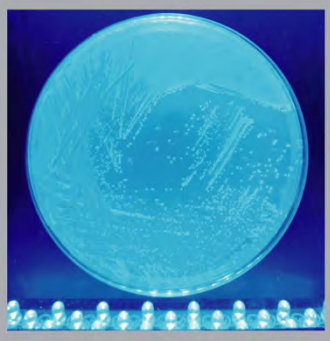
Selection-aided Metabolic Engineering of Microbial Cell Factories for Vitamin Production in Escherichia coli

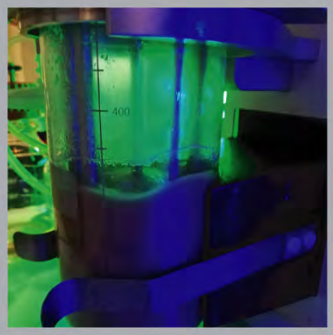
Bali, Anne Pihl

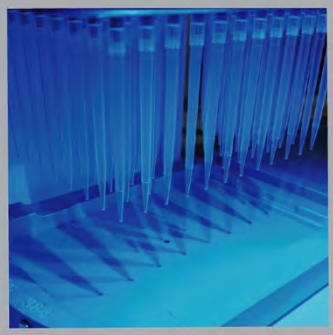
Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Selection-aided Metabolic Engineering of Microbial Cell Factories for Vitamin Production in *Escherichia coli*







Anne Pihl Bali
PhD thesis, October 2019
The Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark

Biosyntia

Selection-aided Metabolic Engineering of Microbial Cell Factories for Vitamin Production in *Escherichia coli*

PhD thesis by Anne Pihl Bali

Principal supervisors: Morten O. A. Sommer and Hans J. Genee

Co-supervisor: Luisa S. Gronenberg

October 2019

PREFACE

This industrial PhD thesis serves as a partial fulfillment of the requirements to obtain a PhD degree from the Technical University of Denmark (DTU). The work was conducted from the 1st of January 2016 to the 1st of October 2019 with a maternity leave between the 11th of June 2018 to the 11th of March 2019. The project was supervised by Dr. Hans J. Genee and Dr. Luisa S. Gronenberg from Biosyntia ApS and Professor Morten O. A. Sommer from Novo Nordisk Foundation Center for Biosustainability (NNF-CFB), DTU. The work was primarily performed at Biosyntia in collaboration with the Bacterial Synthetic Biology Section at NNF-CFB, DTU. In the spring of 2018 part of the work was carried out at an external research stay at the University of Stuttgart, Institute for Technical Biochemistry headed and supervised by Professor Bernhard Hauer, co-supervised by Dr. Jan Klenk. The work was funded partly by Innovation Foundation Denmark, Industrial PhD program, No. 5016-00135B, and partly by Biosyntia ApS.

Anne Pihl Bali, 1st of October 2019
Novo Nordisk Foundation Center for Biosustainability
Technical University of Denmark, Kgs. Lyngby

ABSTRACT

Microbial cell factories provide a promising alternative to the carbon-intensive chemical synthesis of a wide range of products. A key to the development of new, cost-competitive cell factories is the designing, building and testing of high-performing microorganisms and learning from this process. Success and throughput of this process is dependent on efficiently building microbial diversity and subsequent filtering it to identify the most effective design.

In this thesis, metabolic engineering approaches are combined with selection strategies to establish and improve cell factories for vitamin production in *Escherichia coli*. Single mutations in the global transcriptional regulator IscR leading to improvements in a range of cell factories dependent on iron-sulfur (FeS) cluster enzymes are identified. Characterization with proteomics and genetic engineering experiments revealed that the enhanced performance in biotin, thiamine and lipoic acid production strains was centered around FeS cluster formation and repair.

Transport proteins are the gatekeepers between the chemical world and intracellular metabolism. Functional annotation and understanding of sequence-function relationship are important for metabolic engineering, while also increasing our fundamental scientific knowledge. Through auxotrophic selection of metagenomic libraries, transporters with affinity for biotin sulfoxide (BSX) were found. These may improve biotin purity of future cell factories since import of BSX allows reduction of BSX to biotin by biotin sulfoxide reductase, BisC. Similarly, transporter-knowledge was gained in a directed evolution study of the nicotinamide riboside transporter, PnuC. Using random and directed mutagenesis strategies combined with a riboswitch-based selection system, PnuC transporters were engineered to accept thiamine as substrate, with only one to four amino acid changes.

Overall, the results of this thesis illustrate how selection systems can assist in removing production bottlenecks of cell factories, increasing the throughput, and enhancing the success rate of metabolic engineering, hopefully accelerating the development towards a more sustainable future.

DANSK RESUMÉ

Mikrobielle cellefabrikker er en lovende løsning på den øgede brug af fossile brændstoffer. De tilbyder et bæredygtigt alternativt til kemisk syntese af en række produkter. Design, konstruktion og evaluering af eksisterende mikroorganismer er essentielt for at kunne fortsætte og udvide udviklingen af nye konkurrencedygtige cellefabrikker. Succesen af denne proces er ofte afhængig af etablering og efterfølgende filtrering af mikrobiel diversitet for at identificere det mest effektive design.

I denne afhandling kombineres metabolic engineering tilgange med selektionsstrategier til at etablere og forbedre cellefabrikker til vitamin produktion i *Escherichia coli*. Her identificeres enkeltstående mutationer i den globale regulator IscR, der resulterer i forbedringer for en række cellefabrikker med jern-svovl (FeS) gruppe enzymer. Karakterisering med proteomanalyse og genetiske ekspressionsstrategier, viser at forbedret FeS-gruppe syntese og reparation er de væsentligste faktorer til den øgede præstation af biotin, thiamin og liponsyre produktionsstammer.

Transporterer er portnere mellem den kemiske verden og intracellulære metabolisme. Funktionel annotering og forståelse af sekvens-funktion relationen er vigtige redskaber i metabolic engineering samt øger vores grundlæggende videnskabelige forståelse. Auxotrofi-selektioner af metagenomiske biblioteker muliggjorde identificering af transporterer med affinitet for biotin sulfoxid (BSX). Disse kan forhåbentlig øge renheden af produceret biotin fra fremtidige cellefabrikker, da import af BSX muliggør reduktion til biotin via biotin sulfoxid reduktase, BisC. Tilsvarende er viden om transporterer erhvervet gennem et evolutions-studie af nicontinamid ribosid transporterer PnuC. Gennem mutations-strategier kombineret med et biosensor-baseret selektionssystem, blev PnuC-transporterer ændret til at acceptere thiamin som substrat. Substrat-specificiteten blev ændret gennem modificering af én til fire aminosyrer.

Resultaterne i denne afhandling illustrerer hvordan selektionssystemer kan hjælpe med at reducere flaskehalse, øge gennemløbet og højne succesraten af metabolic engineering, hvilket forhåbentligt kan accelerere udviklingen mod en mere bæredygtig fremtid.

ACKNOWLEDGEMENT

After more than three years of studies, countless hours of work, ups and downs, cries and laughter, failures and successes, pipetting, transformations and bioassays, I would like to thank the people who helped me through and accompanied me on the journey.

For the opportunity to conduct a Ph. D. and to work as a scientist, I would like to express my sincere gratitude to my supervisors **Morten**, **Hans** and **Luisa**, for urging and helping me to apply for funding from the Innovation Foundation Denmark. For their guidance, motivation and constructive feedback during the time of the thesis, which has helped me grow as a person as well as developing a strong scientific background.

During my Ph. D. studies I have been fortunate to supervise four gifted students; **Karolina**, **Nina**, **Monika** and **David**. Thank you for your hardworking attitudes, challenging questions and valuable inputs.

I have been extremely privileged to carry out my Ph. D. studies in an industrial environment at Biosyntia. Having the chance to be part of a dynamic and growing startup company has been a great source of motivation and inspiration. Being part of a team of excellent scientists and caring colleagues has help me grow as a person as well as a researcher. Therefore, I would like to thank all current and former employees of Biosyntia; **Bo**, **Lasse**, **Josi**, **Carlos**, **Linda**, **Martin**, **Nemeh**, **Katharina**, **Gijs**, **Daniel**, **Louis**, **Oana**, **Ben, Jing**, **Sofus**, **Line, Paul, Thijs**. Additional thanks to **Dóra**, **David** and **Katie** who provided critical reading of parts of this thesis.

A huge thank goes to **Nils**, for scientific as well as less scientific discussions, endless number of morning teas, and for always being ready with a helping hand.

In parallel with my work at Biosyntia, I have had the privilege to be enrolled as PhD. student at the Novo Nordisk Foundation Center for Biosustainability in the Sommer Lab. Having such a strong network of skilled scientists and accommodating colleagues has been priceless. I

would especially like to thank **Gitte** for helping me with all sorts of administrative struggles, for gentle reminders and quick replies.
I would also like to thank **Bernhard** Hauer and **Jan** for hosting me at the Institute for Technical Biochemistry, University of Stuttgart, in the Biocatalysis group. For their help in settling both outside and in the laboratory, and the opportunity to get hands-on experience with *in vitro* related experiments.

Finally, I would like to thank my entire family for their endless support and love. To **Jonathan** for giving me a new perspective on the world – especially on sleep. To **Asger** for giving me the space and support to develop my love for science, and for countless adventures with many more to come.

LIST OF PUBLICATIONS

The work presented in this thesis has formed the foundation for, or contributed significantly to, the following manuscripts and patent applications

Improved Biotin, Thiamine, and Lipoic Acid Biosynthesis by Engineering the Global Regulator IscR
Bali, A. P., Lennox-Hvenekilde, D., Myling-Petersen, N., Buerger, J., Gronenberg, L.S., Salomonsen, B., Sommer, M. O. A. and Genee, H. J. (*Manuscript submitted to Nature Chemical Biology*)

Metagenomic Mining of Transporters with Affinity for Biotin Sulfoxide
Bali, A. P., Myling-Petersen, N. and Genee, H. J.
(*Manuscript in preparation*)

Directed Evolution of Membrane Transport Using Synthetic Selections
Bali, A. P.*, Genee, H. J.* & Sommer, M. O. A. (2018) *ACS Syntetic Bioogy,* 7 (3), pp. 789–793. doi: 10.1021/acssynbio.7b00407
Note: Main work carried out during master thesis.

Cell Factory Having Improved Iron-Sulfur Cluster Delivery
Bali, A. P., Genee, H. J. and Myling-Petersen, N. (2019) Patent application WO2019/012058 A1

Cell Factories for Improved Production of Compounds and Proteins Dependent on Iron-Sulfur Clusters
Bali, A. P., Genee, H. J., Myling-Petersen, N., Acevedo-Rocha, C. G., Lauridsen, L. H., Gronenberg, L.S., Golabek, M., and Lubrano, P. Patent application EP19152181.4

Additionally, during the Ph. D. studies minor contributions have been provided to the following manuscripts

Functional Mining of Transporters using Synthetic Selections
Genee, H. J., Bali, A. P., Petersen, S. D., Siedler, S., Bonde, M. T., Gronenberg, Kristensen, M., Harrison, S. J. and Sommer, M. O.A. (2016) *Nature Chemical Biology* (12), pp. 1015–1022, doi: 10.1038/nchembio.2189
Note: Additional contributions were provided during master thesis studies.

Accurate Prediction of the Thiamine Biosynthetic Landscape using Gaussian Process Learning
Genee, H. J.*, Riesselman, A. J.*, Petersen, S. D., Nath, S., Gronenberg, L. S., Salomonsen, B., Bali, A. P., Smart, K., Chan, L. J. G., Nhan, M., Baidoo, E. E. K., Wang, G., Oberortner, E., Hillson, N. J., Keasling, J. D., Marks, D. S., Petzold, C. J., Deutsch, S., and Sommer, M. O.A. (*Manuscript in preparation*)

* Denotes equal contribution.

THESIS STRUCTURE

This thesis has been divided into five chapters. The first chapter gives a general introduction to central areas for the thesis work. The following three chapters each present a scientific study written in the form of manuscripts, while the last chapter provide a brief conclusion to the thesis.

Chapter I briefly describe the field of metabolic engineering and offer an introduction to cell factories; two subjects fundamental for the work of this thesis. As biotin and FeS-cluster enzymes are main parts of the studies presented, the introduction chapter continues with an overview of engineering work carried out for microbial production of biotin. It gives a brief summary of FeS-cluster biogenesis and regulation, which serve as an extension of the introduction to the study presented in **Chapter II**.

Chapter II presents the main work efforts of the thesis. It describes the identification and characterization of mutations that reduce growth toxicity observed from overexpression of biotin synthase, BioB. The identified mutations improve biotin production in cell factories. The mutations are hypothesized to improve FeS-cluster supply and repair, and comparisons to alternative strategies are performed. Finally, the beneficial effects are shown to extend to cell factories for thiamine and lipoic acid production.

Chapter III focuses on future improvements of biotin cell factories, specifically handling of oxidized biotin. Through metagenomic mining several proteins, many being transporters, are identified to enhance the BisC mediated conversion of biotin sulfoxide to biotin. Import of oxidized biotin from the fermentation broth is concluded to be an initial bottleneck, and additional work to further decrease the concentration of oxidized biotin is discussed.

Chapter IV describes a directed evolution study, where the substrate recognition site of the nicotinamide riboside transporter PnuC is investigated and altered to accept thiamine by means of synthetic biology. Hopefully, the increased understanding of the sequence-function relationship of the transporters as well as the phenotypic data will be of value for future metabolic engineering efforts. The main contributions to this chapter provided during the timeframe of the thesis work, has encompassed validation experiments, figure generation and manuscript writing.

Chapter V summarizes conclusions and perspectives of the results presented in this thesis.

ABBREVIATIONS

ACM	Acidomycin
$AdoB_{12}$	Adenosyl-cobalamin
AEC	S-2-aminoethylcysteine
ANA	6-aminonicotinamide
ATC	A-type carrier
ATP	Adenosine triphosphate
BSN	Biotin sulfone
BSX	Biotin sulfoxide
DAPA	7,8-diaminopelargonate
DTB	Dethiobiotin
DWP	Deep-well plate
ETN	Ethionine
FeS-cluster	Iron-sulphur cluster
GFP	Green flourescence protein
iBAQ	Intensity-based quantification
IPTG	Isopropyl-β-D-thiogalactoside
KAPA	8-amino-7-oxononanoate
MAGE	Multiplex automated genome engineering
MeB_{12}	Methyl-cobalamin
MS	Mass spectrometry
NADPH	Nicotinamide adenine dinucleotide phosphate
NGT	N-methyl-N'-nitro-N-nitrosoguanidine
OHB_{12}	Hydroxy-cobalamin
PCR	Polymerase chain reaction
RBS	Ribosome binding site
ROS	Reactive oxygen species
SAM	S-adenosylmethionine
TMP	Thiamine monophosphate
TPA	5(2-thienyl)pentanoic acid
TPP	Thiamine pyrophosphate
TVA	5-(2-thienyl)-n-valeric acid
α-HB	α-dehydrobiotin
β-CA	β-chloro-D-alanine
β-HN	β-hydroxy-norvaline

TABLE OF CONTENTS

CHAPTER I
INTRODUCTION

Metabolic Engineering for Establishing Cell Factories

"In any future, metabolic engineering will soon rival and potentially eclipse synthetic organic chemistry."
Jay D. Keasling [1]

The growing world population and globally increasing living standards raises the need for food, energy and manufactured products. The ever-increasing production of these has currently a negative impact on climate and challenges our way of living[2]. The industry is heavily dependent on unsustainable chemical production relying on nonrenewable, fossil-derived energy consumption and catalysts. Often these are associated with environmentally harmful byproducts that are difficult to recycle and degrade. In order to accommodate the increasing demand of chemical synthesis, industrial production needs to adapt sustainable alternatives applying renewable precursors, returning high productivity and reducing the environmental impact of production[3,4].

Cell Factories

The use of cell factories is a potential solution to many of the challenges associated with chemical production and its reliance on high usage of fossil fuels. Cell factories is the use of microbial cells as production facilities. However, microbes has been utilized for production of fermented food and beverages for thousands of years predating the cell factory concept that was established in the beginning of the 20th century[5]. Since then, cell factories based on bacteria, filamentous fungi and mammalian cells have been established to produce a wide range of products including antibiotics, amino acids, immunosuppressants, and vitamins. This production has been strongly accelerated with the

development of genetic engineering in the 1970's [6,7]. Cell factories have a water-based production, avoiding the use of heavy metals, strong acids and bases and organic solvents. Furthermore, cell factories minimize the formation of hazardous waste and chemical pollutants compared to the counterpart; chemical synthesis. The general idea for cell factories is to use renewable feedstocks for production under relatively low temperatures and pressure, which decreases the energy input required compared to chemical synthesis. Besides the potential to be a more environmentally-friendly production method, microbial cell factories are often less likely to produce unwanted byproducts, due to the high substrate specificity of enzymes [3,8].

Traditionally cell factories were native producers with relatively high yields of their natural products, where minimal genetic engineering efforts were needed to establish a reasonable production. However, only few of such native producers are capable of producing sufficient amounts of product to be cost-competitive to chemical synthesis and they are typically slow growing [6]. One very successful example is the production of penicillin from *Penicillium chrysogenum*, which has been improved more than 10.000-fold over time, mainly through mutagenesis and screening [9]. With the introduction of metabolic engineering, "platform organisms" such as *Escherichia coli, Saccharomyces cerevisiae* and *Bacillus subtilis,* are now the preferred route for developing a new cell factory. Numerous advantages are associated with the use of platform cell factories; well-characterized genetics and physiology, extensive genetic toolboxes for genome editing, introduction of foreign DNA and analysis methods. Furthermore, product approval by governmental organizations are typically easier if the host has already been used for a wide product range [1,3,10].

The Emergence of Metabolic Engineering

Introduced in the late 1980's-early 1990's, metabolic engineering is the field within biotechnology exploring the possibilities and limitations of directed improvement of cell metabolism towards specific production goals. An overall goal of metabolic engineering is to enable cost-effective biosynthesis of molecules, which are not possible to produce through chemical synthesis routes. Additionally metabolic engineering seek to replace chemical synthesis by more sustainable, environmentally-friendly production methods [1,11]. The field of metabolic engineering focuses on development of principles and tools that enable construction and optimization of microbial cell factories. This may be done by for example tapping into naturally occurring pathways in specific host organisms, expressing heterologous non-native pathways in well-characterized hosts, or engineering *de novo* biosynthetic pathways for synthesis of various natural and non-natural products [8]. Synthetic biology is a core tool to achieve such constructions. It allows for introduction of heterologous genetic material as well as rewiring of metabolisms and genetic networks to optimize production of specific compounds[12] based on technologies such as PCR [13], sequencing [14], modern molecular cloning [15,16], and genetic editing by CRISPR [17], MAGE [18] and other recombination techniques. Also the use of characterized standard parts is an important tool within Synthetic Biology to accelerate construction of cell factories [19]. Together these developments can facilitate enhanced pathway yields which is critical to reach industrially viable product titers.

The Metabolic Engineering Cycle

A common reference point for establishing a cell factory and a central dogma in the field of metabolic engineering is the "Metabolic Engineering Cycle". Traditionally it has consisted of three main steps; design, test and build, used as foundation to evaluate the metabolic effects of random or (semi-)rational genetic manipulations. With increased

understanding of microorganisms on the molecular level as well as major improvements in automation, modelling of biology, omics techniques, flux measurements and the emergence of systems biology, a fourth element, learn, has generally been included. Evaluation and reflection of results from one iteration has naturally affected the design of the next round of the design-test-build-(learn)-cycle (**Fig. 1**). However, emphasizing the importance of collecting, storing and analyzing data in a more structured way, may likely accelerate improvements in engineering efforts [6,11].

Design

When initiating design of a metabolic pathway in a microorganism, to produce a desired compound, several factors needs to be taken into consideration. The choice of host chassis will affect which basic metabolites are available for pathway construction, the genetic toolbox accessible and the overall "rules" determined by the regulatory network of the cell. Factors like cost and availability of the starting material, product toxicity, as well as processing conditions necessary to produce and purify the desired product will likewise influence the choice of organism. Next, pathway enzymes need to be identified, both native and heterologous alternatives. Regulation, cofactor usage and other enzyme requirements such as pH and temperature are only some of the parameters critical to consider, when designing a pathway. Furthermore, evolving or constructing enzymes *de novo* might be necessary if no specific enzymes can be identified for the desired product. Protein engineering is itself a time-consuming and complex field [6,8]. Because it can be challenging to specify an optimal design from the beginning, often library constructs are designed, leading to a large number of pathway variants which need to be constructed and evaluated[20].

Build

The *build* phase comprises the construction of the designed metabolic pathway or libraries into the desired host, including modulation of the regulatory network and knocking-out competing genes to direct flux toward the product of interest. In practice, establishing a pathway typically includes synthesis and introduction of large DNA constructs, specific genome edits and efficient sequencing for validation of modifications. Great expansions and improvements, especially involving automation, have been achieved in these areas over the past years, as discussed above, [21,22] increasing the throughput of the build phase. Combinatorial pathways and other types of DNA libraries may be part of the *design* and *building* aspects, in order to expand the diversity space. Getting help from nature, which is a fantastic problem solver, with millions of years of practice through evolution is also a possibility. The biodiversity in nature contains near unlimited solutions to complex problems. Through use of metagenomic libraries we may be able to tap into this enormous solution space. However, the larger the library size constructed, the greater the efforts needed to identify the most promising constructs to evaluate [20,23].

Test

Testing metabolic pathways revolves around anything that determines the efficacy of the previous steps including growth and physiological characterization of the engineered cells, transcript, protein and/or product profiles of the constructed pathway and much more. High-throughput methods such as omics-technologies are interestingly used as characterization tools, as these allow for global analysis of the cellular metabolism, and efficiently feed significant amounts of data into the *learning* step [11]. However, throughput constraints often limit the number of tests that can be run on each strain. High-throughput screening or selection strategies to identify high performing strains from the constructed libraries in the build phase, can serve as a valuable filtration step prior to thorough characterization, even though significant efforts

often are needed in order to establish efficient workflows with a low rate of false positive hits[24].

Learn

Even though it has been intuitively used since the metabolic engineering cycle was put into words, having learn as an explicitly defined step in the wheel helps to emphasize the, sometimes neglected, importance. With the steep increase in data generation and collection, learning is slowly moving away from typically being carried out in a nonsystematic way, lacking statistical rigor, relying on ad hoc observations and literature data from individual researchers. Now learning is primarily based on IT-implemented solutions. The elevated use of systems biology and statistical methods to analyze data from engineered

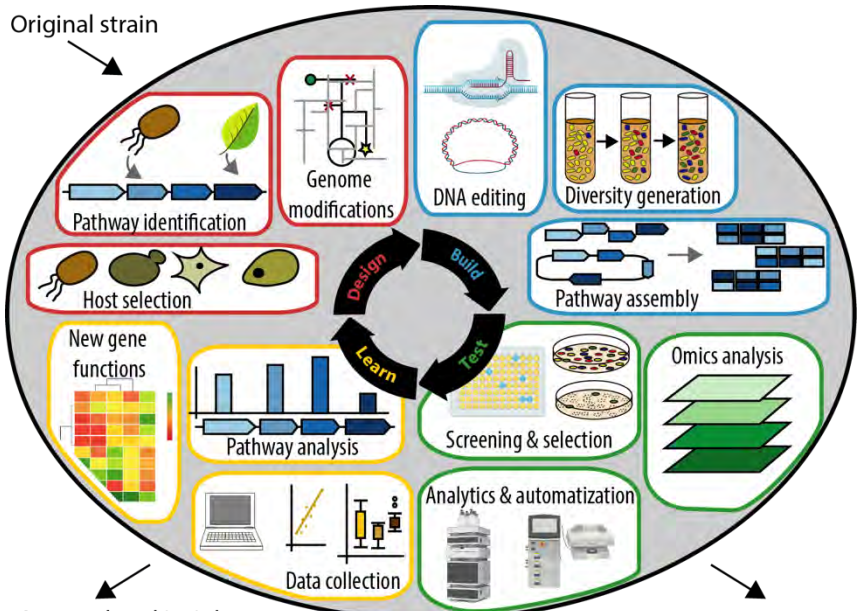


Figure 1: The Metabolic Engineering Cycle is constituted of four parts which can be iterated a number of times to assist in cell factory engineering. From identifying a target molecule, selecting a host and investigating the possible pathway combinations in the design phase. Through building and assembling the parts, generating diversity which can be filtered for improved clones using screening and selection systems and evaluated for specific performance in the test phase. To collecting, analyzing and storing data to learn now and in future setups, the cycle touches upon all aspects needed for evolving an improved strain.

organisms and inform the next round of design, are likely to improve metabolic engineering efforts in the future. Formalization of engineering workflows and data handling to improve the learning process, might help accelerating the efforts in many experienced laboratories [6].

Selection and High-Throughput Screening to Increase Evaluation Throughput

As briefly highlighted above, an important tool for increasing the throughput of metabolic engineering is the application of selection and screening systems, to narrow down which constructs to test and evaluate. Selection systems rely on the fact that only cells overcoming a *specific problem* will grow. This can be done by coupling product formation to antibiotic resistance for example through conditional expression from transcription factors[20,25], or natural[26] or synthetic riboswitches[27] sensing a molecule of interest. Natural requirements for product formation can also be used as a selection approach. Introducing auxotrophy, for example using gene knockouts, represents a relatively straight-forward way to establish a selection system, but it requires that the molecule of interest can become growth-limiting in the selection host [23]. Antimetabolite selections, though also restricted by essentiality of target metabolite, provide an alternative approach to construct dependencies on target molecule in the host strain. Here metabolite analogs inhibiting growth force the cell to elevate production of the target metabolite, for example through elevated enzyme concentrations, increased enzyme activity or minimized use in competing pathways [24,28].

High throughput screenings where product formation is coupled to a measurable output, are other powerful tools as alternatives or supplements to selection systems. Fluorescence is an example output for screening assays, which allows automated sorting on single cell level in FACS[29] or droplet screening[30]. Typically, fluorescent-based screens rely on sensors, which could also be coupled to growth through antibiotic resistance, as

described above. Where selections often provide a “yes” or “no” answer based on growth, high-throughput fluorescence output may offer a deeper insight into the distribution of the metabolite of interest in the library tested [23,24].

However, for most types of selections and screenings, tuning the system to get in the range of interest for industrial purposes and decreasing the false positive rate is a challenging, but highly important job, determining the applicability of the system [24]. In addition to the sensitivity other important factors need to be considered, often complicating the establishment of selection systems. These include but are not limited to the dynamic range and linear range of detection, in which the concentration of the molecule of interest correlates linearly with reporter signal [23].

It is clear that establishing cell factories is not without challenges. Even with advances in the fields adjacent to metabolic engineering, which has helped establishing toolboxes supporting and accelerating metabolic engineering, the development of cost-competitive cell factories is still difficult and lengthy work. Three main parameters are used to evaluate a cell factory performance; titer (final concentration of the product in the fermentation broth), rate (production per unit of time) and yield (units of product synthesized per unit of raw material consumed). Going from the idea conception, through a proof-of-principle cell factory to improve strain performance enough to meet commercial expectations, requires significant metabolic engineering efforts. Such efforts are estimated to take 6-8 years, beyond 100 research years and cost more than $50 million USD[6,7]. The length of time-to-market and the complexity of microbial systems help to explain why we are still lacking microbial cost-competitive production for a wide range of products.

Cell Factories for Biotin Production

Vitamin B7, also called biotin, is commercially produced by an expensive and polluting chemical route [31]. Due to this, engineering of an economical competitive cell factory for sustainable biotin production, has been attempted for several decades, but only few improvements has been achieved over the years [32]. In the following a brief description of the biotin biosynthesis in *E. coli* are given followed by an introduction to the metabolic engineering efforts carried out for a bio-based production of biotin.

Biotin and its Biosynthesis

Biotin is essential for all three domains of life, due to its role as a cofactor in carboxylation reactions. In its active state, biotin is covalently attached to cognate enzymes where it facilitates the transfer of CO_2 equivalents between the carboxylation and carboxyl transfer domains of the enzyme [33]. Biotin synthesis is limited to microbes, fungi and plants, whereas mammals are dependent on obtaining biotin from the diet and/or intestinal microflora. The synthesis pathway can be divided into two stages; synthesis of the pimelate moiety and assembly of the bicyclic rings. The latter part is largely conserved among biotin-producing organism, while the production of the pimelate moiety has quite some diversity.

In *E. coli* six enzymes constitute the biosynthesis pathway of biotin. With assistance from the fatty acid synthetic pathway, *bioC* and *bioH* are responsible for synthesis of Pimeloyl-ACP, whereas *bioFADB* produces biotin through the intermediates 8-amino-7-oxononanoate (KAPA), 7,8-diaminopelargonate (DAPA) and dethiobiotin (DTB) (**Fig. 2A**) [34]. Biotin biosynthesis is an expensive biological process, consuming no less than four equivalents of NADPH, four equivalents of S-adenosylmethionine (SAM) and one equivalent of ATP. It additionally requires regeneration of a [2Fe2S] cluster after each catalytic cycle, as it donates a sulfur atom to DTB to form biotin [35,36]. The DTB to biotin conversion step is itself

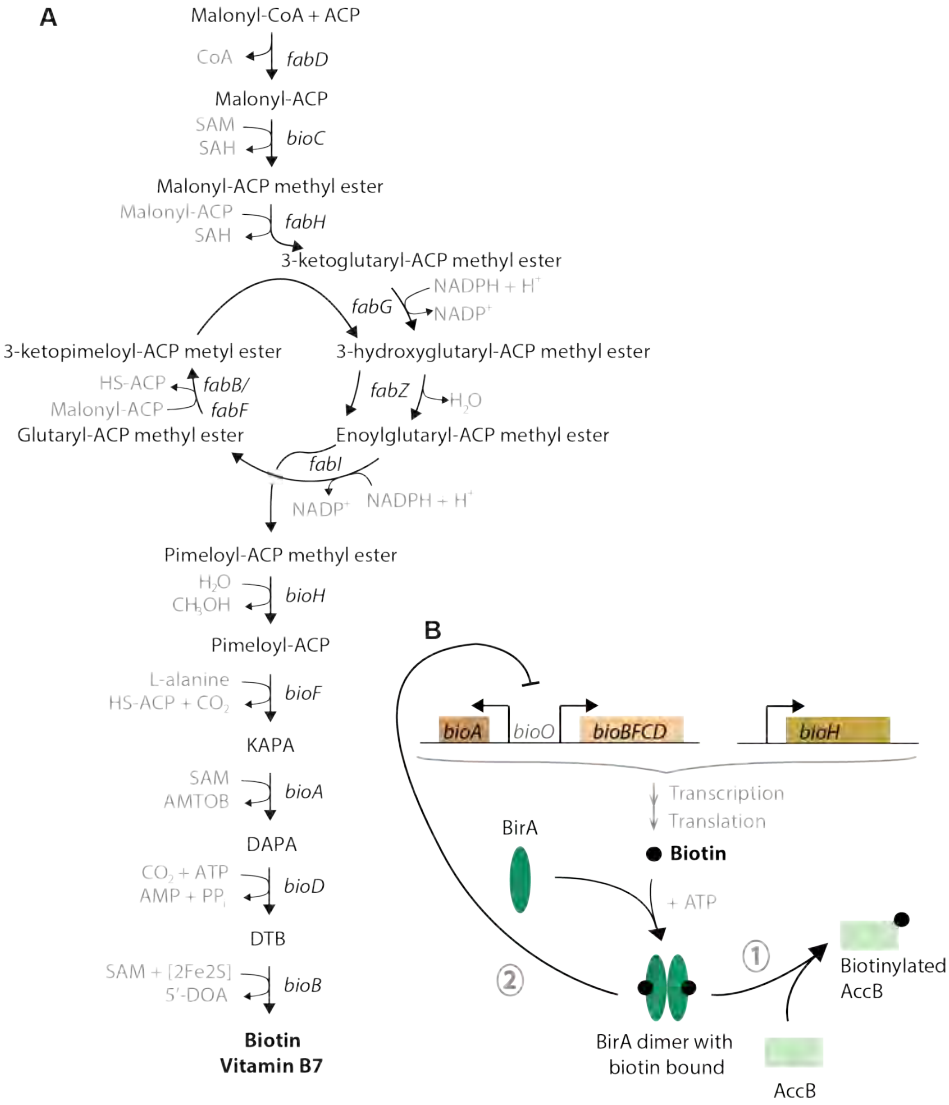


Figure 2: Overview of Biotin Biosynthesis and Regulation in *E. coli*. A) The biosynthetic pathway of biotin in *E. coli* is encoded by *bioABCDFH* and utilizes the fatty acid cycle for two rounds in order to elongate malonyl-ACP methyl ester to pimeloyl-ACP methyl ester. Cofactors are shown in grey, and responsible enzymes in italics. **B)** Schematic drawing of regulation of the biotin biosynthesis genes in *E. coli* [39]. When biotin is produced, it becomes bound by BirA 1) Liganded BirA has first priority to act as a ligase, covalently attaching biotin to biotin acceptor proteins, such as AccB. 2) When biotin acceptor proteins are biotinylated, BirA acts as a transcriptional repressor of the biotin operon.

extremely costly to the cell. Because of this and the low cellular requirement of a few molecules of biotin, expression of the biotin operon is tightly regulated on the transcriptional level by biotin ligase and repressor, BirA. BirA catalyzes the formation of bio-5'-AMP (holo-BirA) from biotin and ATP and binds the *bioO* regulator site, in response to increased biotin concentrations [36] (**Fig. 2B**). The primary role for holo-BirA is to act as a ligase, transferring biotin to the biotin carboxyl carrier protein, AccB, which is part of the essential enzyme, acetyl CoA carboxylase. When biotin is in excess and holo-BirA accumulates, dimerization occurs. The dimerization enable binding to the *bioO* regulator site effectuating feedback inhibition of the biotin operon [37,38].

Metabolic Engineering Efforts for Biotin Cell Factories

For most of the early work towards a bio-based production method for biotin, random mutagenesis followed by selections with drugs or antimetabolites were widely used. This has slowly moved towards a more rational, genetic engineering approach concurrently with the establishment of metabolic engineering and the development of high-throughput tools and techniques such as cheap DNA synthesis and sequencing as discussed previously. Applying several rounds of mutagenesis and selections often comes with the cost of a complex genetic background of the production organisms, which may affect the growth rate and general healthiness of the strains negatively. An overview of selected examples of biotin cell factories including both patented and peer-reviewed work, can be seen in **Table 1**.

Introduction to the Patent Landscape of Biotin

Several of the industry-leading companies in the vitamin industry have worked on establishing a cost-competitive bio-based production process for biotin, forming a significant patent landscape around biotin production from microorganisms. Some of the highest reported biotin

titers to-date, can be found in patented work from the Japanese companies Tanabe and Shiseido from the 1990's. Exposing *Serratia marcescens* for several, consecutive rounds of random mutagenesis and antimetabolite selection combined with overexpression of the native biotin operon enabled Tanabe to reach biotin titers of 600 mg/L, from a 11-day fed-batch

Table 1: Examples of Biotin Cell Factories constructed by either rational engineering and/or selection/screening of mutagenesis libraries. Abbreviations of selection compounds: β-CA; β-chloro-D-alanine, ANA; 6-aminonicotinamide, β-HN; β-hydroxy-norvaline, ACM; acidomycin, TVA; 5-(2-thienyl)-n-valeric acid, ETN; ethionine, AEC; S-2-aminoethylcysteine, TPA; 5(2-thienyl)pentanoic acid, α-HB; α-dehydrobiotin.

Host organism	**Genetic alterations** (chemicals for selection or specific genes manipulated)	**Biotin** (mg/L)	**Reference**
Escherichia coli	ACM, TVA, *β*-CA, ANA, *β*-HN, *bio*-operon with *bioO* mutation	970	Kanzaki *et al,* 1999[44]
Serratia marcescens	ACM, TVA, ETN, AEC, *bio*-operon, *parB*	500	Sakurai *et al*, 1993[56]; Sakurai *et al,* 1994[42]
Putida mutabilis	*Bio*-operon, *E. coli birA, isc*-operon overexpression, *metK* mutant, *B. subtilis bioWI*	272	Xiao *et al,* 2019[55]
Kurthia sp. 538-KA26	*BioB* from *bioH* promoter, ACM, TVA	126	Hoshino *et al*, 1997[48]; Furuichi *et al*, 2005[47]
Agrobacterium	Synthetic *E. coli bio*-operon, *bioB* RBS engineering	110	Shaw *et al,* 1999[53]
Sphingomonas sp.	*Bio*-operon, mutated *bioB*	66	Saito *et al,* 2000[54]
Escherichia coli	*Bio*-operon, *csdA*, *metK*	52	Schröder and Hauer; 1998[45]; Schröder, 1999[46]
Escherichia coli	*Bacillus sphaericus bio*-operon	45	Brown *et al,* 1991[51]
Bacillus subtilis	*Bio*-operon, *bioA*, TPA, α-HB	20	Bower *et al,* 1999[49]; Van Arsdell *et al,* 2005[50]

fermentation process [40–42]. Similarly, Shiseido managed to improve DTB production 7-fold in *E. coli* using random mutagenesis provoked by NGT treatment (N-methyl-N'-nitro-N-nitrosoguanidine) followed by screening for strains with reduced glucose assimilation capabilities. However, the efforts did not lead to improvements in biotin titers [43], suggesting a bottleneck in BioB for these strains. In collaboration with Takeda, Shiseido reported in 1995 the highest biotin titer to date: 970 mg/L from a 3-day fermentation process in rich media. Like the work carried out by Tanabe, Takeda and Shiseido used several rounds of mutagenesis by NGT treatment followed by drug-resistance/antimetabolite selections combined with plasmid-based overexpression of the native biotin operon with deregulated *bioO* site to develop a biotin-producing *E. coli* cell factory [44].

Focusing on the bottleneck of the BioB reaction, rational engineering, rather than random mutagenesis, was used by BASF to establish an *E. coli* strain overexpressing the biotin-related genes along with the expression of the desulfinase *csdA* to overcome sulfur limitations in the DTB to biotin reaction. Overexpression of *csdA* was attempted following a bioinformatic search for *nifS* homologs with a conserved motif, in order to identify a potential accessory enzyme for the BioB catalyzed reaction [45]. This exemplifies how specific considerations may feed into the design of a metabolic engineering strategy. The engineered strain was further improved by overexpression of *metK* to reach final titers of 52 mg/L biotin after 24 hours in a 10 mL batch culture [46]. This is 10- and 20-times less than the strains patented by the Japanese companies Tanabe and Shiseido respectively, however the production conditions are very different making it impossible to compare directly.

