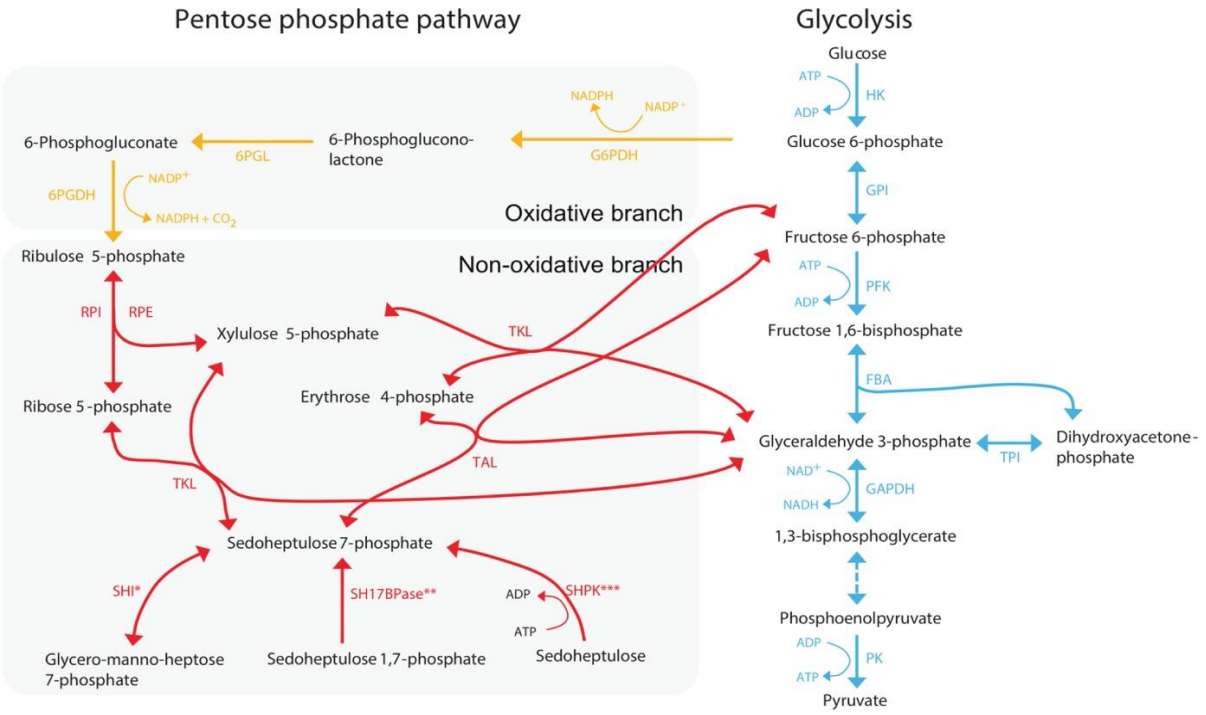



**Figure 4 Metabolic map of central metabolism**

Metabolic map of the pentose phosphate pathway (PPP, yellow and red) and glycolysis (canonical topology) (blue), as part of central metabolism in various organisms (*bacteria; **fungi (*S. cerevisiae*) and plants, ***mammals). PPP is divided into oxidative branch and non-oxidative branch (boxes). Different enzymes were identified catalysing the conversion of sedoheptulose, stars indicated respective organism. Abbreviations: fructose bisphosphate aldolase (FBA); hexokinase (HK); phosphofructokinase (PFK); pyruvate kinase (PK); SH17BPase (SH17BP). Figure adapted from (Stincone et al. 2015).

## Rewiring central metabolisms for $CO_2$ assimilation

Six different natural carbon fixation cycles are known today (Fuchs 2011). The most prominent carbon fixation cycle is the reductive pentose phosphate pathway (rPP) also known as Calvin-cycle (Bassham et al. 1954). Enzymes and reactions in rPP and PPP mostly overlap (Fig 5). Therefore, reverting flux through the PPP could in theory assimilate carbon in a cyclic manner. The problem is that the oxidative branch of the PPP is not reversible. The Calvin cycle bypasses these reactions via ribulose 1,5-

bisphosphate carboxylase oxygenase (Rubisco). Both the non-oxidative PPP and the Calvin cycle interconvert a total of 15 pentose carbon atoms (contained in ru5-p) with 15 glycolytic carbon atoms (in the form of f6-p and g3-p). However, while the classical non-oxidative PPP uses TAL to make sedoheptulose 7-phosphate, the Calvin cycle uses the glycolytic enzyme fructose-bisphosphate aldolase (FBA) to convert erythrose 4-phosphate plus dihydroxyacetone phosphate into sedoheptulose 1,7-bisphosphate, which in turn is hydrolysed by the enzyme SH17BPase to yield sedoheptulose 7-phosphate. This hydrolysis step provides the thermodynamic driving force, pushing the Calvin cycle towards ribulose 5-phosphate. Thus, while the non-oxidative PPP is reversible, the Calvin cycle is not.

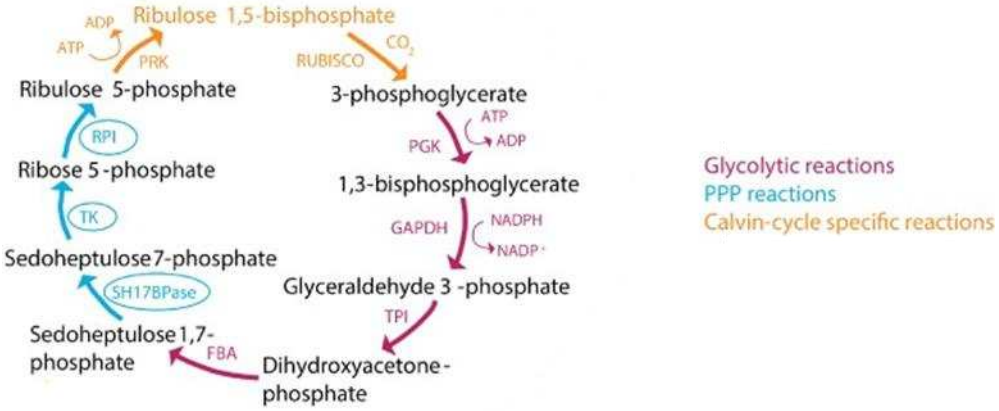



**Figure 5 Calvin cycle overlaps with central metabolism of *E. coli***

Calvin cycle (rPP cycle) share enzymes and reactions with the pentose phosphate pathway (PPP) and glycolysis. Three reactions are Calvin cycle specific (PRK, RUBISCO, SH17BPase). The cycle can only operate in one direction, as indicated with arrows. Abbreviations: phosphoglycerate kinase (PGK); phosphoribulokinase (PRK), transketolase (TK); fructose-bisphosphate aldolase (FBA). Figure adapted from (Stincone et al. 2015).

### Converting *E. coli* to produce all its biomass from $CO_2$

Carbon dioxide ($CO_2$) is one of the major drivers of climate change. Microbes are a powerful tool to convert $CO_2$ into a diverse repertoire of products including food, fuel, bioplastics and other valuable compounds (Cho et al. 2012b, Nielsen and Keasling 2016, Claassens et al. 2019, Gleizer et al. 2020). Unlike heterotrophic organisms that must be supplied with organic substrates for growth (e.g. sugar), autotrophic microorganisms generate biomass from $CO_2$ and derive energy from light (photoautotroph) or an external electron donor (chemoautotroph). Natural autotrophs

already metabolise $CO_2$ efficiently, yet most are challenging to cultivate and manipulate genetically compared to heterotrophic model organisms like *E. coli*.

Recent developments in synthetic biology showed progress using two different approaches. Since the valorisation of $CO_2$ is a great challenge, one approach is to integrate biotic with abiotic processes. One carbon molecules (C1) can be derived from $CO_2$ and renewable energy (Na et al. 2019). Physiochemical methods capture for example wind or sunlight (solar) to energize $CO_2$ into formate or methanol. These C1 molecules are especially interesting since they are miscible in water, under ambient conditions (Yishai et al. 2016). This property can avoid mass transfer limitations that might occur with gases such as carbon monoxide and methane. A recent study showed that formate produced with an energetic efficiency of >40% by direct electrochemical reduction of $CO_2$. Further, methanol can be generated in two steps: i) electrolysis generates hydrogen ii) hydrogen serves as a reactant with $CO_2$ with an efficiency of >50% (Szima and Cormos 2018). Just recently, synthetic biologists converted *E. coli* to generate biomass from formate (Kim et al. 2020) and methanol (Keller et al. 2022).

A second approach successfully converted *E. coli* to directly use $CO_2$ for biomass. That requires a synthetic carbon fixation module. Except for three enzymes (Prk, Rubisco, SH17BPase), reactions overlap with the central metabolism of *E. coli* (Fig 6). By inserting only Rubisco and Prk enzymes from glycolysis and PPP, *E. coli* has all the enzymes to close a functional rPP cycle. Energy and reducing power in this strain is derived from the one-carbon molecule formate ($HCOO^-$), which can be produced electrochemically (Yishai et al. 2016, Hegner et al. 2018).

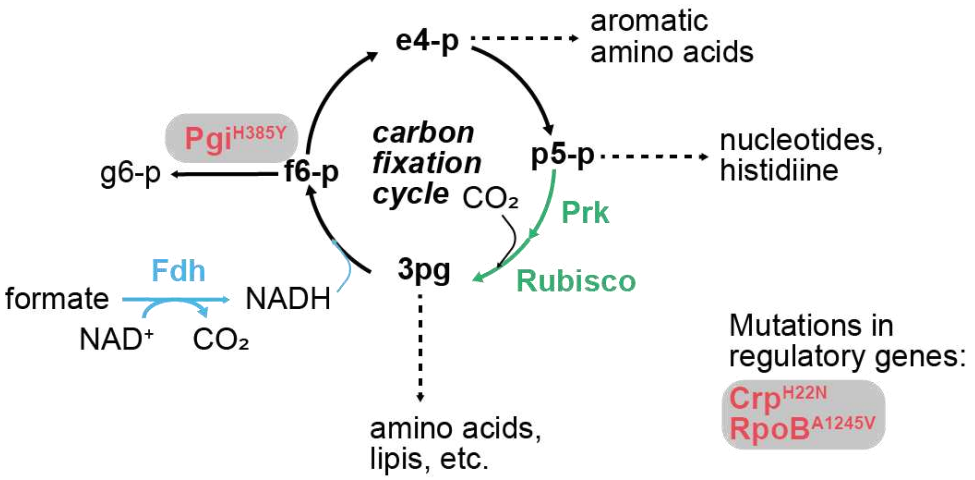


**Figure 6 Metabolic scheme chemoautotrophic *E. coli***

Introduction of carbon fixation module (Prk, Rubisco, carbonic anhydrase not shown), energy module (Fdh) in addition to three mutations ($Pgi^{H385Y}$, $Crp^{H22N}$, $RpoB^{A1245V}$) are sufficient to turn *E. coli* chemoautotrophic (Ben-Nissan et al. 2023).

To convert *E. coli* metabolism to use solely $CO_2$ and formate adaptive laboratory evolution (ALE) was used to rewire metabolism. ALE is a powerful tool to fast acquire phenotypes, one side effect is the accumulation of mutations with unknown functions (Phaneuf et al. 2020). Iterative ALE discovered three mutations are sufficient for *E. coli* to grow autotrophically when introduced alongside non-native energy (formate dehydrogenase) and carbon-fixing (Rubisco, Phosphoribulokinase, Carbonic anhydrase) modules (Ben-Nissan et al. 2023). The mutated genes are involved in glycolysis (Pgi), central-carbon regulation (*crp*), and RNA transcription (*rpoB*). Mutations in Pgi were observed in previously evolved *E. coli* strains (Herz et al. 2017, Gleizer et al. 2020). Pgi catalyses a branchpoint reaction of the rPP cycle and is therefore suspected to limit the efflux of intermediates within the $CO_2$ fixation cycle. Along with the branch-point mutation in Pgi, the chemoautotrophic *E. coli* requires two additional mutations, one in the gene *crp*, a global metabolism regulator and in *rpoB*, the beta subunit of the RNA polymerase. In order to enable *E. coli* to grow autotrophically the cell needs to adapt its catabolism towards $CO_2$ as carbon source and NADH as energy source, due to their global effect I hypothesized that these mutations might be involved in one of these two processes or even combined. For all three mutations, their mechanism is jet to show. In **chapter 3** I show recent advances in characterising their impact on energy and carbon metabolism.

## 3 Metabolic engineering of *E. coli* for amino acid production

Bio-based production of amino acids is a growing market, especially in manufacturing of food and feedstock (Wendisch 2020). The market-size has been estimated at USD 25.21 billion in 2022 and still increasing (Grand View Research 2023). Out of the 20 proteinogenic amino acids that are industrially produced, 19 are currently synthesised through fermentation in microorganisms and only one is produced chemically. These fermentation processes rely on sugar-based substrates that derive from molasses or are generated from starch by enzymatic hydrolysis (Wendisch et al. 2016). Amino acid

production in *E. coli* is highly regulated to balance production and demand with minimal effects on the remaining metabolic network (arginine and proline are shown here as two examples, Fig 7). All 20 amino acids in *E. coli* either inhibit the first step of their synthesis via allosteric regulation or product inhibition, in case of a single-reaction pathway (Chubukov et al. 2014). This mechanism ensures robust amino acid production (Sander et al. 2019a). As the inhibition typically affects only the branch of the pathway that is specific to the particular amino acid, diverging pathways can mostly be tuned independently. In addition to allosteric feedback at least ten amino acids negatively regulate the transcription of their own biosynthesis pathways, either via transcription factors (Cho et al. 2008, 2012a) or transcriptional attenuators (Yanofsky 1981). The regulation of amino acid biosynthesis pathways follows a straightforward logic to decrease pathway flux when amino acid production exceeds supply.

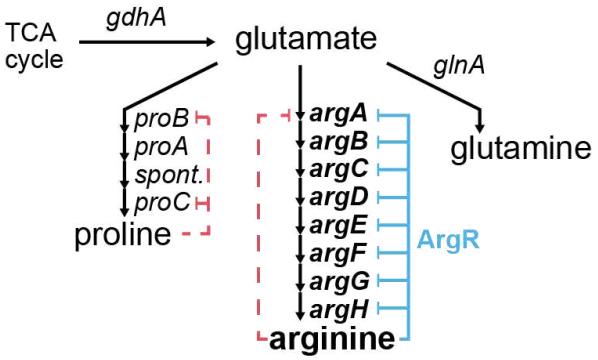



**Figure 7 Amino acid biosynthesis pathway in *E. coli***

Amino acids control their biosynthesis via allosteric regulation (red line) of the first committed step and by transcriptional regulation (blue line) of biosynthetic enzymes. — arginine and proline are shown here as two examples. Map adapted from open source databases: Ecocyc (Keseler et al. 2009) and RegulonDB (Tierrafría et al. 2022). Abbreviation: spont, spontaneous dehydration step; tricarboxylic acid cycle, TCA cycle.

Disabling these cellular control mechanisms haven been effective tools to engineer microbes for overproduction of amino acids (Park et al. 2018, Sander et al. 2019a). Removing cellular control can increase flux into biosynthesis pathways (Nielsen and Keasling 2016, Lee and Wendisch 2017). However, missing regulation can have broad physiological consequences for the host that reduce cellular growth and robustness (He et al. 2016). This problem has been addressed by creating overproduction pathways with new control mechanisms, which sense, for example, the availability of

intracellular metabolites (Farmer and Liao 2000) or the concentration of pathway intermediates by synthetic feedback mechanisms (Zhang et al. 2012) Another approach is called two stage processes by separating growth and production phases by dynamically switching between these states using metabolic valves (Burg et al. 2016, Gupta et al. 2017). Temperature-sensitive (TS) mutations are an alternative method to perturb essential genes. At low (permissive) temperatures, genes with a TS mutation encode a functional product, while at higher (non-permissive) temperatures, the gene product is not functional. TS mutants were used to engineer diverse cellular systems with applications in medical and industrial biotechnology (Weber 2003, Cho et al. 2012b, Lynch et al. 2016, 2019, Piraner et al. 2017, Harder et al. 2018, Schramm et al. 2020, Wang et al. 2021, Kasari et al. 2022). Focused ultrasound, for instance, can control TS proteins *in vivo* (Piraner et al. 2017), and almost all bioreactors are equipped with temperature control. Taken together, temperature sensitive mutations are versatile, controllable and can be reversible tools, used in two-stage processes. In **chapter 4,** I show a proof of concept employing temperature sensitive mutations for the production of amino acids.

## Aim of the Study

The work is presented in four structured chapters showing recent advances in mass spectrometry-based metabolomics to engineer and investigate *E. coli* primary metabolism on a systems-level. Through the application of innovative mass spectrometry-based techniques, we aim to deepen our understanding of bacterial metabolism and provide valuable tools and methodologies for advancing the field of metabolic engineering, metabolomics, and sustainable compound production.

In **Chapter 1**, we introduce a novel protein-based barcoding system designed to facilitate real-time tracking and phenotyping of CRISPR interference (CRISPRi) strains. This system aims to eliminate the need for sgRNA sequencing while enabling simultaneous tracking and phenotyping of CRISPRi strains. The potential implications of this innovation extend to multi-phenotyping CRISPRi screens across diverse organisms.

**Chapter 2** focuses on the quantification of sugar phosphate isomers within *E. coli* central metabolism. Using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS), enabled investigation of gene function within the Pentose Phosphate Pathway (PPP) and glycolysis. The overarching goal is to gain insights into the roles of specific enzymes in sugar phosphate metabolism. This method can be applied in the future to diverse organisms and pathways.

In **Chapter 3**, I study the metabolic adaptations required for *E. coli* to transition from heterotrophy to chemoautotrophy. The core objective is to examine the effects of autotrophic-enabling mutations on energy and carbon metabolism, using metabolomics methods. This investigation holds the potential to contribute to our understanding of sustainable carbon conversion and synthetic autotrophy.

**Chapter 4** introduces a temperature-controlled two-stage process to enhance arginine production in *E. coli* while decoupling growth and overproduction using a temperature switch. Temperature sensitive mutations show great promise as tools for microbial bioproduction strategies.

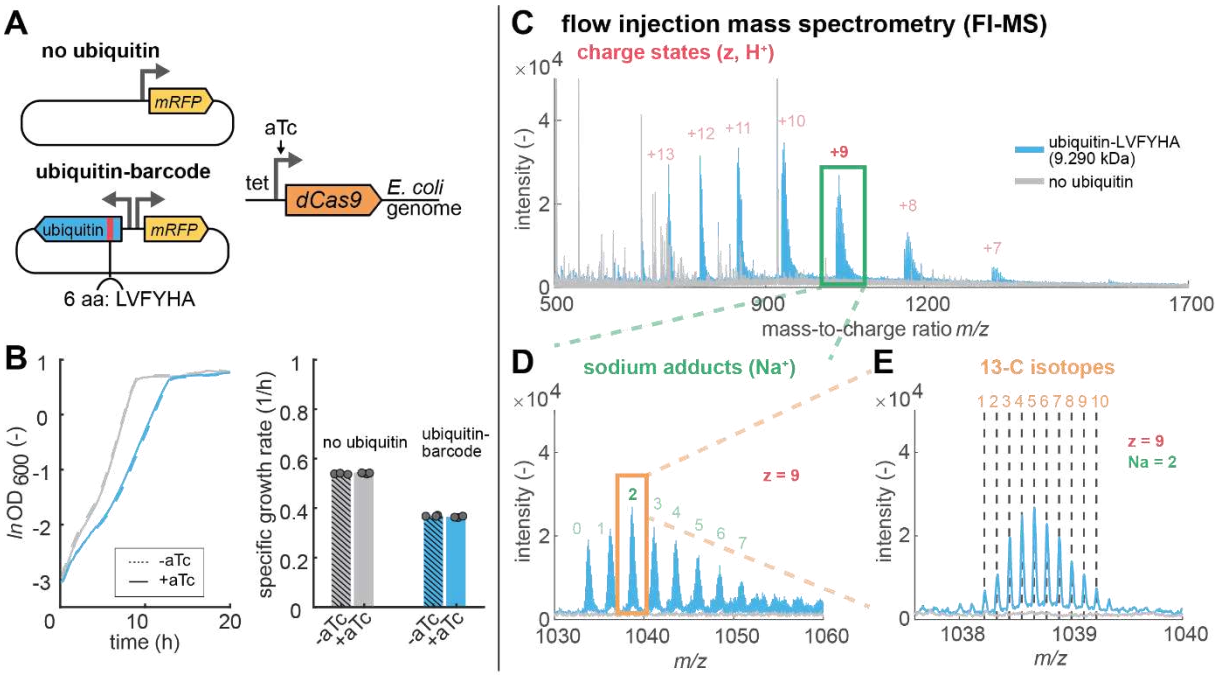
**The scope of the study** encompasses metabolic engineering and systems-level analysis of *E. coli* primary metabolism.

# Results and Discussion

## Chapter 1 A protein-based barcoding system for CRISPRi strains

### Top-down proteomics and FI-MS enable fast detection of protein barcodes in CRISPRi strains

We began with a CRISPRi system comprising an aTc-inducible dCas9 on the chromosome and a plasmid-based sgRNA (pgRNA), as previously described (Qi et al. 2013) (Fig 8A). On the sgRNA plasmid, we introduced an additional gene encoding human ubiquitin (Brzovic et al. 2006), a 76-amino-acid small protein (Fig 8A). Ubiquitin, a widely used reference protein in top-down proteomics (Compton et al. 211, Donnelly et al. 2019), has a mass of 8.6 kDa. We also attached a 6-amino-acid *N*-terminal tag (LVFYHA) to ubiquitin, creating a 9.3 kDa barcode (LVFYHA-Ubiquitin) (Fig 8A). LVFYHA-Ubiquitin was constitutively expressed from the same plasmid as the sgRNA, with a 38 bp separation between their promoter sequences (Fig S1A). To assess expression strength, I introduced the combined expression plasmid into an *E. coli* CRISPRi control strain. Additionally, I cultivated the CRISPRi control strain with the original pgRNA plasmid design, omitting the barcode. The barcoded CRISPRi strain with the new plasmid exhibited a 33% slower growth rate than the same strain with the original plasmid (Fig 8B). This could be attributed to our design of a robust expression system, including a strong ribosomal binding site (T7 G10 RBS) and promoter (J23100). The addition of aTc, initiating dCas9 expression, had no impact on the growth phenotype, underscoring tight control over dCas9 expression in genomic CRISPRi systems. To verify the mass of LVFYHA-Ubiquitin, I developed a protein extraction protocol compatible with 30s time-of-flight mass spectrometry runs (FI-TOF) (Fig S1B). The cells were washed with phosphate buffer (PBS) to decrease salts, followed by freeze-thaw to break the cells and extract metabolites while preserving the integrity of the ubiquitin barcodes.