Using the natural high biotin producing organism *Kurthia sp.* 538-KA26 as a host strain, DSM managed to improve biotin titers from 14 mg/L to 39 mg/L in a 5-day long fermentation, by merely overexpressing the native BioB from the BioH-promoter on a plasmid [47]. This is another

example on rational, metabolic engineering to improve biotin production, addressing the observed bottleneck of BioB, to overcome accumulation of DTB. The strain was additionally improved using antimetabolite selection to produce 126 mg/L biotin [47,48]. DSM further expanded their patent portfolio on cell factories for biotin production in 2003 through the acquisition of Roche Vitamins & Fine Chemicals. Roche had previously reported the development of a *Bacillus subtilis* strain for biotin production. Based on rationally overexpressing the native *B. subtilis bio*-operon from a constitutive, strong promoter on a plasmid as well as process optimizations, they reported biotin titers of 16 mg/L and biotin vitamers titers (KAPA, DAPA, DTB and biotin together) of 200 mg/L in a 34-hour fermentation process fed with pimelic acid and lysine. They further improve the production strain through additional genome engineering increasing BioA expression as well as feeding the identified BioA amino donor, lysine. By doing so, they obtain a strain producing 20 mg/L biotin and >600 mg/L biotin vitamers from a 30-hour fermentation process supplemented with 1g/L pimelic acid and 7.5 g/L lysine [49,50].

Additional Efforts on Bio-Based Production of Biotin

Alongside the industrial interest, shaping the patent landscape for biotin, academic efforts have been published to both elucidate the biotin biosynthesis pathway in various organism and construct biotin overproducing strains.

A recombinant *E. coli* strain overexpressing the *Bacillus sphaericus* biotin operon was able to produce 45 mg/L biotin and 350 mg/L biotin vitamers in a 20 L fermentation process with pimelic acid feeding, after significant process optimization [51]. Another recombinant *E. coli* strain overexpressing the native *E. coli bioA-bioBFCD* using artificial promoters and mutation in the genomic encoded *birA*, inactivating the biotin repressor, enabled production of 190 mg/L/OD biotin and 50 mg/L/OD DTB [52]. In contrast to most other studies which report accumulation of

DTB and a bottleneck involving BioB, Lévy-Schil *et al* discuss the potential of a limiting step upstream of *bioF* [52].

A modified version of the *E. coli bio*-operon has also been used to establish an *Agrobacterium/ Rhizobium* HK4*m* based biotin producing cell factory. The divergently transcribed *bioA-bioBFCD* were reorganized into a single transcriptional unit under the expression of a strong, artificial promoter and RBS engineering was carried out to increase BioB expression further. The resulting strain produced 50 mg/L biotin in a 7-day 2L fed batch fermentation, and 110 mg/L when extending the fermentation time to 20 days [53]. The importance of fermentation conditions is further emphasized in a study using a recombinant *Sphingomonas sp.* as cell factory for biotin production. The biotin titers of *Sphingomonas sp.* overexpressing *bioACDF* and a mutated BioB from *Sphingomonas paucimobilis* were shown to range between 4 mg/L and 66 mg/L biotin in a 5-day fermentation process, for final OD660 ≥15, with only the motor agitation speed changing. They found that an optimum aeration speed was in the middle of their tested interval (400 rpm), for biotin fermentations of *Sphingomonas sp.* [54].

One of the most recently published efforts on biotin producing cell factories, came in 2019, after a long period with limited (published) work in the field. Based on rational metabolic engineering of *Putida mutabilis,* Xiao *et al* report biotin titers of 272 mg/L in an almost 7-day fed-batch fermentation with 1 g/L pimelic acid fed. They work through multiple rounds of the metabolic engineering cycle, overexpressing the *bio*-genes through RBS and promoter engineering, removing feedback inhibition using a BirA homolog from *E. coli*, and improving precursor and cofactor supply through engineering of the *isc*-operon. To further increase cofactor and precursor availability, they introduce a mutation in *metK*, targeting SAM metabolism, and express *bioWI* from *B. subtilis* to enable pimelic acid feeding [55]. Consistently with the accumulation of DTB in several of

the above mentioned examples [47,49,51], Xiao *et al* also report BioB as a difficult bottleneck to solve.

Using heterologous pathways [51], synthetic promoters [52,53], mutations in regulation site [44], and altered regulatory proteins [52,55] are all different approaches of designing a metabolic engineering strategy to overcome the regulatory feedback of the host organism. Despite the impressive development of tools enabling such rational engineering strategies, some degree of random mutagenesis is part of the majority of metabolic engineering efforts for biotin cell factories. The continued use of (semi) random approaches, reflects the complexity involved in the biotin biosynthesis pathway.

While the different methodologies and findings from previous attempts to establish a cost-competitive biotin cell factory, is a valuable source for inspiration, direct comparison of titers to evaluate the performance of the strains needs to be done with care. As fermentation settings including fermentation length, media compositions and supplement additions has a considerable influence on cell factory performance, it is very difficult to compare the different studies directly. Calculating the yield of each process would provide a measurement normalized to the amount of carbon added to the system. However, the use of complex media and lack of transparency of feed-setups makes such calculations extremely challenging based on the published materials. In the end it remains clear that significant work is still needed to turn biotin production from a proof-of-principle cell factory into a commercially viable process.

Iron-Sulfur Cluster Proteins

From the metabolic engineering efforts on biotin cell factories described above, it is evident that BioB is a main bottleneck difficult to overcome [35,39,57,58]. Similar to other enzymes in vitamin biosynthesis pathways such as LipA in lipoic acid biosynthesis [36], ThiC and ThiH in thiamine biosynthesis [59], and NadA in the nicotinamide adenine dinucleotide biosynthesis [60], BioB requires FeS-clusters to function [35]. The most common types of FeS-clusters are [2Fe2S], [3Fe4S], and [4Fe4S] in which iron ions are coordinated by cysteines and linked through sulfide bridges [61]. In radical S-adenosylmethionine (SAM) enzymes, as LipA, ThiC, ThiH and BioB mentioned above, the catalytic cluster for radical chemistry, facilitating cleavage of SAM, is a [4Fe4S] cluster. LipA and BioB accommodate an additional [4Fe4S] cluster and [2Fe2S] cluster respectively, essential for their individual catalysis [62] – the source of the sulfur atoms in lipoic acid and biotin. A brief introduction of FeS-cluster formation and homeostasis will be given below, with focus on the mechanisms in *E. coli*.

FeS-Cluster Biogenesis

FeS-clusters are built up on protein scaffolds by multi-component protein assembly systems. Several different assembly pathways exist, with partially redundant functions and minor specializations. These are widely spread and also highly conserved among bacteria and eukaryotes. Nitrogen fixation (*nif*) operon, FeS-cluster (*isc*) operon and sulfur mobilization (*suf*) operon constitute the three main pathways for FeS-cluster biogenesis, but the number and type vary from one organism to another. In *E. coli* only the *isc*- and *suf*-pathway are present [63]. Under non-stressed conditions these can compensate partly for each other's activity. Neither is essential, but the two form a synthetic lethal pair [64].

Sulfur donation to the FeS-cluster comes from L-cysteine catalyzed mainly by cysteine desulfurases [61]. In *E. coli* the desulfurases includes

IscS, SufS and CsdA, of which both IscS and CsdA are thought to also provide sulfur for pathways unrelated to FeS-cluster biogenesis [65,66]. While IscS transfers the sulfur directly to the scaffold component, SufS and CsdA form complexes with their respective sulfur acceptors, SufE and CsdE, to deliver the sulfur atom to the scaffold protein [67,68] (**Fig. 3**). The identification of the iron source used to build FeS-clusters has lagged behind that of the sulfur source. It is still not completely evident which proteins are responsible, but CyaY, IscA and IscX have been proposed to be involved, likely in combination with iron storage proteins such as ferritins (FtnA, FtnB), bacterioferritin (Bfr) and stress-response DNA binding protein (Dps). However, it is a clear constraint that iron ions must be provided to the assembly system through a shielded pathway preventing leakage of harmful free iron[63,69], which will be discussed later.

Scaffold proteins provide an intermediate assembly site for FeS-clusters or their precursors. They interact closely with a cysteine desulfurase, for sulfur donation, and an iron source, providing a chemical and structural environment that facilitates the formation of FeS-cluster and the transfer to target apo-proteins. The main scaffold proteins encompasses IscU and SufB in *E. coli* [63]. These proteins work together with ATP-hydrolysing components, namely the chaperones HscA and HscB for IscU and SufC for SufB[69,70]. In the *suf*-pathway, SufBCD forms a complex and works as part of a pseudo-ABC-transporter[63,70] (**Fig. 3**). Even though the ATP-hydrolysing components are essential for FeS-cluster biogenesis, the *in vivo* function remains unclear [64,71–73].

Once the FeS-cluster assembly is accomplished, the clusters need to be delivered to apo-proteins, which is referred to as the trafficking step. Currently the most widely accepted view, is that A-type carrier (ATC) proteins are main responsible proteins for carrying the final assembled FeS-clusters from the scaffold to the apo-protein targets. *E. coli* encodes three ATC proteins (IscA, SufA and ErpA), which are interchangeable biochemically speaking, but not fully redundant *in vivo* [74,75]. Additional

proteins are needed for adapting the transfer of ready-made FeS-clusters to the apo-proteins under various conditions and for specific target proteins. These proteins include NfuA, GrxD and Mrp, but the exact mechanisms are yet to be elucidated [63,75].

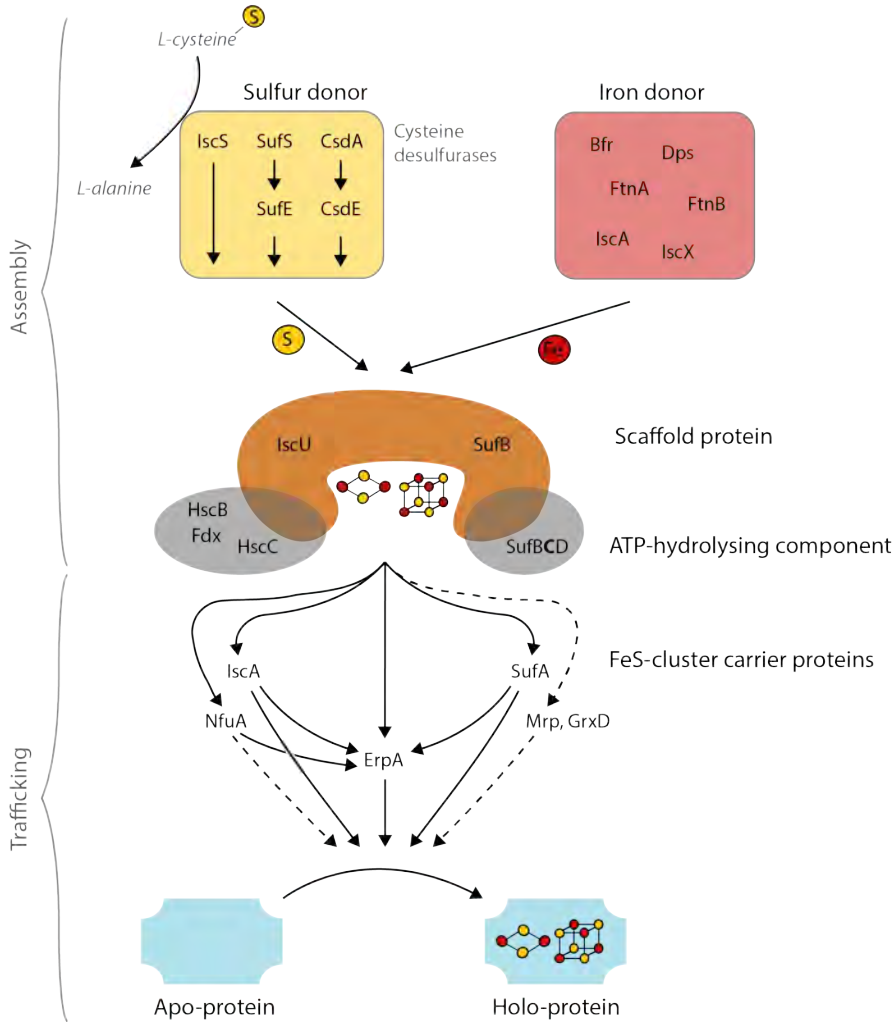


Figure 3: FeS-Cluster Biogenesis in *E. coli*. Two parts constitute FeS-cluster assembly; assembly and trafficking. Cysteine desulfurases transfer sulfur from L-cysteine to scaffold proteins while iron comes from various sources with less described mechanisms. The scaffold proteins finish the assembly step, assisted by ATP-hydrolysing components. FeS-cluster carrier proteins are responsible for the trafficking of the assembled cluster to apo-proteins. Several pathways exist in *E. coli*, unclear mechanisms are represented with dashed lines.

Regulation of FeS-Cluster Formation

It is widely acknowledged that the *isc*-encoded enzymes function as a housekeeping system for FeS-cluster biogenesis, with a broad specificity of target apo-proteins, whereas the *suf*-operon is related to stressed conditions. While being more efficient than the *isc*-system, especially during stress, the *suf*-system may merely meet the minimal requirements of FeS-biogenesis, providing clusters to essential enzymes only. Coordination of the *isc*- and *suf*-pathways to maintain FeS-cluster homeostasis in response to stress, precursor availability and FeS-cluster demand requires regulation on multiple levels [76]. The master regulator of FeS-cluster homeostasis in *E. coli*, IscR, is well characterized. It controls the transcription of both *isc*- and *suf*-operon and is the product of the first gene in the *isc*-operon. IscR is itself an FeS-cluster protein and regulate expression of FeS-cluster biogenesis pathways in response to the FeS-cluster level of the cell through binding of a [2Fe2S] cluster by IscR protein. When cluster synthesis is abundant and IscR is in its holo-form with FeS-cluster bound, it feedback inhibits the transcription of the *isc*-operon[77]. Additionally, IscR can bind and activate the transcription of the *suf*-operon both with and without an FeS-cluster bound (**Fig. 4**). However, under iron replete conditions, iron binds to Fur as a cofactor promoting DNA binding and transcriptional inhibition of the *suf*-operon, preventing binding and subsequent activation by IscR. Transcriptional activation of the *suf*-operon is also mediated by the broad regulator OxyR in combination with the Integration Host Factor (IHF). Full activation of the *suf*-operon requires DNA binding of both IscR and OxyR/IHF, which happens under stressed conditions [78]. Recently CsrA has also been shown to repress *suf*-expression, and potentially affect other areas of FeS-cluster homeostasis, but the mechanism remains to be elucidated [79].

Regulation of the *isc* levels are also controlled on the posttranscriptional level by the small RNA RyhB, which destabilizes the *isc* transcript when iron is limiting. RyhB binds to the *iscS* part of the

transcript and promote cleavage of the downstream RNA by Hfq and the RNA degradosome, leaving only *iscR* RNA intact for translation. IscR can then work as an activator for e.g. *suf*-pathway, while the *isc*-pathway is not expressed[80]. Like the *suf*-operon, *ryhB* is part of the Fur regulon, and the expression is repressed when iron is not limiting[80], preventing degradation of the *isc* transcript (**Fig. 4**). Proteolysis of IscS and IscU by the proteases FtsH and ClpXP [81,82] adds an additional layer of regulation of FeS-cluster biogenesis. The posttranslational control has been proposed to maintain protein quality control as oxidation of FeS-cluster by oxygen or reactive oxygen species (ROS) results in oxidative damage to proteins and inactivates FeS-clusters[76] (discussed later).

CyaY and IscX has been proposed to allosterically regulate FeS-cluster assembly through interactions with IscS. The allosteric binding of CyaY and IscX to IscS is likely to be mutual exclusive, as studies have shown that the binding sites are overlapping [83]. While CyaY binding to IscS has been shown to inhibit *isc*-mediated FeS-cluster assembly *in vitro*, *in vivo* studies point towards CyaY promoting FeS-cluster formation, and so the actual mechanism remains unclear (discussed by Mettert *et al* [76]).

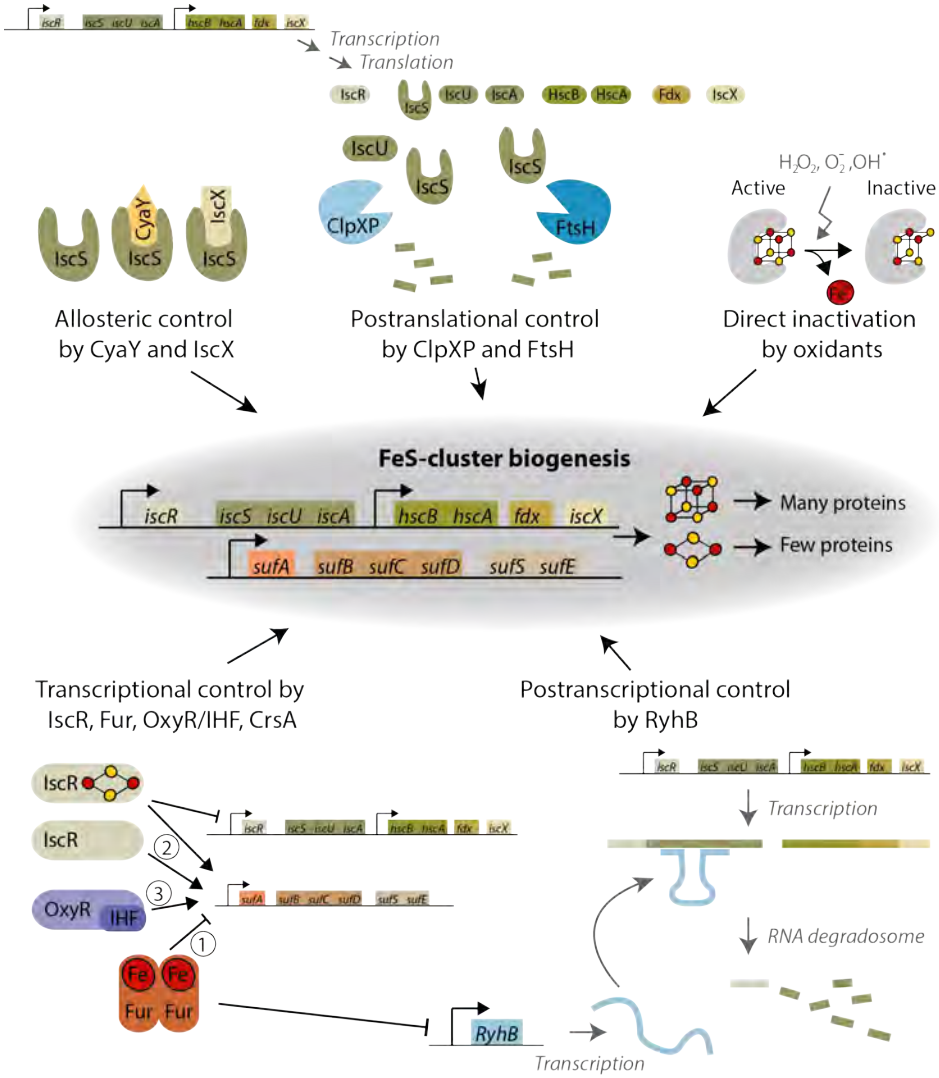


Figure 4: Regulation of FeS-Cluster Biogenesis in *E. coli* Occurs at Multiple Levels. Transcriptional control of FeS-cluster biogenesis encompasses repression of the *isc*-operon by holo-IscR as well as regulation of the *suf*-operon in a hierarchical manner where 1) Fur with iron bound under iron replete conditions has the highest binding affinity, followed by 2) activation by apo- and holo-IscR, which has a binding site overlapping with and thus blocked by Fur. Lastly 3) OxyR in combination with IHF can bind and further activate expression of the *suf*-operon under stressed conditions. Posttranscriptional control is mediated by the mRNA RyhB, promoting degradation of the *iscSUA* transcript. Allosteric binding of CyaY and IscX to IscS prevents IscS action as cysteine desulfurase, while ClpXP and FtsH can promote degradation of IscU and IscS proteins. Finally, FeS-clusters can be directly inactivated by oxidation.

Challenges with FeS-Clusters

FeS-clusters are abundant cofactors, fundamental to proteins involved in numerous biological processes such as respiration, central metabolism, gene regulation, RNA modification and DNA repair and replication. However, they developed in a world without oxygen, which is a destructive threat to FeS-cluster proteins and thus to the organisms relying on them [84]. The biochemical application of FeS-clusters depends on their ability to accept and donate electrons as well as their tendency to bind to acceptor proteins, which often require solvent exposure of the cluster. Reactive oxygen species (ROS) can form intracellularly when oxygen steals electrons from reduced flavins of redox enzymes. ROS rapidly inactivates the exposed [4Fe4S] cluster of dehydratases and potentially other enzymes, including essential protein members of the TCA pathway. The FeS-cluster is converted to an unstable state that releases iron, resulting in an inactive but repairable [3Fe4S] cluster. The damaged [3Fe4S] cluster lacks the catalytic iron atom, leading to inactivation of the enzyme to which it was a cofactor [84,85]. In addition to inactivating the enzyme, inactivation of FeS-clusters by hydrogen peroxide leads to formation of the extremely powerful oxidant hydroxyl radical through the Fenton reaction, which along with other ROS is a significant cause of DNA damage[85,86].

The majority of FeS-cluster proteins, protect their clusters from oxidants by coordinating polypeptides, minimizing the risk of direct contact and oxidation. This is why mostly dehydratases are prone to FeS-cluster damage by oxygen derived species [86]. However, it has been shown that even the cluster of non-dehydratase enzymes are at risk for oxidation, when exposed to hydrogen peroxide for several generations, but the specific mechanisms behind are unclear [72]. In order to overcome the problems with decomposing FeS-clusters upon exposure to oxygen and ROS, some microbes respond by retreating to anaerobic environments. Others have developed sophisticated defense systems and ways to repair

the damaged clusters. In *E. coli* these defense systems include superoxide dismutases (SodA, SodB and SodC), catalases (KatE and KatG) and peroxidases (AhpC and AhpF) [85,86]. Also scavenging proteins such as Dps and ferritins like FtnA take part of the defense systems against oxygen and oxidative stress, by sequestering iron that spills from damaged clusters and the damaged clusters themselves, thus minimizing the formation of hydroxyl radicals [87,88].

Repair of FeS-clusters

Even in wildtype *E. coli* cells under standard aerobic growth conditions, where cellular defense mechanisms work properly, the FeS-cluster damage by ROS is so high, that the half-time for enzyme damage is around 20 minutes [86]. Therefore the cells needs to continuously repair the damaged clusters by reduction and demetallation in order to survive. Contrary to *de novo* assembly of FeS-clusters, which require the entire *isc*-operon, Djaman *et al* suggest that mainly the cysteine desulfurase, IscS, is important for FeS-cluster repair in *E. coli*[71]. Djaman and coworkers show that an *E. coli iscS* mutant grow significantly worse than other single mutants of the *isc*-operon, reflecting the involvement of IscS in processes related to FeS-cluster availability other than cluster assembly. However, the precise mechanism in repair remains unclear[71]. Also the protein YtfE, has been suggested as a likely candidate for repair of iron centers in *E. coli.* The expression of YtfE is induced by treatment with nitric oxide, an enzymatic oxidation product[89], while *ytfE* deletion mutants, has been shown to be hypersensitive to hydrogen peroxide stress[90,91]. YtfE has been proposed to recruit and integrate the iron ion needed for the repair of [3Fe4S] clusters found in proteins damaged by exposure to either hydrogen peroxide or nitric oxide [63]. Other proteins suggested to be involved in FeS-cluster repair include Bfr, FtnA and FtnB [63,87].

With the complexity of FeS-cluster biogenesis and the detrimental effect even partial destruction can have on the cell as well as the specific enzymes, it is not surprising that BioB and other FeS-cluster proteins are causing troubles for cell factory engineering [35,57,58,92]. Increased understanding and focus on cell factories with such enzymes are likely to positively affect further development of commercial feasible and sustainable processes.

References

1. Keasling, J. D. Manufacturing Molecules Through Metabolic Engineering. *Science (80-.).* **330**, 1355–1358 (2010).
2. Godfray, H. C. J. *et al.* Food Security: The Challenge of Feeding 9 Billion People. *Science (80-.).* **327**, 812–819 (2010).
3. Du, J., Shao, Z. & Zhao, H. Engineering microbial factories for synthesis of value-added products. *J. Ind. Microbiol. Biotechnol.* **38**, 873–890 (2011).
4. Gavrilescu, M. & Chisti, Y. Biotechnology - A sustainable alternative for chemical industry. *Biotechnol. Adv.* **23**, 471–499 (2005).
5. Bentley, R. & Bennett, J. W. A Ferment of Fermentations: Reflections on the Production of Commodity Chemicals Using Microorganisms. *Adv. Appl. Microbiol.* **63**, 1–32 (2008).
6. Nielsen, J. & Keasling, J. D. Engineering Cellular Metabolism. *Cell* **164**, 1185–1197 (2016).
7. Hong, K. K. & Nielsen, J. Metabolic engineering of Saccharomyces cerevisiae: A key cell factory platform for future biorefineries. *Cell. Mol. Life Sci.* **69**, 2671–2690 (2012).
8. Himanshu, D. & Kristala, P. Microbial chemical factories : recent advances in pathway engineering for synthesis of value added chemicals. *Curr. Opin. Struct. Biol.* **21**, 488–494 (2011).
9. Thykaer, J. & Nielsen, J. Metabolic engineering of β-lactam production. *Metab. Eng.* **5**, 56–69 (2003).
10. Nielsen, J. Cell factory engineering for improved production of natural products. *Nat. Prod. Rep.* (2019).
11. Burk, M. J. & Van Dien, S. Biotechnology for Chemical Production: Challenges and Opportunities. *Trends Biotechnol.* **34**, 187–190 (2016).
12. Andrianantoandro, E., Basu, S., Karig, D. K. & Weiss, R. Synthetic biology: New engineering rules for an emerging discipline. *Mol. Syst. Biol.* **2**, 1–14 (2006).
13. Saiki, R. K. *et al.* Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science (80-.).* **239**, 487–491 (1988).
14. Metzker, M. L. Sequencing technologies the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
15. Nour-Eldin, H. H., Hansen, B. G., Nørholm, M. H. H., Jensen, J. K. & Halkier, B. A. Advancing uracil-excision based cloning towards an ideal technique for cloning

PCR fragments. *Nucleic Acids Res.* **34**, (2006).
16. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–5 (2009).
17. Pennisi, E. The CRISPR craze. *Science (80-.).* **94**, 38–38 (2013).
18. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–8 (2009).
19. Decoene, T. *et al.* Standardization in synthetic biology: an engineering discipline coming of age. *Crit. Rev. Biotechnol.* **38**, 647–656 (2018).
20. Raman, S., Rogers, J. K., Taylor, N. D. & Church, G. M. Evolution-guided optimization of biosynthetic pathways. *Proc. Natl. Acad. Sci.* **111**, 201409523 (2014).
21. Jinek, M. *et al.* A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* **337**, 816–822 (2012).
22. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: Technologies and applications. *Nat. Methods* **11**, 499–507 (2014).
23. Dietrich, J. a, McKee, A. E. & Keasling, J. D. High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu. Rev. Biochem.* **79**, 563–590 (2010).
24. Buerger, J., Gronenberg, L. S., Genee, H. J. & Sommer, M. O. A. Wiring cell growth to product formation. *Curr. Opin. Biotechnol.* **59**, 85–92 (2019).
25. Dietrich, J. A., Shis, D. L., Alikhani, A. & Keasling, J. D. Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS Synth. Biol.* **2**, 47–58 (2013).
26. Genee, H. J. *et al.* Functional mining of transporters using synthetic selections. *Nat. Chem. Biol.* **12**, 1015–1022 (2016).
27. Schneider, C. & Suess, B. Identification of RNA aptamers with riboswitching properties. *Methods* 1–7 (2015).
28. Fiedurek, J., Trytek, M. & Szczodrak, J. Strain improvement of industrially important microorganisms based on resistance to toxic metabolites and abiotic stress. *J. Basic Microbiol.* **57**, 445–459 (2017).
29. Rogers, J. K. & Church, G. M. Genetically encoded sensors enable real-time observation of metabolite production. *Proc. Natl. Acad. Sci.* **113**, 2388–2393 (2016).
30. Siedler, S. *et al.* Development of a Bacterial Biosensor for Rapid Screening of Yeast p-Coumaric Acid Production. *ACS Synth. Biol.* **6**, 1860–1869 (2017).
31. Goldberg, M. W. & Sternbach, L. H. Debenzylation of benzylated imidazolido-thiophane compounds. (1949). US 2,489,238
32. Eggersdorfer, M. *et al.* One hundred years of vitamins-a success story of the natural sciences. *Angew. Chem. Int. Ed. Engl.* **51**, 12960–90 (2012).
33. Perham, R. N. Swinging arms and swinging domains in multifunctional enzymes : Catalytic Machines for Multistep Reactions. *Annu. Biochem.* 961–1004 (2000).
34. Lin, S., Hanson, R. E. & Cronan, J. E. Biotin synthesis begins by hijacking the fatty acid synthetic pathway. *Nat. Chem. Biol.* **6**, 682–688 (2010).
35. Cramer, J. D. & Jarrett, J. T. *Purification, Characterization, and Biochemical Assays of Biotin Synthase From Escherichia coli*. *Methods in Enzymology* **606**, (Elsevier Inc., 2018).

36. Cronan, J. E. Biotin and Lipoic Acid: Synthesis, Attachment, and Regulation. *EcoSal Plus* **1**, 217–244 (2013).
37. Wang, J. & Beckett, D. A conserved regulatory mechanism in bifunctional biotin protein ligases. *Protein Sci.* **26**, 1564–1573 (2017).
38. Cronan, J. E., Zhao, X. & Jiang, Y. *Function, attachment and synthesis of lipoic acid in Escherichia coli*. *Advances in Microbial Physiology* **50**, (Elsevier Masson SAS, 2005).
39. Cronan, J. E. Biotin and Lipoic Acid: Synthesis, Attachment and Regulation. *Ecosal Plus* **18**, 1199–1216 (2014).
40. Sakurai, N., Imai, Y., Masuda, M., Komatsubara, S. & Tosa, T. Construction of a biotin-overproducing strain of Serratia marcescens. *Appl. Environ. Microbiol.* **59**, 2857–2863 (1993).
41. Komatsubara, S., Imai, Y., Masuda, M. & Sakurai, N. Microorganisms and process for preparing D-Biotin using the same. (1994). US5374554
42. Sakurai, N., Imai, Y., Masuda, M., Komatsubara, S. & Tosa, T. Improvement of a d-biotin-hyperproducing recombinant strain of Serratia marcescens. *J. Biotechnol.* **36**, 63–73 (1994).
43. Kishimoto, J., Haze, S. & Ifuku, O. Process for producing biotin vitamers using novel microorganisms. (1991).
44. Kanzaki, N., Kimura, H., Matsui, J., Nakahama, K. & Ifuku, O. Microorganism resistant to threonine analogue and production of biotin. **1**, (1999). US 6,284,500 B1
45. Schröder, H. & Hauer, B. Method for producing biotin. (1998). doi:10.1016/j.(73) US6436681 B1
46. Schröder, H. Method for producing biotin. (1999). WO1999042591A1
47. Furuichi, Y., Hoshino, T., Kimura, H., Kiyasu, T. & Nagahashi, Y. Biotin biosynthetic genes having biotin synthase activity. **2**, (2005). US 6,955,906B2
48. Hoshino, K. T., Noro, F. A. & Tazoe, Y. M. Process for the production of D-biotin. **2**, 439–443 (1997). US 5,922,581 A
49. Bower, S. G., Perkins, J. B., Youm, R. R. & Pero, J. G. Biotin Biosynthesis in Bacillus Subtilis. (1999). US 6,841,366 B1
50. Van Arsdell, S. W. *et al.* Removing a bottleneck in the Bacillus subtilis biotin pathway: BioA utilizes lysine rather than S-adenosylmethionine as the amino donor in the KAPA-to-DAPA reaction. *Biotechnol. Bioeng.* **91**, 75–83 (2005).
51. Brown, S. W. *et al.* The production of biotin by recombinant strains of Escherichia coli. *J. Chem. Technol. Biotechnol.* **50**, 115–121 (1991).
52. Lévy-Schil, S. *et al.* Biotin biosynthetic pathway in recombinant strains of Escherichia coli overexpressing bio genes: evidence for a limiting step upstream from KAPA. *Appl. Microbiol. Biotechnol.* **38**, 755–762 (1993).
53. Shaw, N. *et al.* Biotin production under limiting growth conditions by Agrobacterium/Rhizobium HK4 transformed with a modified Escherichia coli bio operon. *J. Ind. Microbiol. Biotechnol.* **22**, 590–599 (1999).
54. Saito, I. *et al.* Comparison of biotin production by recombinant Sphingomonas sp. under various agitation conditions. *Biochem. Eng. J.* **5**, 129–136 (2000).
55. Xiao, F. *et al.* Multi-level metabolic engineering of Pseudomonas mutabilis ATCC31014 for efficient production of biotin. *Metab. Eng.* (2019).

56. Sakurai, N., Imai, Y., Masuda, M., Komatsubara, S. & Tosa, T. Construction of a Biotin-Overproducing Strain of Serratia marcescens. *Appl. Environ. Microbiol.* **59**, 2857–63 (1993).
57. Acevedo-Rocha, C. G., Gronenberg, L. S., Mack, M., Commichau, F. M. & Genee, H. J. Microbial cell factories for the sustainable manufacturing of B vitamins. *Curr. Opin. Biotechnol.* **56**, 18–29 (2019).
58. Palmer, L. D. & Downs, D. M. The thiamine biosynthetic enzyme ThiC catalyzes multiple turnovers and is inhibited by S-adenosylmethionine (AdoMet) metabolites. *J. Biol. Chem.* **288**, 30693–9 (2013).
59. Jurgenson, C. T., Begley, T. P. & Ealick, S. E. The structural and biochemical foundations of thiamin biosynthesis. **78**, 569–603 (2009).
60. Ollagnier-de Choudens, S., Loiseau, L., Sanakis, Y., Barras, F. & Fontecave, M. Quinolinate synthetase, an iron-sulfur enzyme in NAD biosynthesis. *FEBS Lett.* (2005).
61. Fontecave, M. & Ollagnier-de-Choudens, S. Iron-sulfur cluster biosynthesis in bacteria: Mechanisms of cluster assembly and transfer. *Arch. Biochem. Biophys.* **474**, 226–237 (2008).
62. Broderick, J. B., Duffus, B. R., Duschene, K. S. & Shepard, E. M. Radical S-adenosylmethionine enzymes. *Chem. Rev.* **114**, 4229–4317 (2014).
63. Py, B. & Barras, F. Building Fe–S proteins: bacterial strategies. *Nat. Rev. Microbiol.* **8**, 436–446 (2010).
64. Outten, F. W., Djaman, O. & Storz, G. A suf operon requirement for Fe-S cluster assembly during iron starvation in Escherichia coli. *Mol. Microbiol.* **52**, 861–872 (2004).
65. Mueller, E. G. Trafficking in persulfides: delivering sulfur in biosynthetic pathways. *Nat. Chem. Biol.* **2**, 185–194 (2006).
66. Mihara, H. & Esaki, N. Bacterial cysteine desulfurases: Their function and mechanisms. *Appl. Microbiol. Biotechnol.* **60**, 12–23 (2003).
67. Urbina, H. D., Silberg, J. J., Hoff, K. G. & Vickery, L. E. Transfer of Sulfur from IscS to IscU during Fe/S Cluster Assembly. *J. Biol. Chem.* **276**, 44521–44526 (2001).
68. Loiseau, L., Ollagnier-de-Choudens, S., Nachin, L., Fontecave, M. & Barras, F. Biogenesis of Fe-S Cluster by the Bacterial Suf System. *J. Biol. Chem.* **278**, 38352–38359 (2003).
69. Mettert, E. L. & Kiley, P. J. How Is Fe-S Cluster Formation Regulated? *Annu. Rev. Microbiol.* (2015).
70. Garcia, P. S., Gribaldo, S., Py, B. & Barras, F. The SUF system: An ABC ATPase-dependent protein complex with a role in Fe-S cluster biogenesis. *Res. Microbiol.* (2019). doi:10.1016/j.resmic.2019.08.001
71. Djaman, O., Outten, F. W. & Imlay, J. A. Repair of oxidized iron-sulfur clusters in Escherichia coli. *J. Biol. Chem.* **279**, 44590–44599 (2004).
72. Nachin, L., Loiseau, L., Expert, D. & Barras, F. SufC: An unorthodox cytoplasmic ABC/ATPase required for [Fe-S] biogenesis under oxidative stress. *EMBO J.* **22**, 427–437 (2003).
73. Tokumoto, U. & Takahashi, Y. Genetic analysis of the isc operon in Escherichia coli involved in the biogenesis of cellular iron-sulfur proteins. *J. Biochem.* **130**, 63–

71 (2001).
74. Vinella, D., Brochier-Armanet, C., Loiseau, L., Talla, E. & Barras, F. Iron-sulfur (Fe/S) protein biogenesis: Phylogenomic and genetic studies of A-type carriers. *PLoS Genet.* **5**, e1000497 (2009).
75. Py, B. *et al.* The ErpA/NfuA complex builds an oxidation-resistant Fe-S cluster delivery pathway. *J. Biol. Chem.* **293**, 7689–7702 (2018).
76. Mettert, E. L. & Kiley, P. J. How Is Fe-S Cluster Formation Regulated? *Annu. Rev. Microbiol.* **69**, 505–526 (2015).
77. Schwartz, C. J. *et al.* IscR, an Fe-S cluster-containing transcription factor, represses expression of Escherichia coli genes encoding Fe-S cluster assembly proteins. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14895–14900 (2001).
78. Lee, K. C., Yeo, W. S. & Roe, J. H. Oxidant-responsive induction of the suf operon, encoding a Fe-S assembly system, through Fur and IscR in Escherichia coli. *J. Bacteriol.* **190**, 8244–8247 (2008).
79. Pourciau, C. *et al.* Regulation of Iron Storage by CsrA Supports Exponential Growth of *Escherichia coli*. *MBio* **10**, 1–18 (2019).
80. Desnoyers, G., Morissette, A., Prévost, K. & Massé, E. Small RNA-induced differential degradation of the polycistronic mRNA iscRSUA. *EMBO J.* **28**, 1551–61 (2009).
81. Westphal, K., Langklotz, S., Thomanek, N. & Narberhaus, F. A trapping approach reveals novel substrates and physiological functions of the essential protease FtsH in Escherichia coli. *J. Biol. Chem.* **287**, 42962–71 (2012).
82. Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T. & Baker, T. A. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol. Cell* **11**, 671–83 (2003).
83. Yan, R. *et al.* Ferredoxin competes with bacterial frataxin in binding to the desulfurase IscS. *J. Biol. Chem.* **288**, 24777–24787 (2013).
84. Imlay, J. A. Iron-sulphur clusters and the problem with oxygen. *Mol. Microbiol.* **59**, 1073–1082 (2006).
85. Imlay, J. A. Cellular Defenses against Superoxide and Hydrogen Peroxide. *Annu. Rev. Biochem.* **77**, 755–776 (2008).
86. Imlay, J. A. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nat. Rev. Microbiol.* **11**, 443–54 (2013).
87. Bitoun, J. P., Wu, G. & Ding, H. Escherichia coli FtnA acts as an iron buffer for re-assembly of iron-sulfur clusters in response to hydrogen peroxide stress. *Biometals* **21**, 693–703 (2008).
88. Ilari, A., Ceci, P., Ferrari, D., Rossi, G. L. & Chiancone, E. Iron incorporation into Escherichia coli Dps gives rise to a ferritin-like microcrystalline core. *J. Biol. Chem.* **277**, 37619–37623 (2002).
89. Justino, M. C., Vicente, J. B., Teixeira, M. & Saraiva, L. M. New genes implicated in the protection of anaerobically grown Escherichia coli against nitric oxide. *J. Biol. Chem.* **280**, 2636–2643 (2005).
90. Justino, M. C., Almeida, C. C., Teixeira, M. & Saraiva, L. M. Escherichia coli di-iron YtfE protein is necessary for the repair of stress-damaged iron-sulfur clusters. *J. Biol. Chem.* **282**, 10352–9 (2007).