## Figure 8 Top-down proteomics and FI-MS enable ubiquitin-based barcode system for CRISPRi strains

**A** Ubiquitin-barcode integrated with the CRISPR interference system. The CRISPRi system consists of a dCas9 expressed from the genome (Lawson et al. 2017) and a single guide RNA (sgRNA) expressed from a plasmid. dCas9 is under the control of an aTc-inducible Ptet promotor. The gene encoding ubiquitin with a *N*-terminal 6 amino acids barcode (LVFYHA) was integrated next to a non-targeting sgRNA on the plasmid pgRNA bacteria (Addgene #44251). A control was a plasmid with the same non-targeting sgRNA (*mRFP*) but without the LVFYHA-Ubiquitin. LVFYHA-Ubiquitin and sgRNA are constitutively expressed with the promoters (J23100, J23199). **B** Growth of CRISPRi strain expressing LVFYHA-Ubiquitin. Growth of CRISPRi strain (gray) and CISPRi strain expressing LVFYHA-Ubiquitin (blue) in microtiter plates with 150 µL culture volume. Cells were grown in M9 minimal medium (5 $gL^{-1}$ glucose) and aTc (200 nM) was added at the start of the cultivation. OD was measured in 10 min intervals in a plate reader. Shown are means of n = 3 cultures. The specific growth rate ($h^{-1}$) is shown as a bar plot and was calculated using linear regression. **C** FI-MS detects protein species of LVFYHA-Ubiquitin. MS1 spectrum of extracts of an *E. coli* strain expressing the LVFYHA-Ubiquitin. Shown is the MS1 spectrum (positive mode) from 500 to 1700 mass-to-charge ratio (*m/z*) acquired with flow injection time of flight (FI-TOF) mass spectrometry. Different charge states arise from varying degrees of protonation (z,$H^+$) indicated in red. **D** Sodium adduct cluster of the +9 charge state. Same MS1 scan as in Fig 8C from 1030 to 1060 *m/z* (green box). Cluster of peaks can be assigned to sodium adducts ($Na^+$) that are indicated in green. **E** 13-C isotope distribution (z=9, 2$Na^+$). Same MS1 scan as in Fig 8C, 8D from 1038 to 1040 *m/z* (orange box 1D). Cluster of peaks can be assigned to 13-C isotopes (orange). Theoretically calculated masses (black dashed lines) overlap with ubiquitin-barcode spectra

FI-MS generated MS1 spectra with seven ion peak clusters spanning the 500-1700 *m/z* mass range (Fig 8C). These ion peak clusters represent various charge states (z=7 - z=13) of the ubiquitin barcode. Multiple charged ($H^+$) protein species are to be expected due to protonation, the most common form of positive charging in ESI-MS. Further analysis revealed subclusters within each charge state cluster, that arise due to sodium adduction (0 - 7 Na+ adducts, Fig 8D) and 13-C isotopes (Fig 8E). Thus, a single protein (LVFYHA-Ubiquitin) resulted in many ion peaks in the MS1 spectrum that result from different protein species: 1) charge states, 2) different numbers of $NA^+$ adducts and 3) 13-C isotopes. Declustering resulted in the correct monoisotopic mass for LVFYHA-Ubiquitin of 9.290 kDa. Ultimately, multiple charging enabled to analyse ubiquitin-barcode in lower *m/z* ranges. MS1 spectra thereby contain multi-layer information, while each ion peak represents a separate mass measurement thereby improving barcode-identification.

**Ubiquitin-barcode to track seven CRISPRi strains**

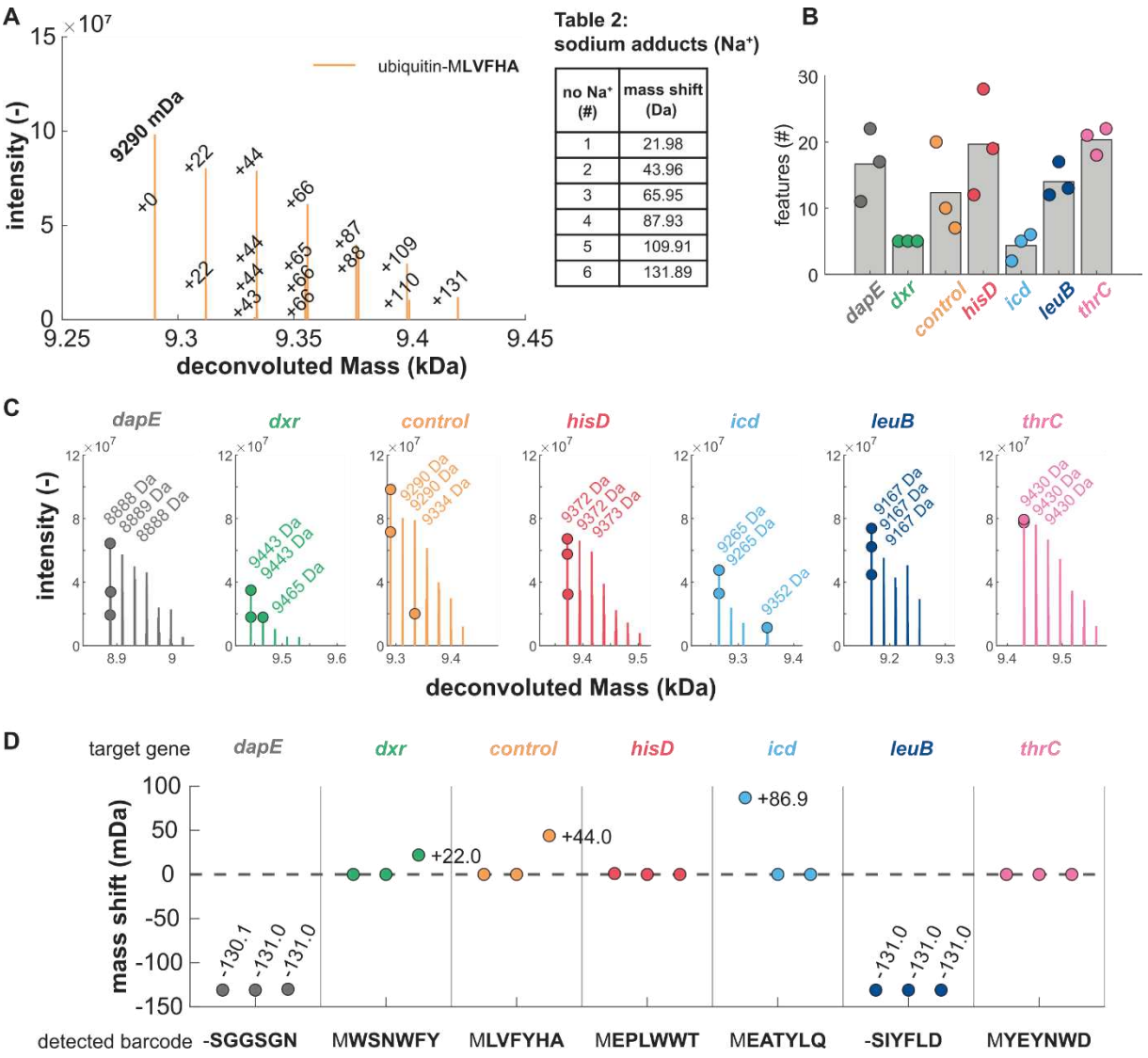
To test the barcoding system, six additional ubiquitin barcodes were designed with monoisotopic masses ranging from 9.018 kDa to 9.443 kDa (Table 1). Concurrently, we selected six distinct CRISPR sgRNAs, targeting genes in primary metabolism (*dapE*, *dxr*, *hisD*, *icd*, *leuB*, *thrC*).

**Table 1 Barcoded CRISPRi strains**

| No | Target gene | sgRNA (20 nt) | *N*-terminal tag (6 aa) | Total barcode mass (kDa) |
|---|---|---|---|---|
| 1 | *dapE* | CCTACCACTTCGATACTCGA | SGGSGN | 9.02 |
| 2 | *dxr* | GAGACGCATCACCTCTTTTC | WSNWFY | 9.26 |
| 3 | *hisD* | TCGCCGGGCGCATTAACAGC | EPLWWT | 9.29 |
| 4 | *leuB* | ACACCCCTTCTGCTACATAG | SIYFLD | 9.30 |
| 5 | *icd* | GTGATCTTCTTGCCTTGTGC | EATYLQ | 9.37 |
| 6 | *thrC* | GGCATCGAAATCGCCGTCGA | YEYNWD | 9.43 |
| 7 | control | CGCATGATAAAACACCAG | LVFYHA | 9.44 |

For each sgRNA, a corresponding plasmid was designed and individually introduced into the *E. coli* YYdCas9 base strain. These barcoded CRISPRi strains were cultured in shake flasks, and samples were collected for protein extraction after 6.5 h. MS1 spectra of the seven strains were measured, with three replicates per strain, using FI-MS. To enable a fast identification of the ubiquitin-barcode monoisotopic masses, feature deconvolution (identification of protein species) was implemented. FLASHDeconv (Jeong et al. 2020), an open-source application for spectral deconvolution in top-down proteomics (decharging and deisotoping). The algorithm transforms MS spectra logarithmically, simplifying the deconvolution problem into a pattern-matching task. FLASHDeconv exhibited superior deconvolution quality and executed significantly faster compared to other available free tools like Agilent MassHunter Protein Deconvolution Software.

Analysing the deconvoluted MS1 data for the control strain expressing LVFHA-ubiquitin, showed 17 mass features closely approximating the 9290 Da monoisotopic mass of LVFHA-ubiquitin (Fig 9A). Among these, the first mass feature matched the monoisotopic mass and exhibited the highest intensity signal. The remaining 16 mass features represented protein species with varying sodium adducts (0 - 6 Na+ adducts), with a mass accuracy of 1 mDa. All observed mass shifts could be attributed to sodium adducts or multiples of sodium adducts. Nevertheless, in some instances, two peaks with masses +1 mDa emerged, indicating that despite successful charge state removal during deconvolution, complete deisotoping had not been achieved.




## Figure 9 Top-down proteomics enables identification of CRISPRi strains

**A** Protein species of LVFYHA-ubiquitin expressed in CRISPRi control strain. Spectral deconvolution of MS1 scan resulted in multiple peaks within a mass range between 9.25 kDa and 9.45 kDa, mass difference to monoisotopic mass (9290 Da) as indicated. **Table2:** Sodium adducts cause mass shift. Table shows respective mass shift for 1 to 7 sodium adducts, with 21.98 Da monoisotopic mass. **B** Number of mass features in deconvoluted spectra from seven different ubiquitin barcodes. Bars indicate number of features as mean of n=3 cultures. Individual measurements are indicated as dots. Strains are named according to their target gene. **C** Deconvolution of spectra from seven different barcoded CRISPRi strains. Overlapp of n = 3 cultures. Smallest identified feature is highlighted with a dot and the respective mass in Da. **D** Matching mass shift to protein species of seven different ubiquitin barcodes. Mass shift (Da) between calculated mass and detected protein species. Barcode was designed as 6 amino acids initiated by fMet (131.04 Da) as indicated in the sub caption. Strains were cultivated in shake flasks in minimal medium supplemented with glucose. Dots show n = 3 cultures.

Deconvolution of all seven MS1 scans identified mass features, and their total count varied across different strains and cultures (Fig 9B). I initially assigned barcode sequences based on the smallest mass feature (Fig 9C), assuming that this mass feature represents the monoisotopic mass. I calculated mass differences to the monoisotopic masses, and most of the deconvoluted masses indeed matched the monoisotopic mass including *N*-formylmethionine (fMet) for initiating of ubiquitin-barcodes translation. In two strains (*dapE*, *leuB*), I observed a mass difference of -131 Da, corresponding to the mass of methionine (Fig 9D). This implies cleavage of methionine from the respective barcodes (SIYFLD-ubiquitin and -SGGSGN-ubiquitin). Repetitive sequences like tryptophan-tryptophan (WW) and glycine-glycine (GG) were functional expression and identified. This allows a simple design of the barcode sequence, specifically the unique six amino acids.

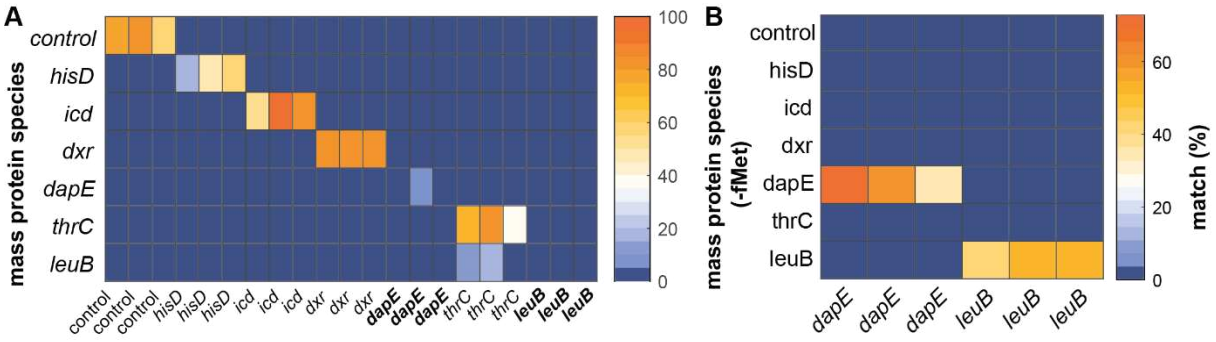



**Figure 10 Matching protein species to barcodes**

**A** Matching mass features identified in deconvoluted spectra to theoretical protein species (1-7 $Na^+$ adducts). Theoretical masses of barcode species including 131 Da for fMet. Heatmap shows in % how many features can be matches with protein species of individual barcodes. **C** Deconvoluted mass features were matched to theoretical mass of barcodes subtracting 131 Da for fMet. Heatmap shows match in %.

For further validation, I matched each mass feature to a reference table of protein species (1-7 $Na^+$ adducts) with a tolerance of 0.03 mDa (Fig 10). Using this approach, I successfully identified five out of the seven ubiquitin-barcodes (Fig 10A). When I corrected the theoretical mass by fMet, the remaining barcodes could be matched (Fig 10B). Matching the theoretical masses and deconvoluted mass features, shows that we can track each barcode individually.

**Ubiquitin-barcode allows high-dimensional phenotyping of CRISPR Screens**

I targeted specific genes involved in amino acid biosynthesis pathways (*hisD*, *dapE*, *thrC*, *leuB*), the TCA cycle (*icd*), and the methylerythritol phosphate (MEP) pathway (*dxr*) (Fig 11B). Knocking down these genes resulted in distinct growth phenotypes for four strains (*thrC*, *hisD*, *dapE*, *leuB*), while the control, *dxr*, and *icd* strains exhibited no growth alterations (Fig 11A); however, even in the absence of growth changes, we can still anticipate metabolic alterations. Our freeze-thaw extraction protocol is well-suited for polar metabolites and FI-MS offers a unique advantage by detecting metabolites within the lower mass range of 50-800 *m/z*. I matched ion peaks to 1,877 *E. coli* metabolites listed in the iML1515 genome-scale model (Monk et al. 2017) and successfully annotated 534 ion peaks to metabolites in the same FI-MS data used for barcode measurement. 157 metabolites were measured in positive mode ($[M+H]^+$) and 376 in negative mode ($[M-H]^-$).

I was able to detect multiple metabolites for all targeted pathways. CRISPRi knockdowns are known to cause local perturbations in metabolism, such as increasing substrates of the reaction. In four strains involved in amino acid biosynthesis, I observed strong accumulations of the substrates. For the *dxr* strain, and *icd* strain, I did not detect substrate changes. For the *dxr* strain, this matches previous measurements of this strain (Donati et al. 2021). The metabolic response of the *icd* strain remains inconclusive.

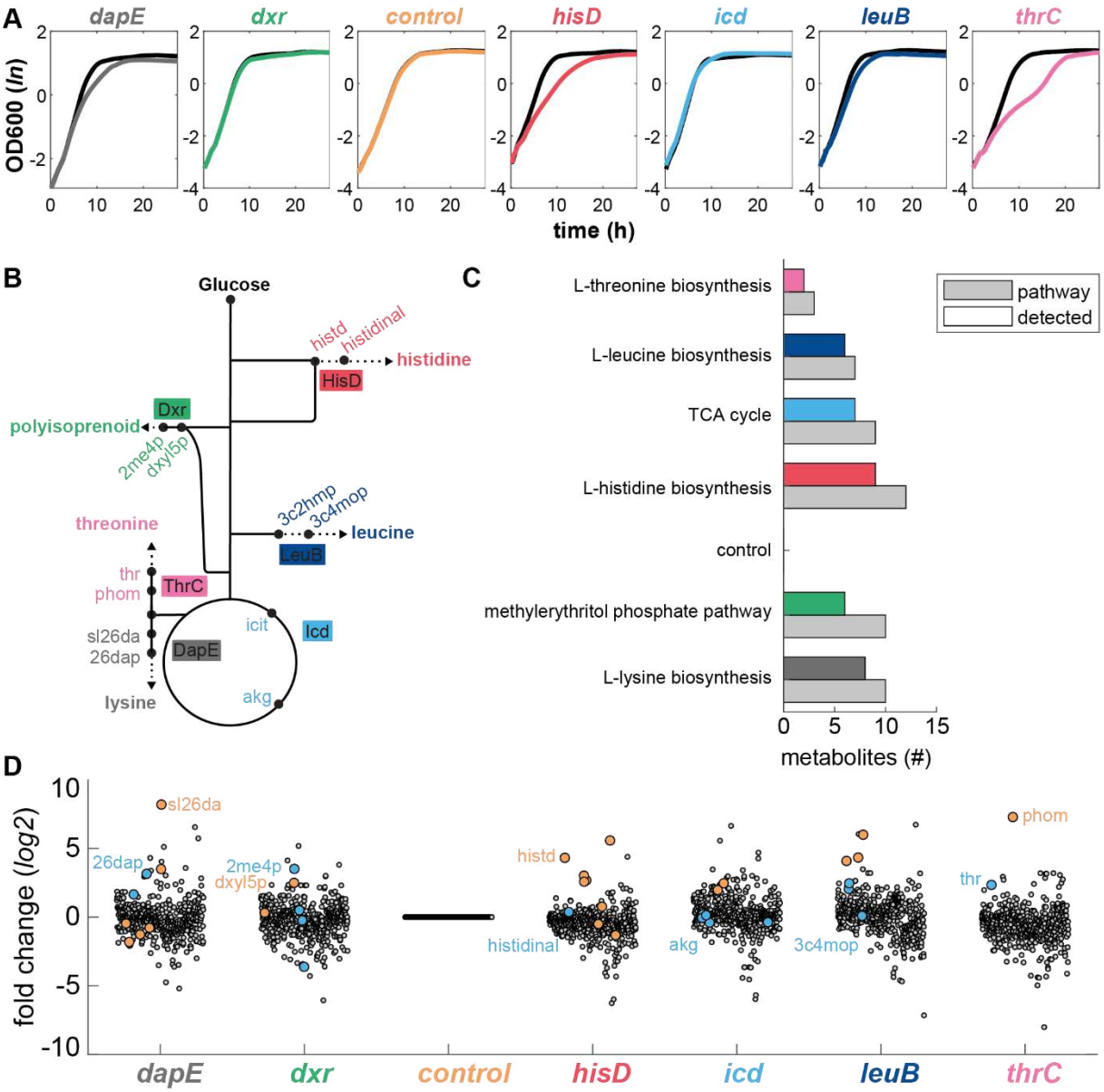


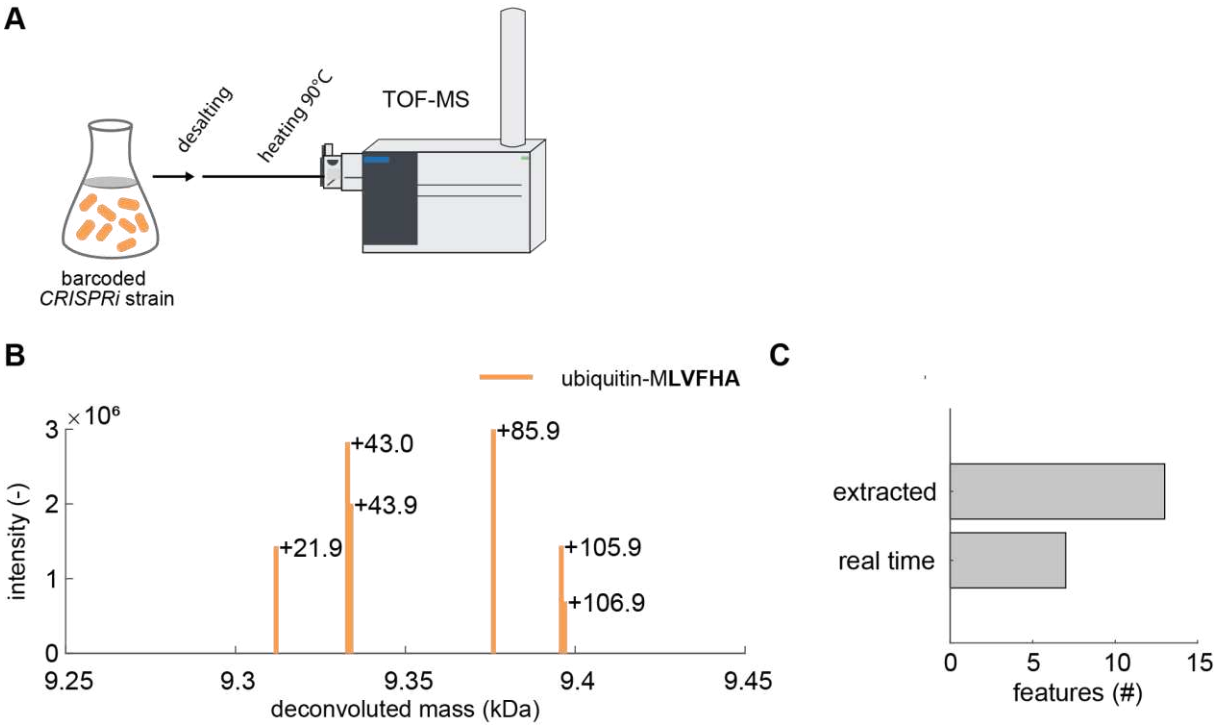

**Figure 11 Ubiquitin-barcodes allow simultaneous analysis of metabolic signature of CRISPRi strains**

**A** Growth curves of the seven CRISPRi strains and a control. aTc was added at the start of the cultivation (200 nM) for induction of CRISPRi mediated knockdown. Strains are indicated with target gene (coloured). Induced cultures are shown in respective colour, uninduced cultures in grey. Strains were cultivated in microtiter plates and optical density was measured every 10 min for 24 h using a plate reader. Lines indicate mean of n = 3 cultures. **B** Metabolic map of 7 target genes (control not shown). Scheme of central metabolism of *E. coli* indicates location of the respective target gene. Product and substrate of the catalysed reaction are coloured according to the respective target gene. Abbreviations: dxyl5p: 1-deoxy-D-xylulose 5-phosphate; 2me4p: 2-C-methyl-D-erythritol 4-phosphate; hisd: histidinol; 3c2hmp: 3-carboxy-2-hydroxy-4-methylpentanoate; 3c4mop: 3-carboxy-4-methyl-2-oxopentanoate; thr: threonine; phom: O-phospho-L-homoserine; sl26da: N-succinyl-LL-2,6-diaminoheptanedioate; 26dap:

LL-2,6-diaminoheptanedioate. Map was designed based on the EcoCyc Database (Keseler et al. 2009) **C** Bar plot shows number of metabolites involved in metabolic pathways targeted by CRISPRi. Coloured bar indicates total number of pathway metabolites that were annotated in MS1 spectra. Target genes are grouped with respective pathway. **D** Intracellular concentration of 534 metabolites of six CRISPRi strains and a control strain. Metabolite levels are shown as log2 fold change in comparison to the control. Data are represented as mean (n = 3). In orange are highlighted upstream metabolites with the direct substrate indicated with the respective name. Metabolites downstream of reaction are indicated in blue with the product highlighted with its respective name. Note that isomers were not separated.

## System for real-time detection of ubiquitin-barcode

Real-time barcode detection can be achieved by the direct injection of living cells (Fig 12).