91. Vine, C. E., Justino, M. C., Saraiva, L. M. & Cole, J. Detection by whole genome microarrays of a spontaneous 126-gene deletion during construction of a ytfE mutant: confirmation that a ytfE mutation results in loss of repair of iron-sulfur centres in proteins damaged by oxidative or nitrosative stress. *J. Microbiol. Methods* **81**, 77–79 (2010).
92. Kriek, M., Peters, L., Takahashi, Y. & Roach, P. L. Effect of iron-sulfur cluster assembly proteins on the expression of Escherichia coli lipoic acid synthase. *Protein Expr. Purif.* **28**, 241–245 (2003).
93. Ifuku, O. *et al.* Sequencing analysis of mutation points in the biotin operon of biotin-overproducing Escherichia coli mutants. *Biosci Biotechnol Biochem* **57**, 760–765 (1993).

CHAPTER II
IMPROVED BIOTIN, THIAMINE, AND LIPOIC ACID BIOSYNTHESIS BY ENGINEERING THE GLOBAL REGULATOR ISCR

Anne P. Bali[1,2], David Lennox-Hvenekilde[1,2], Nils Myling-Petersen[2], Josi Buerger[2], Bo Salomonsen[2], Luisa S. Gronenberg[2], Morten O. A. Sommer[1] & Hans J. Genee[2*]

Affiliations
[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark. [2]Biosyntia ApS, Fruebjergvej 3, DK-2100, Denmark

***Corresponding author:** hjg@biosyntia.com

Manuscript submitted to Nature Chemical Biology

Abstract

Biotin, thiamine, and lipoic acid are industrially important molecules naturally synthesized by microorganisms via biosynthetic pathways requiring iron-sulfur (FeS) clusters. Current production is exclusively by chemistry because pathway complexity hinders development of fermentation processes. For biotin, the main bottleneck is biotin synthase, BioB, a S-adenosyl methionine (SAM)-dependent radical enzyme that converts dethiobiotin (DTB) to biotin. BioB overexpression is toxic, though the mechanism remains unclear. We identified single mutations in the global regulator IscR that substantially improve cellular tolerance to BioB overexpression, increasing *Escherichia coli* DTB-to-biotin biocatalysis by more than 2.2-fold. Based on proteomics and targeted overexpression of FeS-cluster biosynthesis genes, FeS-cluster depletion is found to be the main reason for toxicity. We demonstrate that IscR mutations significantly affect cell viability and improve cell factories for *de novo* biosynthesis of thiamine by 1.3-fold and lipoic acid by 1.8-fold. We illuminate a novel engineering target for enhancing biosynthesis of complex FeS-cluster-dependent molecules, paving the way for industrial fermentation processes.

Introduction

Vitamins are organic compounds essential for both humans and animals. They are widely used in food, feed, cosmetics, and pharmaceuticals. Most vitamins are currently produced by chemical synthesis because of challenges in establishing economically competitive, bio-based production alternatives. For some vitamins, including thiamine (vitamin B1), lipoic acid, and biotin (vitamin B7), slow catalysis of key enzymes is a significant barrier [1–4]. The slow biosynthetic enzymes are phosphomethylpyrimidine synthase (ThiC) for the thiamine pathway, lysosomal acid lipase/cholesteryl ester hydrolase (LipA) for lipoic acid,

and biotin synthase (BioB) for biotin. All are S-adenosyl methionine (SAM) radical enzymes that use iron-sulfur (FeS)-clusters for catalysis. These FeS-clusters are oxygen sensitive and require anaerobic conditions for isolation and *in vitro* studies, making the enzymes challenging to investigate [5].

Production of some vitamins, such as riboflavin, has successfully transitioned from chemical synthesis to a biotechnology-based process, with others in development [1,6]. However, despite decades of strain engineering efforts by industry and academia [7–9], commercial bio-based production of biotin, thiamine and lipoic acid is too inefficient to replace the chemical synthesis processes [10].

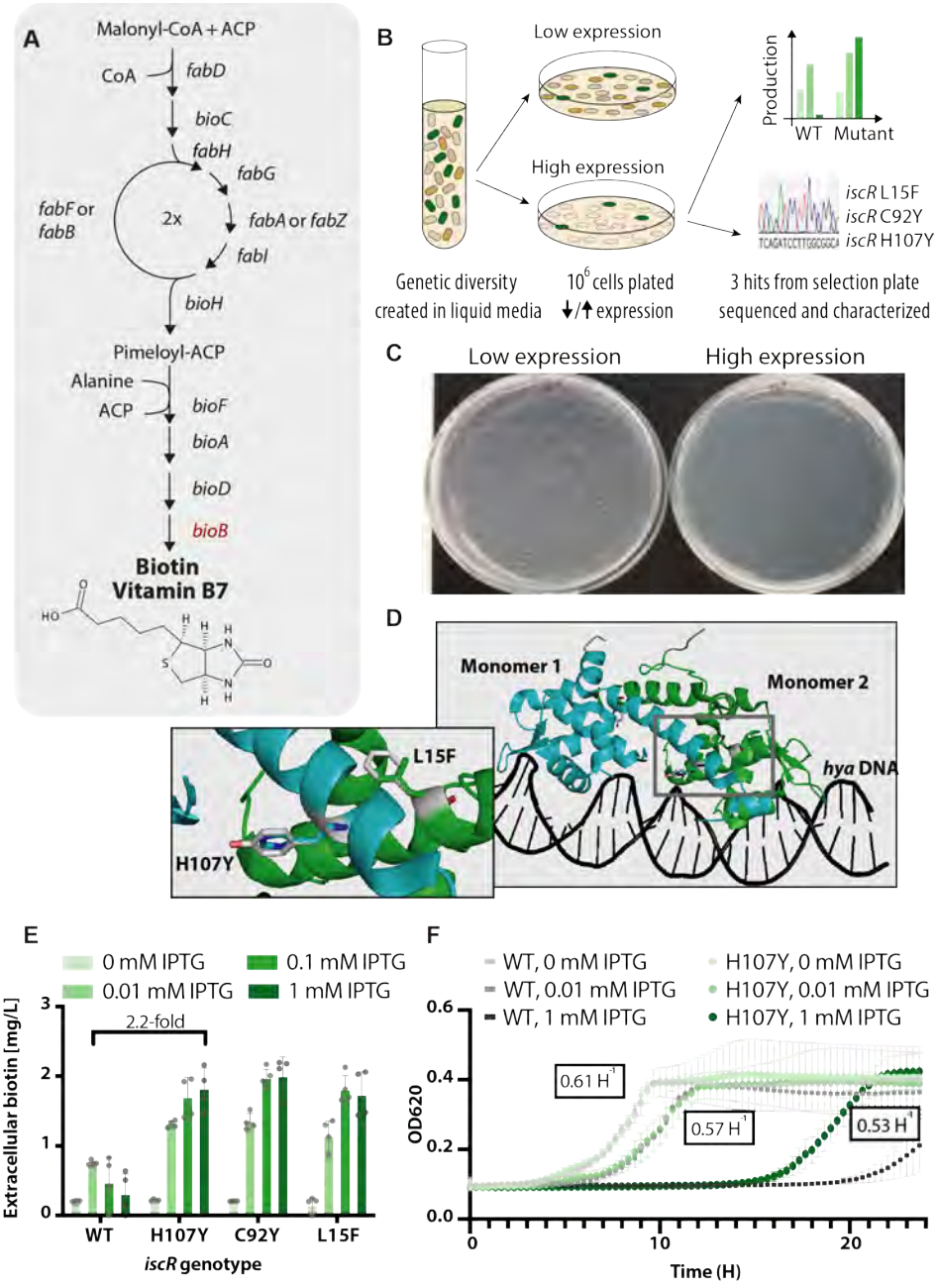
A major bottleneck in biotin production is BioB, which catalyzes the last step in the pathway, converting dethiobiotin (DTB) to biotin (**Fig. 1A**) [8,11]. Evidence for the bottleneck is that engineered biotin cell factories accumulate DTB. For instance, has an engineered *Bacillus subtilis* strain overexpressing native biotin genes been reported to produce 20 mg/L biotin and >600 mg/L DTB in a 30 H fermentation, supplemented with 1g/L pimelic acid and 7.5 g/L lysine [12]. Xiao *et al.* also reports BioB as a bottleneck in *Escherichia coli* cell factories producing 270 mg/L biotin in a fed-batch fermentation of 174 H supplemented with 1 g/L pimelic acid [8]. The highest reported fermentation titer of biotin is 600 mg/L, achieved by multiple rounds of random mutagenesis and antimetabolite selection combined with overexpression of a mutated biotin operon in *Serratia marcescens* [13]. However, also this strain accumulated DTB (200 mg/L) and fermentation time was an infeasible 11 days. Despite these reports of progress, an additional 10-20 fold process performance is still required for an industrially competitive fermentation process.

The complex catalysis of BioB requires two SAM equivalents and sulfur donation from the [2Fe2S] cluster, one of two BioB FeS-clusters. This mechanism presumably requires cluster regeneration after each turnover [3]. BioB has a low turnover number and a slow catalysis when

characterized *in vitro* [5,14]. In addition, when overexpressed *in vivo*, BioB inhibits growth [11], impeding development of biotin cell factories [15,16]. Ifuku *et al* further demonstrated that growth inhibition by BioB overexpression is independent of the biotin-forming activity [11].
Multiple hypotheses exist for the observed toxicity of BioB overexpression including generation of reactive oxygen species (ROS) [12,17], insufficient FeS-cluster supply or limitations in iron availability [18], depletion of SAM/sulfur availability, and product inhibition by 5'-deoxyadenosine [19–21]. The mechanisms of BioB overexpression toxicity are thus numerous and previous rational engineering strategies have had a low success rate.

Using growth selection, we identified specific point mutations that improve cellular tolerance to BioB overexpression and significantly increase biotin biosynthesis. All mutations were in a global regulator involved in FeS-cluster biogenesis. Using proteomics and various engineering strategies we elucidated main cellular effects caused by the mutations. We further show that the mutations improve lipoic acid and thiamine production titers. These results emphasize how synergistic effects might drive metabolic engineering efforts.

Figure 1: Mutant IscR Increases Tolerance to BioB Expression and Enables Higher Biotin Production Titers. Toxicity when overexpressing BioB was used as the foundation for assays to select for colonies with improved tolerance for BioB expression. **A)** Schematic presentation of the *Escherichia coli* biotin biosynthesis pathway utilizing the fatty acid cycle for two rounds (2x) to build pimeloyl-ACP. The *bioB* gene encoding the FeS-cluster enzyme BioB is shown in red. **B)** General workflow for selecting mutant strains with increased BioB tolerance. **C)** Selection behavior on mMOPS plates of strains under low and high BioB expression**:** 10^6 cells were plated and incubated at 37ºC for 24 hours. **D)** Structural representation of dimerized apo-IscR bound to *hya* DNA with identified mutations in gray in stick drawings (L15F and H107Y; C92Y was not part of the crystal structure). **E)** Biotin production titers of three different IscR mutants identified through selection for improved tolerance to BioB expression. N=4, error bars represent standard deviation. Individual replicates are in light gray circles. **F)** Growth rate data on IscR mutant strain H107Y at different induction levels, explicitly stated in black boxes. Darker color means increasing induction. N=4, error bars represent standard deviation.


A
Malonyl-CoA + ACP
CoA
fabD
bioC
fabH
fabG
fabF or fabB
2x
fabA or fabZ
fabI
bioH
Pimeloyl-ACP
Alanine
ACP
bioF
bioA
bioD
bioB
Biotin
Vitamin B7
B
Low expression
High expression
Production
WT Mutant
iscR L15F
iscR C92Y
iscR H107Y
Genetic diversity created in liquid media
10^6 cells plated ↓/↑ expression
3 hits from selection plate sequenced and characterized
C
Low expression
High expression
D
Monomer 1
Monomer 2
hya DNA
L15F
H107Y
E
0 mM IPTG
0.1 mM IPTG
0.01 mM IPTG
1 mM IPTG
2.2-fold
Extracellular biotin [mg/L]
WT
H107Y
C92Y
L15F
iscR genotype
F
WT, 0 mM IPTG
WT, 0.01 mM IPTG
WT, 1 mM IPTG
H107Y, 0 mM IPTG
H107Y, 0.01 mM IPTG
H107Y, 1 mM IPTG
0.61 H-1
0.57 H-1
0.53 H-1
OD620
Time (H)

Results

Mutant IscR Improves Biotin Cell Factories

We exploited the toxic phenotype to select for mutant *E. coli* with increased tolerance for BioB overexpression (**Fig. 1B**). We built an expression construct that allowed BioB expression to be induced with isopropyl-β-D-thiogalactoside (IPTG) to levels that inhibit wildtype growth. At high BioB induction we observed a survival rate of approximately 1 in 10^6 plated cells (**Fig. 1C**). We applied the selection on a production culture grown without induction of BioB expression, choosing 90 surviving colonies from a population of $5 \cdot 10^6$ cells for further investigation, including confirmation of the improved growth phenotype as well as their DTB to biotin conversion rate. We hypothesized that some of these colonies might have spontaneous mutations that enabled them to tolerate higher levels of BioB expression. Five colonies that had improved biotin production levels were whole genome sequenced and their assembled genomes were compared to the parent. The sequencing revealed three unique strains that each harbored a single, unique mutation in the same gene: *iscR*, which encodes a dual regulator of *E. coli* FeS-cluster assembly. The three identified point mutations encode amino acid changes in *E. coli* IscR of L15F, C92Y and H107Y (**Fig. 1D**).

Strains with a mutant IscR were compared to the parent strain (BS1011) and evaluated as potential cell factories based on DTB to biotin conversion and growth profiles at several levels of induced BioB expression. Biotin formation profiles were similar for the three mutants (**Fig. 1E**), reaching maximum titers of almost 2 mg/L in minimal media representing a 2.2-fold increase in production over the parent strain. Biotin production increased with IPTG induction of BioB expression from 0 to 0.1 mM IPTG, after which biotin titers plateaued at ~2 mg/L for the mutant strains. Production of the parent strain, however, started decreasing above 0.1 mM IPTG induction, corresponding with significantly reduced growth

(**Fig. 1E**). The three mutant strains had similar growth profiles (**Fig. 1E** has IscR H107Y only with data on other strains in Supplementary Fig. S3).

IscR is a global regulator in *E. coli*, affecting expression of more than 40 genes [22]. IscR regulates FeS-cluster assembly through transcriptional repression of the *isc*-operon and activation of the *suf*-operon, which are the two main FeS-cluster biosynthesis pathways in *E. coli*. IscR activity depends on intracellular FeS-cluster level, through IscR binding to a [2Fe2S] cluster (**Fig. 2A, B**). For example will IscR only repress transcription of the *isc*-operon when bound to a [2Fe2S]-cluster (holo-form), while activation of the *suf*-operon from IscR may happen both in holo- and apo-form (without a [2Fe2S]-cluster) [23]. Four amino acids coordinate FeS-cluster binding in IscR: C92, C98, C104, and H107 [24]. Two of the three mutated amino acids in the selected strains were part of this binding motif (C92Y and H107Y). The L15F mutation has not yet been described in the literature to our knowledge.

Proteomics Elucidate Cellular Effects of IscR Mutations

Having confirmed that the three strains with IscR mutations produced similar high titers of biotin, we elucidated the global protein changes caused by overexpressing BioB and mutating IscR. Nontargeted label-free proteomics quantification was carried out for mutant strain BS1353 (IscR H107Y) at two BioB induction levels (MutL, 0.025 mM IPTG induction, and MutH, 1 mM IPTG induction). Control strains were IscR wildtype (BS1011, WT) with the BioB expression plasmid and reference strain (BS1013, Ref) with native BioB expression, both at similar induction levels as the lowly induced BS1353.

Based on ANOVA analysis with Bonferroni correction for multiple testing, almost 200 proteins were differentially expressed in the mutant strain compared to a nonmutant IscR strain with wildtype BioB expression (Supplementary Table S4). Proteomics data showed that BioB protein

levels were low in the wildtype strain without the BioB expression plasmid (~0 % of total soluble protein, Ref). BioB levels were high in the IscR wildtype strain with BioB overexpression induced at 0.025 mM IPTG (~1% of total soluble protein, WT). At 0.025 mM IPTG, the strain carrying the IscR mutation (MutL) had BioB levels similar to the wildtype, while at high induction (1 mM IPTG, which is toxic to wildtype cells), BioB levels were more than 2-fold higher than wildtype (~2.5% of total soluble protein, MutH). We further measured biotin production by the same four cultures after feeding with DTB and found that biotin levels correlated linearly with the level of soluble BioB (**Fig. 2C**). Hence, the data showed that the IscR mutation allowed cells to have higher levels of BioB protein and that BioB enzyme activity per protein was constant across induction levels and strain backgrounds.

Next, we considered the FeS-cluster biosynthesis pathways: The protein levels of most proteins from the *isc*- and *suf*-operons increased in the IscR mutant strain compared to the corresponding wildtype strain with BioB overexpression at induction with 0.025 mM IPTG. This result was consistent with the regulatory function of IscR. In the apo-form, without a bound FeS cluster, IscR binds p_{sufA} to activate it and dissociates from p_{iscR} resulting in derepression (**Fig. 2A, B**) [24,25]. In the IscR mutant strain, levels of proteins from the *suf*-operon protein were even higher at high induction of BioB (1 mM IPTG) than at lower induction (0.025 mM IPTG) (**Fig. 2D**). The proteomics data suggested that toxicity observed in IscR wildtype cells overexpressing BioB could be due to FeS-cluster depletion.

To understand the global effects of mutating IscR, we used proteomics data to compare the relative protein levels of wildtype and mutant IscR strains with functional BioB expression plasmids: 15 proteins had significantly changed levels (**Fig. 2E**), of which only SufA, IscS and NfeR (YqjI) are members of the IscR regulon [22]. Comparing protein levels of strains at low induction (0.025 mM IPTG, WT and MutL), we observed that SufA, IscS, NfeR, Ndh, GdhA, MetB, PepB, and GsiA levels

were increased in the IscR mutant strain (MutL), while protein levels of FtnA, NadA, NlpA, BamE and ThiF were decreased. Similarly, we found that levels of SufA, Ndh, GdhA and MetB further increased and FtnA further decreased in IscR mutant strains with high BioB induction (MutH) compared to low BioB induction (MutL). Also, in IscR mutant strains, relative protein levels of YfcF, EntC, BamE and ThiF were higher with high BioB expression than low expression while NfeR and GsiA levels were lower.

We observed an effect on the level of proteins involved in oxidative stress, iron availability and electron transfer potential such as YfcF, EntC and Ndh [26–28]. Also, biosynthesis of methionine, a highly oxidation-sensitive amino acid [29], was affected by mutation in IscR through MetB and NlpA [30,31], which are involved in methionine biosynthesis, although the mechanism is not clear. These results illustrate that the effects of mutating IscR were global and likely extended beyond FeS-cluster biogenesis.

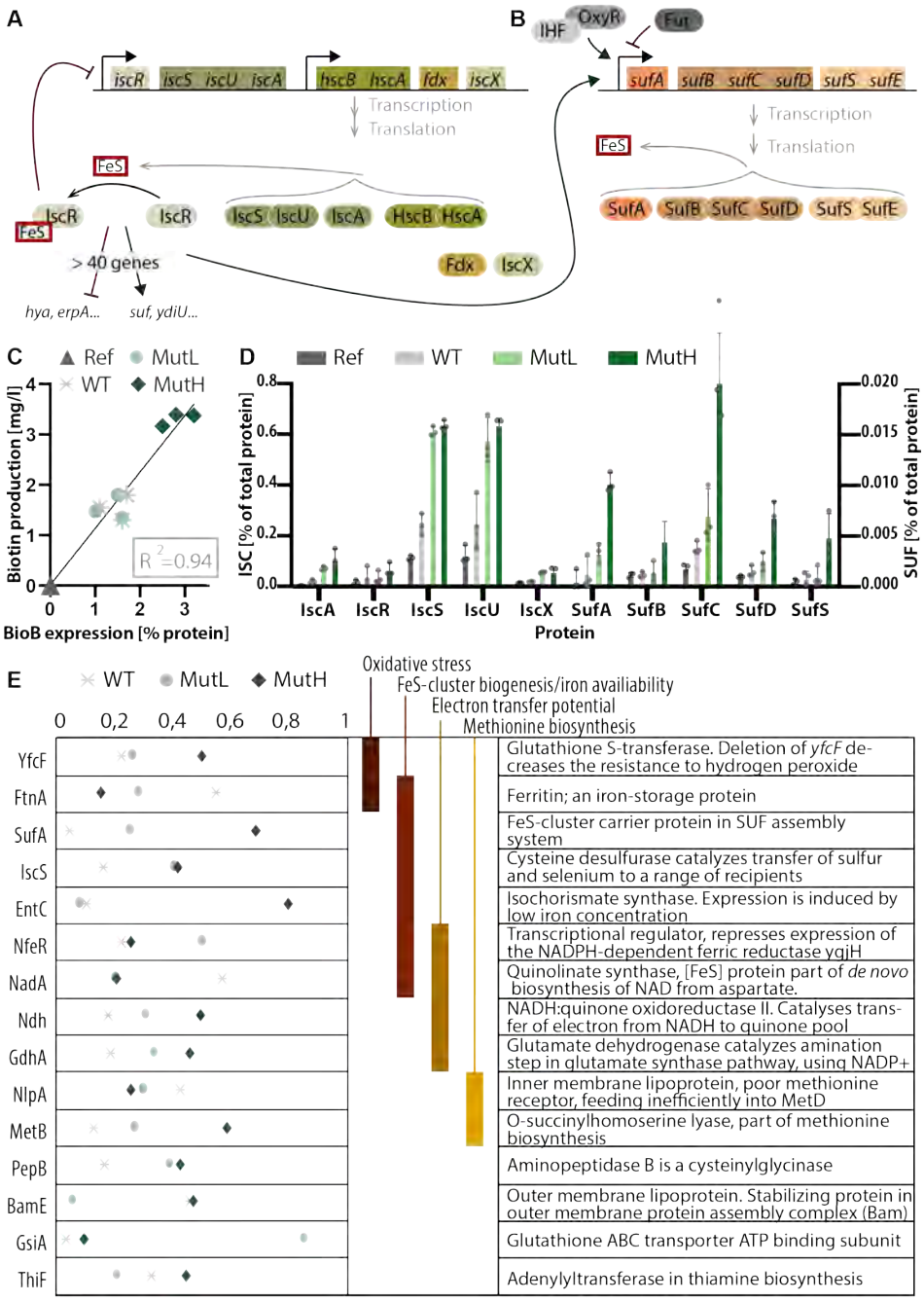


Figure 2: Elucidation of the Effects of Identified IscR Mutations using Proteomics. *Isc*- and *suf*-operons constitute the two main pathways for FeS-cluster biogenesis, both

regulated by IscR. Mutations in IscR were found in strains with elevated biotin production and proteomics was performed to compare protein levels in strains with and without IscR mutations. **A)** Schematic presentation of the *isc*-operon with the IscR regulatory mechanism. **B)** Schematic presentation of the *suf*-operon, including effects on expression from transcriptional regulators IHF, OxyR and FUR. **C)** Relationship between BioB protein expression and measured extracellular biotin produced from DTB for 3 strains at 2 induction conditions, low; 0.025 mM IPTG, high; 1 mM IPTG. Individual replicates are shown as separate dots, n=3. **D)** Percentage protein content of proteins from *isc*- and *suf*-operons. N=3, error bars represent standard deviation; individual replicates are in gray circles. **E)** Relative expression of proteins with significantly changed expression levels based on ANOVA with 95% confidence interval, Bonferroni correction, grouped based on protein function with potential influence on BioB reaction. Abbreviations in the figure; Ref: BS1013 (wildtype IscR) holding pBS430, induced at 0.025 mM IPTG, WT: BS1011 (wildtype IscR) holding pBS412, induced at 0.025 mM IPTG, MutL: BS1353 (IscR H107Y) holding pBS412, induced at 0.025 mM IPTG, MutH: BS1353 (IscR H107Y) holding pBS412, induced at 1mM IPTG.

Combining IscR Mutations was not Additive

We hypothesized that the three identified IscR mutations each disrupted FeS-cluster binding. If binding was completely disrupted by the individual mutations, combining the mutations would have no synergistic effects. However, if binding was only partially disrupted, combining the mutations could lead to full FeS-cluster binding disruption, leading to the phenotypic effect of further increased biotin production. The effect of combinatorial mutations in IscR as well as *iscR* deletion on expression of enzymes in the *isc*- and *suf*-pathways, has been investigated in detail for other processes[24,25]. Substituting alanines for any or all of the three FeS-coordinating cysteines of IscR results in similar changes in expression from the *sufA* and *iscR* promoters [25]. A similar expression profile is seen in strains with an IscR H107A mutation [24]. However, expression profiles from strains combining mutations in H107 with any FeS-cluster coordinating cysteines have not been reported.

We introduced the three mutations we identified in all seven combinatorial possibilities using multiplex genome engineering and tested for biotin production under expression conditions of 0, 0.01 and 0.1 mM IPTG (**Fig. 3A**). Overall, biotin production profiles for the combinatorial

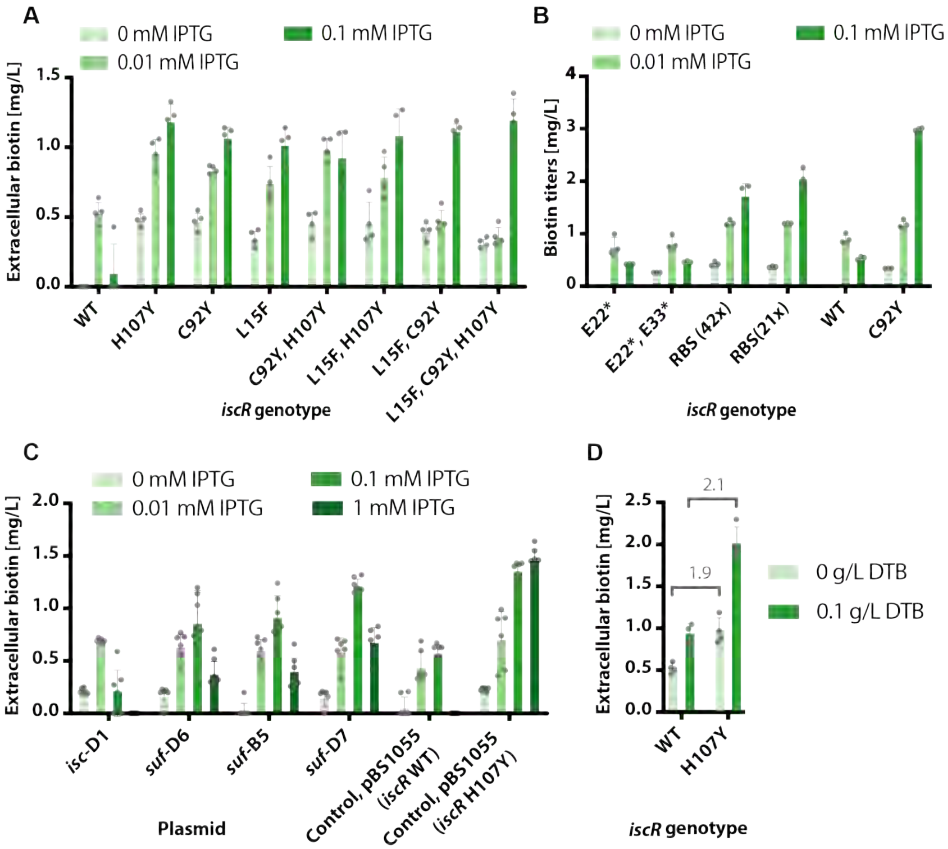


Figure 3: FeS-cluster Engineering in *E. coli*. After identifying three unique mutations in IscR, hypothesized to affect FeS-cluster binding, metabolic engineering strategies were used to evaluate and compare effects on biotin production. Error bars represent standard deviation and individual replicates are in gray circles. All strains expressed BioB from plasmid pBS412. **A)** Biotin production data for a combinatorial IscR mutant strain compared to a single mutant strain and wildtype reference strain (WT) (n=4). Naming is based on IscR genotype. **B)** Biotin production data on *iscR* knockout strains (E22* +/- E33*) and strains with predicted increased expression of the C92Y IscR mutant through RBS, ribosome binding site, engineering. The predicted relative increases are 21x and 42x. Expression is compared to an IscR mutant strain with native RBS and IscR wildtype (n=3). **C)** Biotin production data on IscR wildtype (WT) strains carrying plasmids expressing promoter libraries of *isc*- or *suf*. Mutant and wildtype IscR control strains carried control plasmid expressing GFP instead of *isc*- or *suf*-operon (n=6). **D)** *De novo* biotin production in wildtype and mutant IscR strain with and without 0.1 g/L DTB fed (n=4).

mutant strains were comparable to the single-mutant strains. This result suggested that mutating either of the FeS-cluster coordinating amino acids completely disrupted the IscR FeS-cluster binding ability. Also, mutant strains with the L15F substitution in IscR had a production profile similar to the other strains. From the IscR crystal structure, position L15 is proximal to the FeS-cluster coordinating amino acid H107Y (**Fig. 1D**), but the mechanism of the effects of this mutation are unclear.

Comparing Engineering Strategies for FeS-Cluster Supply

Our results indicated that FeS-cluster assembly was a limiting factor in BioB expression. We tested if previously reported engineering strategies to enhance FeS-cluster availability for other FeS-cluster enzymes through rational modification of *iscR* further improved BioB expression and biotin production. Deleting *iscR* to derepress the *isc*-operon has been shown to improve hydrogenase activity and hydrogen production in *E. coli* [32,33]. In addition to engineering IscR, overexpression of either the *isc*- or *suf*-operons improves expression and/or activity of the FeS-cluster enzymes holo-ferredoxins [34], nitrogenases [35], LipA [36] and IspH [37].

We generated an *iscR* knockout strain by introducing premature stop codons at positions 22 and 33 of *iscR* (**Fig. 3B**) and biotin production was evaluated at three induction conditions (0, 0.01, and 0.1 mM IPTG). The *iscR* knockout strains had production profiles comparable to the IscR wildtype strain and growth toxicity from BioB overexpression was not mitigated. Deleting *iscR* increases p_{iscR} expression and decreases p_{sufA} expression [25]. The finding that *iscR* knockout strains were similar to the IscR wildtype strain suggested that the expected increase in *isc*-operon expression did not supply enough FeS-cluster formation capacity to overcome growth inhibition from BioB overexpression.

The effect on biotin production from overexpressing the *isc*- and *suf*-operons was evaluated using plasmid libraries expressing native *E. coli*

sufABCDSE or *iscSUA-hscBA-fdx-iscX* operons with 16 different constitutive promoters with predicted strength, ranging from 3 arbitrary units (AU) to 866 AU [38]. Cells with the libraries had low viability, with only four unique operon plasmids identified from 96 sequenced colonies. More than half of sequenced colonies had lost the *isc*- or *suf*-plasmid, suggesting that plasmid expression was a burden to the cells. The strongest identified predicted promoter strength was 156 AU for the *isc*-operon and 99 AU for the *suf*-operon (Supplementary Table S2), which was at the lower end of the library range. We did not investigate if host viability correlated with promoter strength.

We evaluated the effect on biotin production of the four uniquely identified constructs (**Fig. 3C**). Co-expression of either the *isc*- or the *suf*-operon constructs with the BioB-overexpressing plasmid improved biotin titers 1.2-fold to 2-fold compared to the wildtype IscR control strain. The best-performing auxiliary construct expressed the *suf*-operon from the apFAB251 promoter with a predicted strength of 30.1 U [39] and had approximately 80% the biotin production of the reference strain with mutated IscR (suf-D7 vs. IscR H107Y, **Fig. 3E**).

Plasmid overexpression of *isc*- and *suf*-operon was hypothesized to be a burden to cells based on low viability with the expression library, which is a drawback for the genetic stability of cell factories [40]. We hypothesized that a stronger ribosome-binding site (RBS) in front of mutant *iscR* would elevate levels of apo-IscR, leading to *suf*-operon activation (**Fig. 4**), while being a lower burden to the cells compared to plasmid overexpression of *isc*- or *suf*-operon. However, engineering to increase RBS strength did not increase biotin titer. Mutant IscR strains with an altered RBS produced more biotin than the IscR wildtype strain but ~30% less than strains with native RBS (**Fig. 3B)**. The decrease in biotin production could be due to predictions about RBS strength, which have high inaccuracy [41], or might be caused by altered expression in parts of the IscR regulon other than *isc* and *suf*.

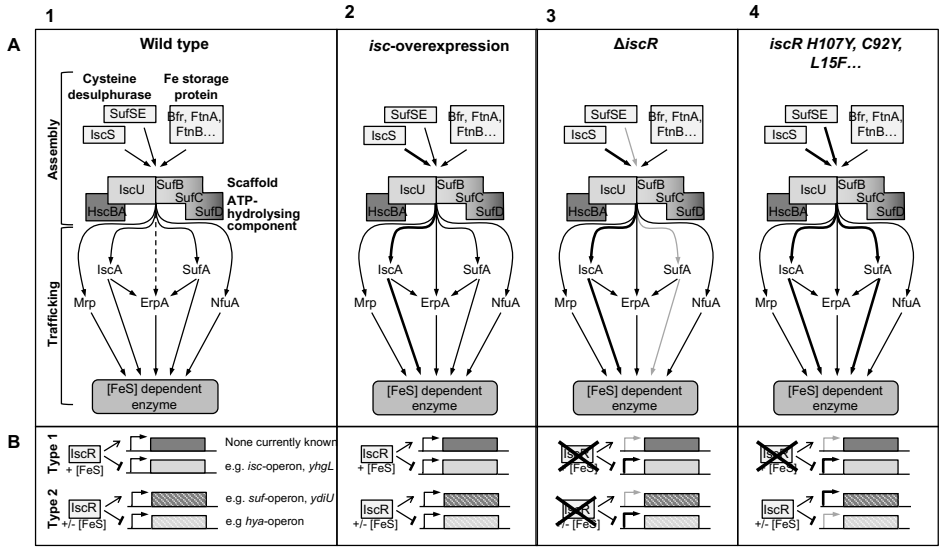


Figure 4: Effect of Isc Mutations on Expression of FeS-cluster Biogenesis Genes and Promoters Regulated by IscR. A) 1, Overview of FeS-biogenesis pathway in wildtype *E. coli*; 2-4, expected relative expression changes in the pathway in indicated *isc*-mutant strains. **B)** 1, Expression from IscR-controlled Type 1 and Type 2 promoters in wildtype *E. coli*; 2-4, Expected relative expression in *isc*-mutant strains compared to wildtype. No data were available on IscR expression and relative abundance of holo-IscR and apo-IscR in the *isc*-overexpression strain, but expression of IscR (and the genomically encoded *isc*-operon) was likely downregulated due to increased FeS-cluster incorporation into IscR.

De novo Production of Biotin Benefit from Mutations in IscR

In order to be an economical feasible cell factory, microbial biotin production needs to be based on a cheap carbon source such as glucose rather than an expensive precursor like DTB. Therefore, we wanted to evaluate whether the identified mutations could improve production from biotin cell factories *de novo*. Simple cell factories were established, overexpressing the native *E. coli* biotin operon with a mutation in the promotor region, to remove feedback inhibition from BirA [42] (pBS936) in an IscR wildtype and IscR H107Y strain. Even though no apparent toxicity from BioB-overexpression is observed in these strains, mutating IscR is found to improve *de novo* production by 1.9-fold and DTB to biotin conversion, in the same strain with fed DTB, by 2.1-fold (**Fig. 3D**). As

feeding of 0.1 g/L DTB increases production in both IscR wildtype and mutant IscR strains, the simple overexpression of the un-regulated biotin operon seems to be limited in DTB production. We observe that mutating IscR reaches similar *de novo* biotin production levels as an DTB-fed wildtype IscR strain (**Fig. 3D**).

IscR Mutations Improve Production from other FeS-Cluster-Dependent Pathways

We investigated if other potential cell factories with FeS-cluster enzymes benefited from IscR mutations. ThiC in thiamine biosynthesis and LipA in lipoic acid biosynthesis are FeS-cluster enzymes that are proposed bottlenecks in their production because of low catalytic rates [2,43]. We hypothesized that, like BioB, increased expression of these enzymes would improve flux towards their products, but with a high risk of FeS-cluster depletion.