## Figure 12 Real-time analysis of ubiquitin-barcode in CRISPRi strain

**A** Scheme real-time analysis using TOF-MS (Agilent 6546). CRISPRi control strain expressing LVFHA-ubiquitin was cultivated in shake flasks. 3 µL of a prepared barcoded CRISPRi strain culture broth was injected for MS analysis. Cells were heated to 90°C and sample was diluted to remove salts prior to MS analysis. **B** MS1 spectral deconvolution of CRISPRi control strain expressing ubiquitin-MLVFYHA barcode. Shown are deconvoluted mass (kDa) and intensities (-). Mass shifts are indicated relative to the monoisotopic mass of the barcode (9290 Da) in Da. Sodium adducts are indicated as multiples of 21 Da. **C** Comparison total amount of mass features in extracted and real-time measurements.

The control strain expressing the LVFHA-ubiquitin-barcode was grown in a shake flask in minimal medium. After 6.5 h, 100 µL culture broth was prepared for real-time analysis, by washing and resuspending the cells in diluted minimal medium (1/8 M9/H2O (v/v)). When I used 30 sec runs for real-time mass spectrometry as previously described (Link et al. 2015) there was no signal for the ubiquitin-barcode. I suspected that the cells did not sufficiently break up to release ubiquitin and indeed additional heating (90°C) resulted in MS1 signals (Fig 12A).

Deconvolution of MS1 scan resulted in six mass features (Fig 12B). All masses can be matched be matched to sodium adducts (1,2,4,5 $Na^+$ adducts). LVFHA-ubiquitin was identified with fewer features than in the purified samples (Fig 12C), which strongly suggests that sample purity impacts barcode detection. I observed again, two peaks with a mass difference of 1 Da emerging, indicating that during deconvolution, complete deisotoping had not been achieved. Nevertheless, we identified the barcode successfully, this shows suitable for high-throughput CRISPR screens.

**Discussion**

CRISPRi libraries serve as powerful tools for systematically investigating gene function on a high-throughput scale. Typically, the identification of individual strains within CRISPRi libraries relies on the sequencing of the 20-25 bp protospacer sequences in single guide RNAs (sgRNAs), which act as strain-specific barcodes. In this study, I introduce a new approach to barcode CRISPRi strains using a small protein, ubiquitin.

A dual expression system was designed, combining sgRNA and ubiquitin, featuring a six-amino-acid tag (Fig 8A). The introduction of this combined expression system resulted in a 33% reduction in growth, indicating successful ubiquitin-barcode expression. In the future, it might be worth considering refining plasmid design and exploring options to reduce expression strength. More importantly, inducing dCas9 expression had no impact on the growth rate, affirming the compatibility of the ubiquitin-barcode with the CRISPRi system (Fig 8B).

An aqueous extraction method, complemented by freeze-draw, was developed as a rapid and gentle method to extract protein barcodes and metabolites (Fig S1). This integrated approach enabled the analysis of MS1 spectra, leading to the identification

of barcode species arising from protonation (Fig 8C), sodium adducts (Fig 8D), and 13C isotopes (Fig 8E). Accurate deconvolution of these features is a crucial step in top-down proteomics data analysis, to identify monoisotopic protein masses. Spectral deconvolution was successfully accomplished with FlashDecon (Jeong et al. 2020), albeit with limitations in resolving sodium adducts (Fig 9A). Seven ubiquitin-barcodes were identified based on sodium adducts in deconvoluted spectra (Fig 9C, 10). All designed barcodes were expressed functionally; however, I observed differences in the presence or absence of formylmethionine (fMet) (Fig 9D). In eukaryotes, protein synthesis starts with methionine, whereas prokaryotes initiate translation with fMet (Wingfield 2017). In *E. coli*, methionine aminopeptidase (MAP1) is responsible for cleaving *N*-terminal methionine from newly synthesized polypeptides. Previous studies have shown that methionine aminopeptidases prefer proteins with small *N*-terminal amino acids (e.g., alanine, cysteine, glycine, proline, serine, threonine, and valine). fMet cleavage was observed for two barcodes starting with serine (Fig 10D). Including fMet cleavage into design and analysis can be valuable when working with ubiquitin-barcodes for future experiments.

The annotation of 534 metabolites in lower mass ranges (~50-500 *m/z*) in CRISPRi strains revealed distinctive metabolic signatures (Fig 11D). Substrates and products were detected for all targeted metabolic pathways (amino acid biosynthesis, TCA, MEP pathway). Amino acid biosynthesis pathways are tightly regulated by negative feedback inhibition. Knockdown of genes involved in histidine, leucine, lysine and threonine biosynthesis resulted in substantial accumulations of substrates, confirming previously reported metabolic perturbations (Sander et al. 2019a, Donati et al. 2021). However, the metabolic profile of two strains remained inconclusive (*dxr*, *icd*). Previous research aligns with the findings for the *dxr* strain (Donati et al. 2021). Isocitrate dehydrogenase (Icd) enables *E. coli* to fast transition between the TCA cycle and the glyoxalate bypass, requiring constant adaptation of enzyme activity (LaPorte et al. 1984, Walsh and Koshland 1985). Investigating this gene is therefore more challenging due to rapid changes. The analysis of metabolic profiles can be a potent tool, not only for gene-function analysis but also for identifying genes involved in antibiotic resistance (Anglada-Girotto et al. 2022) and for metabolic engineering (Li et al. 2020).

To capture rapid changes, real-time metabolome profiling has been demonstrated by injecting living cells. Adaption of previous methods (Bujara et al. 2011, Link et al. 2015) enabled real-time detection of the ubiquitin-barcode (Fig 12B). When comparing extracted and real-time deconvoluted spectra, purification appeared to increase the number of detected features (Fig 12C). Incorporating liquid chromatography columns, such as C-18, can further enhance sample purity and barcode detection. However, flow-injection mass spectrometry (FI-MS) offers distinct advantages over chromatography-MS, primarily in terms of speed (30 sec run time) and method simplicity. This allows for the screening of a large number of strains, which would be impractical with chromatography, typically limited to the analysis of only 20-50 samples per day. Nevertheless, the complex sample matrix, affects the quantitative assessment of annotated metabolites and barcodes due to ion suppression, fragmentation, and the presence of isomers (Alseekh et al. 2021).

One key advantage of utilizing sgRNA sequences for strain validation is the potential to quantify strains within a pooled approach based on their sgRNA copy numbers. To further enhance ubiquitin detection, one avenue for improvement is the incorporation of a quadrupole into the mass spectrometer, transforming it into a QTOF instrument. This enhancement would enable filtering and fragmentation of metabolites (MS2 scans). Analysing fragmented ubiquitin-barcodes for their amino acid distribution represents an exciting opportunity to add an additional layer of ''sequencing'' to our analysis.

## 2 Quantification of sugar phosphate isomers in *E. coli* central metabolism with LC-MS/MS

To achieve separated detection of the isomers I first diluted authentic standards for hexose phosphates namely fructose 6-phosphate (f6-p), glucose 6-phosphate (g6-p), fructose 1-phosphate (f1-p), and glucose 1-phosphate (g1-p) and pentose phosphates including ribose 1-phosphate (r1-p), xylulose 5-phosphate (xu5-p), ribulose 5-phosphate (ru5-p), and ribose 5-phosphate (r5-p) to a final concentration of 10 µM (Table 4). I performed tandem mass spectrometry coupled to liquid chromatography (LC-MS/MS). The LC-column had to be compatible with LC-MS/MS methods for high-throughput quantification of primary metabolites (Guder et al. 2017). I expected that hydrophilic interaction chromatography (HILIC) is particularly suitable for separating sugar phosphates due to their polar and hydrophilic properties (Hemström and Irgum 2006). Therefore, I selected the HILICpak VG-50 2D (shodex) column and applied reversed phase chromatography for separation. I developed two liquid chromatography methods: A for hexose phosphates with long gradients lasting 63 min and Method B for pentose phosphates using 32 min gradients, as summarized in Table 3.

**Table 3 Liquid chromatography methods**

| **column** | **HILICpak VG-50 2D** | |
|---|---|---|
| dimension | 150 x 2 mm | |
| particle size | 5 µM | |
| pH | acidic | |
| temperature | 30°C | |
| method | **A hexose-p** | **B pentose-p** |
| run time | 63 min | 32 min |
| gradient (A/B) | 70/30 | 100/0 |
| flow rate | 0.28 mL $min^{-1}$ | 0.28 mL $min^{-1}$ |

To optimize data acquisition parameters for multiple reaction monitoring (MRM), the MassHunter Optimizer software (Agilent) was used. It automatically optimized collision energy by performing full MS/MS scans to select the best product ions. A minimum distance of 4 min distance was achieved between peaks of hexose-phosphates and 0.8 min for pentose-phosphates, resulting in distinct retention times for various sugar

phosphates with a minimum separation of 4 min for hexose phosphates and 0.8 minutes for pentose phosphates (Fig 12A, Table 4).

**Table 4 Retention time authentic standards**

| Name | Abbreviation | Retention time (min) |
|---|---|---|
| fructose-6-phosphate | f6-p | 45.0 |
| glucose-6-phosphate | g6-p | 49.1 |
| glucose-1-phosphate | g1-p | - |
| fructose-1-phosphate | f1-p | 53.5 |
| ribose-1-phosphate | r1-p | 23.3 |
| ribose-5-phosphate | r5-p | 19.3 |
| ribulose-5-phosphate | ru5-p | 20.5 |
| xylulose-5-phosphat | xu5-p | - |

To increase MS/MS signal intensities, I extended the acquisition time (dwell time) to 300 msec and raised the injection volume from 3 µL to 9 µL. Despite these adjustments, I could not detect any signal for glucose 1-phosphate (g1-p), and the identification of a unique peak for xylulose 5-phosphate (xu5-p) remained elusive (Fig 13A).

To assess detection limits, I tested various dilutions of a pooled standard mix (10 µM, 1 µM, 0.1 µM) containing hexose phosphates. Even at a lower concentration of 0.1 µM, I was able to distinguish individual peaks (see Fig 12B). The pool of hexose phosphates in *E. coli* growing in a glucose condition was estimated to be 8.8*10 µmol $L^{-1}$ (Bennett et al. 2009). To further enhance MS/MS signal intensities, I concentrated *E. coli* extracts tenfold.

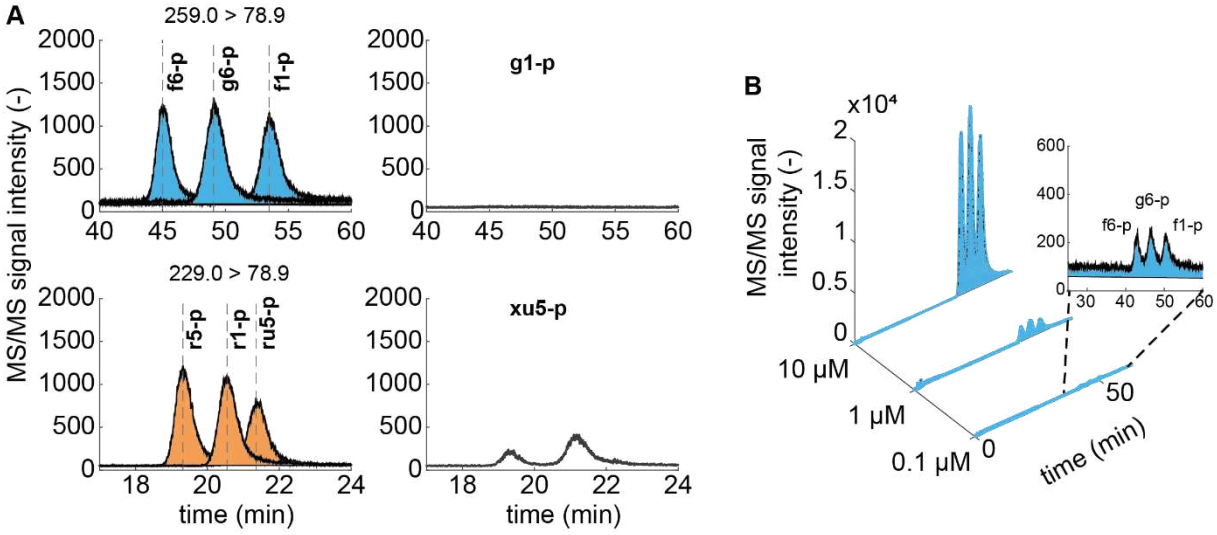



**Figure 13 Separation of authentic sugar phosphate standards by reversed phase chromatography**

**A** Chromatogram of authentic standards. Top: overlay chromatograms of f6-p, g6-p, f1-p, g1-p as authentic standards (10 µM) with method A hexose-p. MRMs used for detection of hexose-phosphates, precursor mass: 259.0 and product mass: 78.9. Dashed line indicates respective retention time as listed in Table 4. bottom: Overlay of chromatograms of r5-p r1-p, ru5-p and xu5-p as authentic standards (10 µM) with method B. MRMs used for detection of the respective compound are given (precursor mass: 229.0, product mass: 78.9). Dashed line indicates respective retention time as listed in Table 4. **B** Method sensitivity. F6-p, g6-p and f1-p were pooled to a final concentration of 10 µM, 1µM and 0.1 µM and measured using method A hexose-p. Chromatogram shows the MS/MS signal intensity for different dilutions.

## Quantification of sugar-phosphates in *E. coli*

Sugar-phosphate isomers serve as crucial intermediates in *E. coli* central metabolism. Hexose phosphates, specifically f6-p, g6-p, f1-p, and g1-p, participate in glycolysis (Fig 14A). Meanwhile, pentose phosphates such as r1-p, xu5-p, ru5-p, and r5-p act as intermediates in the pentose-phosphate pathway. To quantify these sugar-phosphates *in vivo*, I cultivated a CRISPRi control strain in minimal medium supplemented with glucose. This control strain expresses the complete CRISPRi machinery (dCas9 and sgRNA) without targeting any specific gene in the *E. coli* genome. I induced dCas9 expression with aTc at the start of cultivation and collected samples during the exponential growth phase (6.5 h). Utilizing two distinct methods (Method A for hexose-

p and Method B for pentose-p), I identified four peaks (Fig 14B). By comparing their retention times to those of authentic standards, I successfully identified f6-p, g6-p, r5-p, and ru5-p. Detection of f1-p and r1-p was not anticipated due to their limited presence in glucose conditions.

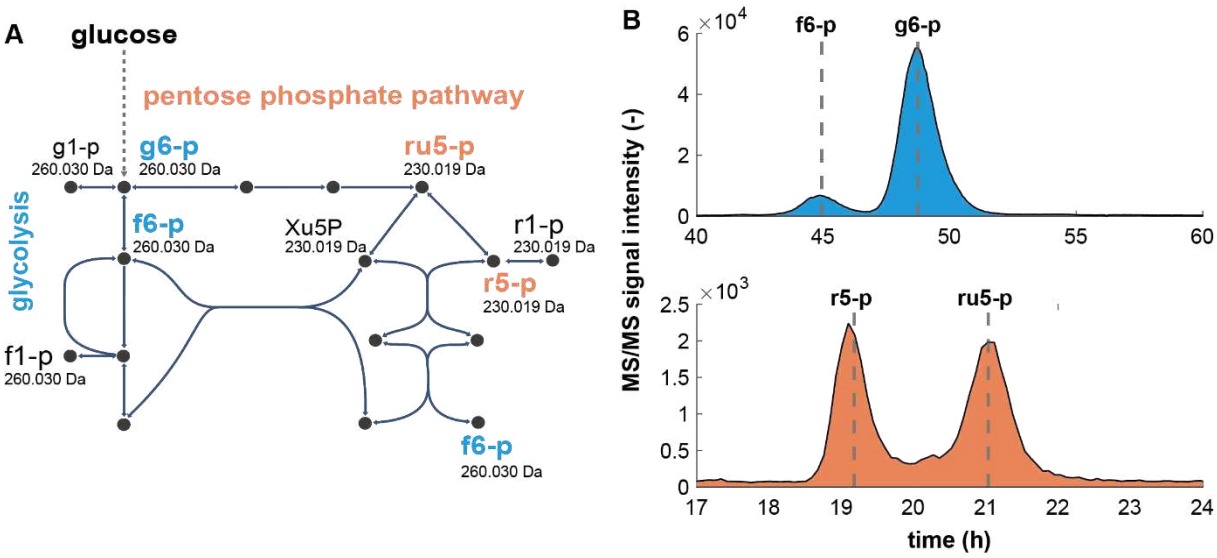



**Figure 14 Separation of sugar phosphate isomers *in vivo***

**A** *E. coli* central metabolism involves sugar phosphates. Hexose-phosphates are involved in glycolysis (blue) and have identical masses as indicated in Da. Pentose-phosphate pathway and pentose-phosphate isomers are indicated in orange. Abbreviations: Ribulose5-phosphate (ru5-p); Ribose1-phosphate (r1-p); Xylulose5-phosphate (xu5-p); Ribose5-phosphate (r5-p); glucose6-phosphate (g6-p); fructose6-phosphate (f6-p); fructose1-phosphate (f1-p); glucose1-phosphate (g1-p). The metabolic map is based on the EcoCyc database (Keseler et al. 2009). **B** Detected sugar phosphate isomers in glucose conditions. Top (blue): Chromatogram of hexose-phosphates of *E. coli* CRISPRi control strain grown in glucose conditions and measured using method A. aTc was added at the start of the cultivation to activate CRISPRi. Retention time of analytical standards (g6-p, f6-p) are indicated as dashed line. Bottom (orange): Chromatogram of pentose-phosphates of *E. coli* CRISPRi control strain grown in glucose conditions and measured using method B. Retention time of analytical standards (r5-p, ru-5P) are indicated as dashed lines and named respectively.

## Targeted gene knockdown to perturb sugar phosphate levels *in vivo*

CRISPR interference enables precise gene suppression, leading to localized alterations in cellular metabolite levels (Donati et al. 2021). In this study, I used 21 distinct CRISPRi strains to target genes in central metabolism, specifically enzymes involved in glycolysis and the pentose phosphate pathway (Fig 15A).

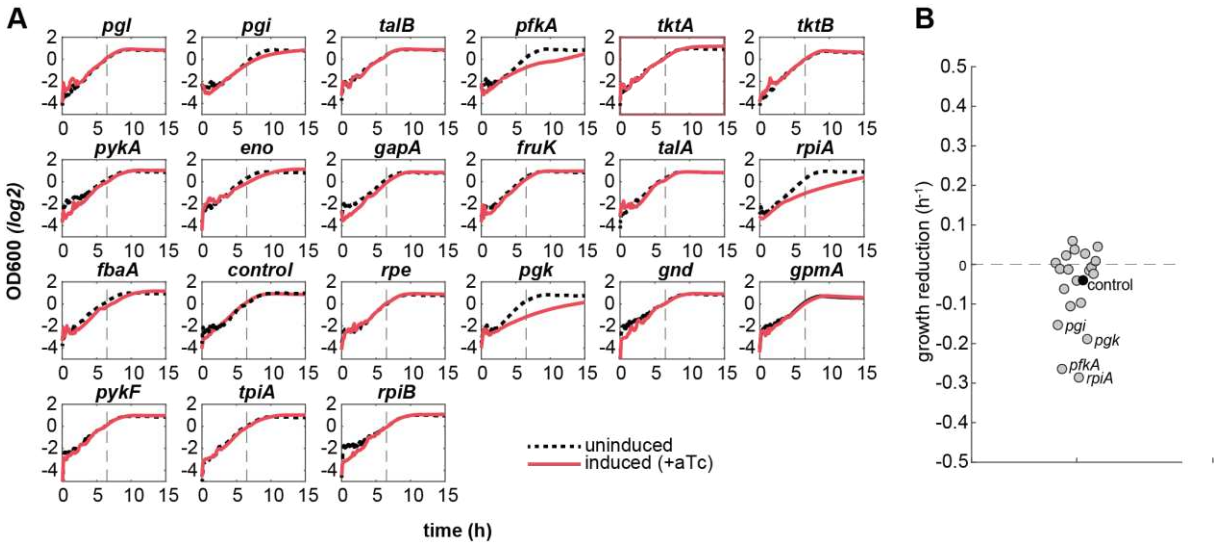



**Figure 15 Knockdown of selected genes caused growth defect**

**A** Growth of the 21 CRISPRi strains. Uninduced cultures are shown in blue. Induced cultures are shown in red (200 nM aTc was supplemented at time = 0 h). Growth curves show means of n = 3 cultures. Background colour indicate the growth rate at 6.5 h (time of sampling). Growth rates were estimated using linear regression with the at least four time points of growth curves. Dashed line indicates 6.5 h, time point of sampling. **B** Growth reduction was determined by subtracting the growth rates of induced cultures from those of uninduced cultures at 6.5 h (see A). Each dot represents one CRISPRi strain, the control strain is indicated in black.

In glucose-rich conditions, eight of these enzymes catalyse reactions utilizing sugar-phosphates as substrates (Rpe, RpiA, RpiB, Pgi, TktA, TktB, PfkA, PfkB), while an additional eight generate sugar-phosphates as products (TalA, TalB, Gnd, Pgi, RpiA, RpiB, TktA, TktB). Furthermore, I included two reactions where sugar-phosphate isomers serve as both substrates and products. The first of these reactions is catalysed by phosphoglucoisomerase (Pgi), which converts g6-p into f6-p. *E. coli* possesses two ribose-5-phosphate isomerases (RpiA, RpiB) that interconvert r5-p into ru5-p, involving pentose-phosphates as both substrates and products. Among the 21 targeted genes, seven are essential for *E. coli* growth on glucose (*pgi*, *tktA*, *eno*, *gapA*, *fbaA*, *rpe*, *pgk*) (Rousset et al. 2018). Knocking down six of these target genes (*rpiA*, *pfkA*, *pgk*, *pgi*, *eno*, *fbaA*) resulted in a significant phenotype, causing a growth reduction exceeding 20% (Fig 15B). However, the knockdown of most other target genes did not affect growth. Substantial changes in substrate and product levels have been previously

observed after 4.5 hours (Donati et al. 2021), regardless of the growth phenotype. Therefore, we anticipate significant alterations in metabolite levels. I cultured the 21 CRISPRi strains in shake flasks (with one replicate) using minimal medium supplemented with glucose for relative quantification of sugar phosphates (Fig 16A).

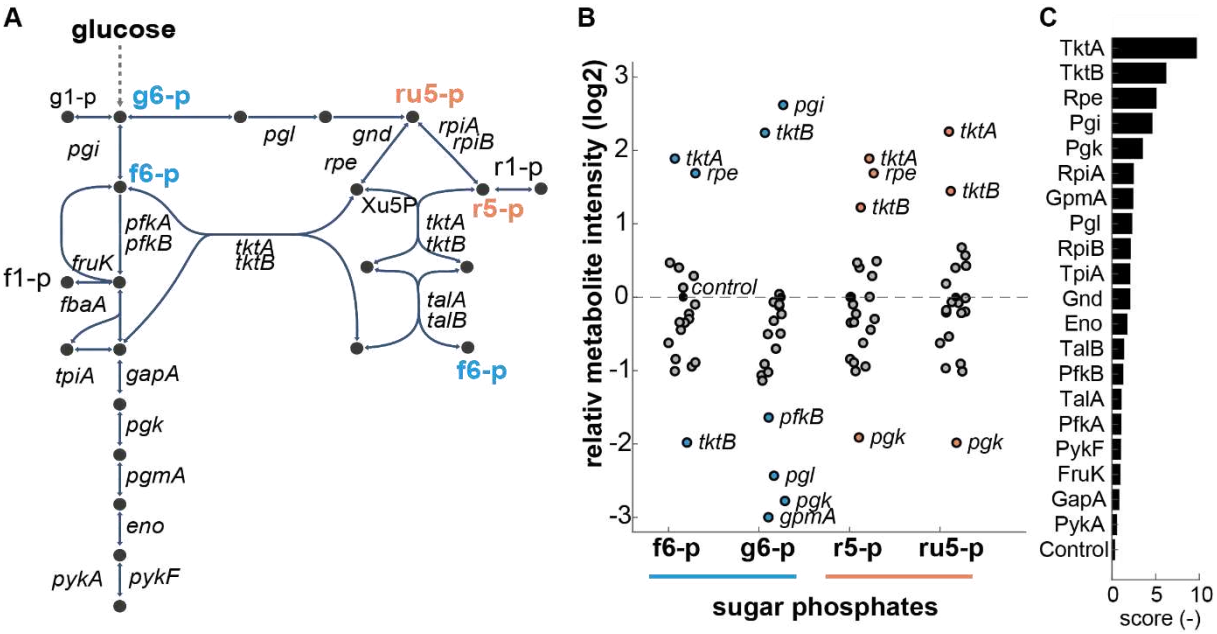



**Figure 16 CRISPRi enforces changes in sugar-phosphates**

**A** Metabolic map shows the target genes of 20 CRISPRi strains. All targets genes are involved in central metabolism namely glycolysis and pentose phosphate pathway. Targeted enzymes involve sugar phosphates as products and or substrates. Hexose-phosphates are highlighted in blue and pentose-phosphates in orange. The metabolic map is based on the EcoCyc database (Keseler et al. 2009) **B** Relative sugar-isomer intensity. Metabolite levels (f6-p, g6-p, r5-p, ru5-p) are shown as log2 fold change compared to the control strain. Samples were collected after 6.5 h of cultivation in shake flasks in the presence of aTc to induce the CRISPRi mediated knockdown. Individual dots represent data of single knockdowns. Black dots show the control strain. CRISPRi strains that induce more than 3fold change in hexose-phosphates levels are highlighted in blue, target-genes inducing more than 3fold changes in pentose-phosphates are indicated in orange. **C** Ranking CRISPRi strains by perturbation. Bar plot shows CRISPRi strains sorted according to their perturbation strength in descending order. Score represents sum of absolute changes in all four isomers in 21 CRISPRi strains.

After 6.5 hours, I collected samples for LC-MS/MS measurements, enabling us to detect four sugar phosphates (f6-p, g6-p, r5-p, ru5-p) (Fig S2). By comparing the relative intensities of these metabolites, I identified eight target genes that induced more than a 3-fold increase or decrease in g6-p or f6-p (*tktA*, *tktB*, *pgi*, *rpe*, *pgl*, *pgk*,

*gpmA*, *pfkB*) (Fig 16B). In the case of r5-p and ru5-p, I identified four target genes that caused changes exceeding 2-fold (*tktA*, *tktB*, *pgk*, *rpe*). I scored the perturbation strength, and TktA, TktB, Rpe, Pgi, and Pgk emerged as the five key enzymes in glycolysis and the pentose phosphate pathway when *E. coli* was grown with glucose as the carbon source. All of these enzymes either utilize sugar-phosphates as substrates or produce them as reaction products.

**Investigating Sugar Phosphate Changes: A Systems-Level Approach with Different Carbon Sources**

I selected key CRISPRi strains from the previous screen: *tktA*, *tktB*, *rpe*, and *pgi*. Additionally, I included the control strain and two strains targeting the bypass reaction of the *rpe* strain (*rpiA*, *rpiB*). These strains were cultured with either glucose or xylose as the sole carbon source (Fig 17A). While glucose ($C_6H_{12}O_6$) enters metabolism through glycolysis, I introduced xylose ($C_5H_{10}O_5$) as an alternative carbon source. Xylose undergoes a two-step conversion into xylulose-5-phosphate (xu5-p) and enters the pentose-phosphate pathway (Fig 17A). The pentose-phosphate pathway (PPP) is the only pathway through which *E. coli* can utilize sugars like d-xylose, d-ribose, and l-arabinose, which cannot be metabolized by other means (Wood, 1985; Sprenger, 1995; Lin, 1996). The strains were grown in triplicates in shake flasks, and after 6.5 hours, I collected samples for LC-MS/MS measurements, detecting four sugar phosphates (f6-p, g6-p, r5-p, ru5-p). Many of the strains did not show significant changes (more than a 2-fold difference) when compared to the control strain (Fig 17B). R5-p is the substrate of transketolase 1 (TktA) and 2 (TktB). TktA was previously identified as responsible for the majority of transketolase activity (Iida et al. 1993). My findings show the only the knockdown of TktA led to the accumulation of the substrate and ru5-p, confirming previous research findings. The enzyme Rpe converts ru5-p to xu5-p, but *E. coli* metabolism harbors a bypass reaction, turning ru5-p into r5-p catalysed by RpiA and RpiB. Knocking down resulted in the accumulation of the ru5-p and r5-p substrates, indicating the activation of the bypass when Rpe was reduced. We observed a similar response in xylose conditions. In turn, knockdown of RpiA and RpiB did not affect intracellular ru5-p indicating low flux trough the reaction in glucose and xylose conditions. In glucose conditions, the *pgi* strain exhibited a more than 4-

fold increase in both g6-p and f6-p. These compounds are both substrates and products of the reaction catalysed by Pgi.

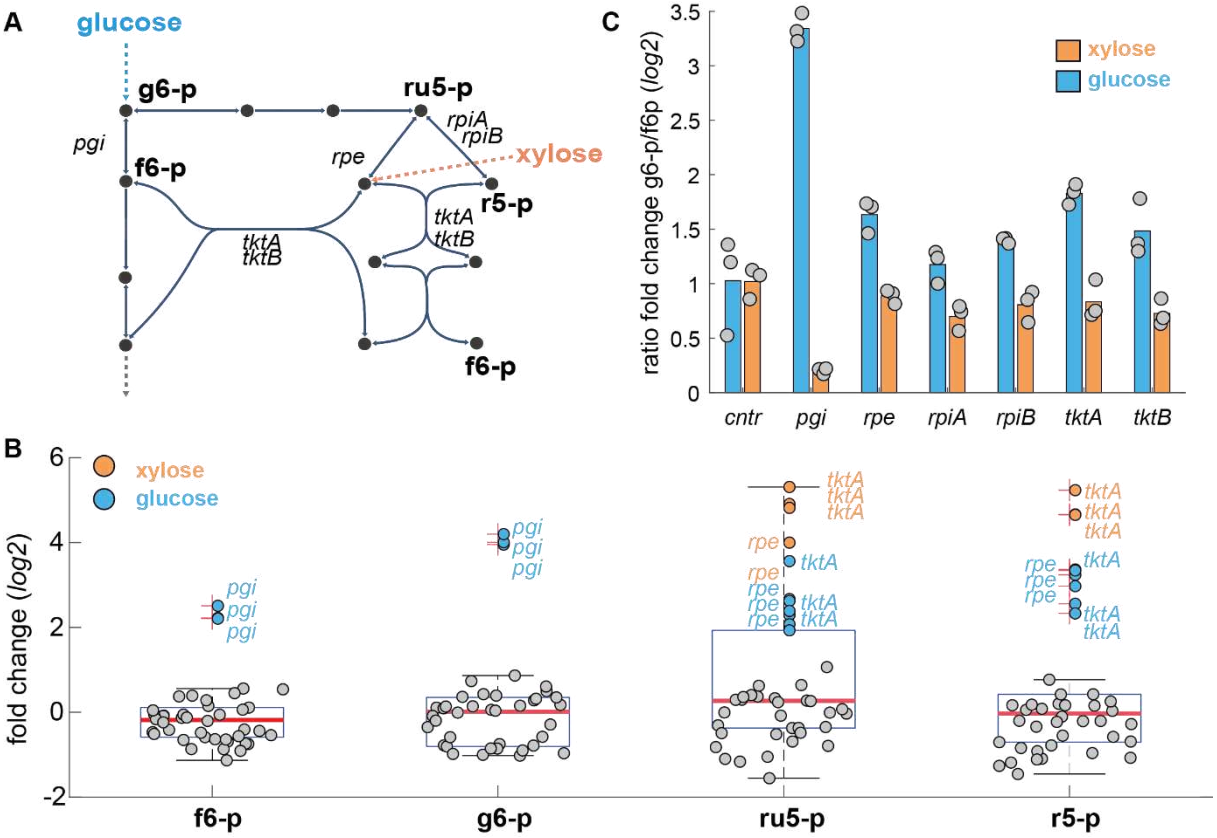



**Figure 17 Carbon source affects metabolic footprint of CRISPRi strains**

**A** Metabolic map of growth with glucose or xylose as carbon source. Metabolic map shows selected CRISPRi target genes and measured sugar phosphates. Measured hexose-phosphates and pentose-phosphates are abbreviated: glucose6-phosphate (g6-p); fructose6-phosphate (f6-p); ribose5-phosphate (r5-p); ribulose5-phosphate (ru5-p). The Map is based on the EcoCyc database (Keseler et al. 2009). **B** Box plots show the fold change (log2) of f6-p, g6-p, ru5-p, and r5-p sugar phosphates in 6 CRISPRi strains (shown in A and B) grown with either xylose or glucose as carbon sources. The fold change was calculated relative to the CRISPRi control strain growing with the respective carbon source. Each boxplot represents the interquartile range (IQR), with the horizontal line denoting the median. Whiskers extend to 1.5 times the IQR, and individual outliers are displayed as crosses beyond the whiskers. Individual samples are shown additionally as dots, with blue dots representing cultures grown with xylose and orange dots representing cultures grown with glucose as the carbon source. Strains are grown 6.5 h in shake flasks with 5 g/L either xylose or glucose in the presence of aTc (200 nM) to activate.

To compare their levels under different conditions, we examined the ratio of f6-p/f6-p (Fig 17C). When glucose was present, g6-p increased, whereas in the presence of

xylose, only the *pgi* strain displayed inverted ratios (with a p-value of less than 0.05). I assessed how different carbon sources, specifically glucose and xylose, influenced the metabolic profiles of CRISPRi strains. While some strains showed significant changes in sugar phosphate levels, others remained relatively unaffected, indicating context-dependent metabolic responses and highly specific metabolic profiles for each knockdown.

**Discussion**

In this study, I present a method for quantifying sugar phosphate isomers within the central metabolism of *E. coli*. Pentose phosphates and hexose phosphates are important intermediates of glycolysis and the pentose phosphate pathway (PPP). Separate quantification of the isomers is challenging as they have the same masses and similar chemical properties. While mass spectrometry-based approaches alone cannot distinguish these isomers effectively, spatial separation through liquid chromatography followed by tandem mass spectrometry enabled their accurate quantification (LC-MS/MS). Authentic standards were used to identify MRMs and to optimize chromatographic step (Fig 13A, 13B).

Previous studies have separated sugar phosphates using LC-MS/MS with the aid of ion-pairing reagents (Buescher et al. 2010). Ion-pairing reagents like 3-tributylamine are added to the mobile phase, and can effectively improve separation and peak shape. One disadvantage to use these reagents are ion suppression issues, affecting the sensitivity and accuracy of mass spectrometric detection and quantification over time (Gustavsson et al. 2001). A second throwback are extensive column equilibrations, prior and after measurement (Buescher et al. 2009, 2010). To circumvent ion pairing, long-gradients had to be used for the separation of hexose phosphates. Therefore, in comparison the previously reported separation the Method A for hexose phosphates was about 25 min longer which resulted in a total run time of 63 min (ion pairing: 25 min, Buescher et al. 2010). Separation of pentose phosphates was achieved in 32 min which is close to the measuring time with ion pairing as shown in previous experiments (25 min, Buescher et al. 2010).

Glycolysis and the PPP are integral components of *E. coli* central metabolism. Therefore, quantifying intermediates in these pathways holds great significance and

interest. Notably, the PPP is challenging to study due to its high turnover rate, resulting in low metabolite abundance. In our investigation, we selected 21 CRISPRi strains to target specific enzymes in glycolysis and the PPP, inducing perturbations in metabolite levels. Screening these strains for changes in sugar phosphates revealed three key enzymes: Pgi, TktA, and Rpe (Fig 16C). We further characterized these enzymes under two different conditions (xylose and glucose) and uncovered three distinct mechanisms.

First, my assessment of transketolase isoenzymes TktA and TktB indicate that TktA is responsible for major transketolase activity (Fig 17B), as confirmed before (Iida et al. 1993). This result highlights the importance of TktA in controlling metabolite flux within the PPP. Secondly, my investigations into the Rpe enzyme were guided by the increasing r5-p levels (Fig 17B). Increased r5-p levels indicated the activation of a bypass reaction, as r5-p is the substrate of RpiA and RpiB in this alternative pathway. Thirdly, when evaluating the impact of Pgi knockdown, I observed an inversion of substrate/product ratio under different carbon source conditions (Fig 17C). This indicates a directional change of the reaction, illustrating the dynamic adaptability of metabolic pathways to changing conditions.

Metabolic systems exhibit robustness against perturbations in enzyme levels, and metabolic changes primarily affect substrates, products, and regulators. As a result, many of the strains did not display significant alterations in sugar phosphate levels. However, the knockdown of the three enzymes (TktA, Rpe, Pgi), led to significant changes in metabolite profiles because they directly involve sugar phosphates as substrates, showing we can use this method to analyse metabolic fingerprints.

These results not only provide valuable insights into the functional roles of specific enzymes but also highlight the versatility of our LC-MS/MS method in characterizing metabolic networks. By quantifying sugar phosphate isomers, I demonstrated the potential of this approach for metabolic profiling, enabling the systematic investigation of cellular responses to genetic perturbations and environmental changes. Such comprehensive metabolic profiling holds great promise for understanding the robustness and adaptability of metabolic systems across diverse organisms. I quantified sugar phosphates to gain insights into the central metabolism of *E. coli*,

specifically PPP and glycolysis. These pathways are evolutionarily conserved, making our findings relevant to a wide range of organisms. Furthermore, our method's applicability extends to other metabolic pathways due to substantial overlap, such as with the Calvin cycle, one of the most common cycles for $CO_2$ fixation in biology.

# 3 Metabolic changes that enable *E. coli* to use $CO_2$ as C-source

*Figures 18 and 20 are part of a collaborative work currently in peer review process:*

Ben-Nissan R, Milshtein E, **Pahl V**, De Pins B, Jona G, Levi D, Yung H, Nir N, Ezra D, Gleizer S, Link H, Noor E, Milo R. 2023. Autotrophic growth of *E. coli* is achieved by a small number of genetic changes. preprint. elife

### Mutation in Phosphoglucoisomerase reduces enzyme activity

To enable chemoautotrophic growth, a mutation in the Phosphoglucoisomerase of *E. coli* was essential. Pgi catalyses a branchpoint reaction of the rPP cycle and we suspected the mutations to decrease enzyme activity. Two strains were designed one with a wildtype Pgi and one containing the $Pgi^{H385Y}$ mutation. The strains were cultivated in semi autotrophic conditions with $CO_2$ and a second carbon source (xylose). The $Pgi^{H385Y}$ mutation is essential for autotrophic growth, therefore feeding xylose was necessary after reverting *pgi* to its wildtype allele. The enzyme catalyses a branchpoint reaction of the $CO_2$ fixation cycle (Fig 18A).

To assess the impact of the mutation for enzyme activity and thereby carbon fixation, the strains were grown in shake flasks to collect samples for targeted metabolomics (LC-MS/MS). Quantification of the substrate (glucose 6-phosphate) and the product (fructose 6-phosphate) of the reaction is challenging since both metabolites have the same masses and similar chemical properties. Here, I use LC-MS/MS with extended gradients to separate the isomers using chromatography, prior to mass spectrometry (Figure 16B). When comparing the Ratio of substrate to product (f6-p/g6-p) the strain with a wildtype *pgi* found that the relative abundance of f6-p is twofold higher than g6-p. However, when introducing the Pgi mutation the ratio increases to 16-fold f6-p to g6-p (Fig 18B). Intermediates of the rPP cycle, namely s7-p and the pool of pentoses (p5p) increased more than 3-fold (Fig 18C). These show the impact of the Pgi mutation on enzyme activity, leading to a significant alteration in the substrate-to-product ratios and the accumulation of crucial intermediates (s7-p, p5p) within the carbon fixation cycle. And give new insights in metabolic adaptions that lead to chemoautotrophy in *E. coli*.

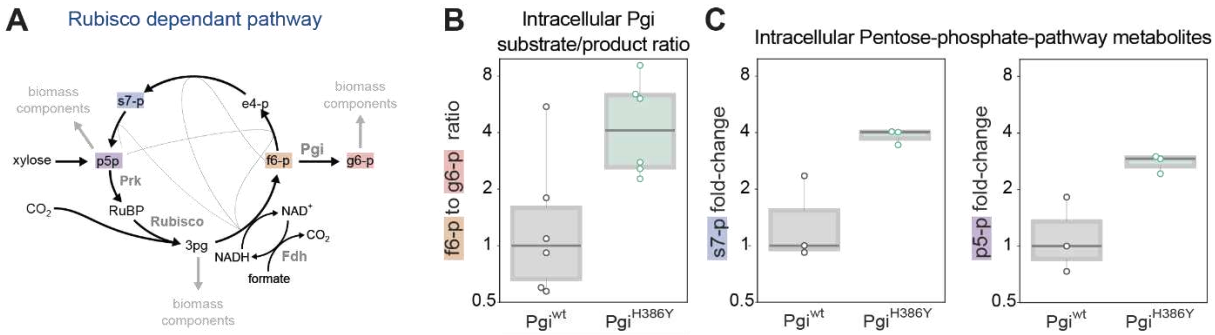



**Figure 18 Pgi$^{H385Y}$ mutation decreases enzyme activity and efflux of intermediates from carbon fixation cycle**

**A** Metabolic scheme of carbon fixating cycle components when growing with $CO_2$ and xylose as second carbon source (rubisco dependent). Pgi catalyses the branchpoint reaction from f6-p to g6-p. **B** Intracellular Pgi substrate/product ratio. Relative intracellular concentration of f6-p and g6-p of the ancestor background harbouring a Pgi$^{H386Y}$ mutation compared to the ancestor with a wild-type Pgi, both grown in rubisco dependent conditions, n = 6 cultures p-value < 0.05. Ratio was calculated. **C** Relative intracellular abundance of pentose-phosphate-pathway metabolites - sedoheptulose-7P (s7-p) and total pool of Pentose-phosphates (Ribulose-5P, Ribose-5P, Ribose-1P, and Xylulose-5P - denoted p5-p) in a Pgi$^{H386Y}$ strain versus the Pgi wild-type strain, both growing in a rubisco-dependent manner, n = 3 cultures. Raw chromatograms are shown in Fig S4.

## Uncoupling formate-driven energy module

Along with the branch-point mutation in Pgi, the compact autotrophic *E. coli* requires two additional mutations, one in the gene *crp*, a global metabolism regulator and in *rpoB*, a component of the transcriptional machinery. I decided to address them in a double mutant. In order to enable *E. coli* to grow autotrophically the cell needs to adapt its catabolism towards $CO_2$ as carbon source and NADH as energy source, due to their global effect I hypothesized that these mutations might be involved in one of these two processes or even combined. To assess the impact of formate-driven energy metabolism, I aimed to decouple the energy module by inserting it into *E. coli* BW25113 (wild type) (Fig 19A). The module consisted of a plasmid expressing formate dehydrogenase (Fdh) from *Pseudomonas sp. 101* (Egorov et al. 1980). The strain was cultivated in shake flasks in minimal medium supplemented with glucose (5 $gL^{-1}$) or xylose (2 $gL^{-1}$). I, tested glucose because it's the preferred carbon source for *E. coli* and xylose because I supplement xylose as second carbon source to autotrophic strains. Formate (30 mM) was supplemented to the culture to test Fdh activity. Strains were grown in microtiter plates in the presence and absence of formate to collect

samples for LC-MS/MS. Targeted metabolomics detected 92 intracellular metabolites including $NAD^+$ (substrate) and NADH (product) of the reaction (Fig 19B).

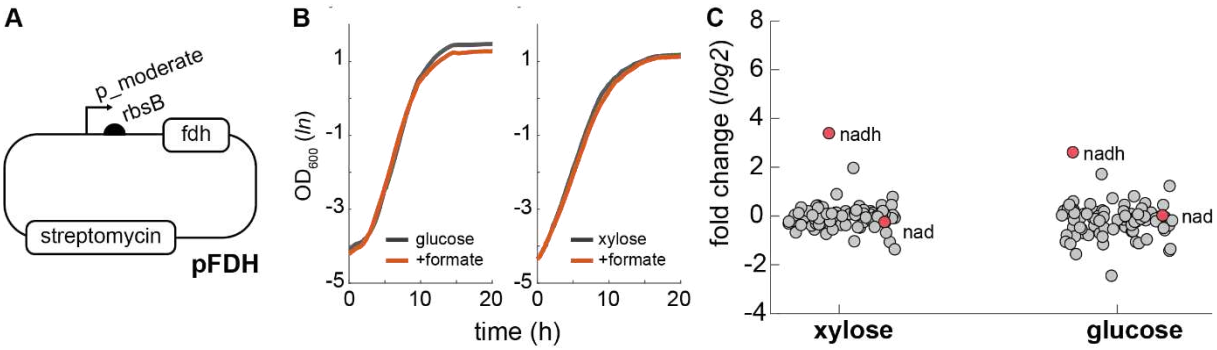



**Figure 19 Format-driven energy module**

**A** Decoupling of energy module from chemoautotrophy by inserting pFDH into *E. coli* BW25113. Formate dehydrogenase (fdh) is expressed constitutively with a synthetic promotor (p_moderate) and a strong rbs (rbsB). **B** Growth of *E. coli* BW25113 expressing pFDH. I cultivated the strains in mirotiterplates minimal medium supplemented with glucose (5 g/L) or xylose (5 g/L) in absence and presence of formate (red). Means of n = 3 cultures are shown. **C** Intracellular concentration of 92 metabolites with 2 different carbon sources. Metabolite levels are shown as log2 fold change of cultures growing with and without formate. Strains were cultivated in 12-well plates. Data are represented as mean (n=3). NADH and NAD are shown as red dots. Raw chromatograms are shown in Fig S3.

The *E. coli* strain expressing the Fdh grew nearly identical in both conditions, the presence and absence of formate (Fig 19B). This suggests that formate supplementation, while serving as an energy source, may not significantly impact the overall growth of the strain under these conditions. Most metabolite levels remained constant (with changes <2 fold) irrespective of the carbon source. Notably, NADH levels increased significantly (>5 fold) in both conditions. NAD and related co-factors (NADP, NADPH) remained unchanged compared to the control condition without formate. These findings underscore the functionality of Fdh in generating NADH, although *E. coli* seems unable to use this excess NADH for enhanced growth. The conversion of NADH to ATP likely hinges on factors such as ATPase activity and respiratory chain efficiency, which may be limiting in this context.

**Mutations in regulatory genes increase NADH and indicates role in formate driven energy metabolism**

$Crp^{H22N}$ and $RpoB^{A1245V}$ mutation were inserted into *E. coli* BW25113 pFdh (decoupled energy module), to investigate NADH levels in a double mutant strain. The strains were cultivated in minimal medium with xylose (2 $gL^{-1}$) in presence and absence of formate (30 mM). The double mutant strain showed a >25% increase in final OD600 compared to the wild type transformed with the pFDH plasmid (Fig 19B). A long (10 h) lag phase is observable when growing the double mutant. Interestingly, when the double mutant was grown exclusively on xylose without formate, the increase in yield was not observed, and the extended lag phase was less pronounced (Fig 20A).

Given that NADH serves as the electron donor for carbon fixation and is a product of formate consumption through Fdh, I used targeted metabolomics to measure the NADH/NAD+ ratio in these strains. All measured ratios were normalized to the wild-type strain (with Fdh) growing without formate. Upon the addition of 30 mM formate, which can act as an electron donor through its oxidation to $CO_2$ by Fdh, the NADH/NAD+ ratio increased by over 20-fold in the double mutant, a significantly greater change than the ≈10-fold increase observed in the wild-type strain. These findings show that regulatory mutations impact NADH levels indicating their roles in energy metabolism. These results indicate that regulatory mutations enable adaption of energy metabolism for chemoautotrophic growth.