In the lipoic acid biosynthesis pathway (**Fig. 5A**), LipA catalyzes the last step converting octanoic acid to lipoic acid. Plasmid overexpression of the *isc*-operon improves expression of soluble LipA in cell lysates [36]. Based on Fe and S measurements in soluble vs. insoluble cell lysate fractions, Kriek *et al.* suggest this increase is due to increases in FeS-cluster-bound holo-LipA, with apo-LipA in the insoluble fraction. Catalysis by LipA is reported to be similar to BioB catalysis, requiring [4Fe4S] cluster regeneration for multiple turnovers [43,44]. Due to these similarities, we hypothesized that the reaction catalyzed by LipA would also benefit from mutations in IscR.

Cell factories for lipoic acid production were established similar to the biotin cell factories, expressing LipA from an inducible plasmid. Because lipoic acid is synthesized *in vivo* as covalently attached lipoylated lipoyl domains to specific enzymes [43] (**Fig. 5A**), we co-expressed a truncated version of a lipoylated pyruvate dehydrogenase subunit encoded by *aceF* [45,46]. Production of free lipoic acid was evaluated in strains with

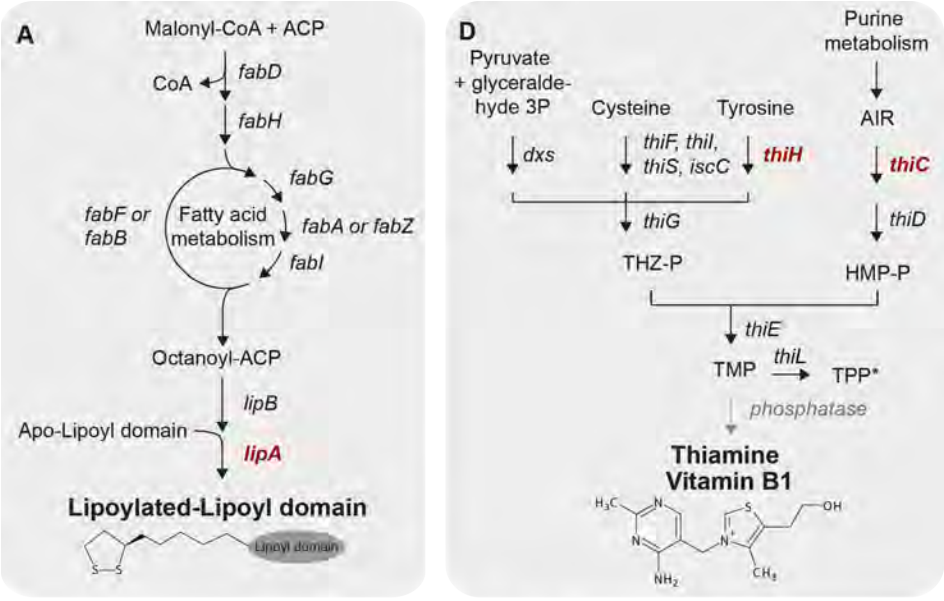
and without IscR mutations (C92Y and H107Y) after 24 hours growth in minimal media supplemented with octanoic acid. Growth profiles for IscR C92Y and wildtype IscR were generated under different LipA induction levels (**Fig. 5B**). The mutant strain had alleviated overexpression toxicity and improved production titers of extracellular, free lipoic acid (**Fig. 5C**). Even though most lipoic acid is expected to be protein bound, a small fraction is released as free lipoic acid by unidentified mechanisms. We assumed this level to correlate with the total amount of protein-bound lipoic acid generated. The mutant strains produced 9.3 µg/L free, extracellular lipoic acid when LipA expression was induced, representing a 1.8-fold increase compared to the wildtype IscR strain[47].

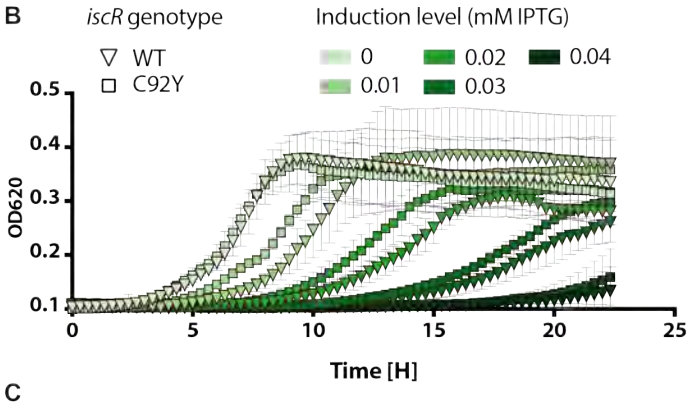
Similar to lipoic acid, the effects of mutations in the FeS-cluster binding site of IscR were tested on thiamine production in bacterial cell factories. Thiamine has a complex biosynthetic pathway with two branches and seven core genes (**Fig. 5D**), of which two encode radical SAM enzymes (ThiC and ThiH) [48]. Improvements in chemical synthesis of thiamine [49] have generated a highly efficient process. Cost-competitive microbial production will require extensive improvements [1] but little metabolic engineering has been reported in the literature on the thiamine pathway [50]. A study using transposon mutants identified several genes related to FeS-cluster homeostasis that were affected in strains with altered thiamine production [51], supporting thiamine as a good fit for testing the IscR mutations.

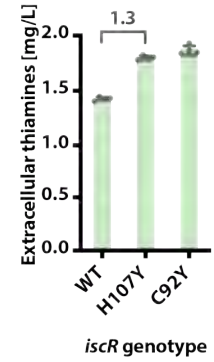
We established an *E. coli* thiamine cell factory by overexpressing the native *thiCEFSGH* operon on a plasmid with a strong constitutive promoter. Based on initial characterization, the thiamine production strain was limited by the SAM radical enzyme ThiC. Each identified mutation in the FeS-cluster binding site of IscR that was beneficial for BioB expression was introduced into the strain. Extracellular thiamine production evaluated after incubation for 24 hours in minimal media showed a 1.3-fold increase in thiamine production by the mutant strains

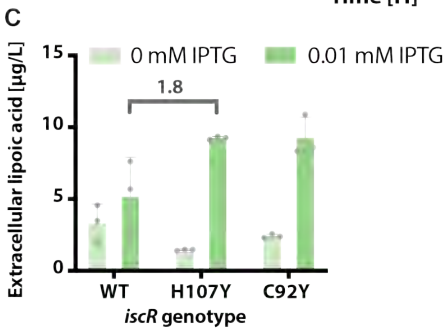
compared to the optimized parent cell factory, for a titer of 1.9 mg/L extracellular thiamine (**Fig. 5E**). This further represents a 1.5 fold improvement over the previously best reported titer of 1.3 mg/L based on engineered *Bacillus subtilis*[50].

Figure 5: Effects of IscR Mutations on other Biosynthetic Pathways with FeS-Cluster Enzymes. A) Lipoic acid biosynthetic pathway in *E. coli*. **B)** Growth rate data on LipA-overexpressing strains with (squares) and without (triangles) mutation in IscR for five induction levels from 0 mM to 0.04 mM IPTG (darker green indicates increasing IPTG) (n = 3, error bars represent standard deviation). **C)** Lipoic acid production titers of strains with and without mutation in IscR (n=3, error bars represent standard deviation, individual replicates are in gray circles). **D)** Thiamine biosynthetic pathway in *E. coli.* **E)** Thiamine production titers of strains with and without mutation in IscR (n=4, error bars represent standard deviation, individual replicates are in gray circles). Red, FeS-cluster enzymes.


A
Malonyl-CoA + ACP
CoA
fabD
fabH
fabG
fabF or fabB
Fatty acid metabolism
fabA or fabZ
fabI
Octanoyl-ACP
lipB
Apo-Lipoyl domain
lipA
Lipoylated-Lipoyl domain
Lipoyl domain
D
Pyruvate + glyceralde-hyde 3P
Cysteine
Tyrosine
Purine metabolism
AIR
dxs
thiF, thiI, thiS, iscC
thiH
thiC
thiG
thiD
THZ-P
HMP-P
thiE
TMP
thiL
TPP*
phosphatase
Thiamine Vitamin B1


B
iscR genotype
WT
C92Y
Induction level (mM IPTG)
0
0.01
0.02
0.03
0.04
OD620
Time [H]


E
Extracellular thiamines [mg/L]
1.3
WT
H107Y
C92Y
iscR genotype


C
Extracellular lipoic acid [µg/L]
0 mM IPTG
0.01 mM IPTG
1.8
WT
H107Y
C92Y
iscR genotype

Discussion

Improving BioB has long been limiting for establishing high-producing biotin cell factories [8,11,12]. We demonstrate that single mutations in the IscR regulator diminish growth inhibition because of BioB overexpression and enable a 2.2-fold increase in biotin production compared to a strain expressing BioB to the limit of toxicity (**Fig. 1E, F**). Based on these results and previous reports that C92, C98, C104 and H107 in IscR are responsible for FeS-cluster binding [24,25], we hypothesize that our mutant strains generate an IscR protein that is always in apo-IscR form regardless of cellular FeS-cluster concentration. FeS-cluster binding determines the regulatory nature of IscR, which has a central role in FeS-cluster biogenesis [23]. IscR in apo-form only would constantly derepress the *isc*-operon and activate the *suf*-operon [24,25], increasing FeS-cluster biogenesis (**Figs. 2A,B,E, 4**). The observation that upregulation of FeS-cluster biogenesis reduced growth toxicity from BioB overexpression suggested that FeS-cluster depletion is a main contributor to the observed toxicity in IscR nonmutant strains. Insufficiencies in FeS-cluster formation can lead to inactivation of other, essential FeS-cluster proteins, including members of the TCA cycle such as those encoded by *acnA*, *acnB*, *fumA* and *fumB* [23].

We evaluated overexpressing the *isc*- or *suf*-operon as an alternative way to improve FeS-cluster supply and affect biotin production. Plasmid-based overexpression of either the *suf*- or *isc*-operon improves catalysis of FeS-cluster enzymes such as nitrogenases [35], LipA [36] and IspH [37]. Although biotin production improvements ranged from 1.2 to 2-fold (**Fig. 3C**) when overexpressing *isc*- and *suf*-operon at different levels, the strains holding the plasmid libraries had low viability. One possible explanation is that the plasmid that highly expressed entire operons was a higher burden to cells than genomically engineered improvements in FeS-cluster supply. Metabolic burdens and toxicities are major drawbacks for cell

factory scale-up [52] and even small differences have a large impact on production over many generations [40,52]. Single point modifications in genomic IscR seemed to burden cells less than plasmid overexpression of *isc*- and *suf*-operons and offer a more elegant solution with greater freedom to quickly test the effect of other plasmid-expressed genes that might benefit biotin production. In summary, the findings supported the hypothesis that FeS-cluster supply was a limiting factor in strains overexpressing BioB, because biotin production improved over wildtype IscR when FeS-cluster biogenesis was improved.

BioB is suggested to be a "suicide enzyme" with one or fewer turnovers per monomer *in vitro* [14,53]. From our proteomic analysis, we observed that extracellular biotin levels correlated linearly with levels of measured BioB protein. These observations neither confirmed nor rejected the hypothesis of BioB as a true suicide enzyme. The catalytic [2Fe2S]-cluster [3] may be regenerated during BioB catalysis, similar to one of the two [4Fe4S]-clusters in LipA. NfuA also efficiently reconstitutes the [4Fe4S] auxiliary cluster during LipA catalysis in a non-rate-limiting step [44]. However, the factors responsible for a potentially similar mechanism in BioB remain to be identified.

When mutating a global regulator such as IscR, which has a regulon of more than 40 genes [22] including some that also have regulatory roles, several pathways are likely to be affected and rationally predicting all changes can be difficult. While limitations in FeS-cluster supply and repair are likely to be main factors explaining BioB toxicity, based on the direct effect of mutating IscR on *isc*- and *suf*-operon expression (**Fig. 2D**), proteomics suggested that other areas might also be affected. These areas include oxidative stress tolerance, iron availability, methionine biosynthesis and redox potential (**Fig. 2E**). Metalloproteins such as FeS-cluster enzymes are prone to damage by ROS leading to oxidative stress

[17]. The decrease in FtnA protein levels in the IscR mutant strain compared to the wildtype strain at low BioB induction (**Fig. 2E**) was likely an effect of improved expression of FeS-cluster biogenesis proteins resulting from mutating IscR [25]. FtnA is a ferritin used to store iron or damaged FeS-clusters to avoid ROS generation [54]. An increased ability to synthesize and repair FeS-clusters would decrease the need for scavenging. In addition to improving FeS-cluster generation, IscR mutations in the FeS-cluster binding site are reported in *E. coli* strains evolved for increased tolerance of oxidative stress using treatment with paraquat [55]. Yang *et al.* explain the increased tolerance to oxidative stress as increased FeS-cluster biosynthesis to overcome predicted mismetallation of IscU, the FeS-cluster scaffold protein of the *isc*-pathway [55]. These observations are consistent with our finding that mutations in IscR affects both FeS-cluster shortage and cellular oxidative stress levels, helping to mitigate the growth inhibition from BioB overexpression. Additional omics analyses such as transcriptomics and metabolomics could further elucidate the cellular effects of mutating IscR and assist in expanding the explanation for BioB toxicity.

Even in biotin-production strains without observed growth inhibition from BioB expression, mutating IscR was observed to improve production (**Fig. 3D**). We found that production from *de novo* biotin cell factories improved 1.9-fold when introducing a mutant IscR, suggesting that the BioB catalysis are limited FeS-cluster supply, which can be improved upon mutating IscR. The *de novo* production of mutant strains we comparable to the DTB-fed of the reference strain with wildtype IscR. This suggests that assisting the cell in modulating its regulation of FeS-cluster biogenesis and potentially other pathways through IscR, redirect resources and increases flux in the biotin biosynthesis pathway upstream DTB.

Finally, we demonstrated improvements of 1.8-fold for lipoic acid titers and 1.3-fold for thiamine production (**Fig. 5**) by introducing the IscR mutations into lipoic acid and thiamine production strains. Since the biosynthesis of lipoic acid and thiamine both depend on FeS-cluster enzymes for key enzymatic steps [2,47], the improvement in titers confirmed that the FeS-cluster(s) in BioB are the most likely source of growth inhibition when BioB is overexpressed. The results further emphasize the likelihood of increased FeS-cluster supply and repair as the main reason for improvements in strains with mutated IscR. The finding that three independent biosynthesis pathways had improved production titers indicated that the mutations in IscR likely disrupted FeS-cluster binding and have broad platform potential. Mutating the FeS-cluster binding site of IscR will likely benefit other cell factories with FeS-cluster enzymes. The results show how the consequences from such mutations can benefit production of different compounds when the mutations are involved in a metabolic hub branching to many different metabolites. Our findings emphasize how tools to improve production of some compounds and their synergistic effects may accelerate the development of cell factories for high-value compounds. Mutating the FeS-cluster binding site of IscR is likely to benefit additional FeS-cluster enzymes.

Acknowledgements

We thank Johan Hekelaar (Molecular Systems Biology, Groningen University, The Netherlands) for performing proteomics. We further thank Gijs Verkleij for assistance with MAGE of combinatorial IscR mutants. A.P.B acknowledges funding from Innovation Foundation Denmark (Industrial Ph.D. program, No. 5016-00135B). L.S.G and H.J.G acknowledge funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 686070, DD-

DeCaf. M.O.A.S. acknowledges funding from The Novo Nordisk Foundation under NFF grant number: NNF10CC1016517.

Author Contributions

A.P.B., H.J.G, and L.S.G conceived the study with input from M.O.A.S. J.B. constructed the selection plasmid. A.P.B. characterized the selection system, carried out selections and characterized evolved colonies. N.M.P. established the biotin bioassay. A.P.B. and D.L.H. established bioassay and cell factory for lipoic acid and evaluated the effect of the identified IscR mutations. B.S. cloned plasmids for the thiamine cell factory. A.P.B. prepared samples for proteomics analysis and carried out downstream data analysis. A.P.B. evaluated the effect of the identified IscR mutations in thiamine cell factories. A.P.B. wrote the manuscript with contributions from all other authors.

Competing Interests

Biosyntia ApS has filed a patent application based on the results of this paper.

Methods

Materials and media

All chemicals were from Sigma Aldrich or Carl Roth unless otherwise stated. Minimal MOPS media was made as described in Supplementary Note 1 and used as standard assay medium unless otherwise stated. Minimal MOPS succinate media was prepared as minimal MOPS media, replacing glucose with 50 mM succinic acid disodium (final concentration). Biotin stock solution was prepared in DMSO to a final concentration of 100 mM and then diluted in mMOPS. Lipoic acid stock solution was prepared in ethanol to a concentration of 1

g/L with dilutions made in minimal MOPS succinate medium. Plasmid and PCR purifications were performed according to manufacturers' guidelines, using E.Z.N.A. Plasmid Mini Kit I, V-spin and E.Z.N.A. Cycle Pure Kit I from VWR.

Isolating iscR Mutant Strains

The biotin selection system was based on a single plasmid encoding an IPTG-inducible *E. coli bioB* gene combined with a biotin auxotrophic BW25113 Δ*bioB* strain. Cell cultures grown for several generations and passaged for 4 days were plated on mMOPS plates with kanamycin (50 μg/mL) and 0.1 mM IPTG to induce BioB expression to toxicity. Cells appearing on induction plates after 24 hours incubation at 37 ºC were characterized as described in the main text. In-depth characterization of the selection system is in Supplementary Note 2.

Strains and Plasmids

See Supplementary tables S1-S3.

Structural Investigations of IscR Mutants

The identified IscR mutations were highlighted by "stick drawings" in the protein structure of apo-IscR bound to *hya* DNA (PDB ID: 4HF1 [56]) using PyMOL v. 1.7 [57]. Each mutated amino acid was changed from the original by copying the residue followed by the "Wizard Mutagenesis" tool. Changing the visualization to stick drawings were done using "sticks" under "show" options.

Construction of Modified IscR Strains

Combinatorial IscR mutant strains and *iscR* knockout and RBS mutant-*iscR* strains were constructed using up to 6 rounds of MAGE [58]. Briefly, BS1011 was grown in 4 mL Luria broth (LB) with 100 μg/mL ampicillin at 37ºC with shaking (250 rpm) in culture tubes. At OD600 0.5,

400 µl 20% arabinose was added to induce the λ-Red system. Incubation continued for 30 minutes, after which cells were placed on ice for 5 minutes, spun down in a precooled centrifuge at 5000G for 5 minutes and washed twice in 2 mL icecold water. Cells were collected by centrifuging at 5000G for 5 minutes and resuspended in 200 µl icecold water. A 45-µl cell suspension was transferred to precooled 1 mm electroporation cuvette and mixed with 5 µl 100 µM oligo stock (or oligo pool) followed by electroporation at 1800 V before 950 µl prewarmed SOC media were quickly added and cells transferred to growth tubes for incubation for 1-2 hours. Three mL LB with 100 µg/mL ampicillin was added, and procedures repeated up to 6 times to OD600 0.5. Successful introduction of mutations was evaluated by colony PCR using oBS1670 and oBS1991, annealing temperature 56ºC, and elongation time 2 minutes with subsequent sequencing of purified PCR constructs. MAGE oligos were designed using MODEST software [59] (Supplementary Table S3).

Construction of Lipoic Acid Cell Factory

BS1912 was constructed by curing the Keio Δ*lipA* strain of its *kanR* marker. MAGE used moBS150, introducing the IscR *C92Y* point mutation into BS1912, as described above, resulting in BS2114. The *lipA* and truncated *aceF* [45] genes were amplified from genomic *E. coli* DNA with PhusionU polymerase using, respectively, oBS2200+oBS2197 and oBS2274+oBS2277. By amplifying pBS0910 with oBS1587+oBS1591 using PhusionU polymerase, *lipA* was USER-cloned behind the IPTG inducible promoter, T5lacO [60] . The resulting plasmid, pBS0992, was amplified with oBS1468+oBS2273 using PhusionU polymerase and truncated *aceF* was USER-cloned behind a strong constitutive promoter, apFAB309. Cloned plasmids were transformed into homemade RbCl chemically competent BS1912 and BS2114. Plasmids were purified and sequenced verified with Sanger sequencing. BS1912 and BS2114 containing the correctly assembled plasmid, pBS1037, with constitutively

expressed, truncated *aceF* and IPTG-inducible *lipA*, were used for lipoic acid production and LipA toxicity experiments.

De novo Production of Biotin

BW25113 Δ*bioB* strains with and without mutation in *iscR* H107Y holding pBS936, were tested for biotin production, as described below, with and without 0.1 g/L DTB added to the mMOPS production media. The native *E. coli bio*-operon with a mutation in the *bioO* regulator site were expressed from a high copy-number plasmid. The *bioO* mutation were reported as a "Type 9" mutation from Ifuku *et al.*[42].

Construction of Thiamine Cell Factory

The two native operons involved in thiamine biosynthesis, *thiC* and *thiM*, were amplified from *E. coli* MG1655 genomic DNA using Phusion High-Fidelity Polymerase (Thermo Scientific, F530L) according to the manufacturer's instructions. Primer pairs oGEN227+oGEN264 and oBS421+oBS422 were used to introduce 25-33 bp of homology overlap with the target backbones and a strong RBS in front of the first genes of *thiC* and *thiM,* respectively. Two empty backbone plasmids carrying strong apFAB promoters [39] pBS100 (apFAB46) and pGEN50 (apFAB71) were linearized by PCR using Phusion polymerase and oligos oGEN265 and oGEN266. The *thiC*-operon was introduced to pBS100 and *thiM*-operon was introduced to pGEN50 using Gibson Assembly Master Mix (New England Biolabs, E2611L) according to the manufacturer's protocol [61] and transformed into One Shot TOP10 chemically competent cells (ThermoFisher Scientific, C404006) according to the supplier's instructions. Single colonies isolated on LB agar plates containing kanamycin (50 mg/mL) were used to inoculate 4-ml cultures grown in liquid LB+kanamycin (50 mg/L) overnight for retrieval of assembled plasmids using NucleoSpin Plasmid miniprep kits (Machery-Nagel, 740588) according to the manufacturer's protocol, omitting the optional

washing step with AW buffer. Resulting plasmids carrying thiamine operons *thiC* and *thiM* were verified by Sanger sequencing (Eurofins Genomics) to ensure correct assembly, and named pBS116 and pBS117, respectively. The entire plasmid carrying the *thiC*-operon (pBS116) was linearized through PCR using Phusion polymerase and primer pair oGEN289+oGEN290. The *thiM*-operon including the promoter was amplified from pBS117 with primer pairs oGEN182+oGEN184. The two linearized operons were assembled into a single plasmid (pBS140) using Gibson Assembly Master Mix, transformed into One Shot TOP10 chemically competent cells, isolated with the NucleoSpin Plasmid miniprep kits and sequenced verified. Finally, the complete thiamine-producing plasmid was introduced into the production strain by electroporation as described above.

Cloning and Evaluation of suf- and isc-libraries

Promoter libraries of native *E. coli suf*- and *isc*-operons (without *iscR*) were made. Plasmid backbone was amplified from pBS1055 using oBS2235-oBS2250 and oBS1468 to introduce promoter variants. The *suf*- and *isc*-operons were amplified from genomic *E. coli* DNA using, respectively, primer pairs oBS2259+oBS2260 and oBS2264+oBS2265. All PCR used PhusionU polymerase, for subsequent USER cloning [60], at annealing temperature 60 ºC and elongation time 7 minutes. Ligations were transformed into chemically competent BS1011 strains with pBS412, plated on selective mMOPS plates, incubated at 37ºC overnight and verified by colony PCR (56ºC annealing temperature, 5 minutes elongation time) and Sanger sequencing using primers oBS666+oBS955 (*isc*) and oBS666+oBS1745 (*suf*) and oBS666. The viability of cells with libraries was very low, resulting in only 4 unique constructs based on sequencing of 96 clones: 1 for the *isc*-operon and 3 for the *suf*-operon.
PCR reactions were in total volume 50 µl, with 1.5 µl each primer (10 µM stock; when several primers were used, 1.5 µl of a pre-mixed pool was

added), 1 µl template (<10 ng total), 1 µl polymerase, 10 µl 5x HF buffer, 1.5 µl dNTP (10 mM) and 34.5 µl MilliQ water. For colony PCR, reactions were scaled to 10 µl using X7 polymerase. For USER cloning, ~1 pmol total fragments in equimolar amounts were used, mixed with 1 µl DpnI enzyme and 1µl DpnI buffer (10x). Reaction volumes were adjusted to 10 µl with MilliQwater and incubated for 15 minutes at 37°C. To reactions, 1 µl USER enzyme was added for incubation for 15 minutes at 37°C, followed by 15 minutes incubation at room temperature.

Production of Biotin, Thiamine and Lipoic Acid

All production tests used a similar protocol. Single colonies were picked for preculturing in 400 µl minimal MOPS with necessary antibiotics (ampicillin, 100 mg/mL; kanamycin 50 mg/mL; spectinomycin, 50 mg/mL; and/or chloramphenicol 30 mg/mL) in deep-well plates (DWPs) sealed with breathable seal and incubated overnight at 37°C with shaking (250 rpm). Precultures were used to inoculate 400 µl production cultures in DWPs to initial OD600 0.01. Production cultures were in minimal MOPS with necessary antibiotics with DTB (0.1 g/L) for biotin production or octanoic acid (0.6 g/L) for lipoic acid production. Production cultures were incubated for 24 hours at 37°C with shaking (250 rpm), sealed with breathable seals. For experiments with growth profiles, 600 µl production cultures were prepared. Before incubation, 200 µl was transferred to a microtiter plate, sealed with transparent breathable seal and OD monitored every 20 minutes using a microplate photometer (Multiskan FC, Thermo Scientific). OD600 of production cultures was measured at the end of incubation, cells were spun at 5000 G for 5 minutes, and supernatants transferred to detect product.

Bioassay for Biotin Detection

We used a modified version of a published liquid bioassay [62] using *E. coli* BW25113 Δ*bioB* (BS1011) with ZeoR plasmid pBS451 as the

bioassay strain that was unable to grow on any amount of DTB. A single colony from an LB plate with zeocin (40 mg/L) was used to inoculate minimal MOPS media with zeocin (40 mg/L) for growth at 37ºC overnight starved for biotin. Bioassay medium was made by diluting the bioassay culture in minimal MOPS medium with zeocin (40 mg/L) to a final OD600 of 0.01. In a microtiter plate, 15 µl sample supernatant was mixed with 135 µl bioassay medium. Separate wells contained a standard curve with known concentrations of biotin in bioassay medium. Plates were sealed with breathable seals and incubated at 37ºC for 20 hours with shaking 250 rpm. OD600 was measured and biotin concentrations calculated. If necessary, sample supernatants were diluted to be in range of the bioassay.

Thiochrome Assay for Thiamine Detection

Thiamine and thiamine derivatives (thiamine monophosphate and thiamine pyrophosphate) in the supernatants of production samples were measured by derivatization to fluorescent thiochromes as described [63]. Briefly, 50 µl supernatant was mixed with 100 µl 4 M potassium acetate in black microtiter plates and 50 µl freshly prepared 3.8 M potassium ferricyanide in 7 M NaOH was added and mixed by pipetting. Quenching was by adding 50 µl fresh 0.06% H_2O_2 in saturated KH_2PO_4. Fluorescence emission was measured at 444 nm after excitation at 365 nm and concentration of thiamines estimated based on comparison to standard curves included in the derivatization plate.

Bioassay for Lipoic Acid Detection

An assay similar to our biotin bioassay was set up based on the *E. coli* BW25113 Δ*lipA* strain (BS1912) with a ZeoR plasmid (pBS451) as the bioassay strain. Liquid bioassays were based on a previous publication [64]. A single colony of the bioassay strain from an LB plate with zeocin (40 mg/L) was used to inoculate minimal MOPS succinate media with zeocin (50 mg/L) for growth at 37ºC overnight, starved for lipoic acid.

Usually no detectable growth was observed, and cultures were used directly as bioassay medium. In microtiter plates, 15 µl sample supernatant was mixed with 135 µl bioassay medium. A standard curve with known concentrations of lipoic acid in bioassay medium was included in separate wells. Plates were sealed with breathable seal and incubated at 37°C for 20 hours with shaking at 250 rpm before measuring OD600 and calculating lipoic acid concentrations. Sample supernatants were diluted if necessary, to be in range of the bioassay.

Proteomics Analysis

Label-free proteomics was carried out at Groningen University by the Molecular Systems Biology group led by Professor Matthias Heinemann as described [65]. Precultures were inoculated with single colonies in mMOPS media with ampicillin (100 µg/mL) and incubated at 37°C with 250 rpm shaking overnight before using to inoculate 50 mL mMOPS with ampicillin and either 0.025 mM IPTG or 1 mM IPTG in a 250 mL baffled shaker flask to initial OD600 0.001. Flasks were incubated at 37°C with 250 rpm shaking for 10 generations (OD600 ~0.5) and 8 x 10^8 cells were harvested in technical triplicates by centrifuging at 4°C at 17,000 G for three minutes. Supernatants were evaluated for production using bioassays and cell pellets were washed twice in 2 mL icecold PBS buffer. Cell pellets were snap-frozen in liquid nitrogen and stored at -80°C until mass spectrometry (MS)-MS-analysis.
Data were normalized as the percentage fraction of each protein to the total amount of protein in each sample, based on intensity-based quantification (iBAQ) values for each strain as

$$Protein\ fraction\ (\%) = \frac{iBAQ\ [specific\ protein]}{iBAQ\ [total\ protein]} \cdot 100$$

Data were analyzed based on ANOVA multiple comparison test with 95% confidence intervals applying Bonferroni-correction using the

MaxQuant software Perseus [66]. Prior to analysis, data with a quality score (measure of false discovery rate) of more than 0.01 were removed, together with proteins undetected in more than 30% of the samples.

References

1. Acevedo-Rocha, C. G., Gronenberg, L. S., Mack, M., Commichau, F. M. & Genee, H. J. Microbial cell factories for the sustainable manufacturing of B vitamins. *Curr. Opin. Biotechnol.* **56**, 18–29 (2019).
2. Palmer, L. D. & Downs, D. M. The thiamine biosynthetic enzyme ThiC catalyzes multiple turnovers and is inhibited by S-adenosylmethionine (AdoMet) metabolites. *J. Biol. Chem.* **288**, 30693–9 (2013).
3. Cramer, J. D. & Jarrett, J. T. *Purification, Characterization, and Biochemical Assays of Biotin Synthase From Escherichia coli*. *Methods in Enzymology* **606**, (Elsevier Inc., 2018).
4. Cronan, J. E. Biotin and Lipoic Acid: Synthesis, Attachment and Regulation. *Ecosal Plus* **18**, 1199–1216 (2014).
5. Broderick, J. B., Duffus, B. R., Duschene, K. S. & Shepard, E. M. Radical S-adenosylmethionine enzymes. *Chem. Rev.* **114**, 4229–4317 (2014).
6. Schwechheimer, S. K., Park, E. Y., Revuelta, J. L., Becker, J. & Wittmann, C. Biotechnology of riboflavin. *Appl. Microbiol. Biotechnol.* (2016).
7. Brown, S. W. *et al.* The production of biotin by recombinant strains of Escherichia coli. *J. Chem. Technol. Biotechnol.* **50**, 115–121 (1991).
8. Xiao, F. *et al.* Multi-level metabolic engineering of Pseudomonas mutabilis ATCC31014 for efficient production of biotin. *Metab. Eng.* (2019).
9. Sakurai, N., Imai, Y., Masuda, M., Komatsubara, S. & Tosa, T. Construction of a Biotin-Overproducing Strain of Serratia marcescens. *Appl. Environ. Microbiol.* **59**, 2857–63 (1993).
10. Brown, S. W. & Komagowa, K. The Production of Biotin by Genetically Modified Micro-organisms. *Biotechnol. Genet. Eng. Rev.* **9**, 295–326 (1991).
11. Ifuku, O. *et al.* Molecular analysis of growth inhibition caused by overexpression of the biotin operon in Escherichia coli. *Biosci Biotechnol Biochem* **59**, 184–189 (1995).
12. Van Arsdell, S. W. *et al.* Removing a bottleneck in the Bacillus subtilis biotin pathway: bioA utilizes lysine rather than S-adenosylmethionine as the amino donor in the KAPA-to-DAPA reaction. *Biotechnol. Bioeng.* **91**, 75–83 (2005).
13. Masuda, M. *et al.* Further Improvement of D-Biotin Production by a Recombinant Strain of Serratia marcescens. *Process Biochem.* **30**, 553–562 (1995).
14. Jarrett, J. T. Biotin Synthase: Enzyme or Reactant? *Chem. Biol.* **12**, 409–410 (2005).
15. Qiu, X. *et al.* Determination of Active Vitamin B12 (Cobalamin) in Dietary Supplements and Ingredients by Reversed-phase Liquid Chromatography: Single-laboratory Validation. *Food Chem.* **12**, 125010 (2019).
16. Sakurai, N., Imai, Y. & Komatsubara, S. Instability of the mutated biotin operon

plasmid in a biotin-producing mutant of Serratia marcescens. *J. Biotechnol.* **43**, 11–19 (1995).
17. Imlay, J. A. Iron-sulphur clusters and the problem with oxygen. *Mol. Microbiol.* **59**, 1073–1082 (2006).
18. Reyda, M. R., Dippold, R., Dotson, M. E. & Jarrett, J. T. Loss of Iron Sulfur Clusters from Biotin Synthase as a Result of Catalysis Promotes Unfolding and Degradation. *Arch. Biochem. Biophys.* **471**, 32–41 (2008).
19. Choi-Rhee, E. & Cronan, J. E. A nucleosidase required for in vivo function of the S-adenosyl-L-methionine radical enzyme, biotin synthase. *Chem. Biol.* **12**, 589–93 (2005).
20. Streit, W. R. & Entcheva, P. Biotin in microbes, the genes involved in its biosynthesis, its biochemical role and perspectives for biotechnological production. *Appl. Microbiol. Biotechnol.* **61**, 21–31 (2003).
21. Parveen, N. & Cornell, K. a. Methylthioadenosine/S-adenosylhomocysteine nucleosidase, a critical enzyme for bacterial metabolism. *Mol. Microbiol.* **79**, 7–20 (2011).
22. Giel, J. L., Rodionov, D., Liu, M., Blattner, F. R. & Kiley, P. J. IscR-dependent gene expression links iron-sulphur cluster assembly to the control of O2-regulated genes in Escherichia coli. *Mol. Microbiol.* **60**, 1058–1075 (2006).
23. Py, B. & Barras, F. Building Fe–S proteins: bacterial strategies. *Nat. Rev. Microbiol.* **8**, 436–446 (2010).
24. Fleischhacker, A. S. *et al.* Characterization of the [2Fe-2S] cluster of Escherichia coli transcription factor IscR. *Biochemistry* **51**, 4453–4462 (2012).
25. Yeo, W. S., Lee, J. H., Lee, K. C. & Roe, J. H. IscR acts as an activator in response to oxidative stress for the suf operon encoding Fe-S assembly proteins. *Mol. Microbiol.* **61**, 206–218 (2006).
26. Nahlik, M. S., Brickman, T. J., Ozenberger, B. A. & McIntosh, M. A. Nucleotide sequence and transcriptional organization of the Escherichia coli enterobactin biosynthesis cistrons entB and entA. *J. Bacteriol.* **171**, 784–790 (1989).
27. Price, C. E. & Driessen, A. J. M. Biogenesis of membrane bound respiratory complexes in Escherichia coli. *Biochim. Biophys. Acta - Mol. Cell Res.* **1803**, 748–766 (2010).
28. Kanai, T., Takahashi, K. & Inoue, H. Three distinct-type glutathione S-transferases from Escherichia coli important for defense against oxidative stress. *J. Biochem.* **140**, 703–711 (2006).
29. Ezraty, B., Gennaris, A., Barras, F. & Collet, J.-F. Oxidative stress, protein damage and repair in bacteria. *Nat. Rev. Microbiol.* **15**, 385–396 (2017).
30. Zhang, Z. *et al.* A transporter of Escherichia coli specific for L- and D-methionine is the prototype for a new family within the ABC superfamily. *Arch. Microbiol.* **180**, 88–100 (2003).
31. Ferla, M. P., Patrick Correspondence, W. M. & Patrick, W. M. Bacterial methionine biosynthesis. (2019).
32. Akhtar, M. K. & Jones, P. R. Deletion of iscR stimulates recombinant clostridial Fe-Fe hydrogenase activity and H2-accumulation in Escherichia coli BL21(DE3). *Appl. Microbiol. Biotechnol.* **78**, 853–862 (2008).
33. Kuchenreuther, J. M. *et al.* High-Yield Expression of Heterologous [FeFe]

Hydrogenases in Escherichia coli. *PLoS One* **5**, e15491 (2010).
34. Nakamura, M., Saeki, K. & Takahashi, Y. Hyperproduction of recombinant ferredoxins in *Escherichia coli* by coexpression of the ORF1-ORF2-*iscS-iscU-iscA-hscB-hs cA-fdx*-ORF3 gene cluster. *J. Biochem.* **126**, 10–18 (1999).
35. Li, X.-X., Liu, Q., Liu, X.-M., Shi, H.-W. & Chen, S.-F. Using synthetic biology to increase nitrogenase activity. *Microb. Cell Fact.* **15**, 43 (2016).
36. Kriek, M., Peters, L., Takahashi, Y. & Roach, P. L. Effect of iron-sulfur cluster assembly proteins on the expression of Escherichia coli lipoic acid synthase. *Protein Expr. Purif.* **28**, 241–245 (2003).
37. Gräwert, T. *et al.* IspH Protein of Escherichia c oli : Studies on Iron−Sulfur Cluster Implementation and Catalysis . *J. Am. Chem. Soc.* **126**, 12847–12855 (2007).
38. Mutalik, V. K. *et al.* Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat.Methods* **10**, 347–353 (2013).
39. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10**, 354–60 (2013).
40. Rugbjerg, P., Myling-Petersen, N., Porse, A. & Sommer, M. Diverse genetic error modes constrain large-scale bio-based production. *Rev.* (2018).
41. Salis, H. M. The ribosome binding site calculator. *Methods Enzymol.* **498**, 19–42 (2011).
42. Ifuku, O. *et al.* Sequencing analysis of mutation points in the biotin operon of biotin-overproducing Escherichia coli mutants. *Biosci Biotechnol Biochem* **57**, 760–765 (1993).
43. Cronan, J. E. Biotin and Lipoic Acid: Synthesis, Attachment, and Regulation. *EcoSal Plus* **1**, 217–244 (2013).
44. McCarthy, E. L. & Booker, S. J. Destruction and reformation of an iron-sulfur cluster during catalysis by lipoyl synthase. *Science (80-.).* **358**, 373–377 (2017).
45. Ali, S. T. & Guest, J. R. Isolation and characterization of lipoylated and unlipoylated domains of the E2p subunit of the pyruvate dehydrogenase complex of Escherichia coli. *Biochem. J.* **271**, 139–145 (1990).
46. Sun, Y., Zhang, W., Ma, J., Pang, H. & Wang, H. Overproduction of α-Lipoic Acid by Gene Manipulated Escherichia coli. *PLoS One* **12**, e0169369 (2017).
47. Cronan, J. E., Zhao, X. & Jiang, Y. *Function, attachment and synthesis of lipoic acid in Escherichia coli*. *Advances in Microbial Physiology* **50**, (Elsevier Masson SAS, 2005).
48. Jurgenson, C. T., Begley, T. P. & Ealick, S. E. The structural and biochemical foundations of thiamin biosynthesis. **78**, 569–603 (2009).
49. Eggersdorfer, M. *et al.* One hundred years of vitamins-a success story of the natural sciences. *Angew. Chem. Int. Ed. Engl.* **51**, 12960–90 (2012).
50. Schyns, G., Geng, Y., Barbosa, T. M., Henriques, A. & Perkins, J. B. Isolation and Characterization of New Thiamine-Deregulated Mutants of Bacillus subtilis. *J. Bacteriol.* **187**, 8127–8136 (2005).
51. Cardinale, S., Tueros, F. G. & Sommer, M. O. A. Genetic-metabolic coupling for targeted metabolic engineering. *Cell Rep.* **20**, 1029–1037 (2017).
52. Rugbjerg, P. & Sommer, M. O. A. Overcoming genetic heterogeneity in industrial fermentations. *Nat. Biotechnol.* 1 (2019).
53. Berkovitch, F., Nicolet, Y., Wan, J. T., Jarrett, J. T. & Drennan, C. L. Crystal

structure of biotin synthase, an S-adenosylmethionine-dependent radical enzyme. *Science (80-.).* **303**, 76–79 (2004).
54. Bitoun, J. P., Wu, G. & Ding, H. Escherichia coli FtnA acts as an iron buffer for re-assembly of iron-sulfur clusters in response to hydrogen peroxide stress. *Biometals* **21**, 693–703 (2008).
55. Yang, L. *et al.* Cellular responses to reactive oxygen species can be predicted on multiple biological scales from molecular mechanisms. *Proc. Natl. Acad. Sci.* 201905039 (2019).
56. Rajagopalan, S. *et al.* Studies of IscR reveal a unique mechanism for metal-dependent regulation of DNA binding specificity. *Nat. Struct. Mol. Biol.* **20**, 740–747 (2013).
57. Delano W. L. The PyMOL Molecular Graphics System. *DeLano Sci. LLC, San Carlos, CA* (2002).
58. Wang, H. H. & Church, G. M. *Multiplexed genome engineering and genotyping methods: Applications for synthetic biology and metabolic engineering*. *Methods in Enzymology* **498**, (Elsevier Inc., 2011).
59. Bonde, M. T. *et al.* MODEST: A web-based design tool for oligonucleotide-mediated genome engineering and recombineering. *Nucleic Acids Res.* **42**, 408–415 (2014).
60. Cavaleiro, A. M., Kim, S. H., Seppälä, S., Nielsen, M. T. & Nørholm, M. H. H. Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. *ACS Synth. Biol.* **4**, 1042–1046 (2015).
61. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–5 (2009).
62. Demoll, E. & Shive, W. Assay for Biotin in the Presence of Dethiobiotin with Lactobacillus plantarum. *Anal. Biochem.* **158**, 55–58 (1986).
63. Genee, H. J. *et al.* Functional mining of transporters using synthetic selections. *Nat. Chem. Biol.* **12**, 1015–1022 (2016).
64. Herbert, A. A. & Guest, J. R. Turbidimetric and Polarographic Assays for Lipoic acid using Mutants of Escherichia coli. *Methods Enzymol.* **18**, 269–272 (1970).
65. Schmidt, A. *et al.* The quantitative and condition-dependent Escherichia coli proteome. *Nat. Biotechnol.* **34**, 104–110 (2015).
66. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).

CHAPTER II - SUPPLEMENTARY

IMPROVED BIOTIN, THIAMINE, AND LIPOIC ACID BIOSYNTHESIS BY ENGINEERING THE GLOBAL REGULATOR ISCR

Anne P. Bali[1,2], David Lennox-Hvenekilde[1,2], Nils Myling-Petersen[2], Josi Buerger[2], Bo Salomonsen[2], Luisa S. Gronenberg[2], Morten O. A. Sommer[1] & Hans J. Genee[2*]

Affiliations
[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark. [2]Biosyntia ApS, Fruebjergvej 3, DK-2100, Denmark

***Corresponding author:** hjg@biosyntia.com

Table S1: Strains used and Constructed in this Study

Strain ID	Description
BS750	BW25113, Δ*thiP, thiL**
BS1011	BW25113, Δ*bioB*
BS1013	BW25113
BS1353	BW25113, Δ*bioB*, *iscR H107Y*
BS1375	BW25113, Δ*bioB*, *iscR C92Y*
BS1377	BW25113, Δ*bioB*, *iscR L15F*
BS1912	BW25113, Δ*lipA*
BS1918	BS1912 with pBS451
BS2114	BW25113, Δ*lipA, iscR C92Y*
BS2615	BW25113, Δ*bioB*, *iscR L15F, H107Y*
BS2616	BW25113, Δ*bioB*, *iscR L15F, C92Y*
BS2617	BW25113, Δ*bioB*, *iscR L15F, C92Y, H107Y*

Table S2: Plasmids used and Constructed in this Study

Plasmid ID	Brief description	Features
pGEN49	Empty backbone plasmid with apFAB46 promoter	KanR, ColE1
pGEN50	Empty backbone plasmid with apFAB71 promoter	KanR, ColE1
pBS116	Constitutive expression of thiCEFSGH from apFAB46 (897.64 AU)	KanR, ColE1
pBS117	Constitutive expression of thiMD from apFAB71 (866.30)	KanR, ColE1
pBS136	MAGE system	AmpR, ColE1
pBS140	Constitutive expression of thiamine biosynthesis genes	KanR, ColE1
pBS412	Inducible *bioB* expression plasmid	KanR, Sc101
pBS430	pBS412 version with mutated *bioB* (early frameshift mutation)	KanR, Sc101

pBS451	Zeocin resistance plasmid	ZeoR, p15A
pBS479	pBS412 with AmpR instead of KanR	AmpR, Sc101
pBS936	Native *bio*-operon from *E. coli* with type 9 mutation in *bioO* regulator site[93]	pBR322, TetR
pBS1037	Inducible *lipA* expression plasmid, constitutive *aceF*	AmpR, p15A
pBS1055	GFP expression plasmid. Backbone for libraries.	KanR, p15A
pBS1183-L	Constitutive expression of *iscSUA-hscBA-fdx-iscX* from promoter library	KanR, p15A
pBS1185-L	Constitutive expression of *sufABCDSE* from promoter library	KanR, p15A
isc-D1	Constitutive expression of *iscSUA-hscBA-fdx-iscX* from apFAB310 (156.73 AU)	KanR, p15A
suf-D6	Constitutive expression of *sufABCDSE* from apFAB206 (99.29 AU)	KanR, p15A
suf-B5	Constitutive expression of *sufABCDSE* from apFAB251 (30.07 AU) with a nucleotide deletion just downstream promoter	KanR, p15A
suf-D7	Constitutive expression of *sufABCDSE* from apFAB251 (30.07 AU)	KanR, p15A

Table S3: Primer and oligos used and Constructed in this Study

Oligo ID	Sequence	Features
moBS19	TTCGGAAATATCAGCCAACGGTACCGGGCC CGCTcaAGAGTTGAGCGCAACGTCAAGCAT TGCGGTCACGGCATAGCGCCCTTTAGATGT	*iscR* KO introduction by MAGE
moBS20	GAAACCAGACCATTTTTACGCAGACGGGAA AACAGTTGTTCCAGATAAGAAAGGGAAATT CCCTGACGTcaGGAAATATCAGCCAACGGT	*iscR* KO introduction by MAGE
moBS134	CATTGCGGTCACGGCATAGCGCCCTTTAGA TGTCAGTCTCATGTATTABTSCGCCTCAAA CTCGCCCCTGCCCGGGGTTTTTTATTGTAA	*iscR* RBS library introduction by MAGE
moBS150	GCATTTATCGCCGCCCTGGCAGCCGCCTTT ACCCTGAtAACGGGTGGCATCTACAGATTC GTCAACGGCGCTAATTACTTCGCCAACGGC	*iscR* C92Y introduction by MAGE
moBS151	TCGCCTAAAGTAATGTTGTTGAGAAAACCG GTGAGACGGTCGCTCAAATCACGCCACAGC GCGTaGGTCAGGCATTTATCGCCGCCCTGG	*iscR* H107Y introduction by MAGE

moBS295	GAAATTCCCTGACGTTCGGAAATATCAGCC AACGGTACCGGGCCCGCTTCAGAGTTGAGC GCAACGTCAAaCATTGCGGTCACGGCATAG	*iscR* L15F introduction by MAGE
oGEN182	AAATTCGCGAGTTCCACTAAGA	Ampl. of pBS117 for assembly of thiamine plasmid
oGEN184	AAACCTCTTTATGTTGCAGTCG	Ampl. of pBS117 for assembly of thiamine plasmid
oGEN227	CCGCTTACCTCCTATTACGTCACTAATAGC TAAGGAGGTAAATATGTCTGCAACAAAACT GACCCGCC	Gibson overlap and Strong RBS for *thiC*
oGEN264	CGGAGGCCTTTCTGACAACACACGAGTAAT AATTCATAGTCTTTGCGAGGCG	Gibson overlap and *thiC* -operon reverse
oGEN265	ATTATTACTCGTGTGTTGTCAGAAAGGCCT CCG	Gibson overlap for opening empty backbones
oGEN266	ATTAGTGACGTAATAGGAGGTAAGCGGTGG G	Gibson overlap for opening empty backbones
oGEN289	AAATCGCTTTGATTTGGGC	Ampl. of pBS116 for assembly of thiamine plasmid
oGEN290	AAATCCTCAGGTGCCCTCACGG	Ampl. of pBS116 for assembly of thiamine plasmid
oBS421	CGGAGGCCTTTCTGACAACACACGAGTAAT AATTCACCACCAGGCGTGGAAG	Gibson overlap and Strong RBS for *thiM*
oBS422	GCTTACCTCCTATTACGTCACTAATAGCTA AGGAGGTAAATATGCAAGTCGACCTGCTGG	Gibson overlap and *thiM* -operon reverse
oBS666	GCAATCCATCTTGTTCAATCATGC	Colony PCR for metagenomic library inserts
oBS955	GCTCAGGTCGATAGGCAGTT	Colony PCR for *isc*-insert in library

oBS1468	AGCTTCAUATGGTCCACAGGACACTCGTTG CTTTCACCATGCGTAAAGCAATC	Ampl. of pBS1055/pBS0992 backbone
oBS1587	acaaGCTUTCCCACTCCCCCTAGAGGC	Ampl. of pBS0910 backbone (*lipA*)
oBS1591	AGAGATACUAGTTAGAGAACCTCCGATTTG AATC	Ampl. of pBS0910 backbone (*lipA*)
oBS1670	GCATCCGACAACAGGTACAA	Ampl. of genomic *iscR*
oBS1745	AAGCttgUCTACTTGTACAGCTCGTCCATG CCG	Colony PCR for *suf*-insert in library
oBS1991	GTACGACCGTGTTTACGGAG	Ampl. of genomic *iscR*
oBS2197	AAGCTTGUTTACTTAACTTCCATCCCTTTC G	Ampl. of *lipA*
oBS2200	AGTATCTCUGGAGCGGCGAAGACTAAGGAG GTAAGAAATGAGTAAACCCATTGTGATGG	Ampl. of *lipA* with synthetic RBS
oBS2225	ATTCAAAUAGGTAATATGGAATTCTTCGGG ATATAAGTAGGGGCGGCATGGCTATCGAAA TCAAAGTACC	Ampl. of truncated *aceF* with synthetic RBS
oBS2234	aCTGCCaUCCACATACAATACGAGCCGGAT GATTAATTAAAAACACCAtTTGCCAGCTGG	Ampl. of pBS1055 backbone
oBS2235	aCTGCCaUCCACACATTATACGAGCCGGAT GATTAAGGAGGAACACCAtTTGCCAGCTGG	Ampl. of pBS1055 backbone
oBS2236	aCTGCCaUCCACAGAACCTAAGAGCCGGAT GATTAATTGTCAACACCAtTTGCCAGCTGG	Ampl. of pBS1055 backbone
oBS2237	aCTGCCaUCCACACACGCTAAGAGCCGGAT GATTAATTGTCAACACCAtTTGCCAGCTGG	Ampl. of pBS1055 backbone
oBS2238	aCTGCCaUCCACACATTATACGAGCCGATG ATTAACAGTAAACACCAtTTGCCAGCTGGG	Ampl. of pBS1055 backbone
oBS2239	aCTGCCaUCCACAGAAGATAGGAGCCGGAT GATTAATTGTCAACACCAtTTGCCAGCTGG	Ampl. of pBS1055 backbone
oBS2240	aCTGCCaUCCACAGACACTACGAGCCGGAT GATTAATTGTCAACACCAtTTGCCAGCTGG	Ampl. of pBS1055 backbone
oBS2241	aCTGCCaUCCACACATTATACGAGCCGATG ATTAAGAGGCAACACCAtTTGCCAGCTGGG	Ampl. of pBS1055 backbone
oBS2242	aCTGCCaUCCACAAACACTACGAGCCGGAT GATTAATTGTCAACACCAtTTGCCAGCTGG	Ampl. of pBS1055 backbone

oBS2243	aCTGCCaUCCACACATTATACGAGCCGATG ATTAAGATTCAACACCAtTTGCCAGCTGGG	Ampl. of pBS1055 backbone
oBS2244	aCTGCCaUCCACAAACATTACGAGCCGGAT GATTAATTGTCAACACCAtTTGCCAGCTGG	Ampl. of pBS1055 backbone
oBS2245	aCTGCCaUCCACACATTATACGAGCCGGAT GATTAATACGCAACACCAtTTGCCAGCTGG	Ampl. of pBS1055 backbone
oBS2246	aCTGCCaUCCACATAAGTTACGAGCCGGAT GATTAATTGTCAACACCAtTTGCCAGCTGG	Ampl. of pBS1055 backbone
oBS2247	aCTGCCaUCCACACATTATACGAGCCGGAT GATTAAATATCAACACCAtTTGCCAGCTGG	Ampl. of pBS1055 backbone
oBS2248	aCTGCCaUGGCTGTAAGTATCCTATAGGTT AGACTTTATGTCAACACCAtTTGCCAGCTG GG	Ampl. of pBS1055 backbone
oBS2249	aCTGCCaUGAATCTATTATACCTATAGGTT AGACTTTATGTCAACACCAtTTGCCAGCTG GG	Ampl. of pBS1055 backbone
oBS2250	aCTGCCaUGAATCTATTATAGGTACAAAAA GATGCGATGTCAACACCAtTTGCCAGCTGG	Ampl. of pBS1055 backbone
oBS2259	AtGGCAGUGGCTAgcggccgcgaagtgagg taaatcgatggacatgcattc	Ampl. of *suf*-operon
oBS2260	ATGAAGCUTttagctaagtgcagcggcttt gg	Ampl. of *suf*-operon
oBS2264	ATGAAGCUTttattcggcctcgtccagcca	Ampl. of *isc*-operon
oBS2265	AtGGCAGUGGCTAgcggccgcAGCTAAGGA GGTAAATatgaaattaccgatttatctcga ctactccgc	Ampl. of *isc*-operon
oBS2273	AGCCCCUGCCTCCACAgACAcTAC	Ampl. of pBS0992 backbone
oBS2274	AGGGGCUAGGTAATATGGAATTCTTCGGGA TATAAGTAGGGGCGGCATGGCTATCGAAAT CAAAGTACC	Ampl. of truncated *aceF*
oBS2277	ATGAAGCUTTACGCAGGAGCTGCC	Ampl. of truncated *aceF*

Table S4: Proteins with Significantly Changed Expression Levels. Overview of relative expression of proteins with significantly changed expression levels based on ANOVA analysis with Bonferoni-correction and 95% confidence interval on proteomics data from IscR wild type reference strain (Ref) with BioB expression from genome, IscR wild type strain with BioB overexpression plasmid induced to toxicity limit (WT, 0.025 mM IPTG), IscR mutant strains with BioB overexpression plasmid induced similar to WT (MutL, 0.025 mM IPTG), IscR mutant strain with BioB overexpression plasmid induced fully (MutH, 1 mM IPTG).

Protein name	Relative expression level			
	Ref	WT	MutL	MutH
AceA				
AckA				
AcnA				
AcnB				
ACS				
AldA				
Amn				
ArcA				
ArgS				
AsnS				
AtpD				
AtpG				
BamE				
Bcp				
BioA				
BioB				
BioD				
BtuB				
CarB				
CirA				
ClpB				
CodA				
CorA				
CpdB				
CycA				
CysA				
CysS				
DapD				
DegP				
DppA				
DppD				
DppF				
Dps				
EcnB				
EfeO				
EntC				
ErpA				
EvgA				
ExbB				
Fdx				
FruB				

Protein name	Relative expression level			
	Ref	WT	MutL	MutH
FtnA				
FumA				
GalE				
GatC				
GatY				
Gcd				
GdhA				
GhrB				
GlgX				
GltA				
GltB				
GltD				
GpsA				
GsiA				
GyrA				
HdeA				
HdeB				
HemL				
HflD				
HisF				
HisJ				
HrpA				
HscA				
HscB				
HtpX				
IbpA				
Icd				
IlvC				
IscA				
IscS				
IscU				
IscX				
IspG				
IaaA				
KdgK				
KdsB				
LdtB				
LeuC				
LivF				
LivK				
LonH				

Protein name	Relative expression level			
	Ref	WT	MutL	MutH
LpcA				
MalE				
MalP				
ManX				
ManY				
MetA				
MetB				
MetF				
MetL				
MetN				
MetR				
MmuP				
MoeA				
MsbA				
MsrB				
MsyB				
NadA				
Ndh				
NhaB				
NlpA				
NuoA				
NuoB				
NuoC				
NuoG				
NuoI				
NusB				
PanB				
Pck				
PepB				
PepT				
Pgi				
PheA				
PpsA				
PrfA				
ProV				
ProW				
ProX				
PstS				
Pta				
PurC				
RapA				

Protein name	Relative expression level			
	Ref	WT	MutL	MutH
RbsB				
RibF				
Rnb				
RodZ				
SdhA				
SdhB				
SdhC				
SdhD				
SecF				
SodB				
SodC				
SseB				
SstT				
StpA				
SucA				
SucB				
SucC				
SucD				
SufA				
SufC				
SufD				
TalA				
TcyP				
TdcE				
TdcF				
TehB				
ThiC				
ThiE				
ThiF				
ThiG				
Tig				
TktB				
TolB				
TopA				
Tsx				
TyrA				
WrbA				
WaaC				
YagU				
YajQ				
YbaK				

Protein name	Relative expression level			
	Ref	WT	MutL	MutH
YbaY				
YbhC				
YbjQ				
YceH				
YceI				
YciI				
YdfZ				
YdhF				
YeaG				
YebV				
YegP				
YcF				
YhgF				
YhjE				
YqjD				
YqjI				
Zwf				

Supplementary Note S1

Minimal MOPS media as outlined in table S5, was used for all production experiments. K_2HPO_4 solution was prepared by autoclaving 23 g of dibasic K_2HPO_4 in 1 L de-ionized water. Glucose solution was prepared by dissolving and autoclaving 100 g anhydrous dextrose in 500 mL of de-ionized water. A 500 X vitamin solution was prepared by mixing individual sterile filtered vitamins as described in table S6 and aliquots were stored at -20° C. KOH solution was prepared by dissolving and autoclaving of 1.247 grams of KOH (90% Reagent Grade, Sigma-Aldrich) in 1 L of de-ionized water. 10 x MOPS was prepared by mixing MOPS buffer and tricine in ~300 mL water and adjusting to PH 7.4 with 10 M KOH in a 1L beaker with stir bar. $FeSO_4$ was added followed by remaining supplements in table S7. Finally, volume was adjusted to 1 L with water.

Table S5. minimal MOPS medium (mMOPS).

Ingredient	Volume (mL) for 1000 mL
10 X MOPS (see below, table S6)	100
0.132 M K2HPO4	10
500 X Vitamin solution (see below, table S5)	2
20% D-glucose	10
Sterile de-ionized water	up to 1000 mL

Table S6. 500 x vitamin solution.

Vitamin	Mass	Volume	
calcium pantothenate (C8731, Sigma-Aldrich)	0.238 g	25 mL	dissolve in de-ionized water

p-aminobenzoic acid (A9878, Sigma-Aldrich)	0.069 g	25 mL	dissolve in 0.02 M KOH
p-hydroxybenzoic acid (240141, Sigma-Aldrich)	0.069 g	25 mL	dissolve in 0.02 M KOH
2,3-dihydroxybenzoic acid (126209, Sigma-Aldrich)	0.077 g	25 mL	dissolve in 0.02 M KOH

Table S7. 10x MOPS stock solution (1 L)

Ingredient	**Mass or molarity**	**Volume (**dissolved in water)
MOPS buffer	83.72 g	
Tricine	7.17 g	
$FeSO_4$	0.028 g	10 mL
NH_4Cl	1.9 M	50 mL
K_2SO_4	0.276 M	10 mL
$CaCL_2 \bullet 2H_2O$	1 M	5 µl
$MgCL_2$	2 M	2.625 mL
NaCl	5 M	100 mL
Micronutrient stock		1 mL
$(NH_4)_6Mo7O_{24} \bullet 4H_2O$	0.00018 g	
H_3BO_3	0.00124 g	
$CoCl_2$	0.00036 g	
$CuSO_4$	0.00012 g	
$MnCl_2$	0.0008 g	
$ZnSO_4$	0.00014 g	

Supplementary Note S2

A Dual Life-Death Selection System for Improved BioB Expression

We combined an inducible expression plasmid of *bioB* with a BioB knockout background strain to establish a robust selection system for improved BioB expression. A T5LacO promoter for *bioB* expression combined with plasmid expressed LacI was used in order to establish a gradient induction of BioB expression through externally added IPTG. Biotin is essential for growth of *E. coli* in trace amounts and by using a *bioB* knockout strain as selection host, we almost eliminated false positives as a result of plasmid rearrangements causing reduced or eliminated expression of BioB from the selection plasmid. Characterization of the dual life-death selection system with various induction levels showed that we were able to significantly reduce the growth rate of our selection host with increased amounts of IPTG (**Fig. S1A**) from a growth rate of 0.60 H^{-1} to almost no growth at all in a timeframe of 20 hours. At 0.01 μM externally added IPTG the growth was slightly hampered with a growth rate of 0.57 H^{-1}, whereas adding 0.1 μM IPTG almost killed the cells. Furthermore, the selection system can be seen to have a leaky expression of BioB, as the cells are able to grow without any induction (0 μM IPTG). To ensure that the toxicity arose from the increased expression of BioB rather than other aspects e.g. the externally added IPTG, we compared the growth to a wild type control strain with a mutated selection plasmid, resulting in expression of a non-functional BioB enzyme due to an early frameshift mutation (**Fig. S1B**) deleting the 6^{th} nucleotide of *bioB*. No differences in growth rate or lag time was observed across a gradient of IPTG, ranging from 0 to 0.1 mM, of the control strain, indicating that the reduced growth in the selection host with the dual life-death selection system, was indeed caused by increased expression of BioB. By comparing the growth of the control strain with the strain harboring the selection system, the leaky, basal BioB expression

from the selection plasmid at no induction, are comparable, indicating a low expression level of BioB without toxic effects on growth.

We further tested whether the dual life-death selection system for improved BioB expression, could function as a solid plate-based selection in addition to differentiating growth in liquid. We plated the *bioB* knock-out strain holding the selection system 10^4 to 10^7 cells plated on minimal media plates with 0.1 g/L DTB on a gradient of IPTG ranging from 0 to 1 mM (**Fig. S2**) to identify optimal selection conditions. It was decided to use 0.1 mM IPTG and 10^5 cells per plate for further experiments.

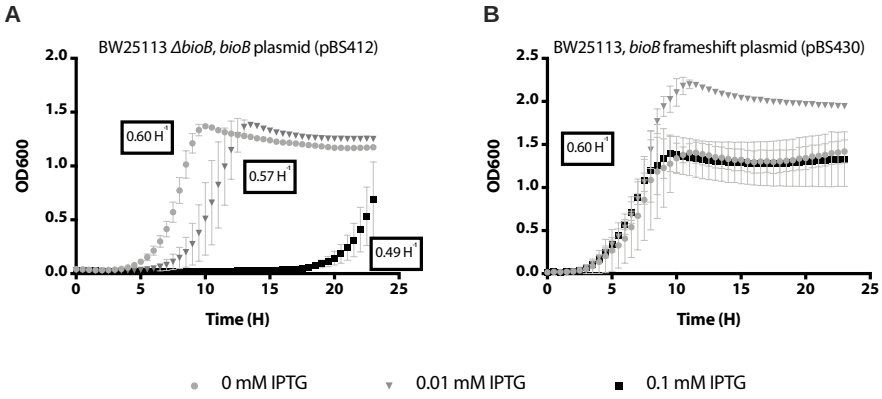


Figure S1: Establishing a Selection System for Strains with Improved Tolerance for BioB Expression. Toxicity observed when overexpressing BioB was used as foundation for a selection assay applied to select for colonies with improved tolerance for BioB expression. **A)** Growth curves for strains expressing a functional BioB protein and **B)** a non-functional BioB protein with an early frame shift mutation in response to IPTG induction

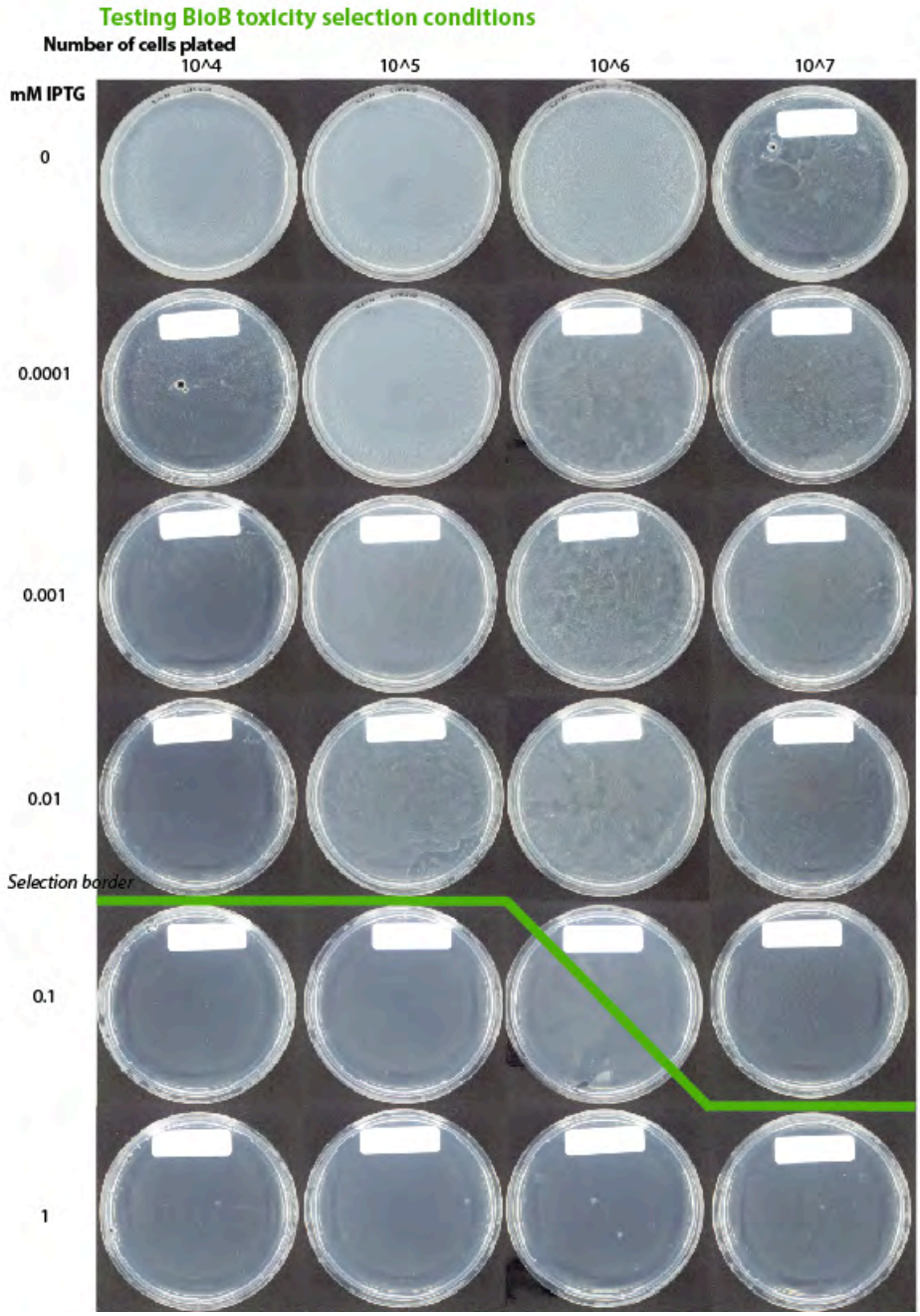


Figure S2: Establishing Plate-Based Selection Conditions. Growth on mMOPS agar plates with kanamycin (50 µg/ml), DTB (0.1 g/L) and IPTG (0, 10^{-4}, 10^{-3}, 0.01, 0.1 or 1 mM) of 10^4-10^7 *E. coli* BW25113 Δ*bioB* cells holding a selection plasmid with IPTG

inducible BioB expression after 17 hours of incubation at 37ºC. The selection border, separating growth from clean plates, is emphasized in green line.

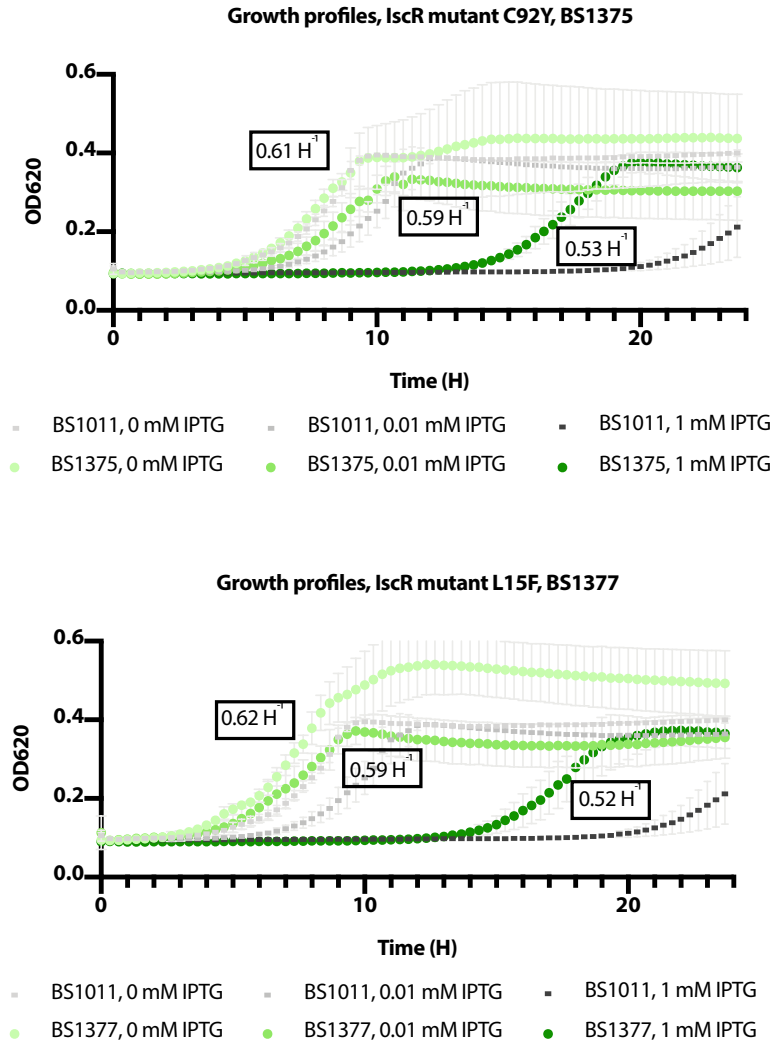


Figure S3: Growth Rate Evaluation of IscR Mutants and Wildtype Strain. Growth rate data on IscR mutant strain C92Y (top) and L15F (bottom) in green circles and IscR wildtype strain in grey lines. Darker color indicates increasing IPTG concentration, n=4 and error bars represents standard deviation.

CHAPTER III
METAGENOMIC MINING OF TRANSPORTERS WITH AFFINITY FOR BIOTIN SULFOXIDE

Anne P. Bali[1,2]**, Nils Myling-Petersen**[2] **& Hans J. Genee**[2*]

Affiliations
[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark. [2]Biosyntia ApS, Fruebjergvej 3, DK-2100, Denmark

***Corresponding author:** hjg@biosyntia.com

Manuscript in preparation

Abstract

Biotin is an industrially important molecule which is synthesized naturally by microorganisms but has the risk of being oxidized into biologically inactive analogs; biotin sulfoxide (BSX) and biotin sulfone (BSN). Currently, biotin is industrially being produced exclusively by chemistry, due to lack of successful developments of economically competitive fermentation processes. High product purity in fermentation broth is of great interest to decrease downstream processing costs. Here, we establish, characterize and use an auxotrophic based selection to facilitate mining of metagenomic libraries for protein factors improving conversion of extracellular BSX to biotin. We find that import of BSX is a limiting factor and identify several transporter proteins with affinity for BSX, enabling cells to grow nearly as efficiently on BSX as biotin when added extracellularly. We suggest that such transporters, if introduced into a biotin cell factory, may help improve the purity of biotin in the fermentation broth, by importing BSX to enable intracellular conversion back into biotin.

Introduction

Biotin is one of the highest priced vitamins and thus a commercially interesting product. It has an estimated market size of 200 mEUR in 2018 [1], and is currently being produced by a polluting, multistep chemical synthesis [2]. Several companies and research groups has explored the possibility of establishing a greener, more sustainable alternative through microbial production of biotin [3–6]. Unfortunately, these previous efforts have not been able to achieve sufficient titers to make an economical feasible cell factory. The many failed attempts to create a cost-competitive cell factory for biotin production reflects the high complexity involved in this pathway, as reviewed previously in the thesis.

In addition to the challenges of engineering a cost-competitive cell factory for biotin production as reviewed previously in the thesis, the

molecule itself is susceptible to oxidation of its ring-bound sulfur (**Fig. 1**). In the first oxidation step, biotin is oxidized to the biologically inactive analog BSX, with oxygen double-bound to the sulfur, while a second oxidation step results in BSN, that has an additional double bonded oxygen to the sulfur atom. The oxidation occurs in the presence of hydrogen peroxide or other types of ROS [7]. In *E. coli*, BSX can be enzymatically salvaged back to biotin by biotin sulfoxide reductase (BisC) and reduced thioredoxin [8] in the presence of a molybdenum derived cofactor [9]. No salvage pathway is reported for BSN and seems to be a dead-end molecule (**Fig. 1**).

When establishing a commercial cell factory, it is in general advisable to produce a pure product in the fermentation broth, as the downstream processing costs can easily constitute up to 70% of the total manufacturing costs [10]. The importance of having a pure product is exemplified in the industrial production of cobalamin, which is mainly produced as a synthetic version; cyano-cobalamin [11]. Cobalamin, or vitamin B_{12}, is the most complex cofactor in nature and is exclusively synthesized by prokaryotes [12]. Microbial production of cobalamin involves approximately 30 enzymatic steps and leads to a mixture of different analogous, some of which are inactive for humans[11,13]. The main active, naturally occurring derivates are adenosyl-cobalamin ($AdoB_{12}$), methyl-cobalamin (MeB_{12}) and hydroxy-cobalamin (OHB_{12}), all sharing the core structure of cobalamin but with different upper ligands. Cyano-cobalamin is a stable and inexpensive synthetic form, commonly used as end-product to achieve a high degree of purity [11,14].

A similar high importance of purity is to be expected for biotin. A Scientific Opinion from European Food Safety Assembly recommend that the purity of biotin should be >98.5% with substance related impurities <0.5% when used as a feed additive [15]. The requirements may be even higher for other sectors, encouraging a high initial purity of biotin in the

fermentation broth. Decreasing impurities of structurally related molecules with similar chemical properties in the fermentation broth may be of particular interest, as downstream processing through filtration and ion exchange chromatography of such molecules can prove extremely challenging. The molecular properties of BSN, which has previously been reported to act as a biotin-antimetabolite [16] emphasizes the importance of avoiding such impurities. It is likely that BSX may have a similar, biotin-antimetabolite effect in organisms without a BisC-homolog.

Here, we evaluate the influence of BisC (over-)expression on the capabilities of *E. coli* cells to grow on oxidized biotin compounds, in order to improve the purity of biotin produced by future cell factories. We find that expression level of BisC does not alter growth abilities on extracellular BSX significantly. We then establish and characterize an auxotrophic based selection and develop and execute a workflow to mine metagenomic libraries for accessory proteins improving growth ability on oxidized biotin vitamers and thus the (re-)conversion into biotin. When applying the selection scheme to metagenomic libraries, we identify a number of different transporters, both known and unknown, from various organisms. The majority of the transporters allowed for improved growth on BSX but not biotin. Finally, we discuss which questions should be addressed including suggestions of specific experimental setups, in order to evaluate the usability of BSX transporters to increase biotin purity in fermentation broth.