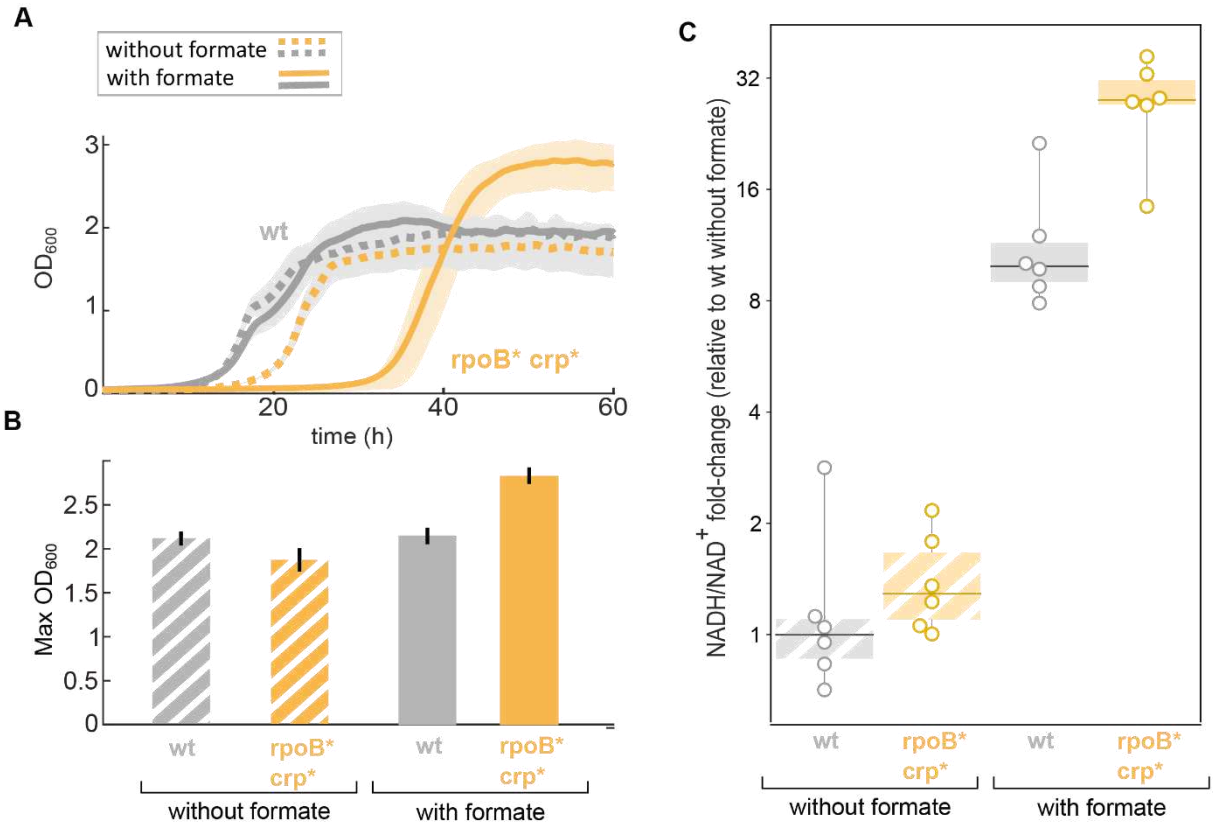



**Figure 20 Mutations in rpoB and crp increase yield and intracellular NADH/NAD+ levels in fdh-expressing *E. coli* in the presence of formate**

**A** Growth of double mutant with and without additional supply of energy (NADH). Strains were grown in microtiterplates (96-Ill plates) in minimal medium supplemented with 2 $gL^{-1}$ xylose and 30 mM formate (solid line) or in 5 the absence of formate (dashed line) with 10% $CO_2$ in the atmosphere, n=6 replicates. Lines represent the mean, light-grey background represents the standard deviation. **B** Maximal OD600 of the wild-type (grey) and mutant (orange) strains grown in 1$gL^{-1}$ xylose and 40 mM formate or in the absence of formate. Bar heights represent averages (n=6) of the median of the top 10 OD600 measurements of each replicate (end point measurements). Error bars represent standard deviation. **C** $NADH/NAD^+$ ratios as fold-changes relative to wild-type without formate. Intracellular $NADH/NAD^+$ ratio of the wild-type (grey) was compared to the mutant (orange). The strains were grown in 2$gL^{-1}$ xylose and 30 mM formate, or in the absence of formate. The y-values are. Boxes represent 25-75 percentile ranges and dark lines represent median values. Raw chromatograms are shown Fig S5.

## Discussion

Autotrophic growth in microorganisms, such as *E. coli*, has gained increasing attention due to its potential applications in biotechnology and bioengineering. In this study, we investigate the metabolic changes that enable *E. coli* to utilize carbon dioxide ($CO_2$) as a sole carbon source, investigating the genetic adaptations required for

chemoautotrophic growth. The achievement of autotrophy in *E. coli* necessitates a limited set of genetic modifications. These modifications include mutations in phosphoglucoisomerase (Pgi), cyclic AMP receptor protein (Crp), and RNA polymerase beta-subunit (RpoB).

One mutation was identified in the Pgi enzyme, which catalyses a key branchpoint reaction in the ribulose-5-phosphate (rPP) cycle. Here I compare two strains: 1) with a wild-type Pgi and 2) including the $Pgi^{H385Y}$ mutant. When grown under semi-autotrophic conditions with $CO_2$ and xylose, the mutant strain showed changes in the substrate-to-product ratios of g6-p to fructose f6-p. The $Pgi^{H385Y}$ mutation resulted in a 16-fold increase in the f6-p to g6-p ratio, accompanied by the accumulation of key intermediates of the rPP cycle, such as s7-p and pentose-phosphates (Fig 18). In a previous study, mutations in Pgi was found to be essential for the function of the rPP autocatalytic cycle and was already part of a minimal set for hemi-autotrophic growth (Herz et al. 2017). As our results suggest, the mutated enzyme shows a significant reduction in activity, which decreases the efflux from the rPP cycle to biosynthesis. This indicates the Pgi mutation might generate a new, higher, steady-state concentration for the rPP cycle metabolites, thus stabilising the cycle.

In addition to the Pgi mutation, two other mutations, namely $Crp^{H22N}$ and $RpoB^{A1245V}$, were identified as essential for autotrophic growth. To investigate the role of these mutations in energy metabolism and carbon source adaptation, we constructed a double mutant strain incorporating both $Crp^{H22N}$ and $RpoB^{A1245V}$ mutations. Our experiments revealed that these mutations significantly increase the NADH/$NAD^+$ ratio (Fig 20C).

NADH and NAD are important cofactors for the reaction catalysed by GAPDH in the non-native rPP cycle. The net flux of the reaction is reversed in the direction of D-glyceraldehyde-3P compared to native glycolytic conditions. Increasing NADH could facilitate the reverse and offer one solution to thermodynamic and kinetic constraints. Increasing cofactor levels can help to thermodynamically shift the reaction. Since the reaction rate depends on substrate and enzyme levels, increasing the substrate pool, as the ratio suggests, can additionally increase flux in the desired direction. This hypothesis is in line with prior research indicating that maintaining high cofactor levels

is critical for the rPP cycle's stability (Haverkorn Van Rijsewijk et al. 2011, Janasch et al. 2018). These results contribute to the groundwork for future research and improvements in this cycle and offers an avenue for future rational designs. Since both Crp and RpoB affect gene expression globally, the role they play in the autotrophic phenotype likely extends beyond controlling the NADH level. In the future, this could be explored with various methods, such as proteomics, transcriptomics, ChIP-seq, or other assays for transcription factor activity. To transition to chemoautotrophic growth, metabolism has to adapt, changing the activity of many cellular components. Mutations in regulatory genes such as *crp* and *rpoB* could perturb the activity of hundreds of components simultaneously and might be the only viable solution to implement several required changes simultaneously.

In conclusion, our research provides a new insight in the genetic changes that enable *E. coli* to grow chemoautotroph. By dissecting the roles of the Pgi, Crp, and RpoB mutations, we elucidate the intricate interplay between enzyme activity, substrate-to-product ratios, and energy metabolism. This knowledge opens avenues for the rational design of autotrophic strains with enhanced capabilities for bioproduction and environmental applications.

# 4 Temperature controlled two-stage arginine production in *E. coli*

Data for Figure 21 is part of a collaborative work and published:

Schramm T, Lubrano P, **Pahl V**, Stadelmann A, Verhülsdonk A, Link H 2023. Mapping temperature-sensitive mutations at a genome scale to engineer growth switches in *Escherichia coli*. Molecular Systems Biology e11596

### Temperature to switch from growth to arginine production

Temperature sensitive valves can be applied to overproduce value added compounds. The temperature sensitive DnaX$^{L289Q}$ mutation affects a subunit of a DNA polymerase III. At 30°C the mutant can grow, while at 42°C growth is inhibited (Schramm et al. 2023). Therefore, I selected the DnaX$^{L289Q}$ mutation to control growth by temperature. The arginine overproduction strain was constructed by removing transcriptional feedback of the arginine repressor (Δ*argR*), which results in overexpression of enzymes in the arginine biosynthesis pathway (Sander et al. 2019b). Additionally, to remove allosteric feedback inhibition by arginine a point a point mutation was inserted into the first enzyme of the arginine pathway (ArgA$^{H15Y}$). To facilitate arginine export, the arginine transporter ArgO was overexpressed from a plasmid (Fig 21A). The final arginine production strain harboured four modifications: 1) a point mutation in DnaX$^{L289Q}$ 2) deletion of the Δ*argR* gene 3) a point mutation in ArgA$^{H15Y}$ 4) overexpression of ArgO.

To assess arginine production in this strain, precultures were prepared at 30°C and the final cultivation of the strain was performed at 30°C in minimal medium supplemented with glucose (5 gL$^{-1}$). Cultivation at 42°C was performed in shake flasks for 56 h and optical density was measured using a spectrophotometer. Arginine concentration was determined using LC-MS/MS and authentic arginine standard.

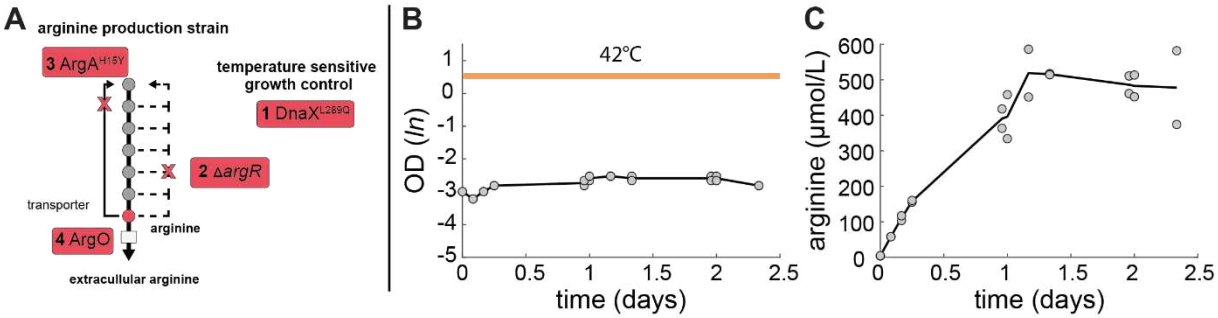



**Figure 21 Decoupling growth from arginine production by DnaX$^{L289Q}$**

**A** Scheme arginine production strain. DnaX$^{L289Q}$ mutation was inserted for temperature sensitive growth control. To dysregulate the arginine pathway *argR* was deleted (removes transcriptional feedback) the ArgA$^{H15Y}$ mutation was inserted (removes allosteric feedback). Additionally, the arginine exporter ArgO was strongly expressed from a plasmid. **B** Temperature to switch growth. Optical density (ln) of arginine overproduction strain cultivated for 56 h at 42°C. The strain was cultivated in minimal medium supplemented with glucose in shake flasks. Dots indicate cultures of n=2, line connects the mean. **C** Arginine production during growth arrest. Arginine concentration (µmol/L) during the same cultivation (shown in B). Arginine levels were quantified in the whole culture broth by LC-MS/MS and calibrated with an authentic arginine standard (Supplementary Figure S7). Dots show samples from two replicates (n = 2), and the line connects the means. Raw chromatograms and calibration curve are shown in Fig S7A, S7B.

I started the culture at OD600 0.05 and during 56 h of cultivation at 42°C the optical density remained stable (Fig 21B). During the first 24 h of growth arrest, arginine concentration increased and a constant specific arginine production of 1.4 mmol $g_{DW}^{-1}$ $h^{-1}$ can be observed in the first 6 h of cultivation (Fig S6). After 28 h, the absolute arginine concentration levelled up and the culture achieved a maximum concentration of 518 µmol/L.

### Testing long-term cultivation

Next, I addressed reasons that could stop arginine production, focusing on the accumulation of waste products and nutrient depletion in the culture. The strain was cultivated for 56 h at 42°C, mirroring the conditions of the previous cultivation. After 24 h, I washed the cells and transferred them to new medium to extend the cultivation for an additional 32 h at 42°C. Optical density was measured over the 56-hour cultivation period, and arginine levels were absolutely quantified in the while culture broth with LC-MS/MS (Fig 22).

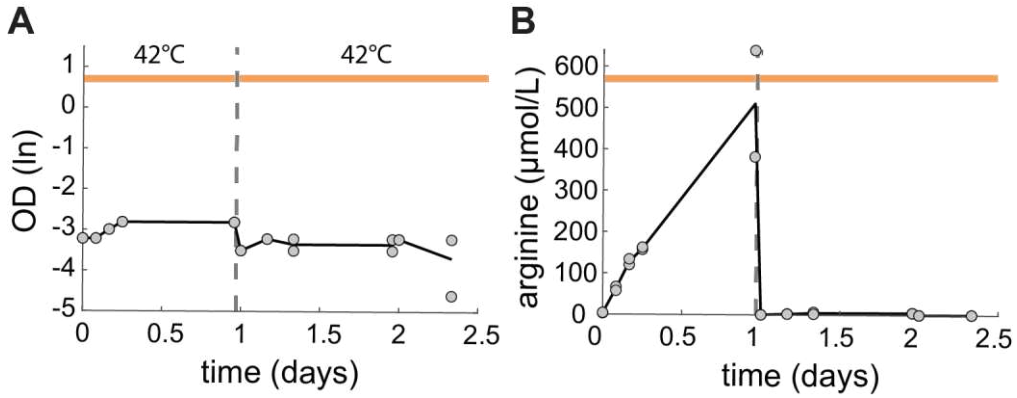



**Figure 22 Arginine production is achieved for 24 h**

**A** Temperature to control growth. OD (ln) of arginine production strain at 24 h at 42°C and after re inoculation 32 h at 42°C. Cells grew on minimal glucose medium in shake flasks. Dots are samples from two replicates (n = 2), and the line connects the means. **B** Arginine production stops after 24 h. Arginine concentration (µmol/L) during the same cultivation (shown in A). Arginine levels were quantified in the whole culture broth by LC-MS/MS and calibrated with an authentic arginine standard. Dots show samples from two replicates (n = 2), and the line connects the means. Raw chromatograms and calibration curves are shown in Fig S7A, S7B.

Main cultivation was started at OD600 0.05 and remained constant during the first 24 h of cultivation at 42°C. As observed before, arginine was produced for 24 h while cells enter growth arrest (Fig 22A). After 24 h, I washed the cells and the optical density decreased by 4%. During additional 32 h of cultivation at 42°C growth remained still constant. However, new medium did not awaken arginine production and washing of the cells removed all arginine in the culture (Fig 22B).

### Testing continuous cultivation – switching back from production to growth

To investigate the reversibility of the temperature switch, I aimed to transition from arginine production back to growth. I initiated the cultivation at 42°C for 24 h and subsequently shifted the temperature to 30°C for an additional 24 h. In between these temperature changes (after 24 h), arginine was removed by washing the cells (Fig 23).

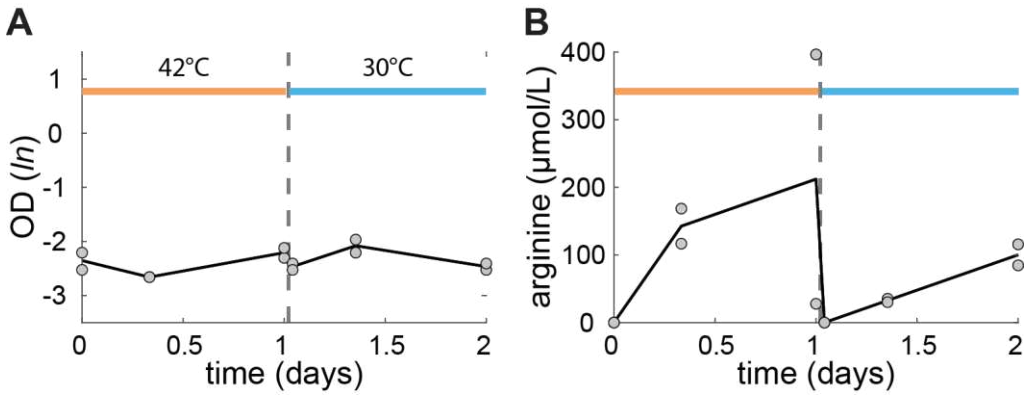



**Figure 23 Decreasing temperature to 30°C rescues arginine production**

**A** Growth when switching between 42°C and 30°C. OD (ln) of arginine production strain at 24 h at 42°C and after re inoculation 24 h at 30°C. Temperature as indicated, dashed line indicates washing steps after 24 h. Cells grew on minimal glucose medium in shake flasks. Dots are samples from two replicates (n = 2), and the line connects the means. **B** Arginine concentration (µmol/L) during the same cultivation (shown in C). Arginine levels were quantified in the whole culture broth by LC-MS/MS and calibrated with an authentic arginine standard. Dots show samples from two replicates (n = 2), and the line connects the means. Re inoculation in fresh medium was done after 24h (dashed line). Raw chromatograms and calibration curve are shown in Fig S8A, S8B.

As expected, no significant growth was observed during the first 24 h of cultivation. However, decreasing the temperature to 30°C did not affect growth during the subsequent 24 h of cultivation (Fig 23A). Washing the cells effectively removed the arginine produced at 42°C. Decreasing the temperature to 30°C reactivated arginine production, ultimately yielding an absolute concentration of 116 µmol$L^{-1}$ after 24 hours (Figure 23B). These findings suggest the feasibility of using temperature control to dynamically regulate arginine production and growth in engineered *E. coli* strains.

## Discussion

The $DnaX^{L289Q}$ mutation was recently identified in a screen for temperature sensitive mutation to affect growth (Schramm et al. 2023). Here I show the application of this temperature-sensitive mutation to decouple growth from arginine production. Prior research has identified a mutation in DnaE that inhibits cell division in *Caulobacter crescentus* (Lo et al. 2004), hinting at the broader potential for utilizing various DNA polymerase variants as growth-inhibiting mutations.

I absolutely quantified arginine concentration in the whole culture broth during the cultivation at 42°C and observed that the arginine production pathway remained active for a 24 h (Fig 21C). Within the initial 6 hours of growth arrest, the arginine production rate was constant 1.35 mmol $g_{DW^{-1}}$ $h^{-1}$, (Fig S6) which is 59% of the production rate (2.3 mmol $g_{DW^{-1}}$ $h^{-1}$) achievable without the $DnaX^{L289Q}$ mutation (Sander et al. 2019a). Transitioning the culture from 42°C to a lower temperature (30°C) in theory allows the expression of intact DnaX protein, that over-time dilute the temperature-damaged copies. However, lower temperatures did not reinstate growth (Fig 23A). Nonetheless, it did reactivate arginine production, indicating persistent metabolic activity in the cells. Further investigations, including extended recovery times after temperature shifts, hold promise for potentially resuming growth

Production of arginine with non-growing *E. coli* cells, has several advantages such as higher product yields, because non-growing bacteria can shift available resources form biomass production towards product synthesis. Another advantage is, that nongrowing bacteria are less likely to undergo mutations that impact product formation, thus making non-growing cells genetically more stable (Kolter et al. 2022). Finally, process control of non-growing cells in two stage processes can be simplified, since there is no need to control feeding rates are other growth-related parameters.

Temperature-based growth control creates new possibilities to increase robustness and productivity of two-stage bioprocesses. These mutations can not only improve the production of native metabolites, such as various amino acids but also holds great promise for strains with heterologous production pathways. In conclusion, our study demonstrates the effectiveness of the $DnaX^{L289Q}$ mutation in decoupling growth from arginine production, highlighting its potential in temperature-controlled bioprocesses. This research opens doors to further explore temperature-sensitive mutations and their applications in enhancing the efficiency and sustainability of chemical production processes.

# Conclusion and Outlook

### Key Findings

**1)** Novel barcoding system for CRISPRi strains with small protein (ubiquitin)

- Flow injection mass spectrometry enable tandem-readout of barcode and metabolic fingerprint
- Time-efficient workflow for ubiquitin-barcode identification using top-down proteomics methods (spectral deconvolution)
- Real-time detection of protein barcodes

**2)** Intracellular quantification of hexose and pentose phosphates using targeted metabolomics

- Hexose and pentose phosphates can be separated using chromatography without additional reagents
- Perturbation of hexose and pentose phosphates achieved through targeted gene knockdown (CRISPR interference) of enzymes in *E. coli* central metabolism
- CRISPRi-induced accumulations of hexose and pentose phosphates show, particularly when metabolites acted as substrates in reactions
- Potential for broad applicability across various organisms and pathways due to the conservation of central metabolism

**3)** Metabolic Adaptations for Chemoautotrophic Growth

- A Mutation in branch point reaction (catalysed by Pgi) reduced efflux from carbon fixation cycle
- Mutations in regulatory genes (*crp*,*rpoB*) increased the abundance of the electron donor (NADH)
- Chemoautotrophic conversion of *E. coli* requires metabolic adaptations in both carbon and energy metabolism.

**4)** Decoupling of growth from arginine overproduction with temperature switch

- Cultivation at high temperatures achieved robust growth arrest
- Arginine production maintained for 24 h
- Cultivation at lower temperatures rescued arginine production after long-term growth arrest, indicating reversibility

**Outlook**

In this work, I show new advances in metabolic engineering and metabolomics to improve our understanding of bacterial metabolism, using the model organism *E. coli*.

In **chapter 1**, I introduce a novel protein-based barcode system for CRISPR interference strains, enabling real-time tracking and phenotyping while avoiding the need for sequencing sgRNAs. This method offers a cost-effective solution for simultaneous tracking and phenotyping of CRISPRi Libraries. The protein-based barcode system for CRISPRi strains holds promise for further development and integration with mass spectrometry-based metabolomics. Future work could focus on improving quantification methods and expanding its applications to other screening techniques.

In **chapter 2** I developed a liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) approach to quantify sugar phosphate isomers in *E. coli*, providing insights into the central metabolism. By perturbing metabolic pathways through CRISPR interference, I revealed dynamic changes in hexose and pentose phosphates, underscoring the potential of this method for various organisms. The quantification of sugar phosphate isomers and the insights gained into glycolysis and the PPP are transferable to a wide range of organisms and pathways. This research paves the way for similar studies in other microbes, including cyanobacteria and pathways like the Calvin cycle.

In **chapter 3**, I investigated the metabolic adaptations of *E. coli* during its transition to chemoautotrophy, shedding light on the roles of autotrophic-enabling mutations in energy and carbon metabolism. This has implications for the conversion of $CO_2$ into valuable compounds in sustainable bioprocesses. The understanding of metabolic adaptations for chemoautotrophic growth has implications for harnessing microorganisms in sustainable carbon conversion processes. Further research can

explore the potential of *E. coli* and other organisms as platforms for synthetic autotrophy, extending to natural autotrophs, cell-free systems, and environmental applications.

In **chapter 4**, I introduced a temperature-controlled two-stage process to produce arginine in *E. coli*, demonstrating the decoupling of growth and overproduction using a temperature switch. This strategy holds promise for microbial production of amino acids and valuable compounds. Temperature-controlled bioprocesses offer one avenue for enhancing chemical production. Future investigations can explore different temperature-sensitive mutations and their applications in optimizing the production of various metabolites, both native and heterologous.

This work extends the toolbox for genetic and metabolic engineering, metabolomics, and sustainable compound production, offering innovative methods and valuable insights to address challenges in metabolic engineering and bioproduction in *E. coli*. The previously presented results have the potential to contribute to future advancements in the fields of synthetic biology, biotechnology, and sustainable compound production.

# Material and Methods

## 1 A protein-based barcode for CRISPR interference strains

### Media

Cultivations were performed with SOC medium (Cat#15544034, Invitrogen), LB medium (Cat#L3522, Sigma-Aldrich) or M9 minimal medium with glucose (5 g/L) as sole carbon source. The M9 medium consists of (per liter): 7.52 g $Na_2HPO_4 \cdot 2\ H_2O$, 5 g $KH_2PO_4$, 1.5 g $(NH_4)_2SO_4$, 0.5 g NaCl. Additionally, the following components were sterile filtered and added separately (per liter M9 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 consists of (per liter): 180 mg $ZnSO_4 \cdot 7\ H_2O$, 120 mg $CuCl_2 \cdot 2$ H2O, 120 mg $MnSO_4 \cdot$H2O, 180 mg $CoCl_2 \cdot 6\ H_2O$.For strains transformed with pgRNA-ubiquitin plasmid or variants, 100 µg/mL carbenicillin (Carb) was added to the media. To induce the expression the dCas9 protein aTc was supplemented to a final concentration of 200 nM.

### Strains and culture

*E. coli* YYdCas9 strain (Lawson et al. 2017) was the wild-type strain used in this study. MegaX DH10B competent *E. coli* (Cat#C640003, Invitrogen) cells were used for cloning. All strains in this study are derived from the YYdCas9 strain.

### Cloning of ubiquitin open reading frame into pgRNA

To assemble the ubiquitin open reading frame into pgRNA-bacteria (Qi et al. 2013) (Cat#44251, Addgene), three DNA fragments were used. The ubiquitin sequence was supplied from the plasmid Ubiquitin WT (Brzovic et al. 2006) (Cat#12647, Addgene). The coding sequence flanked by a T7 terminator and a strong ribosome binding site (T7 G10 RBS) and the backbone pgRNA carrying a protospacer for mRFP2 were amplified. The ubiquitin wild-type amplicon and the backbone were assembled using Golden gate cloning (Engler and Marillonnet 2014). The promotor PJ23100 was amplified from a gene fragment that was ordered from Twist Bioscience (San

Francisco, USA). To assemble pgRNA-Ubi-mRFP IVA cloning (García-Nafría et al. 2016) was used.

### Cloning of 7 CRISPRi strains

Oligonucleotides of 241 bp long single stranded DNA were ordered from Twist Bioscience (San Francisco, USA). Each oligo differs in sgRNA and barcode sequence. sgRNAs sequences were chosen from (Donati et al. 2021). Barcode sequence was designed based on 21 essentials amino (Table 5) acids using custom MATLAB scripts. The mass of 6 amino acids was systematically combined to receive barcodes of masses between 9.018 kDa and 9.443 kDa with a minimum difference of 1 Da. Each plasmid was cloned individual and transformed into *E. coli* YYdCas9.

**Table 5 Essential amino acids**

| 1-letter code | 3-letter code | Name | Chemical formula | Monoisotopic |
|---|---|---|---|---|
| A | Ala | Alanine | C3H5ON | 71.03711 |
| R | Arg | Arginine | C6H12ON4 | 156.10111 |
| N | Asn | Asparagine | C4H6O2N2 | 114.04293 |
| D | Asp | Aspartate | C4H5O3N | 115.02694 |
| C | Cys | Cystein | C3H5ONS | 103.00919 |
| E | Glu | Glutamine | C5H7O3N | 129.04259 |
| Q | Gln | Glutamate | C5H8O2N2 | 128.05858 |
| G | Gly | Glycine | C2H3ON | 57.02146 |
| H | His | Histidine | C6H7ON3 | 137.05891 |
| I | Ile | Isoleucine | C6H11ON | 113.08406 |
| L | Leu | Leucine | C6H11ON | 113.08406 |
| K | Lys | Lysine | C6H12ON2 | 128.09496 |
| M | Met | Methionine | C5H9ONS | 131.04049 |
| F | Phe | Phenylalanine | C9H9ON | 147.06841 |
| P | Pro | Proline | C5H7ON | 97.05276 |
| S | Ser | Serine | C3H5O2N | 87.03203 |
| T | Thr | Threonine | C4H7O2N | 101.04768 |
| W | Trp | Tryptophane | C11H10ON2 | 186.07931 |
| Y | Tyr | Thyrosine | C9H9O2N | 163.06333 |
| V | Val | Valine | C5H9ON | 99.06841 |

**Flask cultivation for barcode and metabolite extraction**

For metabolite and ubiquitin-barcode analysis, strains were inoculated from a glycerol stock in 3 mL rich medium (LB) with carbenicillin (100 µg/mL). And cultivated for 8 h at 37°C. Cells were then transferred in 10 mL M9 minimal medium with glucose and carbenicillin for overnight cultivation at 37°C in 100 mL shake flasks. M9-pre-culture were adjusted to a starting OD600 of 0.05 into 10 mL M9 medium. The strains were cultivated for 6.5 h in triplicates with and without the addition of aTc.

**Flow-injection mass spectrometry (FI-MS)**

OD600 was measured and an equivalent of 1 OD600 was sampled into a 1.5 mL microcentrifuge tube. Cells were then centrifuged for 10 min at maximum speed and 4°C. Supernatant was discarded and cell pellets were resuspended in 1 mL of ice-cold Phosphate Buffer Saline (PBS, Cat#10722497, Invitrogen). Centrifugation and PBS washing was then repeated. Supernatant was discarded and cell pellets were stored at -80°C for 1 h. Finally, pellets were resuspended in 100 µL of ammonium carbonate ($(NH_4)_2CO_3$, Cat#10361-29-2, Honeywell). Extracted samples were centrifuged 10 min at 4°C and either stored at -80°C or directly used for FI-MS analysis. 15 µL of sample was transferred to MS vial for measurements.

Flow-injection mass spectrometry (FI-MS) was performed as described before [10,14]. Electrospray ionization was applied (ESI) and ESI-MS experiments were performed on an Agilent 6546 Series quadrupole time-of-flight mass spectrometer (Agilent Technologies, USA). Mass spectrometer scans (MS1) were performed in negative and positive ionisation mode and the injection volume was set to 3 µL. As mobile phase was used 60:40 2-propanol:water buffered with 10 mM ammonium carbonate $(NH_4)_2CO_3$ and 0.04% (v/v) ammonium hydroxide ($NH_4OH$) for both ionization modes. The flow rate was set to 0.15 mL/min. Online mass correction was performed using a reference mix with purine and HP-921 for positive mode and 2-propanol (in the mobile phase) and HP-921 for negative mode. Mass spectra were recorded from 50 to 1700 *m/z* with a frequency of 1.4 spectra/s for 0.5 min using 10 Ghz resolving poIr. Source temperature was set to 225°C, with 1 L/min drying gas and a nebulizer pressure of 20 psi. Fragmentor, skimmer, and octupole voltages were set to 120 V, 65 V, and 650 V, respectively. Capillary voltage was set to 3,500 V.

**Real-time metabolome profiling**

OD600 was measured and an equivalent of 1 OD600 was sampled into a 1.5 mL microcentrifuge tube. Cells were then centrifuged for 10 min at 4,500 rpm and 4°C. Supernatant was discarded and cell pellets were resuspended in 1/8 M9 medium:water to decrease salt concentrations. 10 µL of the sample was transferred fast to MS vials for measurements.

Samples were analysed using an Agilent 6546 Series quadrupole time-of-flight mass spectrometer (Agilent Technologies, USA) and an Agilent 1290 Infinity II UHPLC system (Agilent Technologies). The automated liquid sampler was preheated to 37°C and the in-line filter was removed to enable real-time measurements. Real-time metabolome profiling was adapted from (Link et al. 2015) The heat-exchanger was set to 90°C to lyse the cells. The injection volume was changed to 3 µL. Mass spectrometer scans (MS1) were performed in negative and positive ionisation mode. As mobile phase was used 60:40 2-propanol:water buffered with 10 mM ammonium carbonate $(NH_4)_2CO_3$ and 0.04 % (v/v) ammonium hydroxide ($NH_4OH$) for both ionization modes. The flow rate was 0.15 $mLmin^{-1}$. Online mass correction was performed using a reference mix with purine and HP-921 for positive mode and 2-propanol (in the mobile phase) and HP-921 for negative mode. Mass spectra were recorded from 50 to 1700 *m/z* as described for FI-MS. For real-time mass spectrometry I set capillary voltage to 5,000 V and skimmer voltage to 750 V.

**Data processing**

For metabolomics raw agilent files were converted into ‘’mzXML’’ format using MSConvert (Version: 3.0.21082-dcbcfcd2c, Chambers et al. 2012). Further analysis was done using custom MATLAB scripts. For each sample, an average spectrum was calculated from the 10 scans with the highest ion counts. The spectra were resampled to 1.7 x $10^6$ data points to align *m/z* values of all samples. Ion peaks were picked with the “findpeaks” function of MATLAB (Version: R2021a), using a peak height and prominence cut off (1,000 a.u.). Hierarchical clustering with a tolerance of 7.5 mDa was used to bin peaks. For each peak bin, I calculated a centroid *m/z* value from the individual peak *m/z* values. Peaks were annotated to metabolites using the centroid *m/z* value with a tolerance of 3 mDa as described before (Farke et al. 2023). For

barcode detection raw Agilent files were converted into ‘’mzML’’ format using MSConvert (Version: 3.0.21082-dcbcfcd2c, Chambers et al. 2012). mzML files were used for Flash daconv (Version: 2.0 BETA+, Jeong et al. 2020).

### Spectral Deconvolution

To effectively decrease the processing time further I adapted the mass range 1 kDa – 10 kDa and the charge states +1 to +100.

## 2 Quantification of sugar phosphate isomers in *E. coli* using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS)

### Chemicals and Materials

LC eluents were water form an ultra-pure water system (MilliQ) and ultra-pure acetonitrile (#cat 14261-1l, Honeywell). LC additives were formic acid (#cat 56302-50ml-F, Merck) and ammonium formate (#cat 70221-25g-F, Merck). As LC column was used a Shodex HILICpak VG-50 2D (2.0 x 150 mm). Column dimension and particle size are listed in Table 3.

### Metabolite standards

Authentic metabolite standards were purchased with purity ≥ 99% (Table 4). Stock solutions of metabolites were prepared by dissolving standards in water. From stock solution, a 10µM standard mixture was prepared to a final dilution of 40:40:20 (v/v) methanol/acetonitrile/water. Aliquots of the mix was stored at -80°C. Parameters for MRM assays in positive and negative ionization mode were determined using single standards at a concentration of 1 µM and single compound optimization. To determine the retention time using LC-MS/MS of each metabolite I injected authentic standards with a concentration of 10 µM.

**Table 2 Analytical standards**

| Name | Abbreviation | #Cat | Company |
|---|---|---|---|
| D-Fructose-6-phosphate | f6-p | F1502-500MG | Sigma Aldrich |
| D-Fructose 1-phosphate | f1-p | MF03840 | Biosynth |
| D-Glucose-6-phosphate | g6-p | G7879-500MG | Sigma Aldrich |
| D-Glucose-1-phosphate | f1-p | G6750-50MG | Sigma Aldrich |
| D-Ribose-1-phosphate | r1-p | 83866-1MG | Sigma Aldrich |
| D-Ribose-5-phosphate | r5-p | 83875-250MG | Sigma Aldrich |
| D-Ribulose-5-phosphate | ru-5p | 83899-5MG | Sigma Aldrich |
| D-Xylulose-5-phosphat | xu-5p | 15732-1MG | Sigma Aldrich |

### Liquid chromatography

An Agilent 1290 Infinity II UHPLC system (Agilent Technologies) was used for liquid chromatography. Temperature of the column oven was 30°C. A 0.3 µM inline filter was used (Agilent Technology) and no guard column. LC solvents were kept in Teflon FEP bottles (Nalgene, Thermo Scientific) with PFTE inlet filters (Vici Jour). The injection volume for method A and B was set to 9 µL. LC solvent in channel A was water with 10 mM ammonium formate and 0.1% formic acid (v/v), and in channel B was acetonitrile with 0.1% formic acid (v/v). The LC column was a Shodex HILICpak VG-50 2D. The gradient was: 0 min 90% B; 58 min 70 % B; 3 min 90 % B. The 12C peak heights were used to quantify metabolites. Retention time of hexose-phosphates were determined with authentic standards of glucose-6-P, fructose-phosphate Flow rate was set to 0.28 mL min-1.

### Mass spectrometry

An Agilent 6495 triple quadrupole mass spectrometer (Agilent Technologies) was used for mass spectrometry. Source gas temperature was set to 200°C, with 14 $Lmin^{-1}$ drying gas and a nebulizer pressure of 24 psi. Sheath gas temperature was set to 300°C and flow to 11 $Lmin^{-1}$. Electrospray nozzle and capillary voltage were set to 500 V and 2500 V respectively. Dwell time of 300 msec was used.

**Media**

Cultivations were performed with LB rich medium (Cat#L3522, Sigma-Aldrich) or M9 minimal medium with glucose (5 $gL^{-1}$) or xylose (5 $gL^{-1}$) as sole carbon source. The M9 medium consists of (per liter): 7.52 g $Na_2HPO_4 \cdot 2\ H_2O$, 5 g $KH_2PO_4$, 1.5 g $(NH_4)_2SO_4$, 0.5 g NaCl. Additionally, the following components were sterile filtered and added separately (per liter M9 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 consists of (per liter): 180 mg $ZnSO_4 \cdot 7\ H_2O$, 120 mg $CuCl_2 \cdot 2$ H2O, 120 mg $MnSO_4 \cdot$H2O, 180 mg $CoCl_2 \cdot 6\ H_2O$. To induce the expression the dCas9 protein aTc was supplemented to a final concentration of 200 nM. I used carbenicillin with a final concentration of 50 µg $mL^{-1}$.

**Strains and culture**

*E. coli* YYdCas9 strain (Qi et al. 2013) was the base strain used in this study. All strains in this study are borrowed from an inhouse curated arrayed CRISPRi library (LinkLab 2020). Strains ae isolated from pooled Library generated in (Donati et al. 2021). Each strain contains a 25 bp long unique sgRNA sequence to target genes in *E. coli* metabolism. All sgRNA sequences are derived from iML1515 model (Monk et al. 2017).

**Cultivation conditions for growth data**

Strains were inoculated form cryo culture and transferred to 3 mL LB supplemented with ampicillin (100 µg/mL) liquid cultures. The LB pre-cultures were used to inoculate a second preculture in M9 minimal medium that was incubated overnight in 13 mL culture tubes under shaking at 37°C. Starting OD600 was adapted to 0.1 for main cultivation. Main culture was prepared in microtiter plates (96-well). Every strain was cultures in triplicates with and without the addition of aTc to the M9 main culture (aTc was not added to preculture). Growth rate was calculated based on linear regression with 10 min intervals using a plate reader (Tecan, Spark).

**Cultivation for metabolome sampling**

Strains were inoculated from a cryo stock to 3 mL liquid LB with ampicillin (100 ng$\mu L^{-1}$) and re-inoculated in M9 minimal medium overnight in 10 mL in shake flasks under

shaking at 37°C. For metabolomics, M9 preculture were adjusted to a starting OD600 of 0.05 into 10 mL shake flasks. Strains were cultivated with aTc. After 6.5h cultures were rapidly transferred to a thermostatically controlled hood at 37°C to collect samples. 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 min. Subsequently, metabolite extracts were centrifuged for 15 min at 13,000 rpm at -9°C and the supernatant was stored at -80°C until analysis. For mass spectrometry 100 µL of sample was evaporated using a speed vac and resuspended in 10 µM to concentrate the sample 10x.

**Data analysis**

Data processing and analysis was performed with Matlab (Version: MatlabR2021a). LC-MS/MS raw data were converted to text using MSconvert (Chambers et al. 2012). The highest peak in authentic standards was used to determine the retention time. Further detection of sugar-phosphates in *E. coli* samples was based on finding the highest peak with a tolerance of 1 min close to retention time of analytic standards.

## 3 Metabolic changes that enable *E. coli* to use $CO_2$ as C-source

**Media**

M9 minimal medium with glucose (5 $gL^{-1}$) as sole carbon source. The M9 medium consists of (per liter): 7.52 g $Na_2HPO_4 \cdot 2\ H_2O$, 5 g $KH_2PO_4$, 1.5 g $(NH_4)_2SO_4$, 0.5 g NaCl. Additionally, the following components were sterile filtered and added separately (per liter M9 medium): 1 mL 0.1 M $CaCl_2$, 1 mL 1 M $MgSO_4$, 0.6 mL 0.1 M $FeCl_3$ and 10 mL trace salts solution. The trace salts solution consists of (per liter): 180 mg $ZnSO_4 \cdot 7\ H_2O$, 120 mg $CuCl_2 \cdot 2$ H2O, 120 mg $MnSO_4 \cdot H2O$, 180 mg $CoCl_2 \cdot 6\ H_2O$. Media was supplemented with appropriate antibiotics dependent on the plasmid (see Plasmids and strains).

**Investigation of PGI strains and cultivation**

P1 transduction (Thomason et al. 2007) was used to transfer knockout alleles from the KEIO strain collection (Baba et al. 2006) to our engineered strain, and to knock out the

genes phosphofructokinase (*pfkA* and *pfkB*), and 6-phosphate-1-dehydrogenase (*zwf*). Following the transduction of each knockout allele the KmR selection marker was removed by using the FLP recombinase encoded by the pCP20 temperature-sensitive plasmid (Cherepanov and Wackernagel 1995). Loss of the selection marker and the temperature sensitive plasmid were validated by replica-plating the screened colonies and PCR analysis of the relevant loci. The cells were then transformed with the pCBB plasmid (Antonovsky et al. 2016, #Cat KX077536) and a pFDH plasmid with a constitutive promoter controlling the expression of the *fdh* gene that resulted from Gleizer et al. 2019. The growth test experiments were conducted in 96 well-plates. The final volume of each well was 200 µL (50 µL of mineral oil and 150 µL culture). The media consisted of M9 media supplemented with varying concentrations of the relevant carbon source, and trace elements (without addition of vitamin B1). Bacterial cells were seeded from a culture tube. Growth temperature was set to 37°C, and either aerated with ambient air or air with elevated $CO_2$ (10%). OD600 measurements were taken every 10-30 min using a Tecan Spark plate reader. For the G6P and F6P, P5P, and S7P measurements: Single colonies were transferred into liquid 3 mL M9 media containing 2 $gL^{-1}$ xylose, 30 mM formate, Chloramphenicol (30 $mgL^{-1}$) and streptomycin (100 $mgL^{-1}$) from fresh plates with the same medium. The M9 pre-cultures were adjusted to a starting OD600 of 0.05 into 10 mL culture in 250 mL shake flasks. The flasks were placed inside an elevated $CO_2$ (10%) incubator (New Brunswick S41i $CO_2$ incubator shaker, Eppendorf, Germany) at 37°C.

**Investigation of the role of *crp* and *rpoB* for energy metabolism strains and cultivation**

I used *E. coli* BW25113 to introduce the pFDH plasmid. The pFDH plasmid, an *E. coli* codon optimised DNA sequence of formate dehydrogenase from the methylotrophic bacterium *Pseudomonas* sp. 101 (Popov and Lamzin 1994) was cloned with an *N*-terminal His-tag into a pZ plasmid (Expressys, Germany) under a constitutive promoter with and a strong ribosome binding site rbsB of (Zelcbuch et al. 2013). The plasmid has a pMcB1 origin of replication and therefore is present in high copy number. The KmR selection marker on the plasmid was replaced with a StrepR marker. An 8 bp deletion appeared in the promoter region of the first evolved clone. This plasmid was isolated and was the plasmid used for all consequential evolution experiments and

autotrophic strains (Gleizer et al. 2019). Details regarding the pCBB plasmid are reported in (Antonovsky et al. 2016). $NAD^+$/NADH measurements: Single colonies were transferred into 10 mL M9* xylose streptomycin from fresh M9* xylose streptomycin plates and cultivated for 24 h while shaking at 37°C. M9 pre-cultures were adjusted to a starting OD600 of 0.1 into 12-well plates, with 2 mL of medium in each well. Strains were cultivated in triplicates with or without formate (30 mM), added at the beginning of the culture. Optical density at 600 nm was monitored every 10 min using a plate reader (Tecan, Spark). Plates were then rapidly transferred to a thermostatically controlled hood at 37°C and kept shaking during the sampling procedure.

**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 min. Subsequently, metabolite extracts were centrifuged for 15 min at 13,000 rpm at -9°C and the supernatant was stored at -80°C until analysis. For measurements of NADH, NAD, s7-p and the pool of pentose-P, extracts were mixed with a 13C-labelled internal standard in a 1:1 ratio. LC-MS/MS analysis was performed with an Agilent 6495 triple quadrupole mass spectrometer (Agilent Technologies) as described previously (Guder et al. 2017). An Agilent 1290 Infinity II UHPLC system (Agilent Technologies) was used for liquid chromatography. Temperature of the column oven was 30°C, and the injection volume was 3 µL. LC solvents in channel A were either water with 10 mM ammonium formate and 0.1% formic acid (v/v) (for acidic conditions, NADH/NAD), or water with 10 mM ammonium carbonate and 0.2% ammonium hydroxide (for basic conditions, s7-p and pool of Pentoses). 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 12C and 13C peak heights was used to quantify metabolites. 12C/13C ratios were normalized to OD at the time point of sampling. For the measurements of intracellular glucose-6-P and

fructose-6-P, metabolic extracts were 10x concentrated using a vacuum evaporation (Eppendorf Concentrator plus) and resuspended in 40:40:20 (v-%) acetonitrile/methanol/water extraction solution. Temperature of the column oven was 30°C, and the injection volume was 5 µL. LC solvent in channel A was water with 10 mM ammonium formate and 0.1% formic acid (v/v), and in channel B was acetonitrile with 0.1% formic acid (v/v). The LC column was a Shodex HILICpak VG-50 2D (2.0 x 150 mm). The gradient was: 0 min 90% B; 58 min 70 % B; 3 min 90 % B. The 12-C peak heights were used to quantify metabolites. Retention time of hexose-phosphates were determined with authentic standards of glucose-6-P, fructose-6-P.