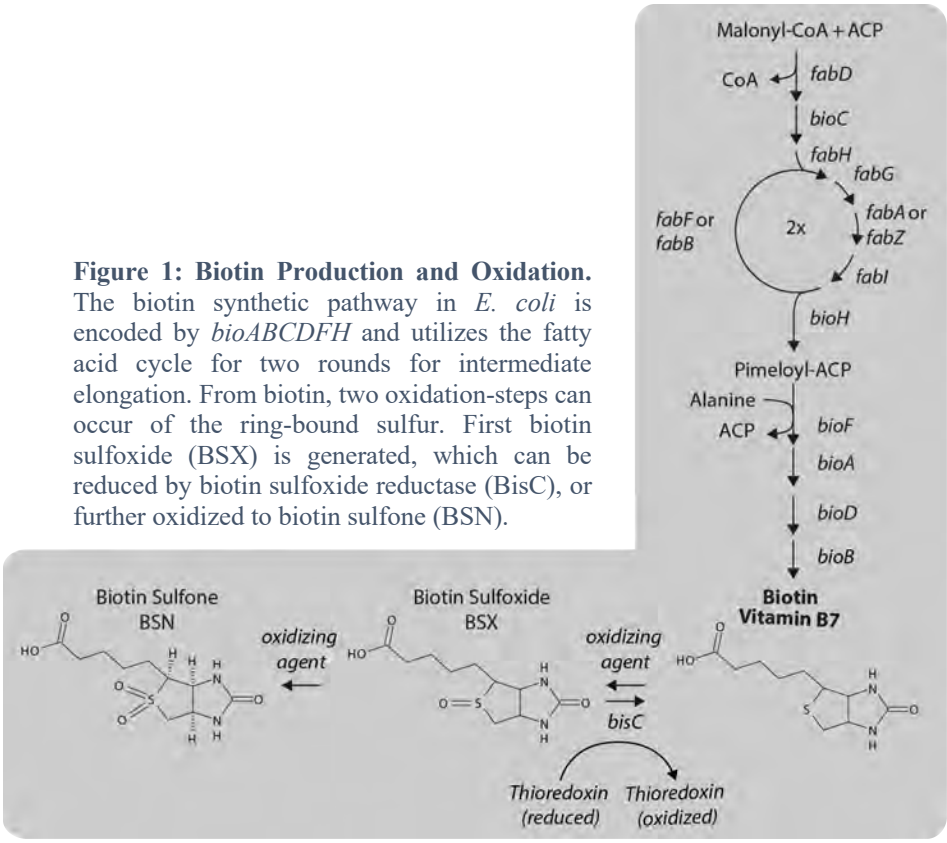


Figure 1: Biotin Production and Oxidation. The biotin synthetic pathway in *E. coli* is encoded by *bioABCDFH* and utilizes the fatty acid cycle for two rounds for intermediate elongation. From biotin, two oxidation-steps can occur of the ring-bound sulfur. First biotin sulfoxide (BSX) is generated, which can be reduced by biotin sulfoxide reductase (BisC), or further oxidized to biotin sulfone (BSN).

Results

Overexpression of BisC does not Improve Growth on Biotin Sulfoxide

With a goal of eliminating oxidized biotin compounds in the fermentation broth of biotin cell factories, we investigated how BisC overexpression affected the growth of *E. coli* on extracellularly added BSX and BSN. We hypothesized that overexpression of the native *E. coli* BisC could improve catalysis of BSX to biotin. Two plasmids with different origins of replication, expressing the native *E. coli bisC* from a strong constitutive promoter, were introduced into a biotin auxotrophic *E. coli* strain. The growth phenotypes of the resulting strains and a reference strain without additional BisC expression were evaluated in the presence

of biotin and oxidized vitamers. As expected, none of the strains were able to grow on BSN (**Fig. 2,** purple lines with squares). Comparable responses for growth on biotin were observed for all strains (**Fig. 2**, blue lines with circles). Growth on extracellular BSX between strains with BisC expression (**Fig. 2B, 2C** and **2D**, green lines with triangles) was also largely similar. We had expected that at optimal expression level of BisC the growth response on BSX would be the same as that on biotin. The fact that this was not the case could indicate that either BisC expression has an optimum for BSX to biotin conversion that was not found in our rough setup or something else is limiting. Tuning the BisC expression through a promoter or RBS library, or using an inducible promoter, would facilitate a finer expression grid and potentially reveal an ideal expression level for BisC.

As BisC requires a molybdenum cofactor [9] and a small, thioredoxin-like protein, for its biotin sulfoxide reductase activity [17], increased formation of the cofactor or expression of the coenzyme might be needed in order to explore the full potential of BisC overexpression. Yet another potential limiting factor to explain the lack of effect from overexpressing BisC is insufficient import of extracellularly added BSX. If the native level of BisC is already enough to keep up with the uptake of BSX, no additional effect would be expected from increased protein expression of BisC.

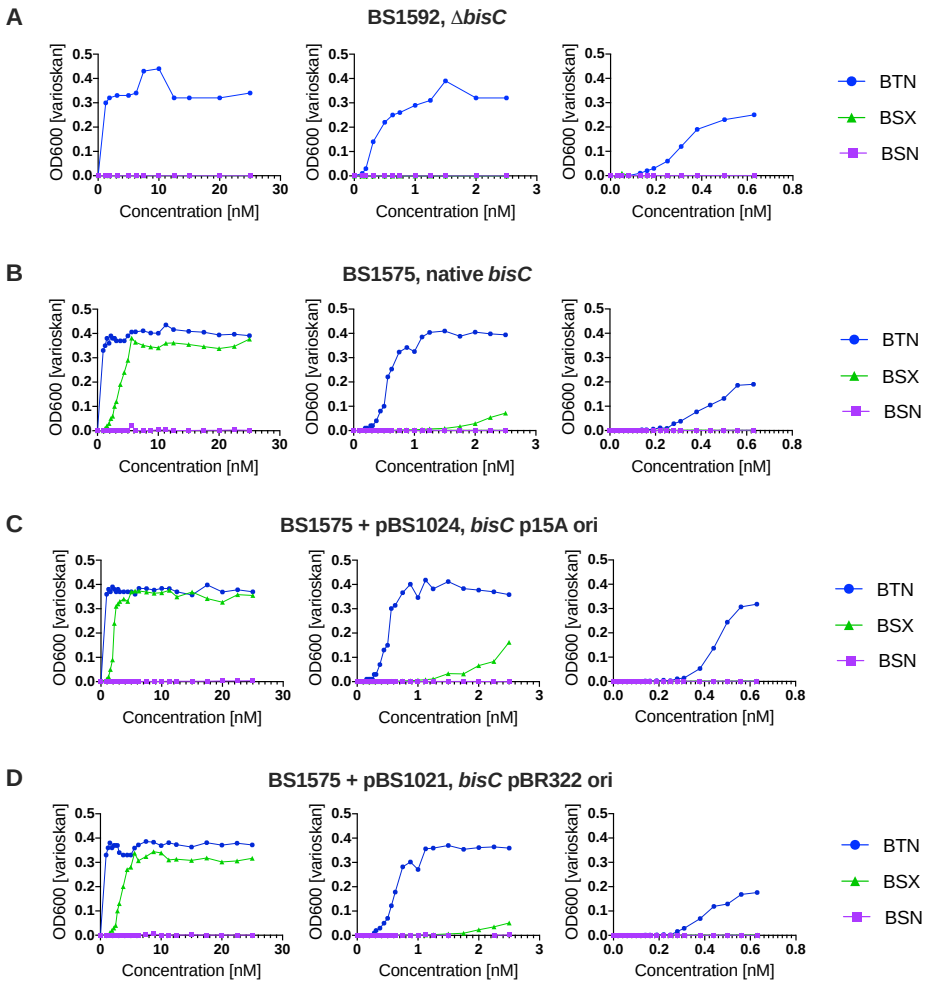


Figure 2: Phenotypic Growth Characterization of Strains with Various BisC Expression. Growth response in mMOPS media with a concentration gradient of biotin and oxidized vitamers for strains with different expression of BisC. **A)** BS1592: No BisC expression (knock-out strain). **B)** BS1575: Wildtype BisC expression. **C)** Additional BisC expression from medium copy-number plasmid (p15A, pBS1024). **D)** Additional BisC expression from high copy-number plasmid (pBR322, pBS1021). Growth response on biotin is shown in blue (BTN), biotin sulfoxide in green (BSX) and biotin sulfone in purple (BSN).

Establishing, Characterizing and Applying an Auxotrophic Selection for Metagenomic Mining of BisC "Accessory" Proteins

Since the overexpression of BisC alone did not give the expected efficient conversion of BSX to BTN we set out to elucidate if any protein-encoded factors might limit this reaction. We established an auxotrophy based selection. The goal was to apply the selection to mine metagenomic libraries for bottleneck-proteins (**Fig. 3A**). Working concentrations of BSX were chosen based on previous data (**Fig. 2A** and **2B)**. Selections for growth on BSN were also performed to allow for discovery of any potential biotin sulfone reductases in the libraries simultaneously.

Based on the initial requirement of biotin vitamer concentrations for growth in liquid minimal media, ranging from ~0.2 nM to ~0.6 nM for extracellular biotin (**Fig. 2A** and **2B**), characterization was carried out on mMOPS agar plates. The threshold concentrations were broadly similar to the ones identified in liquid media (small selection of plates can be seen in **Fig. 3B**).

Mining of four different metagenomic libraries using the plate-based selection, resulted in 19 potential hits. Some of these colonies grew on top of a faint lawn, due to the high number of cells plated. To confirm improved growth behavior of the selected colonies, they were re-streaked on plate conditions similar to where the colonies were first observed, along with a control strain harboring the backbone plasmid for the metagenomic library. Of the initial 19 colonies selected 11 managed to form single colonies when re-streaked.

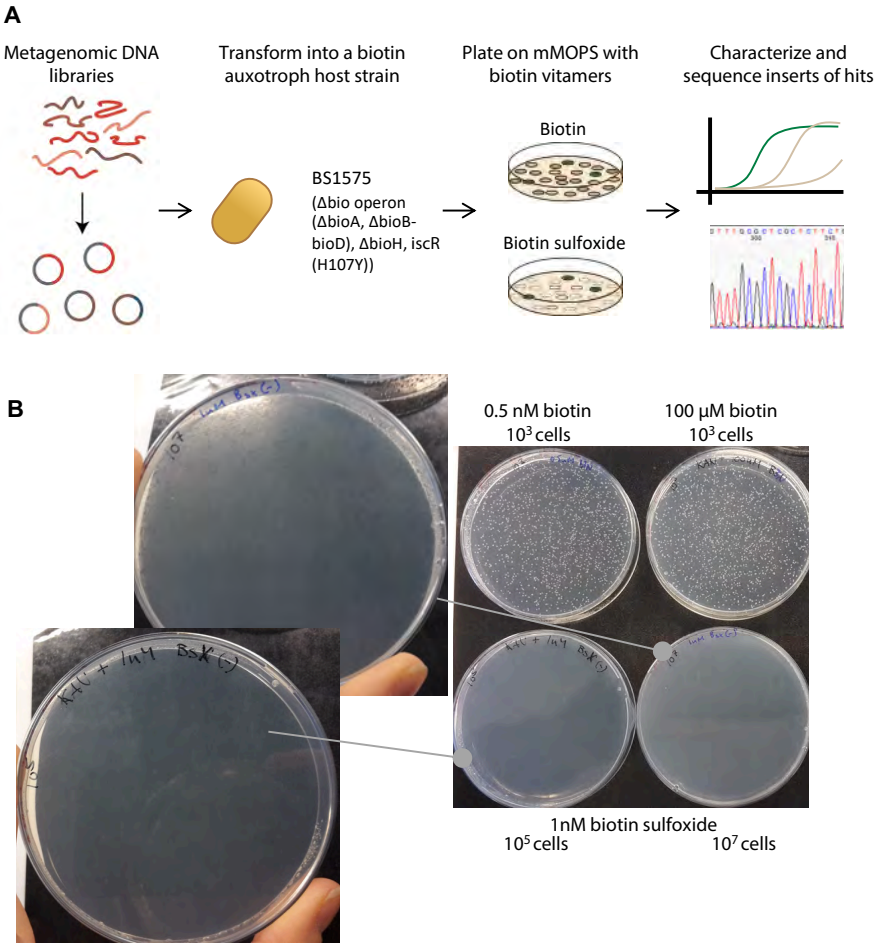


Figure 3: Auxotrophic Selections for Protein Factors Improving BSX to Biotin Conversion. A) General workflow for auxotrophic selections used to mine metagenomic libraries for protein factors improving BSX to biotin conversion. Metagenomic libraries are transformed into a selection host and plated on minimal media agar plates with low amounts of BSX. Selected hits are characterized, and the inserts sequenced. **B)** Phenotypic growth characterization of *E. coli* strain with genomic encoded BisC on mMOPS agar plates with different amounts of cells plated using biotin or biotin sulfoxide as biotin source.

Evaluation of Metagenomic Mining Hits

After confirming the growth phenotype on agar plates, the 11 identified hits were characterized for growth on biotin and oxidized vitamers in liquid minimal media. The characterization was done for two

biological replicates, followed by sequencing of the 9 inserts that repeatedly grew superior to the control on oxidized biotin vitamers. Sequencing revealed that two of the identified inserts were identical (col 6 and col 16. See supplementary **List S1** for all consensus sequences). Identification of open reading frames (ORF's) in the sequenced inserts were done using DeFUME [18]. From sequencing and BLASTp analysis, half (4 out of 8) of the metagenomic library inserts improving growth on oxidized biotin vitamers were found to be known or hypothetical transporters (**Fig. 4A**). The characterization of these hits in liquid media confirmed the improvement in growth on extracellular BSX and/or BSN over a control strain holding an empty plasmid (**Fig. 4B**, all data are shown in supplementary **Fig. S1**). Together these findings strongly suggest that the conversion of BSX to biotin was limited in uptake of extracellular BSX. This likely explains why changing the expression level of BisC did not affect growth response on extracellular BSX (**Fig. 2**).

Growth was observed on extracellular BSN for some strains expressing metagenomic inserts encoding hypothetical transporters (**Fig. 4B**). This suggests that the commercial BSN used contains an impurity of BSX, enabling strains with an efficient transporter and functional BisC to grow.

For some of the transporter inserts identified, for example the insert in col 6 (**Fig. 4B**), the growth response on extracellular BSX is close to the growth response on extracellular biotin. This observation suggests that the native expression levels of BisC as well as availability of the molybdenum cofactor and thioredoxin-like protein, are sufficient to efficiently convert all available BSX into biotin at the low concentrations required for growth complementation [19]. Alternative evaluation methods will be needed in order to determine to what extent the native BisC expression can efficiently convert BSX at higher concentrations, or whether import of BSX again will become a bottleneck.

A

Contig name	Significance replicate 1	Significance replicate 2	Insert description from best BLASTp hit	Coverage (%)	Hit identity (%)	Insert size (bp)	Coding region size (bp)
Col 1			**Likely a transporter** Hypothetical protein, *Sutterella wadsworthensis*	96	89.2	1892	231
			Hypothetical protein, *Sutterella wadsworthensis*, GO terms: transmembrane, transporter activity	99.8	99.7		1350
Col 6 **Col 16**			**Likely a transporter** Membrane protein, *Serratia liquefaciens*	99.7	99.7	1520	864
Col 10			**Likely a transporter** Transcriptional regulator *Candidatus Stoquefichus sp. KLE1796,*	93.3	33	1727	561
			ABC-type multidrug transport system, ATPase component, *Firmicutes bacterium*	96.7	66.4		921
Col 11			Ribokinase, *Geobacillus stearothermophilus*	82.7	41.3	1568	840
Col 14			Many weird mutations on plasmid				
Col 15			Hypothetical protein, unknown organism	93	46.8	1710	1399
Col 17			NlpC/p60 family protein, peptidase , *Pseudomonas*	100	99.3	894	411
			Hypothetical protein , *Mycobacterium tuberculosis*	98.1	84.9		183
Col 18			**Transporter** ABC transporter ATP-binding protein, *Achromobacter piechaudii.*	62.1	100	1542	441
			Branched-chain amino acid ABC transporter permease *Achromobacter piechaudii*	99.7	99.3		882

B

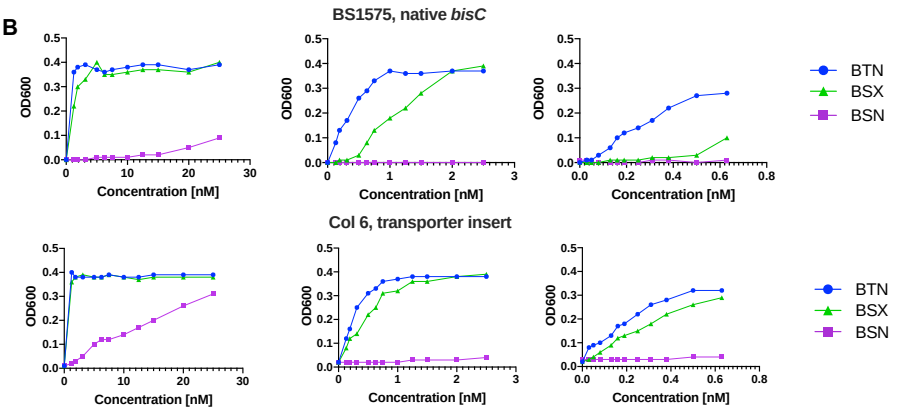


Figure 4: Characterization of Hits from Metagenomic Mining for BisC Auxiliary Proteins. A) Overview of contigs in metagenomic hits for protein factors improving BSX to biotin conversion based on DeFUME results. Color-coded growth response compared to control strain with empty plasmid; green: high, yellow: medium, red: low. **B)** Phenotypic growth characterization of control and a hit-example in liquid minimal media. Growth response on extracellular biotin is shown in blue (BTN), biotin sulfoxide in green (BSX) and biotin sulfone in purple (BSN).

Phylogenetic Analysis of Metagenomic Inserts

Since salvage of biotin from the environment through uptake is metabolically cheaper than *de novo* synthesis, biotin transport is widespread in prokaryotes despite the fact that most of them have the ability to synthesize biotin *de novo*. The best characterized and most prevalent type of biotin transporters in prokaryotes are BioY; energy coupling factor (ECF) proteins – a subclass of ATP-dependent importers [20,21]. The YigM biotin transporter in *E. coli* belongs to the carboxylate/amino acid/amine family of secondary transporters and is not structurally related to BioY [20]. In an attempt to further understand the selected metagenomic hits, we evaluated their phylogenetic relation to prokaryotic biotin transporters. In addition to YigM from *E. coli*, the phylogenetic context of the selected metagenomic hits were evaluated based on 108 BioY protein sequences from prokaryotes used for a genome context analysis study [21]. Predicted or confirmed BioY protein sequences represented prokaryotes and archaea from the major taxonomic groups of *Firmicutes, α-Proteobacteria, Actinobacteria, Cyanobacteria* and *Archaea*.

The phylogenetic analysis confirmed the distant relationship between YigM (blue text) and BioY (black text) (**Fig. 5,** supplementary **Fig. S2**). Additionally, it became evident that the selected hypothetical or confirmed transporters (green text) are not likely related to BioY transporters, based on their phylogenetic grouping far from the known and predicted BioYs. A similar trend is observed for the remaining metagenomic hits (magenta text), which are also located distantly from the BioY transporters.

Col 6, encoding a membrane protein from *Serratia liquefaciens*, is the identified insert clustering closest to a characterized sequence, being YigM from *E. coli*. A pairwise alignment between the two protein sequences reveals a sequence identity of 72.3% and a sequence similarity of 82.7% (see supplementary **List S2** for sequences). Amongst the list of

organisms found to encode the identified transporters with affinity for BSX, BioY annotated transporters were found in the genomes of all but *S. liquefaciens*, *S. wadsworthensis* and *C. stoquefichus*. The search for BioY proteins in the specific genomes were based on a BLASTp search with a representative BioY protein sequence from *Rhodobacter capsulatus* (supplementary **List S2**). It is interesting that *S. liquefaciens* does not have a BioY annotated in the genome while the metagenomic insert encoding a membrane transporter from this organism cluster closely with YigM from *E. coli*. However, whether this or any of the transporters identified in this study may constitute YigM-like proteins remains unclear. Further characterization of the transporters is needed, but it seems likely that the identified proteins have not previously been known for their ability to transport biotin sulfoxide.

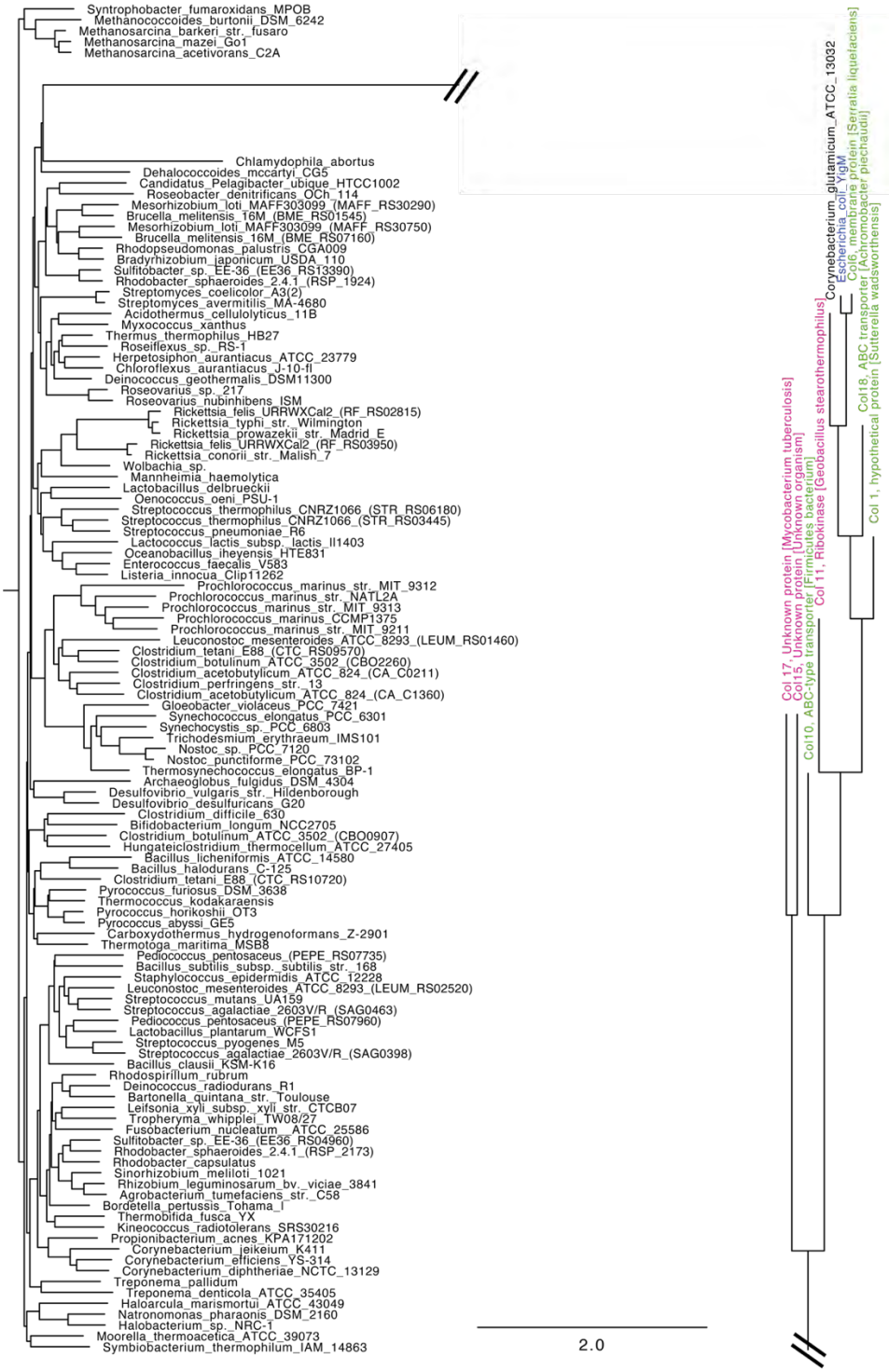

Syntrophobacter_fumaroxidans_MPOB
Methanococcoides_burtonii_DSM_6242
Methanosarcina_barkeri_str._fusaro
Methanosarcina_mazei_Go1
Methanosarcina_acetivorans_C2A
Chlamydophila_abortus
Dehalococcoides_mccartyi_CG5
Candidatus_Pelagibacter_ubique_HTCC1002
Roseobacter_denitrificans_OCh_114
Mesorhizobium_loti_MAFF303099_(MAFF_RS30290)
Brucella_melitensis_16M_(BME_RS01545)
Mesorhizobium_loti_MAFF303099_(MAFF_RS30750)
Brucella_melitensis_16M_(BME_RS07160)
Rhodopseudomonas_palustris_CGA009
Bradyrhizobium_japonicum_USDA_110
Sulfitobacter_sp._EE-36_(EE36_RS13390)
Rhodobacter_sphaeroides_2.4.1_(RSP_1924)
Streptomyces_coelicolor_A3(2)
Streptomyces_avermitilis_MA-4680
Acidothermus_cellulolyticus_11B
Myxococcus_xanthus
Thermus_thermophilus_HB27
Roseiflexus_sp._RS-1
Herpetosiphon_aurantiacus_ATCC_23779
Chloroflexus_aurantiacus_J-10-fl
Deinococcus_geothermalis_DSM11300
Roseovarius_sp._217
Roseovarius_nubinhibens_ISM
Rickettsia_felis_URRWXCal2_(RF_RS02815)
Rickettsia_typhi_str._Wilmington
Rickettsia_prowazekii_str._Madrid_E
Rickettsia_felis_URRWXCal2_(RF_RS03950)
Rickettsia_conorii_str._Malish_7
Wolbachia_sp.
Mannheimia_haemolytica
Lactobacillus_delbrueckii
Oenococcus_oeni_PSU-1
Streptococcus_thermophilus_CNRZ1066_(STR_RS06180)
Streptococcus_thermophilus_CNRZ1066_(STR_RS03445)
Streptococcus_pneumoniae_R6
Lactococcus_lactis_subsp._lactis_Il1403
Oceanobacillus_iheyensis_HTE831
Enterococcus_faecalis_V583
Listeria_innocua_Clip11262
Prochlorococcus_marinus_str._MIT_9312
Prochlorococcus_marinus_str._NATL2A
Prochlorococcus_marinus_str._MIT_9313
Prochlorococcus_marinus_CCMP1375
Prochlorococcus_marinus_str._MIT_9211
Leuconostoc_mesenteroides_ATCC_8293_(LEUM_RS01460)
Clostridium_tetani_E88_(CTC_RS09570)
Clostridium_botulinum_ATCC_3502_(CBO2260)
Clostridium_acetobutylicum_ATCC_824_(CA_C0211)
Clostridium_perfringens_str._13
Clostridium_acetobutylicum_ATCC_824_(CA_C1360)
Gloeobacter_violaceus_PCC_7421
Synechococcus_elongatus_PCC_6301
Synechocystis_sp._PCC_6803
Trichodesmium_erythraeum_IMS101
Nostoc_sp._PCC_7120
Nostoc_punctiforme_PCC_73102
Thermosynechococcus_elongatus_BP-1
Archaeoglobus_fulgidus_DSM_4304
Desulfovibrio_vulgaris_str._Hildenborough
Desulfovibrio_desulfuricans_G20
Clostridium_difficile_630
Bifidobacterium_longum_NCC2705
Clostridium_botulinum_ATCC_3502_(CBO0907)
Hungateiclostridium_thermocellum_ATCC_27405
Bacillus_licheniformis_ATCC_14580
Bacillus_halodurans_C-125
Clostridium_tetani_E88_(CTC_RS10720)
Pyrococcus_furiosus_DSM_3638
Thermococcus_kodakaraensis
Pyrococcus_horikoshii_OT3
Pyrococcus_abyssi_GE5
Carboxydothermus_hydrogenoformans_Z-2901
Thermotoga_maritima_MSB8
Pediococcus_pentosaceus_(PEPE_RS07735)
Bacillus_subtilis_subsp._subtilis_str._168
Staphylococcus_epidermidis_ATCC_12228
Leuconostoc_mesenteroides_ATCC_8293_(LEUM_RS02520)
Streptococcus_mutans_UA159
Streptococcus_agalactiae_2603V/R_(SAG0463)
Pediococcus_pentosaceus_(PEPE_RS07960)
Lactobacillus_plantarum_WCFS1
Streptococcus_pyogenes_M5
Streptococcus_agalactiae_2603V/R_(SAG0398)
Bacillus_clausii_KSM-K16
Rhodospirillum_rubrum
Deinococcus_radiodurans_R1
Bartonella_quintana_str._Toulouse
Leifsonia_xyli_subsp._xyli_str._CTCB07
Tropheryma_whipplei_TW08/27
Fusobacterium_nucleatum_ATCC_25586
Sulfitobacter_sp._EE-36_(EE36_RS04960)
Rhodobacter_sphaeroides_2.4.1_(RSP_2173)
Rhodobacter_capsulatus
Sinorhizobium_meliloti_1021
Rhizobium_leguminosarum_bv._viciae_3841
Agrobacterium_tumefaciens_str._C58
Bordetella_pertussis_Tohama_I
Thermobifida_fusca_YX
Kineococcus_radiotolerans_SRS30216
Propionibacterium_acnes_KPA171202
Corynebacterium_jeikeium_K411
Corynebacterium_efficiens_YS-314
Corynebacterium_diphtheriae_NCTC_13129
Treponema_pallidum
Treponema_denticola_ATCC_35405
Haloarcula_marismortui_ATCC_43049
Natronomonas_pharaonis_DSM_2160
Halobacterium_sp._NRC-1
Moorella_thermoacetica_ATCC_39073
Symbiobacterium_thermophilum_IAM_14863
2.0
Corynebacterium_glutamicum_ATCC_13032
Escherichia_coli_K12
Col18, ABC transporter [Achromobacter piechaudii]
Col 1, hypothetical protein [Sutterella wadsworthensis]
Col 17, Unknown protein [Mycobacterium tuberculosis]
Col15, Unknown protein [Unknown organism]
Col10, ABC-type transporter [Firmicutes bacterium]
Col 11, Ribokinase [Geobacillus stearothermophilus]

Figure 5: Phylogenetic Tree of Biotin Transporters and Metagenomic Hits. Phylogenetic tree showing the protein sequence similarity of BioY from a wide range of prokaryotes (black text), as well as YigM from *E. coli* (blue text) and selected metagenomic hits encoding hypothetical or confirmed transporters (green text). Remaining metagenomic hits are shown in magenta. The tree is cut and rotated around the two black lines for better readability. The scale bar represents branch length as amino acid substitutions per site.

Discussion

Our results show that introduction of a transporter with affinity for BSX into an *E. coli* strain lacking the ability to synthesize biotin *de novo,* can improve growth response on BSX to almost mimic that of biotin at low, extracellular concentrations (**Fig. 4B**). We suggest that such a transporter, if introduced into a biotin cell factory, may help improve the purity of biotin in a fermentation broth by importing extracellular BSX for BisC to convert back into biotin. Especially interesting is the possibility, raised by the likely presence of both BioY and the transporters with affinity for BSX in a number of organisms, that the newly identified transporters are specific to BSX over biotin. A high affinity for BSX vs. biotin would be an important feature for a purity-improving importer, needing to transport low amounts of BSX in the presence of high concentrations of biotin. Having a high purity of the product of interest in the fermentation broth is important for the overall manufacturing costs [10]. Furthermore, impurities from BSX could potentially be a problem for the consumers, as there may be a risk of BSX acting as an antimetabolite in cells lacking BisC, similar to the antimetabolite activity reported for BSN [16].

Following auxotrophic selections of metagenomic libraries, several different transporter proteins improving growth on mMOPS plates supplemented with BSX were identified (**Fig. 4A**). These results demonstrate BSX uptake as a bottleneck for evaluating the effect of BisC expression in *E. coli* strains lacking the ability to synthesize biotin *de novo* (**Fig. 2**). Since some of the identified transporters allow growth responses

with only minor differences between extracellular BSX and biotin (**Fig. 4B,** Col 6), there is no room for detecting potential beneficial effects from BisC overexpression in these strains based on phenotypic growth readout. Overexpression of BisC will only have the potential to improve BSX to biotin conversion if BSX is accumulating intracellularly. As the biotin growth requirement for *E. coli* is extremely low, only around 100-200 molecules per cell [19], the concentration range of BSX affecting the growth response will be equally low. Increasing the biotin requirements of the cell, e.g. through *birA* modifications reducing the ligase ability [22–24], could potentially shift the response range sufficiently to reveal a new bottleneck point, diverging BSX and biotin growth responses, as initially observed (**Fig. 2B**). Alternatively, analytical methods such as HPLC or LC-MS could be used to measure the concentration of intra- and extracellular BSX in strains with and without BisC overexpression, to determine whether BSX accumulates at concentrations above the growth requirement and if increased BisC expression eliminates such accumulation.

BisC has been reported to act as a defense against oxidative damage, e.g. by reducing oxidized products such as BSX and methionine sulfoxide [8,25]. In addition to potentially increasing the conversion of BSX to biotin to reduce the extracellular BSX in the fermentation broth, overexpression of BisC might decrease the amount of BSX released to the broth in the first place. To evaluate the effect of BisC overexpression on intracellular generated BSX, a biotin cell factory producing significant amounts of BSX as a byproduct would be needed.

The transporter proteins with affinity for BSX identified in this study were found to partly match proteins from both beneficial microorganisms found in the human gut, such as *Firmicutes sp.,* and pathogens belonging to the *Serratia* genus (**Fig. 4A**). Phylogenetic analysis of the transporters suggest that the selected hits are not typical BioY biotin transporters (**Fig.**

5). However, at least some of the identified transporter proteins seems to be relatively closely related to YigM. It should be kept in mind that the phenotypic comparison between YigM and the metagenomic inserts are distorted as the metagenomic inserts are expressed from a high-copy number plasmid while YigM is expressed from the genome. Nevertheless, further characterization could improve our understanding of biotin scavenge pathways. Bioinfomatic analysis has shown that biotin biosynthesis is present in only 40% of the human gut microbiota genomes [26]. Since biotin is essential for all living organism, scavenge of biotin and intermediates likely play a significant role for the remaining 60%.

Future Directions

The above results describe only initial metabolic engineering efforts to improve the purity of microbial produced biotin in fermentation broth. Further characterization of the BSX transporters identified in this study is required to evaluate whether BSX uptake and conversion will be beneficial in a fermentation setting.

Would Biotin Concentrations Affect Import of BSX?

In order to contribute to the production of pure biotin the desired BSX transporter should have high affinity for extracellular BSX, but not for biotin. To evaluate if the identified transporters fit these criteria, the affinity for biotin should be characterized. Evaluating the biotin-transporting ability of the identified transporters could be carried out by comparing the growth response on biotin of Δ*yigM* strains expressing the transporters to a strain expressing YigM from a comparable expression vector. Subsequently, characterization of the transporter proteins affinity for BSX in the presence of biotin should also be evaluated. This could be done by assessing the growth response of a Δ*yigM* strain, harboring one of

the identified transporters, with characterized limited ability to transport biotin, on a substrate matrix of different biotin and BSX concentrations.

Growth characterization of *E. coli* strains unable to synthesize biotin *de novo* and lacking YigM requires >1000 times more extracellular biotin to grow compared to a strain expressing YigM [27]. Therefore, passive diffusion or uptake by other native transporters with less affinity for biotin than YigM are not believed to have a significant influence on biotin import and does not need to be considered.

Once a BSX transporter fitting the requirements has been identified, the effect of overexpressing BisC should again be evaluated, similar to the experiments presented at the beginning of the chapter. However, in order to answer whether BSX accumulates intracellularly, analytical methods such as HPLC or LC-MS enabling detection of quantitative BSX concentrations above the growth requirement reflected in the growth characterization experiments might be needed.

Would BisC Expression or Catalysis be Affected by Biotin Concentrations?

Similar to the transporters with affinity for BSX, which should ideally enable import of BSX in the presence of high concentrations of biotin, BisC conversion of BSX to biotin needs to be efficient to improve product purity in a biotin cell factory. BisC has been suggested to work as a scavenging pathway, allowing cells to use BSX as a biotin source, in addition to it acting as a defense against oxidative damage [8,25]. Thus, it is likely that the regulation of BisC would be affected by biotin concentrations and oxidative stress. Biotin concentrations could be problematic in a biotin cell factory where concentrations are intended to be as high as possible. However, to our knowledge, regulation of *bisC* has only been reported on the transcriptional level. Here the expression is affected by the transcriptional regulator *nac*, which in turn is regulated by

nitrogen starvation [28]. Nevertheless, it would be advisable to evaluate the influence of high amounts of biotin on BSX to biotin conversion by BisC, due to the lack of published data. Such characterization could for example be carried out in cell free systems, which among others has been used to investigate the effect of mutations in the biotin operon on the biotin-forming activity of BioB [29]. If biotin is shown to inhibit native expression of BisC, introduction of a non-regulated, synthetic promoter could be considered. Similarly, introducing a BisC homolog, as discussed below, might be useful if *E. coli* BisC turns out to be product inhibited by biotin.

Searching for BisC Homologs as an Alternative to Overexpression

Even though native BisC expression and activity has been sufficient at the low, growth-complementing levels of BSX evaluated here, limitations might be faced when elevated amounts of BSX needs to be reduced to biotin. Overexpression of a protein can be a burden for cells for example due to increased use of resources and demand of cofactor supply [30]. An alternative to overexpression of the native protein is to use a functional homolog which may have increased activity and in some cases different precursor and/or cofactor requirement[31,32]. Identification of functional BisC homologs could be performed using the auxotrophic based selection system presented in this study. Performing metagenomic mining in a Δ*bisC* Δ*bioABCDH* strain, selecting for growth on BSX, should theoretically result in hits encoding BisC homologs. Whether the selection host should 1) express a BSX transporter and 2) have increased biotin requirements through *birA* modifications (as discussed above) needs to be considered. Theoretically, host modifications would shift the biotin requirements and potentially allow for BSX accumulation within the new selection range of the growth response, depending on the efficiency of BSX import and BisC catalysis. If such an accumulation can be achieved, the setup could be used to select for a BisC homolog with higher

conversion rate of BSX to biotin as well as evaluating the effect of BisC overexpression.

The described selection will be used to identify BisC homologs, if we find that simple overexpression of *E. coli* BisC cannot handle the conversion of the necessary amounts of BSX. Evaluating enzyme kinetics data of BisC could provide indications to the concentrations of BSX that would lead to BisC enzyme expression or activity becoming limiting. However, *E. coli* BisC has not been characterized thoroughly regarding enzyme kinetics due to limited availability and absence of a direct and convenient assays[33]. The lack of solid data makes it difficult to predict the approximate concentrations of BSX leading to limitations in BisC.

Is Molybdate Limiting for BisC catalysis?

To function as a reductase, BisC requires the molybdenum cofactor; bis-MGD, in which the chemical element molybdate is activated by members of the *moa, moe, mog* and *mob* operons in *E. coli* [9]. Cofactor limitation of BisC does not seem to be a problem in the characterization experiments of the BSX transporters from metagenomic selections, at the low concentration range of biotin required for growth. Here, growth characterizations of strains with transporters on extracellular BSX and biotin resulted in comparable response profiles (**Fig. 4B**). In the case of future experiments with increased expression of BisC, molybdate might become limiting, and it is therefore recommended to supplement molybdate as a standard component in the growth media.

Acknowledgements

We thank Biosyntia for providing strains, plasmids, primers and metagenomic libraries not developed in this study. David Lennox-Hvenekilde, Dóra Vitae and Luisa S. Gronenberg are thanked for critical reading of the manuscript. A.P.B. acknowledges financial support from Innovation Foundation Denmark (Industrial Ph.D. program, No. 5016-00135B).

Author Contributions

A.P.B., N.M.P. and H.J.G. conceived the study. A.P.B. established and characterized the auxotrophic selection assay, constructed and characterized the BisC overexpressing strains and performed the metagenomic selections. A.P.B. sequenced and evaluated the library hits improving growth on BSX. A.P.B. designed and performed the phylogenetic and bioinformatic analysis of the identified metagenomic hits. A.P.B. wrote the manuscript with contributions from all authors.

Methods

Media and General Considerations

All chemicals were bought at Sigma Aldrich or Carl Roth unless otherwise stated.

BSX ((+)-Biotin (+)-Sulfoxide, Cas # 10406-89-0, ≥98%) and BSN (Biotin Sulfone, Cas # 40720-05-6, ≥98%) were bought from Santa Cruz Biotechnology.

Minimal MOPS media made as described in supplementary note S1, Chapter 2. Used as standard assay medium unless otherwise stated.

Plasmid and PCR purifications were performed according to manufactures guidelines, using E.Z.N.A.® Plasmid Mini Kit I, V-spin and E.Z.N.A.® Cycle Pure Kit I from VWR

Growth Characterization on Extracellular Biotin and Oxidized Vitamers

Characterization in Liquid

A single colony of each strain to be characterized were picked from LB plates with antibiotic (kanamycin (50 mg/l), zeocin (40 mg/L) or ampicillin (100 mg/L)) and inoculated in minimal MOPS media with antibiotic. The cultures were grown at 37°C o/n to be starved for biotin. Optical density was measured at 600 nm (OD600), and assay medium made by diluting each of the strains in minimal MOPS (mMOPS) medium with antibiotic to a final OD600 of 0.01. 24 dilutions of biotin, BSX and BSN ranging from 0 to 250 nM were made in mMOPS media and diluted 10 and 40 times in mMOPS. 15 µl of each dilution set of standards were mixed with 135 µl of assay culture, sealed with breathable seal, and incubated at 37°C for 20 hours, shaking 250 rpm, after which OD600 was measured to evaluate growth requirement for the strains on biotin and oxidized vitamers.

Characterization on Agar Plates

1.5 % agar mMOPS plates with kanamycin (50 mg/l) and various amounts of biotin and oxidized vitamers (see table below for concentrations tested) were made. 3% agar were melted and cooled to 40°C mixed with preheated ~40°C 2x mMOPS, kanamycin, to a final concentration of 50 mg/l, and biotin vitamers (see concentrations below). 15 mL were transferred to agar plates (Ø = 7.5) left to dry in laminar flow bench for 30 minutes before storing in fridge for maximum 10 days. For evaluating growth response of background strain for metagenomic mining, BS1575 holding pBS0018, were grown o/n in LB with kanamycin (50 mg/l). The cells were spun down at 5000G for 4 minutes, and washed three times in 5 mL mMOPS, resuspending, centrifuging (5000 G, 4 minutes) and removing supernatant in each wash, in order to remove biotin from the media. At the final resuspension, OD600 were measured, and 10^3, 10^5 and 10^7 cells were plated using 100 µl and sterile glass beads for equal

distribution. Cell culture dilutions were made in mMOPS based on the conversion factor of 1 OD600 = $8 \cdot 10^8$cell/mL[1]. The plates were incubated at 37°C and evaluated continuously.

Table 1: Concentration of Biotin and Oxidized Vitamers Evaluated.

Biotin	BSX	BSN
100 nM	2.5 nM	100 nM
0.5 nM	1 nM	20 nM
0.05 nM	0.5 nM	10 nM
		5 nM
		1 nM
		0.5 nM

Mining Metagenomic Libraries for BisC "Accessory" Proteins

Four different metagenomic libraries, derived from soil and fecal microbiota samples, were used as the diversity foundation to search for proteins which could improve conversion of the extracellular BSX to biotin. Each plasmid library was transformed into a selection strain, auxotrophic for biotin, which was then plated on mMOPS agar plates with various concentrations of BSX and BSN stated above. The plates were evaluated continuously for 64 hours, comparing growth to control plates comprising both mMOPS plates with biotin as well as LB plates. Transformation efficiencies above $3 \cdot 10^7$cfu/µg were achieved for all four libraries.

[1] https://www.chem.agilent.com/store/biocalculators/calcODBacterial.jsp

Sequencing and ORF Identification of Hits from Auxotrophic Selections

Selected hits from auxotrophic selections of metagenomic libraries were grown to stationary phase in 4 mL LB media at 37°C with shaking at 250 rpm and plasmids were purified. Initial sequencing of metagenomic inserts were performed using oBS0036 and oBS0041, followed by primer walking for inserts not covered fully (**Table 4**). Identification of ORF's in the sequenced inserts were done using DeFUME web tool[18]. Briefly, DeFUME generates an interactive table showing the sequence reads, assembled contigs, predicted ORF's, BLASTp hits and InterPro functional data from the sequencing files. Default settings was used and are described here[18].

Plasmids and Strains used in this Study

Table 2: Overview of Strains used and Generated during this Study

Strain ID	Description
BS1575	BW25113, Δ*bioA,* Δ*bioBCD,* Δ*bioH, iscR H107Y*
BS1592	BW25113, Δ*bioA,* Δ*bioBCD,* Δ*bioH, iscR H107Y,* Δ*bisC*
BS1615	BW25113, Δ*bioA,* Δ*bioBCD, iscR H107Y*
Col 1	BS1575 with metagenomic library insert col 1, pBScol1
Col 5	BS1575 with metagenomic library insert col 5, pBScol5
Col 6	BS1575 with metagenomic library insert col 6, pBScol6
Col 10	BS1575 with metagenomic library insert col 10, pBScol10
Col 11	BS1575 with metagenomic library insert col 11, pBScol11
Col 14	BS1575 with metagenomic library insert col 14, pBScol14
Col 15	BS1575 with metagenomic library insert col 15, pBScol15
Col 16	BS1575 with metagenomic library insert col 16, pBScol16
Col 17	BS1575 with metagenomic library insert col 17, pBScol17
Col 18	BS1575 with metagenomic library insert col 18, pBScol18
Col 19	BS1575 with metagenomic library insert col 19, pBScol19

Table 3: Overview of Plasmids used and Generated during this Study

Plasmid ID	Brief description	Features
pBS0018	Backbone plasmid for metagenomic libraries	KanR, ColE1
pBS0451	Zeocin resistance plasmid	ZeoR, p15A
pBS1021	*bisC* expression plasmid, high copy-number	AmpR, pBR322
pBS1024	*bisC* expression plasmid, medium copy-number	AmpR, p15A
pBScol1	Metagenomic insert from selected hit col 1	KanR, ColE1
pBScol5	Metagenomic insert from selected hit col 5	KanR, ColE1
pBScol6	Metagenomic insert from selected hit col 6	KanR, ColE1
pBScol10	Metagenomic insert from selected hit col 10	KanR, ColE1
pBScol11	Metagenomic insert from selected hit col 11	KanR, ColE1
pBScol14	Metagenomic insert from selected hit col 14	KanR, ColE1
pBScol15	Metagenomic insert from selected hit col 15	KanR, ColE1
pBScol16	Metagenomic insert from selected hit col 16	KanR, ColE1
pBScol17	Metagenomic insert from selected hit col 17	KanR, ColE1
pBScol18	Metagenomic insert from selected hit col 18	KanR, ColE1
pBScol19	Metagenomic insert from selected hit col 19	KanR, ColE1

Table 4: Overview of Primers and Oligos used in this Study

Oligo ID	Sequence	Description
oBS0036	GCCCTTTCGTCTTCACCTCG	FW sequencing primer for all metagenomic contigs
oBS0041	CGTTCACCGACAAACAACAG	RV sequencing primer for all metagenomic contigs
oBS1904	ACGGCGGCAATAAGATAAACACGTTTC	FW primer for genomic *bisC* amplification
oBS1905	TTTACCAGGCACAGGGTTTTCACATTG	RV primer for genomic *bisC* amplification
oBS2335	GCCTTTGAGGAGCCAA	RV sequencing primer for col 1 primer walking

oBS2336	TTCTCCACGGTCTCA	FW sequencing primer for col 1 primer walking
oBS2337	AACCAACAGTCCTGC	RV sequencing primer for col 6 primer walking
oBS2338	TATCTGTTCAGCTTCC	FW sequencing primer for col 6 primer walking
oBS2339	CCGAAAGCGTTGCGG	RV sequencing primer for col 10 primer walking
oBS2340	CACGAGCTATATCAGC	FW sequencing primer for col 10 primer walking
oBS2341	ACCAAAACTATCGTCC	RV sequencing primer for col 11 primer walking
oBS2342	AGCAATTAAAGCCCAA	FW sequencing primer for col 11 primer walking
oBS2343	ATTTATTGATGCTTCAT	RV sequencing primer for col 15 primer walking
oBS2344	TGCCAACGCTGTGTTA	FW sequencing primer for col 15 primer walking
oBS2345	AAAGTACCGGCATCCA	RV sequencing primer for col 16 primer walking
oBS2346	TTCCTGCGCTGGCGCA	FW sequencing primer for col 16 primer walking
oBS2347	GCGCGAGAGATAGGCT	RV sequencing primer for col 17 primer walking
oBS2348	TGGTTATTGTCAGCCTA	FW sequencing primer for col 17 primer walking
oBS2349	GTCTCCCCGGGCCGGT	RV sequencing primer for col 18 primer walking
oBS2350	GACGCCGACGCGCGTT	FW sequencing primer for col 18 primer walking

Phylogenetic Analysis of Transporter Inserts from Metagenomic Hits

Multiple alignment of protein sequences of the ORF's with highest coverage identified in metagenomic hits using BLASTp together with 108 BioY protein sequences[21] and the protein sequence of YigM from *E. coli* were carried out using CLC Bio having a gap opening cost of 10.0, a gap

extension cost of 1.0 and end gap cost set to "cheap". Next a maximum likelihood phylogenetic tree was made based on the Neighbor Joining method, using the WAG protein substitution model without rate variation. Bootstrap analysis with 1000 replicates were included. Subsequent pairwise alignment between YigM and Col 6 insert were carried out using the global alignment Needle (EMBOSS) method provided by EMBL (https://www.ebi.ac.uk/Tools/psa/). Blosum62 matrix was used with the following settings; gap open cost = 10, gap extension cost = 0.5, end gap penalty cost = FALSE, end gap open cost = 10 and end gap extension cost = 0.5. Additional BLASTp searches were carried out using NCBI's website (https://blast.ncbi.nlm.nih.gov/Blast.cgi) using default parameters (Max target sequences = 100, parameters for short queries are automatically adjusted, Expected threshold = 10, Word size = 6, Max matches in a query range = 0, Matrix = BLOSUM62, Gap Cost = Existence: 11 Extension: 1, Compositional adjustments = Conditional compositional score matrix adjustments.)

References

1. Center, Q. C. & M. R. *Global Biotin Market Research Report 2017*. (2017).
2. Eggersdorfer, M. *et al.* One hundred years of vitamins-a success story of the natural sciences. *Angew. Chem. Int. Ed. Engl.* **51**, 12960–90 (2012).
3. Komatsubara, S., Imai, Y., Masuda, M. & Sakurai, N. Microorganisms and process for preparing D-Biotin using the same. (1994). US5374554
4. Kanzaki, N., Kimura, H., Matsui, J., Nakahama, K. & Ifuku, O. Microorganism resistant to threonine analogue and production of biotin. **1**, (1999). US 6,284,500 B1
5. Brown, S. W. *et al.* The production of biotin by recombinant strains of Escherichia coli. *J. Chem. Technol. Biotechnol.* **50**, 115–121 (1991).
6. Xiao, F. *et al.* Multi-level metabolic engineering of Pseudomonas mutabilis ATCC31014 for efficient production of biotin. *Metab. Eng.* (2019).
7. Melville, B. Biotin sulfoxide. *J. Biol. Chem.* **208**, 495–502 (1954).
8. Pierson, D. E. & Campbell, A. Cloning and nucleotide sequence of bisC, the structural gene for biotin sulfoxide reductase in Escherichia coli. *J Bacteriol* **172**, 2194–2198 (1990).
9. Leimkühler, S. The Biosynthesis of the Molybdenum Cofactor in Escherichia coli and Its Connection to FeS Cluster Assembly and the Thiolation of tRNA . *Adv. Biol.* **2014**, 1–21 (2014).
10. Stanbury, P. F., Whitaker, A. & Hall, S. J. Chapter 10: The recovery of prurification

of fermentation products. *Princ. Ferment. Technol.* 725–775 (2016).
11. Marie Sych, J., Lacroix, C. & Stevens, M. J. A. Vitamin B12-Physiology, Production and Application. in *Industrial Biotechnology of Vitamins, Biopigments, and Antioxidants* 129–160 (Wiley-VCH Verlag GmbH & Co. KGaA, 2016). doi:10.1002/9783527681754.ch6
12. Moore, S. J. & Warren, M. J. The anaerobic biosynthesis of vitamin B12. *Biochem. Soc. Trans.* **40**, 581–586 (2012).
13. Králová, B. & Rauch, P. The effect of antimetabolites on the microbial biosynthesis of vitamin B12 and related compounds. *Appl. Microbiol. Biotechnol.* **21**, 331–335 (1985).
14. Obeid, R., Fedosov, S. N. & Nexo, E. Cobalamin coenzyme forms are not likely to be superior to cyano- and hydroxyl-cobalamin in prevention or treatment of cobalamin deficiency. *Mol. Nutr. Food Res.* **59**, 1364–1372 (2015).
15. European Food Safety Authority (EFSA). Scientific Opinion on the safety and efficacy of biotin a feed additive for all animal species based on a dossier submitted by VITAC EEIG. *EFSA J.* **11**, 3211 (2014).
16. Melville, B. Biological properties of biotin D- and L-sulfoxides. *J. Biol. Chem.* **208**, 503–512 (1953).
17. Campillo-Campbell, A. Del & Campbell, A. Molybdenum Cofactor Requirement for Biotin Sulfoxide Reduction in Escherichia coli. *J. Bacteriol.* **149**, 469–478 (1982).
18. Van Der Helm, E., Geertz-Hansen, H. M., Genee, H. J., Malla, S. & Sommer, M. O. A. DeFUME: Dynamic exploration of functional metagenomic sequencing data. *BMC Res. Notes* **8**, 1–3 (2015).
19. Jarrett, J. T. Biotin Synthase: Enzyme or Reactant? *Chem. Biol.* **12**, 409–410 (2005).
20. Azhar, A., Booker, G. W. & Polyak, S. W. Mechanisms of Biotin Transport. *Biochem. Anal. Biochem.* **04**, (2015).
21. Hebbeln, P., Rodionov, D. A., Alfandega, A. & Eitinger, T. Biotin uptake in prokaryotes by solute transporters with an optional ATP-binding cassette-. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2909–2914 (2007).
22. Campbell, a, Del Campillo-Campbell, a & Chang, R. A mutant of Escherichia coli that requires high concentrations of biotin. *Proc. Natl. Acad. Sci. U. S. A.* **69**, 676–80 (1972).
23. Campbell, A., Chang, R., Barker, D. & Ketner, G. Biotin regulatory (bir) mutations of Escherichia coli. *J. Bacteriol.* **142**, 1025–1028 (1980).
24. Chakravartty, V. & Cronan, J. E. The wing of a winged helix-turn-helix transcription factor organizes the active site of BirA, a bifunctional repressor/ligase. *J. Biol. Chem.* **288**, 36029–36039 (2013).
25. Denkel, L. a, Rhen, M. & Bange, F.-C. Biotin sulfoxide reductase contributes to oxidative stress tolerance and virulence in Salmonella enterica serovar Typhimurium. *Microbiology* **159**, 1447–58 (2013).
26. Magnúsdóttir, S., Ravcheev, D., De Crécy-Lagard, V. & Thiele, I. Systematic genome assessment of B-vitamin biosynthesis suggests cooperation among gut microbes. *Front. Genet.* **6**, (2015).
27. Finkenwirth, F., Kirsch, F. & Eitinger, T. A versatile Escherichia coli strain for identification of biotin transporters and for biotin quantificatio. *Bioengineered* **5**,

129–132 (2014).
28. Switzer, A., Brown, D. R. & Wigneshweraraj, S. New insights into the adaptive transcriptional response to nitrogen starvation in Escherichia coli . *Biochem. Soc. Trans.* **46**, 1721–1728 (2018).
29. Ifuku, O. *et al.* Sequencing analysis of mutation points in the biotin operon of biotin-overproducing Escherichia coli mutants. *Biosci Biotechnol Biochem* **57**, 760–765 (1993).
30. Bolognesi, B. & Lehner, B. Reaching the limit. *Elife* **7**, 1–3 (2018).
31. Van Arsdell, S. W. *et al.* Removing a bottleneck in the Bacillus subtilis biotin pathway: BioA utilizes lysine rather than S-adenosylmethionine as the amino donor in the KAPA-to-DAPA reaction. *Biotechnol. Bioeng.* **91**, 75–83 (2005).
32. Himanshu, D. & Kristala, P. Microbial chemical factories : recent advances in pathway engineering for synthesis of value added chemicals. *Curr. Opin. Struct. Biol.* **21**, 488–494 (2011).
33. Campillo-Campbell, A. Del, Dykhuizen, D. & Cleary, P. P. Enzymic reduction of d-biotin d-sulfoxide to d-biotin. **62**, 379–385 (1979).

CHAPTER III - SUPPLEMENTARY

METAGENOMIC MINING OF TRANSPORTERS WITH AFFINITY FOR BIOTIN SULFOXIDE

Anne P. Bali[1,2]**, Nils Myling-Petersen**[2] **& Hans J. Genee**[2*]

Affiliations
[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark. [2]Biosyntia ApS, Fruebjergvej 3, DK-2100, Denmark

***Corresponding author:** hjg@biosyntia.com

Supplementary table S1: Detailed overview of contigs in selected hits from metagenomic mining of accessory proteins for increased BisC activity, based on improved growth on biotin sulfoxide and biotin sulfone as biotin source.

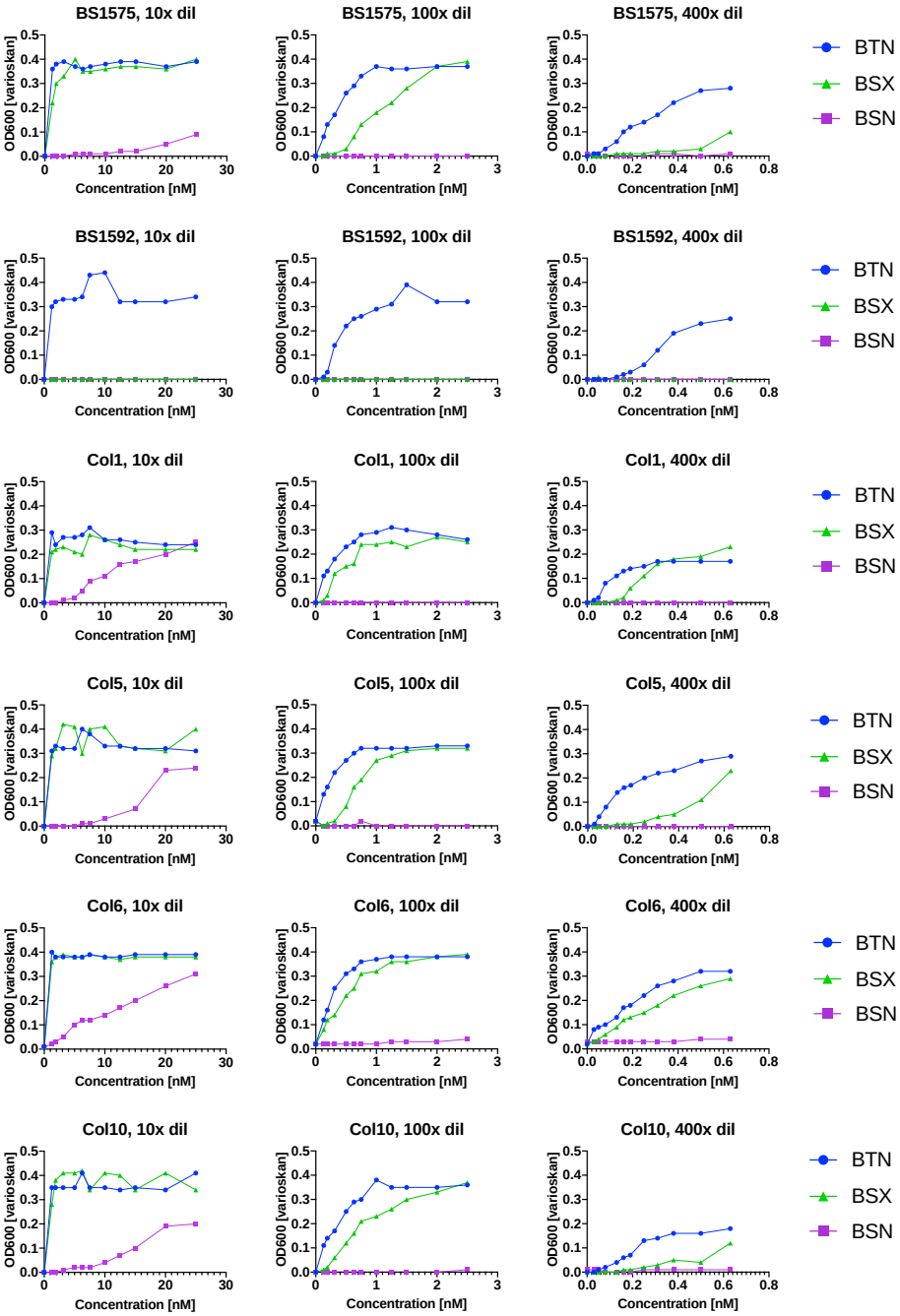
Isolation Source	Contig	Region on Contig	Hit description	Associated GO term	% Identity	% Coverage
	Con1	<2..232	Hypothetical protein [*Sutterella wadsworthensis*]		89.2	96.0
		291..1640	Hypothetical protein [*Sutterella wadsworthensis*]	Transporter activity, transmembrane	98.7	99.8
		1789..>1890	Phosphoglycerate mutase (2,3-diphosphoglycerate-independent) [*Sutterella wadsworthensis*]		100	6.47
	Con6	<1..225	Hypothetical protein SLIQ_21330 [*Serratia liquefaciens FK01*]		92.5	73.2
		352..1215	Membrane protein [*Serratia liquefaciens]*	Membrane, integral component of membrane	99.7	99.7
		1248..>1520	LysR family transcriptional regulator [*Serratia grimesii*]		100	28.2
	Con10	<3..140	Trehalose utilization protein [*Firmicutes bacterium CAG:272*]		47.6	16.5
		245..805	Transcriptional regulator [*Candidatus Stoquefichus sp. KLE1796*]	DNA binding	33.0	93.3
		805..1725	ABC-type multidrug transport system,		66.4	96.7

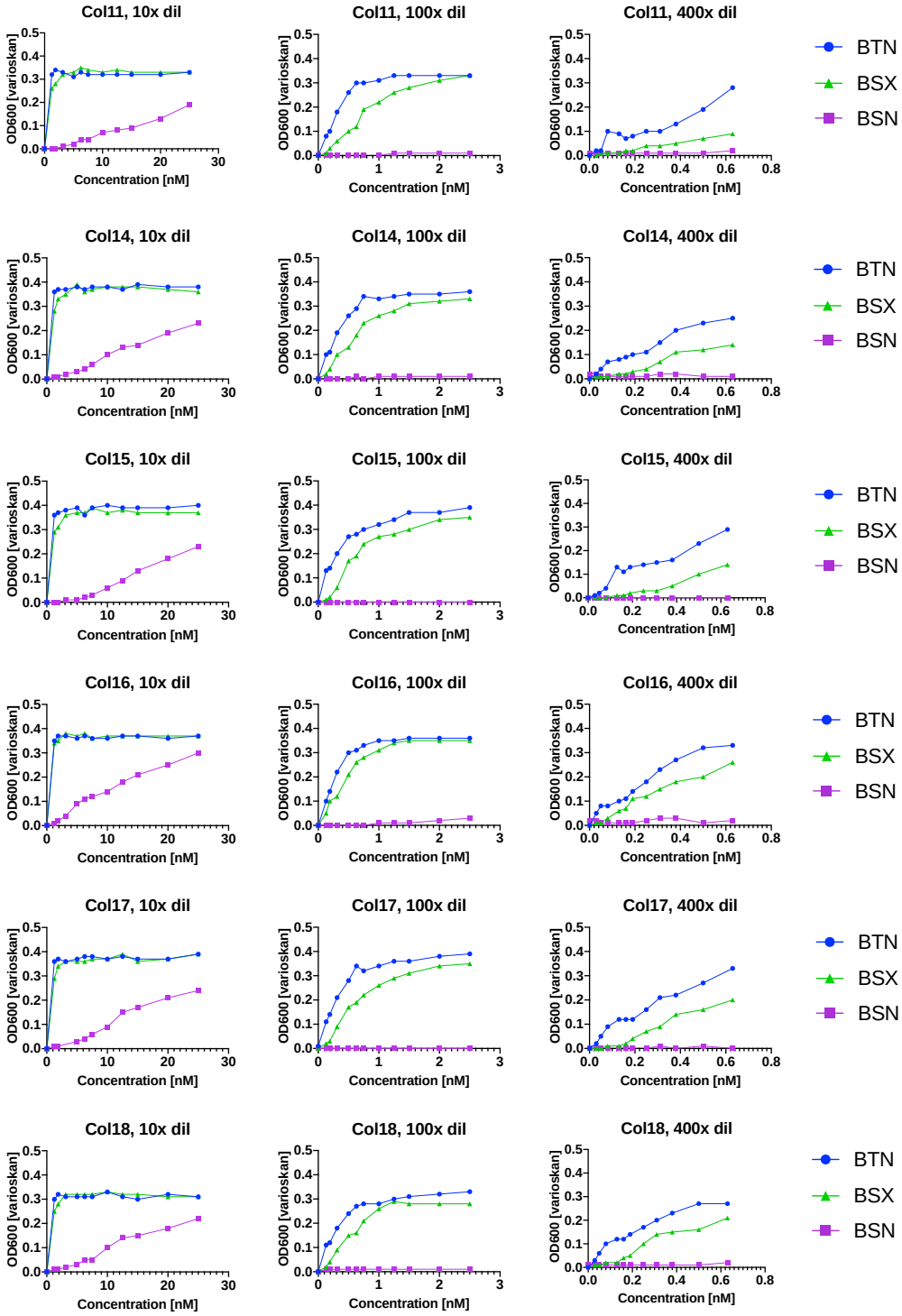
			ATPase component [*Firmicutes bacterium CAG:238*]			
	Con11	<2..841	Ribokinase [*Geobacillus stearothermophilus*]	D-ribose metabolic process, ribokinase	41.3	82.8
		1009..1344	Putative uncharacterized protein [*Firmicutes bacterium CAG:552*]		73.9	22.2
	Col15	<1..1398	Hypothetical protein F083_859 [*bacterium F083*]		46.8	93.0
	Col16 (identical to Con6)	<1..273	LysR family transcriptional regulator [*Serratia grimesii*]		100	28.2
		306..1169	Membrane protein [*Serratia liquefaciens]*	Membrane, integral component of membrane	99.7	99.7
		1296..>1520	Hypothetical protein SLIQ_21330 [*Serratia liquefaciens FK01*]		92.5	73.2
	Col17	<2..412	Multispecies: NlpC/p60 family protein [*Pseudomonas*]		99.3	100
		662..844	Uncharacterized protein [*Mycobacterium tuberculosis*]		84.9	98.1
	Col18	<2..442	ABC transporter ATP-binding protein [*Achromobacter piechaudii*]		100	62.1
		464..1345	Branched-chain amino acid ABC transporter permease [*Achromobacter piechaudii*]	Transporter activity, transport, membrane	99.3	99.7

		1355..>1540	Branched-chain amino acid ABC transporter permease [*Achromobacter sp.*]		98.4	18.2

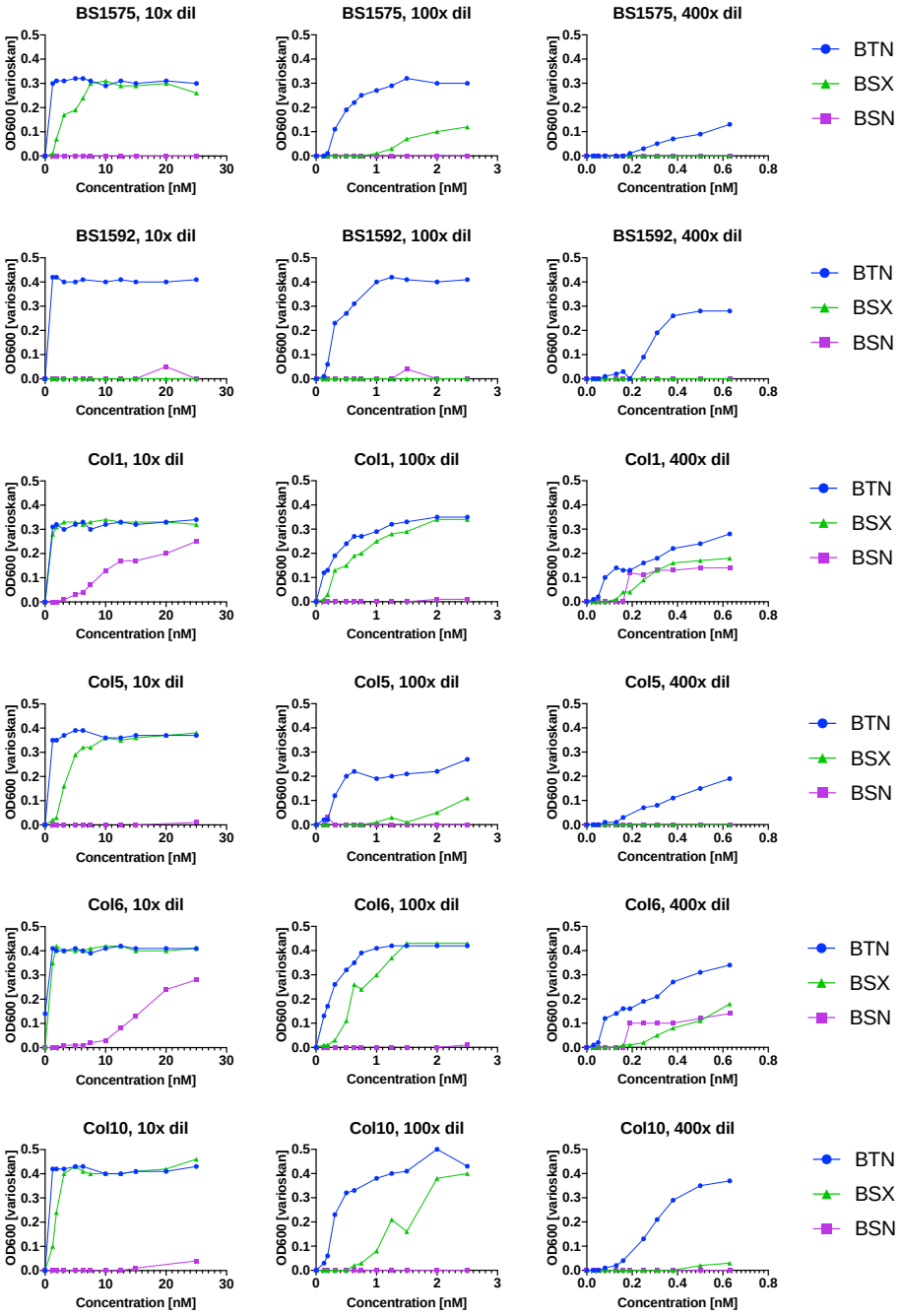
Supplementary Figure S1: Phenotypic growth characterization of hits from auxotrophic selections in liquid mMOPS with added biotin sources as well as control strain (BS1575) holding an empty background plasmid (n=2). Growth response on biotin shown in blue (BTN), biotin sulfoxide in green (BSX) and biotin sulfone in purple (BSN).

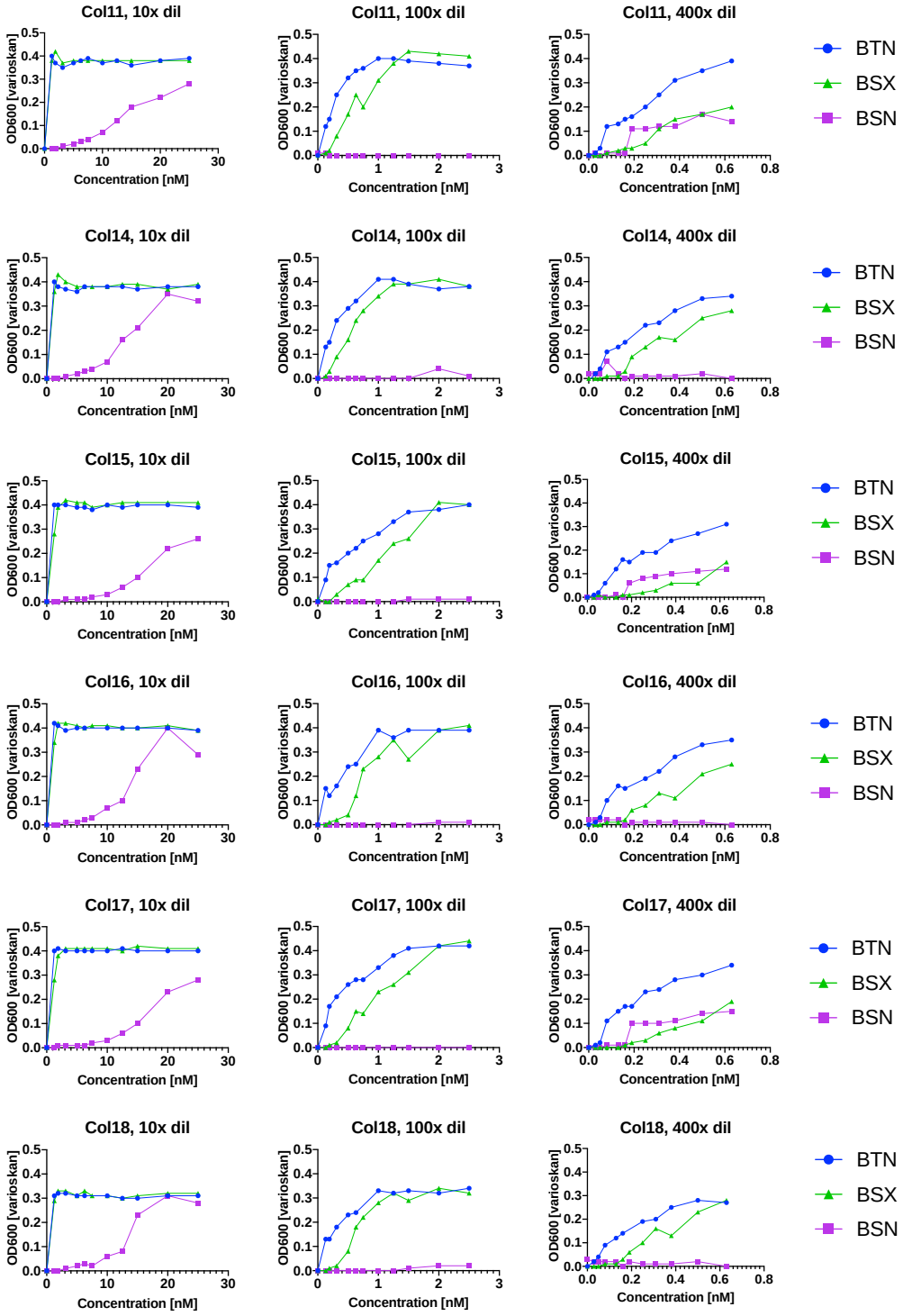
First replicate

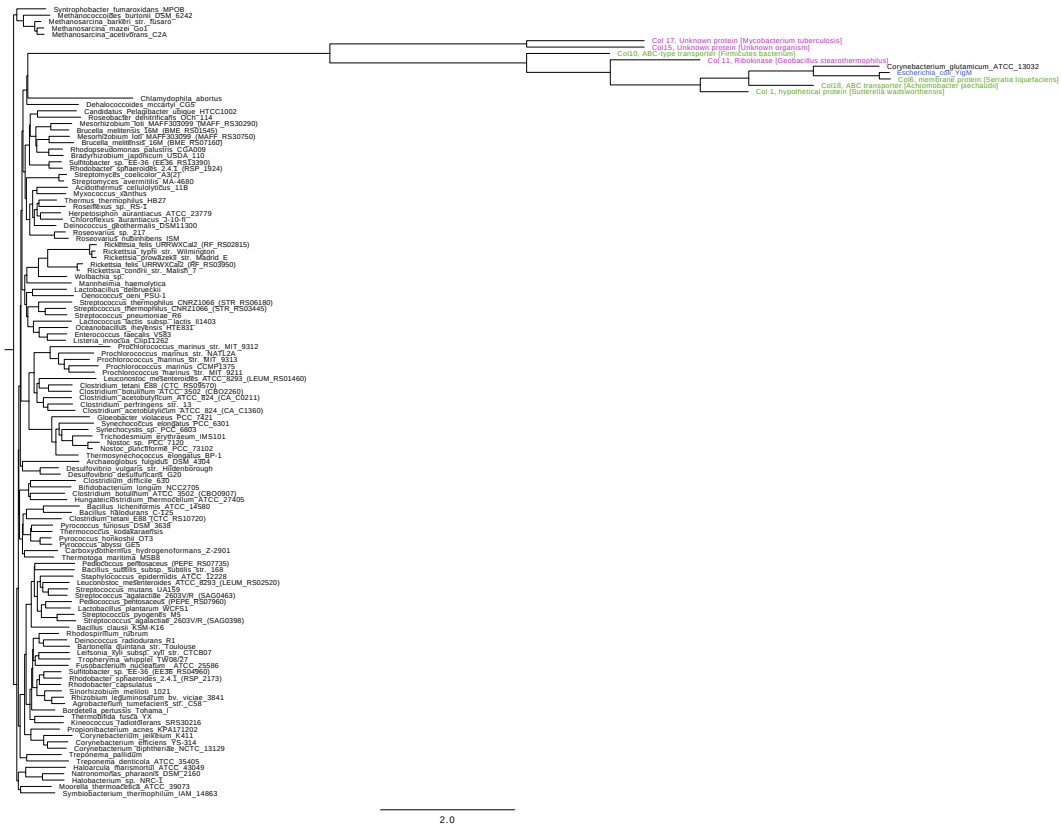



Col11, 10x dil
Col11, 100x dil
Col11, 400x dil
Col14, 10x dil
Col14, 100x dil
Col14, 400x dil
Col15, 10x dil
Col15, 100x dil
Col15, 400x dil
Col16, 10x dil
Col16, 100x dil
Col16, 400x dil
Col17, 10x dil
Col17, 100x dil
Col17, 400x dil
Col18, 10x dil
Col18, 100x dil
Col18, 400x dil
OD600 [varioskan]
Concentration [nM]
BTN
BSX
BSN

Second replicate




Col11, 10x dil
Col11, 100x dil
Col11, 400x dil
Col14, 10x dil
Col14, 100x dil
Col14, 400x dil
Col15, 10x dil
Col15, 100x dil
Col15, 400x dil
Col16, 10x dil
Col16, 100x dil
Col16, 400x dil
Col17, 10x dil
Col17, 100x dil
Col17, 400x dil
Col18, 10x dil
Col18, 100x dil
Col18, 400x dil
OD600 [varioskan]
Concentration [nM]
BTN
BSX
BSN



Supplementary Figure S2: Overview of phylogenetic tree of biotin transporters. Phylogenetic tree showing the protein sequence similarity of BioY from a wide range of prokaryotes (black text), as well as YigM from *E. coli* (blue text) and selected metagenomic hits encoding hypothetical or confirmed transporters (green text). Remaining metagenomic hits are shown in magenta. The scale bar represents branch length as amino acid substitutions per site.