## 4 Temperature controlled two-stage arginine production in *E. coli*

### Construction of the DnaX$^{L289Q}$ arginine overproduction strain

A *E. coli* MG1655 Δ*argR* ArgA strain was used to introduce the DnaX$^{L289Q}$ mutation for growth control. For genome editing a CRISPR-Cas9 system was modified from the CREATE method (Garst et al. 2017). Two plasmids are used (pTS040 and pTS041). pTS040 had the p15A origin of replication and carried a chloramphenicol resistance gene, a cassette with the homology arm for recombination, and the guide RNA of the CRISPR system under control of a constitutive promoter (PJ23119). pTS041 had the pSC101 origin of replication and carried a kanamycin resistance gene, a gene for the anhydrotetracycline (aTc)-sensitive repressor tetR, cas9 under control of the aTc controlled PLtetO1 promoter, the arabinose-sensitive repressor *araC*, and the Escherichia virus lambda genes red under control of the arabinose-controlled promoter ParaBAD. First, I used electroporation to introduce pTS041. The strain *E. coli* MG1655 Δ*argR* ArgA-H15Y harbouring pTS041 was used for transformation with pTS040 containing the repair cassette with DnaX$^{L289Q}$ mutation. To perform the genome editing a 50 mL LB culture was started from a cryo stock and incubated for 22h at 30°C under 220 rpm of shaking. After diluting the culture 1:10,000 in fresh LB, the cells were further incubated over night at 30°C under shaking of 220 rpm. A fresh culture was started in the morning by 1:50 dilution in the same medium and conditions. After 2 h, Lamda Red was induced for 30 min by addition of L-arabinose (7.5 gL$^{-1}$). Cells were subsequently transformed by electroporation with pTS040 (DnaX$^{L289Q}$). The genomic edit was confirmed by Sanger sequencing. Subsequently, the strain was transformed with

pTS056 that was used to overexpress the arginine exporter ArgO. It was previously described that basal expression of ArgO was sufficient (Sander et al. 2019a) such that aTc was not added to subsequent cultures. The pTS056 had a p15A origin of replication, an ampicillin resistance gene, the anhydrotetracycline(aTc)-sensitive repressor *tetR*, and *argO* encoding for an arginine exporter under the aTc controlled PLtetO1 promoter. pTS056 was based on a plasmid from (Sander et al. 2019b).
If not stated otherwise, Minimal medium (M9) was used for the experiments and contained 42.2 mM $Na_2HPO_4$, 11.3 mM $(NH_4)_2SO_4$, 22 mM $KH_2PO_4$, 8.56 mM NaCl, 1 mM $MgSO_4$ x 7 $H_2O$, 100 µM $CaCl_2$ x 2 $H_2O$, 60 µM $FeCl_3$, 6.3 µM $ZnSO_4$ x 7 $H_2O$, 7 µM CuCl2 x 2 H2O, 7.1 µM, $MnSO_4$ x 2 $H_2O$, 7.6 µM $CoCl_2$ x 6 $H_2O$, and 2.8 µM thiamine-HCL. 5 $gL^{-1}$ glucose was used as carbon source. 30 $\mu gL^{-1}$ chloramphenicol, 50 $\mu gmL^{-1}$ kanamycin, 100 $\mu gmL^{-1}$ carbenicillin, and 50 $\mu gmL^{-1}$ spectinomycin was added to the media when required.

**Arginine overproduction experiment with a DnaX$^{L289Q}$ mutant**

10 mL LB cultures were started from cryo stock. After ca. 6 h at 30°C under shaking at 220 rpm, 10 mL M9 (5 $gL^{-1}$ glucose) overnight cultures (30°C, 220 rpm) were started using 50 µL of the LB culture for inoculation. Overnight M9 cultures were washed and the cultures were pelletized by 5 min centrifugation at 4,000 rpm and 40°C. After removing the supernatant, 10 mL of fresh M9 glucose medium was added for resuspending cells. This step was repeated an additional 2-times. The final 50 mL cultures were started in 500 mL shaking flasks at an OD of 0.05 and incubated for 24 h under shaking of 220 rpm and 42°C. At the start of the cultivation and as indicated, the OD600 was measured, and whole culture broth samples for LC-MS/MS analysis were taken.

**Metabolite Measurements**

Shake flask cultivations on M9 glucose were performed as described above. Cells were grown to an optical density (OD600) of 0.5 mL and 2 mL culture aliquots were vacuum-filtered on a 0.45 µm pore size filter (HVLP02500, Merck Millipore). Filters were immediately transferred into 40:40:20 (v-%) acetonitrile/methanol/water at -20°C for extraction. Extracts were centrifuged for 15 min at 13,000 rpm at -9°C. Centrifuged extracts were mixed with 13C-labeled internal standard and analysed by LC-MS/MS,

with an Agilent 6495 triple quadrupole mass spectrometer (Agilent Technologies) as described previously (Guder et al. 2017). An Agilent 1290 Infinity II UHPLC system (Agilent Technologies) was used for liquid chromatography. Temperature of the column oven was 30°C, and the injection volume was 3 µL. LC solvents A were water with 10 mM ammonium formate and 0.1% formic acid (v/v) (for acidic conditions); and water with 10 mM ammonium carbonate and 0.2% ammonium hydroxide (for basic conditions). LC solvents B were acetonitrile with 0.1% formic acid (v/v) for acidic conditions and acetonitrile without additive for basic conditions. LC columns were an Acquity BEH Amide (30 x 2.1 mm, 1.7 µm) for acidic conditions, and an iHILIC-Fusion(P) (50 x 2.1 mm, 5 µm) for basic conditions. The gradient for basic and acidic conditions was: 0 min 90% B; 1.3 min 40 % B; 1.5 min 40 % B; 1.7 min 90 % B; 2 min 90 % B. Absolute concentrations of amino acids in the 13C-labeled internal standard were determined with authentic standards. Quantification of intracellular metabolite concentrations was based on the ratio of 12-C and 13-C peak heights, and a specific cell volume of 2 $\mu L mg^{-1}$ was used to calculate the cell volume.

# References

Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, Chen Y, Villalta JE, Gilbert LA, Horlbeck MA, Hein MY, et al. 2016. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* 167: 1867-1882.e21.

Alseekh S, Fernie AR 2018. Metabolomics 20 years on: what have we learned and what hurdles remain? *The Plant Journal* 94: 933–942.

Alseekh S, Aharoni A, Brotman Y, Contrepois K, D'Auria J, Ewald J, C. Ewald J, Fraser PD, Giavalisco P, Hall RD, et al. 2021. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nature Methods* 18: 747–756.

Anglada-Girotto M, Handschin G, Ortmayr K, Campos AI, Gillet L, Manfredi P, Mulholland CV, Berney M, Jenal U, Picotti P, et al. 2022. Combining CRISPRi and metabolomics for functional annotation of compound libraries. *Nature Chemical Biology* 18: 482–491.

Antonovsky N, Gleizer S, Noor E, Zohar Y, Herz E, Barenholz U, Zelcbuch L, Amram S, Wides A, Tepper N, et al. 2016. Sugar Synthesis from $CO_2$ in *Escherichia coli*. *Cell* 166: 115–125.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* 2: 2006.0008.

Bassham JA, Benson AA, Kay LD, Harris AZ, Wilson AT, Calvin M 1954. The Path of Carbon in Photosynthesis. XXI. The Cyclic Regeneration of Carbon Dioxide Acceptor. *Journal of the American Chemical Society* 76: 1760–1770.

Beckmann M, Parker D, Enot DP, Duval E, Draper J 2008. High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray mass spectrometry. *Nature Protocols* 3: 486–504.

Ben-Nissan R, Milshtein E, Pahl V, De Pins B, Jona G, Levi D, Yung H, Nir N, Ezra D, Gleizer S, et al. 2023. Autotrophic growth of *E. coli* is achieved by a small number of genetic changes. preprint. elife.

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

Bjerrum JT (ed.) 2015. *Metabonomics: Methods and Protocols*. New York, NY: Springer New York.

Bock C, Datlinger P, Chardon F, Coelho MA, Dong MB, Lawson KA, Lu T, Maroc L, Norman TM, Song B, et al. 2022. High-content CRISPR screening. *Nature Reviews Methods Primers* 2: 8.

Brzovic PS, Lissounov A, Christensen DE, Hoyt DW, Klevit RE 2006. A UbcH5/Ubiquitin Noncovalent Complex Is Required for Processive BRCA1-Directed Ubiquitination. *Molecular Cell* 21: 873–880.

Buescher JM, Czernik D, Ewald JC, Sauer U, Zamboni N 2009. Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism. *Analytical Chemistry* 81: 2135–2143.

Buescher JM, Moco S, Sauer U, Zamboni N 2010. Ultrahigh Performance Liquid Chromatography−Tandem Mass Spectrometry Method for Fast and Robust Quantification of Anionic and Aromatic Metabolites. *Analytical Chemistry* 82: 4403–4412.

Bujara M, Schümperli M, Pellaux R, Heinemann M, Panke S 2011. Optimization of a blueprint for in vitro glycolysis by metabolic real-time analysis. *Nature Chemical Biology* 7: 271–277.

Burg JM, Cooper CB, Ye Z, Reed BR, Moreb EA, Lynch MD 2016. Large-scale bioprocess competitiveness: the potential of dynamic metabolic control in two-stage fermentations. *Current Opinion in Chemical Engineering* 14: 121–136.

Cajka T, Fiehn O 2016. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Analytical Chemistry* 88: 524–545.

Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, et al. 2012. A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* 30: 918–920.

Chen B, Brown KA, Lin Z, Ge Y 2018. Top-Down Proteomics: Ready for Prime Time? *Analytical Chemistry* 90: 110–127.

Cherepanov PP, Wackernagel W 1995. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* 158: 9–14.

Cho B-K, Barrett CL, Knight EM, Park YS, Palsson BØ 2008. Genome-scale reconstruction of the Lrp regulatory network in *Escherichia coli*. *Proceedings of the National Academy of Sciences* 105: 19462–19467.

Cho B-K, Federowicz S, Park Y-S, Zengler K, Palsson BØ 2012a. Deciphering the transcriptional regulatory logic of amino acid metabolism. *Nature chemical biology* 8: 65–71.

Cho H-S, Seo SW, Kim YM, Jung GY, Park JM 2012b. Engineering glyceraldehyde-3-phosphate dehydrogenase for switching control of glycolysis in *Escherichia coli*. *Biotechnology and Bioengineering* 109: 2612–2619.

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

Claassens NJ, Cotton CAR, Kopljar D, Bar-Even A 2019. Making quantitative sense of electromicrobial production. *Nature Catalysis* 2: 437–447.

Compton PD, Zamdborg L, Thomas PM, Kelleher NL 2011. On the Scalability and Requirements of Whole Protein Mass Spectrometry. *Analytical Chemistry* 83: 6868–6874.

Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, Schuster LC, Kuchler A, Alpar D, Bock C 2017. Pooled CRISPR screening with single-cell transcriptome readout. *Nature Methods* 14: 297–301.

Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, et al. 2016. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* 167: 1853-1866.e17.

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

Donnelly DP, Rawlins CM, DeHart CJ, Fornelli L, Schachner LF, Lin Z, Lippens JL, Aluri KC, Sarin R, Chen B, et al. 2019. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nature Methods* 16: 587–594.

Du D, Roguev A, Gordon DE, Chen M, Chen S-H, Shales M, Shen JP, Ideker T, Mali P, Qi LS, et al. 2017. Genetic interaction mapping in mammalian cells using CRISPR interference. *Nature Methods* 14: 577–580.

Egorov AM, Popov VO, Berezin IV, Rodionov UV 1980. Kinetic and structural properties of Nad-Dependent bacterial formate dehydrogenase. *Journal of Solid-Phase Biochemistry* 5: 19–33.

Engler C, Marillonnet S 2014. Golden Gate Cloning. In: Valla S, Lale R (eds), *DNA Cloning and Assembly Methods* (1116 vol.). Totowa, NJ: Humana Press. pp 119–131.

Famiglini G, Palma P, Termopoli V, Cappiello A 2021. The history of electron ionization in LC-MS, from the early days to modern technologies: A review. *Analytica Chimica Acta* 1167: 338350.

Farke N, Schramm T, Verhülsdonk A, Rapp J, Link H 2023. Systematic analysis of in-source modifications of primary metabolites during flow-injection time-of-flight mass spectrometry. *Analytical Biochemistry* 664: 115036.

Farmer WR, Liao JC 2000. Improving lycopene production in Escherichia coli by engineering metabolic control. *Nature Biotechnology* 18: 533–537.

Fuchs G 2011. Alternative Pathways of Carbon Dioxide Fixation: Insights into the Early Evolution of Life? *Annual Review of Microbiology* 65: 631–658.

Fuhrer T, Heer D, Begemann B, Zamboni N 2011. High-Throughput, Accurate Mass Metabolome Profiling of Cellular Extracts by Flow Injection–Time-of-Flight Mass Spectrometry. *Analytical Chemistry* 83: 7074–7080.

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

García-Nafría J, Watson JF, Greger IH 2016. IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. *Scientific Reports* 6: 27459.

Garst AD, Bassalo MC, Pines G, Lynch SA, Halweg-Edwards AL, Liu R, Liang L, Wang Z, Zeitoun R, Alexander WG, et al. 2017. Genome-wide mapping of mutations

at single-nucleotide resolution for protein, metabolic and genome engineering. *Nature Biotechnology* 35: 48–55.

Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* 6: 343–345.

Gleizer S, Bar-On YM, Ben-Nissan R, Milo R 2020. Engineering Microbes to Produce Fuel, Commodities, and Food from $CO_2$. *Cell Reports Physical Science* 1: 100223.

Gleizer S, Ben-Nissan R, Bar-On YM, Antonovsky N, Noor E, Zohar Y, Jona G, Krieger E, Shamshoum M, Bar-Even A, et al. 2019. Conversion of *Escherichia coli* to Generate All Biomass Carbon from $CO_2$. *Cell* 179: 1255-1263.e12.

Glish GL, Vachet RW 2003. The basics of mass spectrometry in the twenty-first century. *Nature Reviews Drug Discovery* 2: 140–150.

Grand View Research 2023. https://www.grandviewresearch.com/industry-analysis/amino-acids-market (accessed: 4th February 2023).

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

Gupta A, Reizman IMB, Reisch CR, Prather KLJ 2017. Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nature Biotechnology* 35: 273–279.

Gustavsson SÅ, Samskog J, Markides KE, Långström B 2001. Studies of signal suppression in liquid chromatography–electrospray ionization mass spectrometry using volatile ion-pairing reagents. *Journal of Chromatography A* 937: 41–47.

Harder B-J, Bettenbrock K, Klamt S 2018. Temperature-dependent dynamic control of the TCA cycle increases volumetric productivity of itaconic acid production by *Escherichia coli*. *Biotechnology and Bioengineering* 115: 156–164.

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

Hegner R, Rosa LFM, Harnisch F 2018. Electrochemical CO2 reduction to formate at indium electrodes with high efficiency and selectivity in pH neutral electrolytes. *Applied Catalysis B: Environmental* 238: 546–556.

Hemström P, Irgum K 2006. Hydrophilic interaction chromatography. *Journal of Separation Science* 29: 1784–1821.

Herz E, Antonovsky N, Bar-On Y, Davidi D, Gleizer S, Prywes N, Noda-Garcia L, Lyn Frisch K, Zohar Y, Wernick DG, et al. 2017. The genetic basis for the adaptation of *E. coli* to sugar synthesis from $CO_2$. *Nature Communications* 8: 1705.

Huang L, Yuan Z, Liu P, Zhou T 2015. Effects of promoter leakage on dynamics of gene expression. *BMC Systems Biology* 9: 16.

Iida A, Teshiba S, Mizobuchi K 1993. Identification and characterization of the tktB gene encoding a second transketolase in *Escherichia coli* K-12. *Journal of Bacteriology* 175: 5375–5383.

Jeong K, Kim J, Gaikwad M, Hidayah SN, Heikaus L, Schlüter H, Kohlbacher O 2020. FLASHDeconv: Ultrafast, High-Quality Feature Deconvolution for Top-Down Proteomics. *Cell Systems* 10: 213-218.e6.

Kasari M, Kasari V, Kärmas M, Jõers A 2022. Decoupling Growth and Production by Removing the Origin of Replication from a Bacterial Chromosome. *ACS Synthetic Biology* 11: 2610–2622.

Keller P, Reiter MA, Kiefer P, Gassler T, Hemmerle L, Christen P, Noor E, Vorholt JA 2022. Generation of an *Escherichia coli* strain growing on methanol via the ribulose monophosphate cycle. *Nature Communications* 13: 5243.

Kemmeren P, Sameith K, van de Pasch LAL, Benschop JJ, Lenstra TL, Margaritis T, O'Duibhir E, Apweiler E, van Wageningen S, Ko CW, et al. 2014. Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors. *Cell* 157: 740–752.

Keseler IM, Bonavides-Martinez C, Collado-Vides J, Gama-Castro S, Gunsalus RP, Johnson DA, Krummenacker M, Nolan LM, Paley S, Paulsen IT, et al. 2009. EcoCyc: A comprehensive view of *Escherichia coli* biology. *Nucleic Acids Research* 37: D464–D470.

Kim S, Lindner SN, Aslan S, Yishai O, Wenk S, Schann K, Bar-Even A 2020. Growth of *E. coli* on formate and methanol via the reductive glycine pathway. *Nature Chemical Biology* 16: 538–545.

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

Kolter R, Balaban N, Julou T 2022. Bacteria grow swiftly and live thriftily. *Current Biology* 32: R599–R605.

LaPorte DC, Walsh K, Koshland DE 1984. The branch point effect. Ultrasensitivity and subsensitivity to metabolic control. *The Journal of Biological Chemistry* 259: 14068–14075.

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

Lee J-H, Wendisch VF 2017. Production of amino acids – Genetic and metabolic engineering approaches. *Bioresource Technology* 245: 1575–1587.

Li S, Jendresen CB, Grünberger A, Ronda C, Jensen SI, Noack S, Nielsen AT 2016. Enhanced protein and biochemical production using CRISPRi-based growth switches. *Metabolic Engineering* 38: 274–284.

Li S, Jendresen CB, Landberg J, Pedersen LE, Sonnenschein N, Jensen SI, Nielsen AT 2020. Genome-Wide CRISPRi-Based Identification of Targets for Decoupling Growth from Production. *ACS Synthetic Biology* 9: 1030–1040.

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

Lo T, Van Der Schalie E, Werner T, Brun YV, Din N 2004. A Temperature-Sensitive Mutation in the *dnaE* Gene of *Caulobacter crescentus* That Prevents Initiation of DNA Replication but Not Ongoing Elongation of DNA. *Journal of Bacteriology* 186: 1205–1212.

Lu W, Su X, Klein MS, Lewis IA, Fiehn O, Rabinowitz JD 2017. Metabolite Measurement: Pitfalls to Avoid and Practices to Follow. *Annual Review of Biochemistry* 86: 277–304.

Lynch M, Louie M, Copley S, Spindler E, Prather B, Lipscomb M, Lipscomb T, Liao H, Hogsett D, Evans R 2019. Microorganisms and methods for the production of fatty acids and fatty acid derived products.

Lynch MD, Gill RT, Lipscomb TEW 2016. Method for producing 3-hydroxypropionic acid and other products.

Madalinski G, Godat E, Alves S, Lesage D, Genin E, Levi P, Labarre J, Tabet J-C, Ezan E, Junot C 2008. Direct Introduction of Biological Samples into a LTQ-Orbitrap Hybrid Mass Spectrometer as a Tool for Fast Metabolome Analysis. *Analytical Chemistry* 80: 3291–3303.

Markley JL, Brüschweiler R, Edison AS, Eghbalnia HR, Powers R, Raftery D, Wishart DS 2017. The future of NMR-based metabolomics. *Current Opinion in Biotechnology* 43: 34–40.

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

Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY 2013. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nature Biotechnology* 31: 170–174.

Na J, Seo B, Kim J, Lee CW, Lee H, Hwang YJ, Min BK, Lee DK, Oh H-S, Lee U 2019. General technoeconomic analysis for electrochemical coproduction coupling carbon dioxide reduction with organic oxidation. *Nature Communications* 10: 5193.

Nielsen J, Keasling JD 2016. Engineering Cellular Metabolism. *Cell* 164: 1185–1197.

Palsson BO, Abrams M 2011. *Systems Biology: Simulation of Dynamic Network States*. Cambridge: Cambridge University Press.

Park SY, Yang D, Ha SH, Lee SY 2018. Metabolic Engineering of Microorganisms for the Production of Natural Compounds. *Advanced Biosystems* 2: 1700190.

Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, Hawkins JS, Lu CHS, Koo B-M, Marta E, et al. 2016. A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell* 165: 1493–1506.

Phaneuf PV, Yurkovich JT, Heckmann D, Wu M, Sandberg TE, King ZA, Tan J, Palsson BO, Feist AM 2020. Causal mutations from adaptive laboratory evolution are outlined by multiple scales of genome annotations and condition-specificity. *BMC Genomics* 21: 514.

Piraner DI, Abedi MH, Moser BA, Lee-Gosselin A, Shapiro MG 2017. Tunable thermal bioswitches for in vivo control of microbial therapeutics. *Nature Chemical Biology* 13: 75–80.

Popov VO, Lamzin VS 1994. $NAD^+$-dependent formate dehydrogenase. *Biochemical Journal* 301: 625–643.

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152: 1173–1183.

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

Sander T, Farke N, Diehl C, Kuntz M, Glatter T, Link H 2019a. Allosteric Feedback Inhibition Enables Robust Amino Acid Biosynthesis in *E. coli* by Enforcing Enzyme Overabundance. *Cell Systems* 8: 66-75.e8.

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

Santos-Moreno J, Tasiudi E, Stelling J, Schaerli Y 2020. Multistable and dynamic CRISPRi-based synthetic circuits. *Nature Communications* 11: 2746.

Schmidt R, Steinhart Z, Layeghi M, Freimer JW, Bueno R, Nguyen VQ, Blaeschke F, Ye CJ, Marson A 2022. CRISPR activation and interference screens decode stimulation responses in primary human T cells. *Science* 375: eabj4008.

Schramm T, Lempp M, Beuter D, Sierra SG, Glatter T, Link H 2020. High-throughput enrichment of temperature-sensitive argininosuccinate synthetase for two-stage citrulline production in *E. coli*. *Metabolic Engineering* 60: 14–24.

Schramm T, Lubrano P, Pahl V, Stadelmann A, Verhülsdonk A, Link H 2023. Mapping temperature-sensitive mutations at a genome scale to engineer growth switches in *Escherichia coli*. *Molecular Systems Biology* e11596.

Shalem O, Sanjana NE, Zhang F 2015. High-throughput functional genomics using CRISPR–Cas9. *Nature Reviews Genetics* 16: 299–311.

Stincone A, Prigione A, Cramer T, Wamelink MMC, Campbell K, Cheung E, Olin-Sandoval V, Grüning N, Krüger A, Tauqeer Alam M, et al. 2015. The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biological Reviews* 90: 927–963.

Szima S, Cormos C-C 2018. Improving methanol synthesis from carbon-free $H_2$ and captured $CO_2$: A techno-economic and environmental evaluation. *Journal of CO2 Utilization* 24: 555–563.

Tang D-Q, Zou L, Yin X-X, Ong CN 2016. HILIC-MS for metabolomics: An attractive and complementary approach to RPLC-MS: HILIC-MS FOR METABOLOMICS. *Mass Spectrometry Reviews* 35: 574–600.

Thomason LC, Costantino N, Court DL 2007. *E. coli* Genome Manipulation by P1 Transduction. *Current Protocols in Molecular Biology* 79.

Tierrafría VH, Rioualen C, Salgado H, Lara P, Gama-Castro S, Lally P, Gómez-Romero L, Peña-Loredo P, López-Almazo AG, Alarcón-Carranza G, et al. 2022.

RegulonDB 11.0: Comprehensive high-throughput datasets on transcriptional regulation in *Escherichia coli* K-12. *Microbial Genomics* 8.

Walsh K, Koshland DE 1985. Branch point control by the phosphorylation state of isocitrate dehydrogenase. A quantitative examination of fluxes during a regulatory transition. *The Journal of Biological Chemistry* 260: 8430–8437.

Wang X, Han J-N, Zhang X, Ma Y-Y, Lin Y, Wang H, Li D-J, Zheng T-R, Wu F-Q, Ye J-W, et al. 2021. Reversible thermal regulation for bifunctional dynamic control of gene expression in *Escherichia coli*. *Nature Communications* 12: 1411.

Weber W 2003. Conditional human VEGF-mediated vascularization in chicken embryos using a novel temperature-inducible gene regulation (TIGR) system. *Nucleic Acids Research* 31: 69e–669.

Wendisch VF 2020. Metabolic engineering advances and prospects for amino acid production. *Metabolic Engineering* 58: 17–34.

Wendisch VF, Brito LF, Gil Lopez M, Hennig G, Pfeifenschneider J, Sgobba E, Veldmann KH 2016. The flexible feedstock concept in Industrial Biotechnology: Metabolic engineering of *Escherichia coli, Corynebacterium glutamicum, Pseudomonas, Bacillus* and yeast strains for access to alternative carbon sources. *Journal of Biotechnology* 234: 139–157.

Wingfield PT 2017. N-Terminal Methionine Processing. *Current Protocols in Protein Science* 88.

Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, et al. 2007. HMDB: the Human Metabolome Database. *Nucleic Acids Research* 35: D521–D526.

Wroblewska A, Dhainaut M, Ben-Zvi B, Rose SA, Park ES, Amir E-AD, Bektesevic A, Baccarini A, Merad M, Rahman AH, et al. 2018. Protein Barcodes Enable High-Dimensional Single-Cell CRISPR Screens. *Cell* 175: 1141-1155.e16.

Yanofsky C 1981. Attenuation in the control of expression of bacterial operons. *Nature* 289: 751–758.

Yao L, Shabestary K, Björk SM, Asplund-Samuelsson J, Joensson HN, Jahn M, Hudson EP 2020. Pooled CRISPRi screening of the cyanobacterium *Synechocystis* sp PCC 6803 for enhanced industrial phenotypes. *Nature Communications* 11: 1666.

Yishai O, Lindner SN, Gonzalez De La Cruz J, Tenenboim H, Bar-Even A 2016. The formate bio-economy. *Current Opinion in Chemical Biology* 35: 1–9.

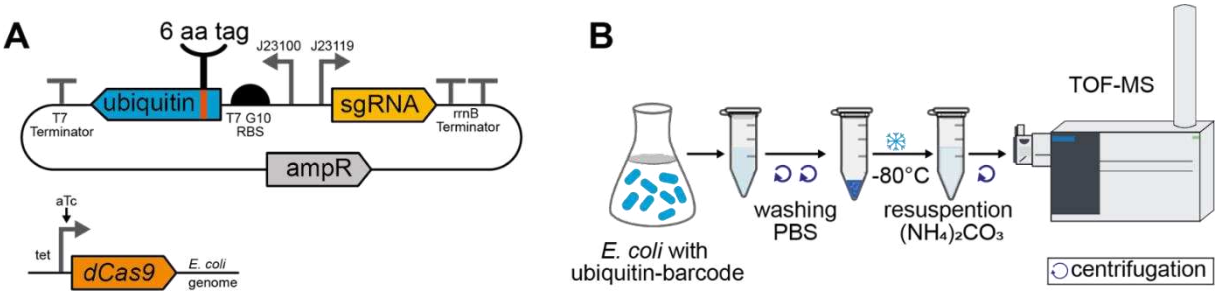
Yuan M, Breitkopf SB, Yang X, Asara JM 2012. A positive/negative ion–switching, targeted mass spectrometry–based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. *Nature Protocols* 7: 872–881.

Zelcbuch L, Antonovsky N, Bar-Even A, Levin-Karp A, Barenholz U, Dayagi M, Liebermeister W, Flamholz A, Noor E, Amram S, et al. 2013. Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Research* 41: e98–e98.

Zhang F, Carothers JM, Keasling JD 2012. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology* 30: 354–359.

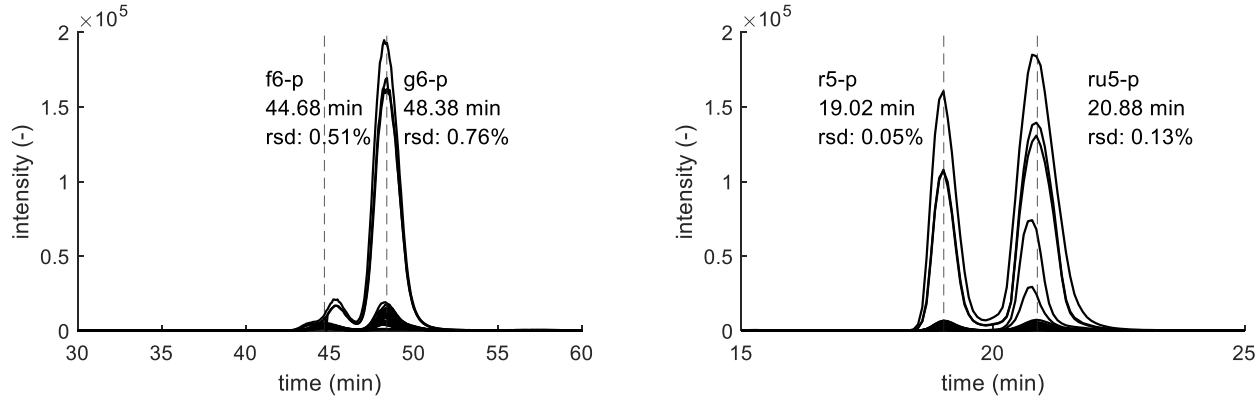
# Supplementary data

## 1 Supplementary data



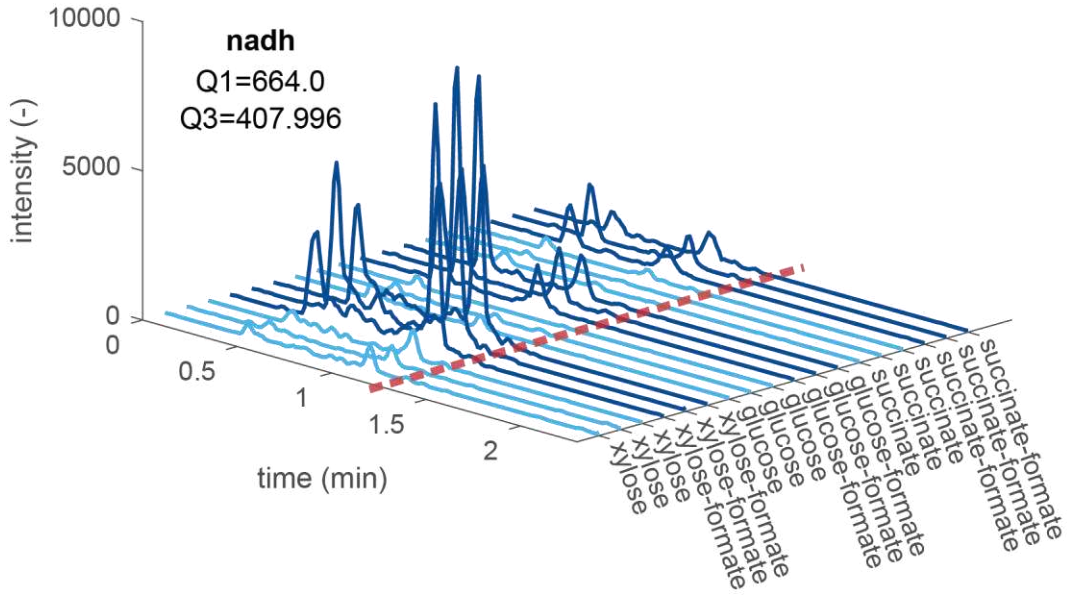

### Fig S1

**A** Combined expression system of sgRNA and ubiquitin (extended plasmid map). Ubiquitin-barcode and sgRNA are constitutively expressed (J23100, J23119). Ubiquitin-Barcode sequence was governed by a strong ribosomal binding site (G10 RBS) and a strong terminator (T7). **B** Short workflow for simultaneous extraction of metabolites and ubiquitin-barcode. Cells are washed twice with PBS. I used freeze draw to break the cells and water-based extraction and storage solution (NH4)2CO2 prior to injection into TOF MS (Agilent).



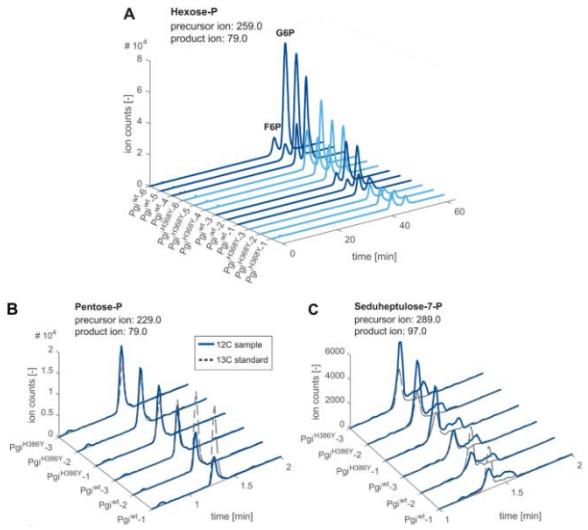

### Fig S2

**A** Chromatogram of hexose-phosphates measured using method A. f6p and g6p Retention time of authentic standards are indicates as dashed lines. Mean retention time (rt) and standard deviation (rsd) of peaks as indicated. **B** Chromatogram of hexose-phosphates measured using method B. r5-p and ru5-p Retention time of authentic standards are indicates as dashed lines. Mean retention time (rt) and standard deviation (rsd) of peaks as indicated.




## Fig S3

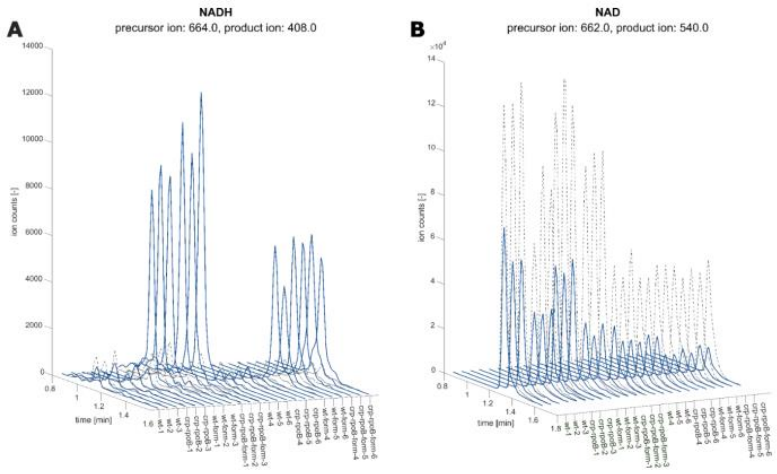
12C NADH chromatogram of *E. coli* BW25113 pFDH for different carbon sources. Precursor ion and product ion used for $^{12}C$ detection are indicated. Strains are grown with 30 mM formate (dark blue) or without the addition of formate (light blue). Cultures were grown in 12-Ill plates with n = 3 cultures.




## Figure S4

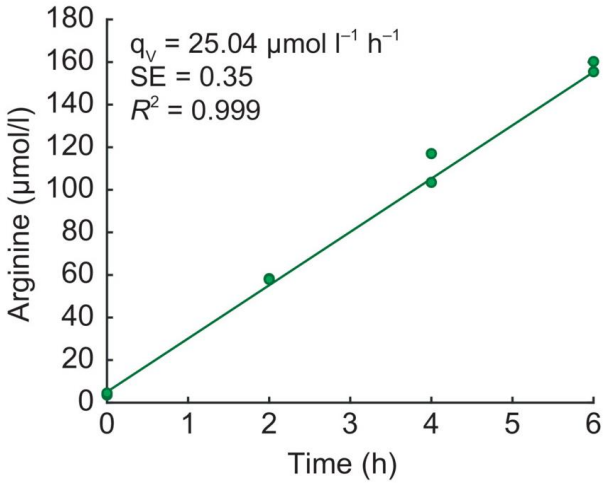
**A** Chromatogram of fructose-6-phosphate (F6P) and glucose-6-phosphate (G6P). Precursor ion (259.0) and 5 product ion (79.0) were used for 12C detection. Strains (n=6) are grown in rubisco dependent conditions. Samples were measured using a Shodex HILICpak VG-50 2D. The retention time of hexose-phosphates were determined with authentic standards of G6P and F6P and the 12C peak heights were used to quantify metabolites. Experiments were performed in two batches of n=3 cultures. First batch samples (replicates 1-3) were used for measurements in S4A, S4B and S4C. **B** Chromatogram of total pool of Pentose-phosphates (Ribulose-5P, Ribose-5P, Ribose-1P, and 10 Xylulose-5P - denoted P5P).

Strains were grown in a rubisco-dependent manner (n=3) cultures. Overlay of 12C chromatograms (continuous line) and 13C chromatograms (dashed line) with precursor ion and product ion used for 12C detection as indicated. Samples were measured using an Acquity BEH Amide column and the ratio of 12C (sample) and 13C (internal standard) peak heights was used to quantify metabolites. **C** Sedoheptulose-7P (S7P) chromatogram of the same samples and conditions as indicated in B.



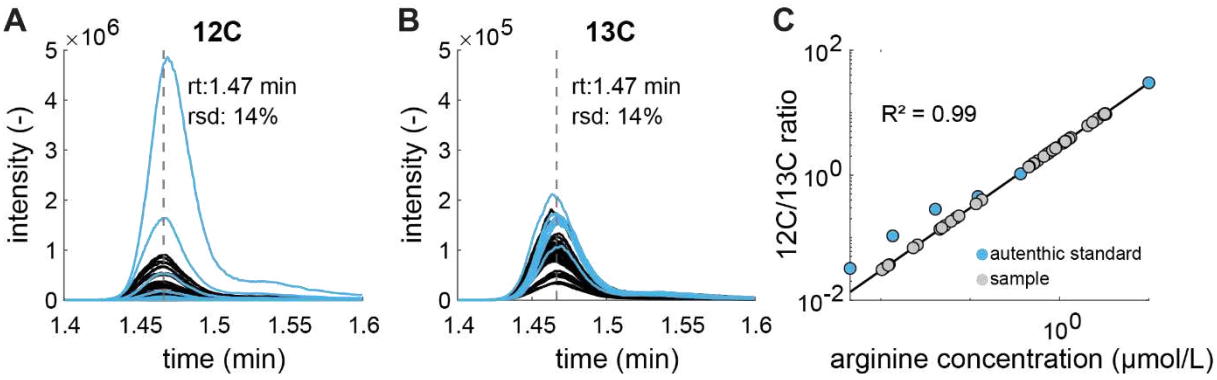

## Figure S5

**A** NADH chromatogram of *E. coli* BW25113 wild-type compared to a $Crp^{H22N}$ $RpoB^{A1245V}$ mutant. Experiments were performed in two batches of n=3 cultures. Chromatograms are sorted 5 accordingly. Overlay of 12C chromatograms (continues line) and 13C chromatograms (dashed line). Precursor ion and product ion used for 12C detection are indicated. Strains are grown with 30 mM formate (blue) or without the addition of formate (gray). **B** NAD chromatogram, same samples as in A. The ratio of 12-C (sample) and 13-C (internal standard) peak heights was used to quantify metabolites.



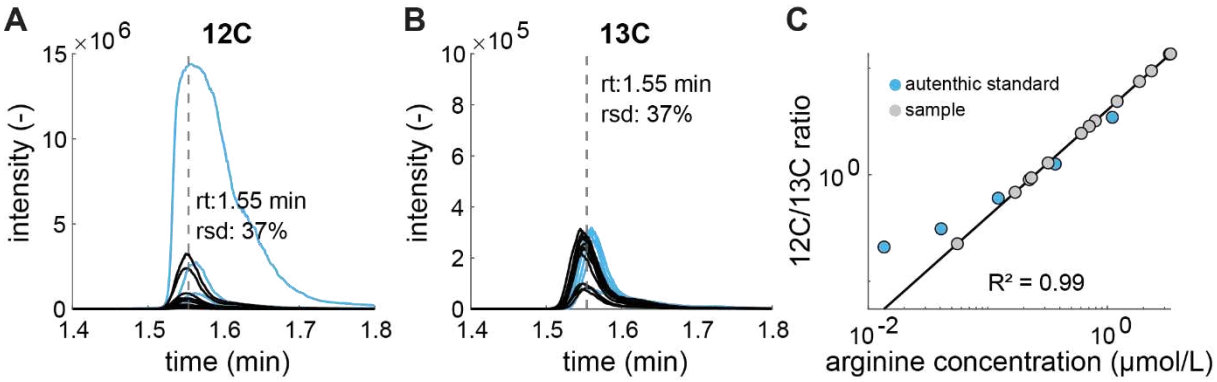

## Figure S6

The volumetric arginine production rate qV in the initial 6 h of cultivation was determined using a linear regression model (R2 = 0.99, SE: standard error of the slope qV).




## Figure S7

**A** Overlap of arginine chromatograms. 12C mass of arginine was used for MRM (Q1: 175 Q3: 60). Mean retention time (rt: 1.47 min) and relative standard deviation (rsd:14%) as indicated. Blue lines indicate 7 authentic arginine standards with concentrations between 10 µM and 5 nM. **B** Overlap of arginine chromatograms. 13C mass of arginine was used for MRM (Q1: 181 Q3: 61). Mean retention time (rt: 1.47 min) and relative standard deviation (rsd:14%) as indicated. Blue lines indicate 7 authentic arginine standards with concentrations between 10 µM and 5 nM. **C** Arginine calibration for absolute concentrations. Linear regression was used to calculate absolute sample concentrations based on 12C/13C ratio.




## Figure S8

**A** Overlap of arginine chromatograms. 12C mass of arginine was used for MRM (Q1: 175 Q3: 60). Mean retention time (rt: 1.47 min) and relative standard deviation (rsd:37%) as indicated. Blue lines indicate 6 dilutions of authentic arginine standards with concentrations between 10 µM and 5 nM. **B** Overlap of arginine chromatograms. 13C mass of arginine was used for MRM (Q1: 181 Q3: 61). Mean retention time (rt: 1.47 min) and relative standard deviation (rsd:37%) as indicated. Blue lines indicate 6 dilutions of authentic arginine standards with concentrations between 10 µM and 5 nM. **C** Arginine calibration for absolute concentrations. Linear regression was used to calculate absolute sample concentrations based on 12C/13C ratio.

# Acknowledgments

This thesis and the last three years have been quite a journey and an important part of my life. It would not have been possible without the people helping and supporting me over the years.

**Thus, I warmly thank...**

... my **supervisor**, Hannes Link, for the last three years. Thank you for giving me the opportunity to work on this research and for your supervision.

... my **thesis committee members**, Karl Forchhammer and Paul Hudson, for the feedback during the thesis advisory meetings.

... I want to thank the **members of my working group**. Moving the lab at the start of my PhD was stressful, but I am glad I got to work with great people I would not have met otherwise. I want to thank the two students I supervised, Nina and Victoria. I hope you learned and enjoyed the time as much as I did. And lastly, I want to especially thank **Ale**, **Ying**, **Johanna**, **Felicia**, **Selly**, and **Andy** for the scientific but mainly non-scientific conversations that made the stressful times much lighter.

... I want to thank our **colleagues from the Petras Lab** for a lot of cake, conversations, and great times in the big office.

... I want to thank **Eliya** and **Roee**, who are part of this thesis, but also the extended Milo Lab (**Benoit**, **Lior**, **Elad**, **Ron**, **Merav**, **Hadas**) for scientific discussions and collaborations but also for great holidays, visits, weddings, friendship, and more to come…

... **my friends**, first of all, **Sina** und **Korbi**, I would sail around the world with you anytime, without thinking twice. To all the Tübinger: **Stelio**, **Laura**, **Sina**, **Yazze**, **Leo**, **Nemo**, for so much more than just kniffeln, climbing, beer, holidays, movies or taking a walk.

... dir, **Steffi**, ganz speziell, wir wohnen schon so lange zusammen, dass ich es mir anders fast gar nicht mehr vorstellen kann (und will). Danke, dass du immer für mich da warst!

... meiner **Familie**: meinen Geschwistern **Jana**, **Jonas** und **Leon**, meiner **Oma** Ingrit, und vor allem meinen Eltern **Sylvia** und **Manfred**, dafür dass ihr immer an mich geglaubt habt und mit mir durch alle Hochs und Tiefs gegangen seid!

# Erklärung

Ich möchte hiermit bestätigen, dass die vorliegende Promotionsarbeit von mir eigenständig verfasst wurde. Ich habe ausschließlich die angegebenen Quellen und Hilfsmittel verwendet. Beiträge meiner Kooperationspartner, die in meiner Arbeit Verwendung finden, sind angemessen gekennzeichnet. Eine zusätzliche ausführliche Abgrenzung meiner eigenen Beiträge von den Beiträgen meiner Kooperationspartner habe ich im Abschnitt „Contributions and Publications“ vorgenommen.

_____ _____

Tübingen, den 02.10.2023 Vanessa Pahl

# Contributions and Publications

The herein presented work was prepared by myself. Part of the presented data are involved in a collaborative work with the contribution of other Scientists as indicated:

**Chapter 1 A protein-based barcode for CRISPR interference strains**

The top-down proteomics approach was initiated by Daniel Petras and Hannes Link. Paul Lubrano cloned the combined expression plasmids and cultivated the 8 CRISPRi strains. I developed the mass spectrometry and extraction methods, performed experiments and cloned strains. I curated MS1 spectra for metabolite and barcode analysis. I investigated, visualized the data. Finally, I wrote this chapter. Hannes Link designed 6 amino acid tags, conceived the study, discussed results and provided funding.

**Chapter 2 Quantification of sugar phosphate isomers in *E. coli* central metabolism with LC-MS/MS**

I developed the LC-MS/MS method, performed experiments and investigated intracellular metabolite concentrations and visualised the data. Finally, I wrote this chapter. Hannes Link conceived the study, discussed results and provided funding.

**Chapter 3 Metabolic changes that enable *E. coli* to use $CO_2$ as C-source**

Data for Figure 18 and 20 are part of a collaborative publication, currently in review. Eliya Milshtein and Roee Ben Nissan performed genomic insertion of Pgi, Crp and RpoB mutations. Eliya Milshtein further cultured and sampled the Pgi mutant strain. Roee Ben Nissan, Eliya Milshtein and I performed data curation, formal analysis, investigation, visualization. I developed and applied all metabolomics methods and wrote this chapter. Hannes Link and Ron Milo conceived the study, discussed results and provided funding.

Ben-Nissan R, Milshtein E, **Pahl V**, De Pins B, Jona G, Levi D, Yung H, Nir N, Ezra D, Gleizer S, Link H, Noor E, Milo R. 2023. Autotrophic growth of *E. coli* is achieved by a small number of genetic changes. preprint. elife

**Chapter 4 Temperature controlled two-stage arginine production in *E. coli***

Data for Figure 21 is part of a collaborative publication. Thorben Schramm and myself performed cloning of the Arginine overproduction strain. I performed experiments, curated, analysed, investigated, visualization the data. Finally, I wrote this chapter. Hannes Link performed calculation of absolute arginine concentrations, visualisation, conceived the study, discussed results and provided funding.

Schramm T, Lubrano P, **Pahl V**, Stadelmann A, Verhülsdonk A, Link H 2023. Mapping temperature-sensitive mutations at a genome scale to engineer growth switches in *Escherichia coli*. Molecular Systems Biology e11596