Supplementary List S1

Nucleotide sequences of assembled contigs from metagenomic inserts in selected hits, improving growth on BSX.

>Col1

TCTCCTAAAGCGTTTGAATGGACTGCCGCAGTTCAATGGGTGCTGTCCGTCAGAGGACAACGGCTTAAACGTG
CCTGCAGTCGTGTTGACCGATTTATTTTCGCCATTCCTGTCAACAAAGGAAATATCCATTTCTCTCAAAGATG
TGCCCAAATCGGAACTGCAGAAAAACGCCGGCACTGGTCCTTTTCACGGGATAGCGTTCAGCCGAAAGATCAA
GCGTAAACCCTAAATCGGCTGCGGCAGGCGCGTAAGATCATGAACTTTCGAATTCTTCTCAGAGCCTTTCGAT
GAACTATCTCACTCTGCTTGTTCTGGCCTACCTCGCGGCACTCCTCTGGTTCGGCATCCGAGGCGGGACCCGC
GAACGTGAACTCAACGCGTTTCTGACGGCCGGCGGATCGGCCGGCGCGCTGCTCTGCGCCTTCTCGCTTGTGA
GTACGATCATCGGCGGCTCAGCCACGCTCGGCATGGGCGCACTCGCTCAGAAGGTCGGCCCGGCGGCTTTTTG
GTGGCTGGGCGTCGGCGCTATCGGACTTTTTCTCCACGGTCTCATCATTGCGCCCGTTGTCCGCAAGCTCCCC
GTCGTGACGCTCCCGGATGTGCTCAGACACCTTGCGGGCCCGGCGGCAGAAAAATGGGCAGGCGCCATTATCG
CGCTCTCCTGGGCAGCGGTGACGGCCGCGCAGTTCACTGCACTGCGCGCACTCCTTCATTCCATTTCGGGCGG
CATGACAGCTGAGGTGCTCTATGTGCTTCTCGCGGCCGGCATTGTGCTTCATACGGCTTTAGGCGGACAGCGC
GGCATTCTGCGGGCCGACTTTCTGCACACGATTTTGCTGCTCGGAGGCTTCTCCGCCGCAGCGCTCTGGTGCG
TCATGGAGCGAAGCGCCGACATCGCGGCGCTCCCTGTGATGCCCTTTACTGAGACCTTTGGGCTCGGCGACTG
GGTGAAGATGATGCTCCTCGTCGGCATCACTTACCTGATCGGCCCAGACATGTTCTCGCGCACCTTCGCGGCC
AAGAATGCGCAGACCGCGCGCCGCGCGGCCTGGACAGCTGCGCCGCTCCTCGTTTTCTTCGGCGTAATCGTGA
CGATGCTCGCACTGCTCAACATCAAGGCTGAGCAGCCGATTTCTGACTGGCTCGCGGTCGGAAGCCCGCTCCC
GGCCGCTGCGGCCGCAGCGCTTTCCATGGGGCTTATCAGTGCGCTTGCCGGCTCGGCCGATACGATTCTGCTC
TCCGCGGCCGGCATCATTGAGAAGGACCTCTTTGGCGGCGCCCGTGCATGCGCCGTGCGTCTCTGGGCTGCAG
CCGTCGGCATTTTGGCCGCACTGGCCGCTTATACCTCGGGAGACATCATCGGTTGGCTCCTCAAAGGCTACGC
GCTCTTCGTCCCGGGCGTTGCCGTGCCGCTCCTCGTCCTCGTCGCCGGCCGCGCCCGACACGCGGATCCCAAG
CTCTGGACTATTGGAGCCGCCGCGGGCGGTCTGGGAGGGCTTGTCGGCAGCCTTACCGGCATCACCCTCTGGA
CTTACGGCGGCATCGCTGCTGCCGCCCTTTTTGCGGTCTTTGCCGTACGGCGCGCCGCCCAAACCCGCCGAAC
GGCCGCCGCGTCACCGCTTCAGCCCCTCTCATAAATCTGAAGCCTTAATGTCGTCATCGGGCTTTTTGGAGCT
CGATCCCGACACGCAGATCCTTTATGCTTACGGGCAGTTTTTTCGGAGCGCCGCCAGTTTTTATCCGACAAAG
CCGGGCGCCGCTCCGCTGCATCTGCGTTATTAAGCAATGACCAAAGTTCTTCTCATGATTCTCGACGGCTGGG
GCATCGGCGGCGGTGCCGCAGACCCGGCCGACGCCATCAGCGCGGCGCACCCGGCCTTCATTGAAAG

>Col6

CTGCTGGTTATTATCGAACTGCCGCTGTTGTTGCAGGCGGCGCTGATCCAGCTGCTGATTATTCAACATACGC
TGTTGTTGCGTCTGCTGAGTGCGTTGGTTGAGAAACTGCTGTTGCGAGCTGAGCACGGTGACGCCGTTGGCTG
CGGTGGCCAATCCGGACAGCAGCAACAGGCCATAGAGCAGATGATAACTTTTCATGATGTTCTCCTCGTCTTT
TATTAAGCCTAACCCAGCGGTGCAAATTTGCTGAAAGTTTGTATTTTGCTCAGATGTAAACCTTCTTTTCGCT
ACACCCCTTTCTGCCTTCCGTTACAATCGCCCTTTCTGAATCTTTAGGGTGAACCCATCGTGCTGCTTCTGGT
GATTACTACCATCCTGTGGGCATTTTCTTTCAGCCTGATTGGCGAATACCTGGCGGGCCAGGTCGACAGCTGG
TTCTCGGTGCTGATGCGCATCGGCCTGGCGGCGGCGGTGTTCCTGCCTTTCCTGCGCTGGCGCAACATCAAGC
CAAAAGTGATCCTGTTGTACATGCTGGTGGGTGCTTGCCAGCTTGGCATCATGTATCTGTTCAGCTTCCGCGC
GTACATGTACCTGACGGTGCCGGAATTCCTGCTGTTTACCGTGCTGACGCCGCTGTACGTTACGCTGATTTAT
GACTTGATCAGCGGACAGCGCCTGCGCTGGGGTTATGTGTTCAGTGCGCTATTGGCAGTGCTGGGCGCTGCGA
TTATTCGATATCATCAATTGAGCGAGCATTTTTGGTGGGGGCTGTTGCTGGTGCAGGCGGCGAACATCAGCTT
TGCCATCGGCATGGTCGGTTACAAACGGTTGATGGAAGTGCATCCGCTGCCGCAACACACCGCATTTTCCTGG
TTCTATCTGGGGGCCTTGGTCGTGGCGGTGGTCGCCTGGTGCCTGTGGGGCAATCCGCAGCAACTGCCAACCA
CTAACCTGCAGTGGGGCATTCTGGTCTGGCTGGGCGTAGTGGCTTCCGGGCTGGGTTATTTCATGTGGAACTA
CGGCGCAACGCAGGTGGATGCCGGTACTTTGGGCATCATGAATAACATGCATGTGCCGGCAGGACTGTTGGTT
AATCTGGCCATCTGGCAGCAACAGCCGCATTGGCCAAGTTTCATTATCGGCGGGGCGGTGATTGTCGCTTCGC
TGTGGGTGCATCGCCGTTGGGTGGCGCCTAAAATAGTCAGCAAATAAACGGGCAGCCCAAGCGCGAGCTGCCC
TGATATTCAGGGTTGCGAAGCCAGCGGTTTGACGCTGGCGACACCGGCGCTTGGGCGTGCAGCATCACGCACG
AATGGCAGATGGTTGCAGGCATGTTGCCGAGCCGAGCGGATAAAGGCTTCGGTGACGGCCTGGCGCTGTTCAC
CGTCGCGTACCGCCGCGTACAGGCGGCTCCAAAGCCCATCCCCCAGGGTTTTGGTCACCACTAACCCCTGGCG
TTCAAAACTTTCCACCACCCAGTGCGGCAACGCCGCGATCCCCATGCGGGCCGATACCAT

>Col10

TCCCCATATATATGATGCCTGAGGTACAGACCGTAATTAAAAATGCTGTAAGATATCTTGCGCCTACATACAG
AGTAAGAGTTGGATGTCCCAAGGTTAATACAGTAAACGACGAAAAATACGTCGAGAAGGATATTTAAAAGAAA
AAAGATGTCAATTTCCGAAAAAGGACTTGACATCTTTTTCTTTTTATGATAAAATAACAGGTGTTATATAACA

ACTGTTATTAAAAAAGGAGGAGCAAATGCCACCTAAATCTATATTGACAAAAGAGCGGATAGCGATACAAGCG
TTTGAAATAATTCGCAAGGATGGCTACGATGCCCTGAACACAAGACGACTTGCCTCGGAGCTTGGGGTTTCCA
CCATGCCGCTTTTCAAGCACTACAAAAGTATGGACGAAATAAAGAGCGCCGCCGTTAATCTCGGAGTTGAGGA
ATATACGAAATATATGCGATCTGGCGCTGAAGATCCAATACCCTTTAAGGGGATAGGAAGAGCGTATATAAGG
TTTGCCAAGAATGAGCCGAGGCTTTTTGAGATATTTTATATGCGACCTACCGAAAGCGTTGCGGGGCTTGACA
AGGTAGATCCCAACACCGCCGCTGCACTTAATATGGCATCGAACATATTGAAGGGAAATGCCGACAGTGGTGA
AAAGCTGTTAATGGATATGTGGGTCGTCGTTCACGGTATAGCGACACTTGAGGCTACGGGGAAATATAGTTTT
TCCGACGATGAGATAGGTGAGATATTAAGTACCATCTTCCTTGGATTGAAAAAGCAGTATGGAGGGGATTTAT
AATGGAAAACGTTATTGAAATATCGGGACTTTATAAGAGCTTTAAGGATATAAAGGCGGTGCAGAACTTGAGC
TTTTCGGTAGAGCGGGGAAGCCTGTTTGCCTTTCTTGGCGTAAACGGAGCTGGAAAAAGCACTACTATTTCGA
TGATGTGCGGTCAGCTTACTCCCGATAGCGGCAGAATTTTTATTGACGGAAACGATATAGTAAACGCTCCTGA
TGCGGTAAAGTCGTCTATTGGCGTTGTGTTTCAGGGATCGGCTCTCGATAAGCCCTTGAGCGTTTTTGATAAT
CTTAAATGTCGCGCGGCTCTGTACGGAATATGCGGGGATGCCTTTGAAAAAAGGGCGGCAGAGCTTACGGAGC
TTCTGGATATGTCGGATATCTTAAAGCGCCCCTACGGAAAGCTCTCGGGAGGGCAGAAGCGTAAGGCTGATAT
AGCTCGTGCGCTTTTGCACAAGCCGAGAATACTGATACTTGACGAGCCCACGACGGGACTTGATCCACAGACC
AGAAGAATGGTTTGGGAGGTTATTGACAGCCTGAGGCGTACAGAGAGCATGACGGTCTTTCTCACAACACACT
ATATGGAGGAGGCGGCTGATGCCGACAAGGTAATAATTCTCGATGGCGGAAGGATCGTGGCGGAAGGTACTCC
CGTAGAGCTGAAAAACAAATATACGGCAGACTATATTACGTTATACGGTAAGAGTGAGGAAGAGGTTGCGGCT
ATTGGAATGCCCTATGAACGGGTAGCCGACGGATACCAACTGAAGATTCAAAGCACGGCAGAGGCAACAAGAC
TTATAGAGAGTGCTCCGGAGCTCTTTACCGACTATGAGATAGTTAAGGGGAAAATGGACGACGTATTTATTGC
CGCAACGGGCAAGAGTCTTAATACGGACGAAAACACGGGCAAGAAGGG

>Col11

AGACACGCTAATAGTCCGCCACAAAAAGTGTCCCCCGCAGAAGTGGTATCTACCGCTTTAACTTTTATGCAAG
GATAAATTTTGTCTATATTTTCGTTTATTATTTCGTAGCCTTTACTACCTAGCGTTATAATGATATTTTTAAT
ATATTTTTTAAGTTTTCTTTACCGCCAAGTATTTCCGTTTCGGTTTCGTTTGGGATAACGTAGTCGCAATAT
TGTAAATAAGGTAAAATATCTGCGTTTGCAGGGGCAGGGTTTAAAACTACTATAAGCCCTTTTTCTTTTGCTT
TTTTAATGCCGTAACCTATTATATCAATAGGGTTTTCTAGTTGAGTTAAATAATAGTCGCCCGACTTTGCATT
AGATAAAAAGTTGTCTATATCGGTTTTAGATAAGCAATTGTTTGCTCCTTTATCTAAAATTATACGGTTGTCG
CCGTTTTCGATAATAATAACCGCCGTGCCCGTTGGAGTGTTTTTTACTATCCTAACGCAAGAAGTGTTTACCC
CAGCGTTTTTTAGGCTATTTATTGCGTCTATACCAAAACTATCGTCCCCAACGCAACCGCACATAAAAACTTT
TCCGCCAAGTTTGCTACAAGCCACCGCTTGATTTGCACCTTTTCCGCCGTGGGCAGAGAAAAAGTTACTACCC
GTTAAGGTTTCGCCTTTTTTAGGTATATAAGGGGTTGTGATAACTAAATCGGTATTAATGCTACCCGTAATAA
AAACGTTTTTCATTTTGTTTCCTTATGTTGTTATTATGTTTTTTAATGTTCAAAAAACATTTAAAATAAATAA
AGATGGATTAAAGTTATAGTCCATCTTTGTTTTTAAGTTTGTAATAAATAAACCGAAAAACTAATAGGTTATA
AAAAATCCCTAACGAAATATATTAGGGATATTTTTTACTTTTGATATAGTTTTTTAGTTTGATATTTGGCTTC
GCTAGTGTAGTCGATACCAAGTTTTTTAAGCACGTTTATATCTACTTGAGAAAGTATAACGGTAGAGTGTACT
TGCGCGCCTTTTAGTTTTGGCAATTGCTCTAACGCTTTTTTAGCCATTGGGTTTGTAGTTGCGCTAATAGATA
AAGCAATTAAAATTTCGTCGGTATGTAGTCTTGGGTTTTTATTGCCTAAAATTTTAACTTTTAACTCTTGTAT
AGGCTCTATTACGCTTGGCGGAAGTAAAATGACTTTGTCGTCTATCTCCGCTAAATATTTTAAAGAATTTAAT
AGCATTGCGCTAGTTGAGCCAAGTAGGTCGCTAGTTTTACCAGTAATTATAGTGCCGTCGTTTAACTCTATTG
CGGTTGCGGGTTGTCCGCCAGTTTTTGCAGATTTGTTTAGAGCAACGTTTACAACCTTTCTTTCGCTAACGGA
AATAGAGTATTTTTTCCATTAACAACCAAATTTTGTTTATGGTTTCTTCCGTAGTTTTTTTTCTTTAATTTTG
TCGCAAGGGG

>Col15

TGTGCCGTAAGTTCTTGAGTAGATTTCGTTGGTCCAACGGAAGTTGTTAAGTTCGCCGATAGCATCGGAGTGA
ATGAGTTGGAAGTCGTTGGTGACGATATCTCTTGTCATGAAGACAAATTTCTTATTTCTTTGATTATCTAGAG
TATAGTAGTGCCAAATTTCGTAAGGATAAGCTCCTGGCTCATTATGGTTTTGAACTATTCTGTCAGGTGAGCC
GTATTTCAAGAACACGCGACCTCTGTCTGTCAAGTAACCTTTAACTCTTTGTGATGTGAATGATGCATTAACG
CGTTTTACCTGTCCATAATAGTTTTCCCATTCAATTTGAGGTGATAAAGGATTGCGGCTTGCCCAGAAACTAT
AGAAATATTGTTGCATTGTTGTGATGTCGTCTGTTGCTATCAGGTTTTTGGCATATTCGCTTTCAACTTGTGA
TGAAATAGGACTTAATGTTCTTATATATTCTCTAATAGTATCGATGTTGTCTATATTTCCGGAGAATACATTT
TCGGGATTGATAGATGCCAACGCTGTGTTATCTATCTGATAGTTTGAATTGCTTCTTTGGAAAAAGAATTGGT
TGTATCCTATAATCTCGTTATTTCTGTTGCGAGCTTCGAGTACCAAATAATAGTTGCCCGAAGGAAGATTGCT
AATGTCCATTGTATTGATAACTACTTCAGTGTCTTTTACATTCATTCTTTTTGTGAAGTAGTAATTATTGATT
AATGTGTTGCTTTCAAAGGTTTTGAGATAAGTGTTGAGAAGGTATTTCTCATTTTCGCCGAGAGTTTTCTTTG
AGTTATAGATTTCTGCATAGAATGTCAGTCTGTTGATTGACTCAGGATAATAAGGCATTACCATAGGAATAAG
GTCATATCCGTTTTTTGAATTTGCAGATTCTGTTGTGCTAGGTTTATAATCTTCGAGGGCTATGATGCTTGAG
AAGCAAATGTCTTCAGGAAAATCAACTAATATGGTTTCTGAAACGTTTAATGGTTTTTCGGTTTTGTTGTTTG

TATCTTCAATTGTGACTTCCATCTGATATTCGCCATTCGCTAATGAATAACGTTGAACGTCCATGAAGAAACC
TTCTAGATTTGTCGTATCGCTCACTTTAGGGCTTATAACAGTATATTTGTCGTAGTTCTTTATTGCCTCACCT
TGTCTGAATATAACTGTCAAATTTATTGATGCTTCATATTGACCGTTTGCATCTTTTGTGTACATCAGGCTAT
TGCAGTCAAATGTGATGTATGTTTCTAAATAAGGTGTGGCATTCTCACTTGTAATGTTAAATGTGCCGTATGA
AATAAATGCGTTGATATTTTTCTCTTGTGCAACAACCTGAGTTGCTATAAAACTGAGTAATAAAAATGTTAAT
AATTTTTTCATCGTTCTTAAAATTTATTTGCAATGCAAAGATATGCATACTTTAAGAAAAATCAACATTTTTT
ACCTTTGAATTTTATAGAAATGTATCGACAAATTTTTGGTAAATTATTGTAATTTAATAAAATATAGAAAAAA
TTTCTCTAGAATTTCGACTTGATATATGAAGCAATAAACTGGTTTAAATCTTCATTTATGTCAAAAATATTTT
TGATTTTGTTACTTCTTTTGTTTGTAGCGCTATAAGTCAGACCTAAAAGAACTGCTATCTCAACGTTATTAAA
GTTCATCAAATTAAGAATGATGAACTTAAAG

>Col16

ATGGTATCGGCCCGCATGGGGATCGCGGCGTTGCCGCACTGGGTGGTGGAAAGTTTTGAACGCCAGGGGTTAG
TGGTGACCAAAACCCTGGGGGATGGGCTTTGGAGCCGCCTGTACGCGGCGGTACGCGACGGTGAACAGCGCCA
GGCCGTCACCGAAGCCTTTATCCGCTCGGCTCGGCAACATGCCTGCAACCATCTGCCATTCGTGCGTGATGCT
GCACGCCCAAGCGCCGGTGTCGCCAGCGTCAAACCGCTGGCTTCGCAACCCTGAATATCAGGGCAGCTCGCGC
TTGGGCTGCCCGTTTATTTGCTGACTATTTTAGGCGCCACCCAACGGCGATGCACCCACAGCGAAGCGACAAT
CACCGCCCCGCCGATAATGAAACTTGGCCAATGCGGCTGTTGCTGCCAGATGGCCAGATTAACCAACAGTCCT
GCCGGCACATGCATGTTATTCATGATGCCCAAAGTACCGGCATCCACCTGCGTTGCGCCGTAGTTCCACATGA
AATAACCCAGCCCGGAAGCCACTACGCCCAGCCAGACCAGAATGCCCCACTGCAGGTTAGTGGTTGGCAGTTG
CTGCGGATTGCCCCACAGGCACCAGGCGACCACCGCCACGACCAAGGCCCCCAGATAGAACCAGGAAAATGCG
GTGTGTTGCGGCAGCGGATGCACTTCCATCAACCGTTTGTAACCGACCATGCCGATGGCAAAGCTGATGTTCG
CCGCCTGCACCAGCAACAGCCCCCACCAAAAATGCTCGCTCAATTGATGATATCGAATAATCGCAGCGCCCAG
CACTGCCAATAGCGCACTGAACACATAACCCCAGCGCAGGCGCTGTCCGCTGATCAAGTCATAAATCAGCGTA
ACGTACAGCGGCGTCAGCACGGTAAACAGCAGGAATTCCGGCACCGTCAGGTACATGTACGCGCGGAAGCTGA
ACAGATACATGATGCCAAGCTGGCAAGCACCCACCAGCATGTACAACAGGATCACTTTTGGCTTGATGTTGCG
CCAGCGCAGGAAAGGCAGGAACACCGCCGCCGCCAGGCCGATGCGCATCAGCACCGAGAACCAGCTGTCGACC
TGGCCCGCCAGGTATTCGCCAATCAGGCTGAAAGAAAATGCCCACAGGATGGTAGTAATCACCAGAAGCAGCA
CGATGGGTTCACCCTAAAGATTCAGAAAGGGCGATTGTAACGGAAGGCAGAAAGGGGTGTAGCGAAAAGAAGG
TTTACATCTGAGCAAAATACAAACTTTCAGCAAATTTGCACCGCTGGGTTAGGCTTAATAAAAGACGAGGAGA
ACATCATGAAAAGTTATCATCTGCTCTATGGCCTGTTGCTGCTGTCCGGATTGGCCACCGCAGCCAACGGCGT
CACCGTGCTCAGCTCGCAACAGCAGTTTCTCAACCAACGCACTCAGCAGACGCAACAACAGCGTATGTTGAAT
AATCAGCAGCTGGATCAGCGCCGCCTGCAACAACAGCGGCAGTTCGATAATAACCAGCAG

>Col17

GCGCACACGCGTACGACTCGTCCGTCCTCGGCCTGTAGGAATTCGCCATCGTGCAGCGGTGGCTGGCCGCGCT
CCAGGAAAAGGCCGACGTCTTCACCGTCGGCACTGAAACAGCGCAGGCGGCTTTTGCTGCGGGCTTCGAAGTT
CAGCAGCAACTCGGCGGCCCAGAGGGCTTGGGGGGCGATTCGGCGGTGGATCACCAGCATCAGGAAGCTTCCA
GCTATGAGCGATGAAGCAGCTAGAGCAAGGGGCTTGCCAACCTCGATGGGGAGTAGGAATTGCCTGTAGGCAA
GCGCGGCATGCCACTAAACAGGGCGAGGAAAAATGCTCGACGTGCGCCAAAGTGGTGCGCAATGTTGCGTGTT
GCACGCGTGATGGGCATTTTCGGATTTTCCCTACACGTTATGGGGAATCGGCGTTCATCACCTTACCTGCGAG
ATCATTAGGCTGTGCGCCGTGTTTAATTCACTGCGACGTTCATCATGTTCAAATCATTTTGTTGCCTGGTTAT
TGTCAGCCTATCTCTCGCGCTTTCATCGGCTTCCGCGAATTTGAAGTCCCATCCGTCCTTCAGCACTTCGCTG
AAGCAAGCGTCTATCGAGGATGTGGTCGACCGCGCCCACGAACTGCTGGGTACGCCTTACAAGTGGGGTGGCA
CCTCGGCGGACCAGGGTTTTGATTGCAGCAGCCTGCTGGTCTATCTGTTCAAGACCGAAGCCAATATCCAGTT
ACCGCGCACCACTGCGGCCATGCAACGGTCAACAGCAGCCACTATCACGCGCAATGCGCTCAAGGCGGGTGAT
GCGGTGTTTTTCAAAGGAAATGGGGGCGGACAAGTCAGCCATGTGGGCCTTTATATAGGCGAGGGCAAATTTA
TCCATTCACC

>Col18

TGCGCGCCAGCTCCACCAGCTTCTGCTGGCCGCCCGACAGTTCCTCGGCGTAGTTGTGGCGCAGGCGATGCAG
CTCCACCTTGTGCAGCCAGTACTCGGACTGCGCATAGCGTTCGTCCGCGCTGCGGCCCGCCTGCTGCTGCGCC
ACGTCCAGGTTGTCCAGCAGGCTGAGCTGTCGGAAGATGCGCGGCACCTGAAAAGTACGCCCGACGCCCAGCC
GGGCCACTTCATGAACCTGCCGCCCGCCGATCTCGCGGCCGTCCAGTTCGATGCGGCCTTCGTCCGGCGTGTA
GATGCCATTGATGCAATTGAGCATCGTGGTCTTGCCTGACCCGTTCATGCCGATGATGCCCAGGATGGACCCC
TTCGGCACCTCCAGATCCACGCCCTGCAGCGCCACCAGGCCGCCGAAGCGCTTCTTGAGATTCGTGACTTTCA
ACATGCGTCCGTCTCCCCGGGCCGGTTAACGCAGGGCCTGGCGGCCGAACAAGCCGCTGGGCCGGATGAAGAT
CGTGGCGACCATCAGCGCGAAGCCGCAGATGGTGGCGATGGAAGGGCTGAAGAACACCACCGACACCTGTTCC
GTCACTCCGAACAACAGCGCCCCGACCAGGGCGCCCACCGGATGGCCCAGGCCGCCCAGGACGATGACGATGA
AGCCGATCAGCGTGTAGTCGTTGCCCATGAACGGCGAGAACGCCGGGAACACCATGGCGACCAGCACGCCGCT

GGCGGCCGCCAGGCCCACGCCCAGGCCGAACGCCACGGCCGACAGCCGTTCCGCGTTGATGCCCACGATGCGG
ACCGCATCGCGGTTCATGATGATGGCGCGCAGCGCGCGGCCAAGCTGCGTGCGGTACAGGAACACCGTCAGCG
CGGCGATGATCGCCAGCGCGATGACCAGCGCCAGCAGCCGCACCGTGGGGATCAGCACCGGACCGAACTTCAG
GAAGACCGGTTCGATCTCGTAGGGCAGCGAACGCGCGTCGGCGTCGAACGCCAGCAGCATCAGCGCGCCCAGC
ATGATGGATACGCCGAACATCAGCAGCAGCGAACTGATCTCGGGGTCGTCCGCCTTGGACAGCCGCGGAATCA
GGACGAAGTACAGCAGGTAGCCGGCCAGGAAGAAGACCGCGAACGCCAGCGGCAGCGCAATCAGCGGGTTGAT
GCCCGCGTACTCCATCAGGTAGAACGCGAAGTACGCGCCCAGCACCAGGAATTCCCCGTGCGCGAAGTTCACC
ACGCGCAGCACGCCGAAGACGACGTTCAGGCCCACGCCGATCAGGCCGTAGATGGCGCCGAGGATCAGGCCGT
TGATCAGGCCTTGGATAAGCAGTCCGGTCATGGCGGGTCCTCAGGCGATGTAGCGGCGCAGTTGCGGCTTGCG
TTTGAGCAGGCTGCCGACGATGCCCCGCGGCAGGAACAGCATCAGCAGGACGATGATGAGGCCCAGCAGCAGC
AGGTTCACCACCGGGAATTTCTGCCAGATCATGTGGTCGATGCCCAGCAGCGCCAGCGTGCCGACCACGGGGC
CGAGGATGG

Supplementary List S2

Protein sequences of ORF from metagenomic hit Col6 identified as membrane protein through BLAST, YigM from *E. coli* and BioY from *Rhodobacter capsulatus*

>Col6_membrane_protein

VLLLVITTILWAFSFSLIGEYLAGQVDSWFSVLMRIGLAAAVFLPFLRWRNIKPKVIL
LYMLVGACQLGIMYLFSFRAYMYLTVPEFLLFTVLTPLYVTLIYDLISGQRLRWGYVF
SALLAVLGAAIIRYHQLSEHFWWGLLLVQAANISFAIGMVGYKRLMEVHPLPQHTAFS
WFYLGALVVAVVAWCLWGNPQQLPTTNLQWGILVWLGVVASGLGYFMWNYGATQVDAG
TLGIMNNMHVPAGLLVNLAIWQQQPHWPSFIIGGAVIVASLWVHRRWVAPKIVSK*

>E_coli_YigM

VALLIITTILWAFSFSFYGEYLAGHVDSYFAVLVRVGLAALVFLPFLRTRGNSLKTVG
LYMLVGAMQLGVMYMLSFRAYLYLTVSELLLFTVLTPLYITLIYDIMSKRRLRWGYAF
SALLAVIGAGIIRYDQVTDHFWTGLLLVQLSNITFAIGMVGYKRLMETRPMPQHNAFA
WFYLGAFLVAVIAWFLLGNAQKMPQTTLQWGILVFLGVVASGIGYFMWNYGATQVDAG
TLGIMNNMHVPAGLLVNLAIWHQQPHWPTFITGALVILASLWVHRKWVAPRSSQTADD
RRRDCALSE*

>Rhodobacter_capsulatus_BioY

MERNVTLIGLFAALIVALGFVPAIPLGFGVPITLQSLGVMLAGAVLGSWRGALAVLLV
QALVAIGLPVLAGGRGGLGIFVGPTAGFLIGWLPAAFVTGLIVERLRRVPVALAAGIG
ATLGGIIIMYALGILGFWLVKNAGLKPEDAPISLWAATAIMAPFIPGDLVKVVVTGLV
ARTIAQYRPSALLARG

CHAPTER IV
DIRECTED EVOLUTION OF MEMBRANE TRANSPORT USING SYNTHETIC SELECTIONS

ACS SyntheticBiology
Cite This: ACS Synth. Biol. 2018, 7, 789–793
Letter
pubs.acs.org/synthbio

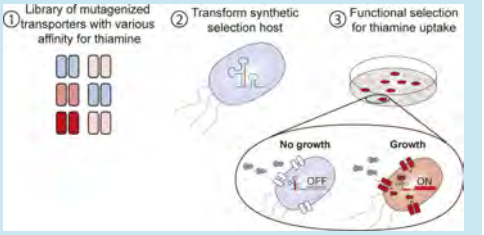
Directed Evolution of Membrane Transport Using Synthetic Selections

Anne P. Bali,[†,‡,#] Hans J. Genee,[‡,#] and Morten O. A. Sommer*[,†]

[†]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[‡]Biosyntia ApS, Fruebjergvej 3, DK-2100, Østerbro, Denmark

Supporting Information

ABSTRACT: Understanding and engineering solute transporters is important for metabolic engineering and the development of therapeutics. However, limited available experimental data on membrane transporters makes sequence-function relationships complex to predict. Here we apply ligand-responsive biosensor systems that enable selective growth of *E. coli* cells only if they functionally express an importer that is specific to the biosensor ligand. Using this system in a directed evolution framework, we successfully engineer the specificity of nicotinamide riboside transporters, PnuC, to accept thiamine as a substrate. Our results provide insight into the molecular determinants of substrate recognition of the PnuC transporter family and demonstrate how synthetic biology can be deployed to engineer the substrate spectrum of small molecule transporters.



KEYWORDS: *synthetic biology, biosensors, selection, directed evolution, structural biology, membrane transport*

The *in vivo* characterization of solute membrane transporters typically involves labor-intensive experimental protocols and complex radiolabeled substrates.[1] Unfortunately, *in silico* predictions of transporter function, in particular, substrate specificity, are also challenging, and this is, in part, due to a lack of experimental data.[1,2] Consequently, functional data and accurate annotations of membrane transporters and their activities are lagging behind that of enzymes. This gap in our knowledge of transporters challenges the engineering of microbial cell factories and limits the development of therapeutics that target transporters. We previously developed a synthetic selection system for transporters using ligand-specific RNA biosensors that control the expression of selectable markers (Figure 1A). Using this system, we robustly coupled ligand uptake to bacterial growth enabling the functional mining of novel vitamin transporters from the metagenome.[3] Here, we extend this method and show that synthetic selections can be applied to probe the sequence-function relationship of small molecule transporters and engineer functionalities, such as substrate acceptance. Using this approach we functionally identified key residues that control the substrate recognition of the nicotinamide riboside transporter (PnuC) and expand its substrate recognition to include thiamine (Figure 1B).

The Pnu family of transporters is broadly distributed among bacteria and facilitates the uptake of structurally diverse B-vitamins, as experimentally demonstrated for riboflavin (PnuX),[4] nicotinamide riboside (PnuC),[5] and thiamine (PnuT).[3,6] Similarly to PnuX and PnuC it was recently shown that PnuT mediates thiamine uptake by facilitated diffusion.[6] Directionality is achieved by metabolic trapping as thiamine is immediately phosphorylated in the cell and as a consequence can no longer serve as substrate for PnuT. The pnu transporters provide an important route to bypass the biosynthesis of essential B-vitamins through direct import to save energy and resources. Furthermore, Pnu transporters are essential for many bacteria and higher organisms, in which a complete *denovo* pathway does not exist. For example does PnuT appear to be the only thiamine transporter in several pathogens, including *Bacteroides fragilis and Helicobacter pylori,* of which the latter is an experimentally validated thiamine auxotroph.[3] Similarly, PnuC is essential for *Haemophilus influenzae* growth and virulence, as *H. influenzae* does not have the enzymes necessary for *denovo* synthesis of NAD^+.[7] For biological production of B-vitamins, the Pnu-family of transporters has also shown to be important: overexpression of PnuX in an industrial production strain of *Bacillus subtilis* facilitated riboflavin export and led to increased riboflavin titers during fermentation.[8]

The crystal structure of PnuC from *Neisseria mucosa* provides structural insight into this important class of transporters.[9] Guided by this, a recent study by Jaehme et al. experimentally identified residues involved in substrate binding and gating of PnuT.[6] To extend this analysis, we performed a sequence and structural comparison of 5 experimentally validated PnuC transporters and 26 experimentally validated PnuT transporters (Figure 2A,B). The structural homology modeling in Figure 2A

Received: November 13, 2017
Published: February 23, 2018

DOI: 10.1021/acssynbio.7b00407
ACS Synth. Biol. 2018, 7, 789–793

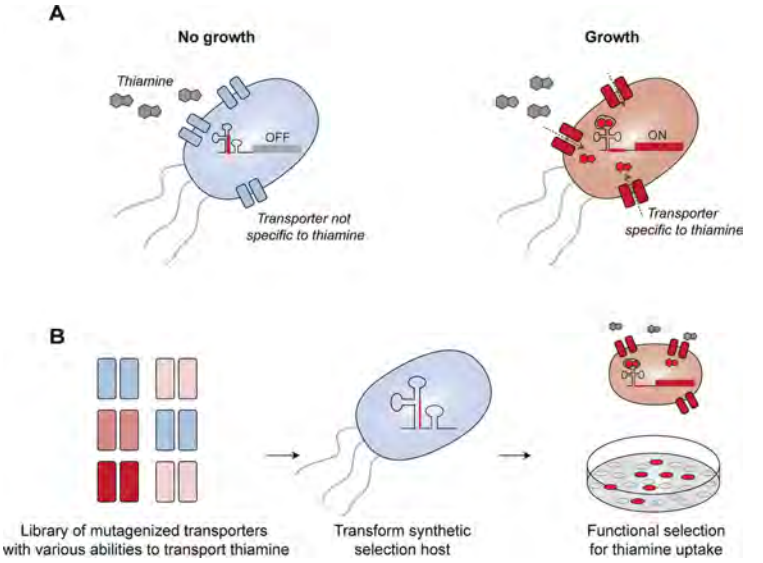


Figure 1. Synthetic selections for elucidating the sequence-function relationship of solute transporters. (A) The principle for the selection system of thiamine transport. A thiamine pyrophosphate (TPP) dependent riboswitch controls the expression of an antibiotic marker gene. When thiamine cannot be transported into the cell (left), the ribosome binding site (RBS) (red line) is inaccessible for the ribosome, and the strain will not be resistant toward antibiotic selection, resulting in no growth for strains lacking thiamine transport (blue transporter and blue cells). When the cell is capable of transporting (red transporter), thiamine will be phosphorylated to TPP (red structure) inside the cell and bind to the riboswitch. The binding triggers a conformational change making the RBS accessible for the rioboswitch, leading to expression of the antibiotic resistance gene and allowing growth under antibiotic selection conditions (red cell). (B) Selection strategy based on the functional selection for thiamine uptake among established libraries of mutagenized PnuC transporters.

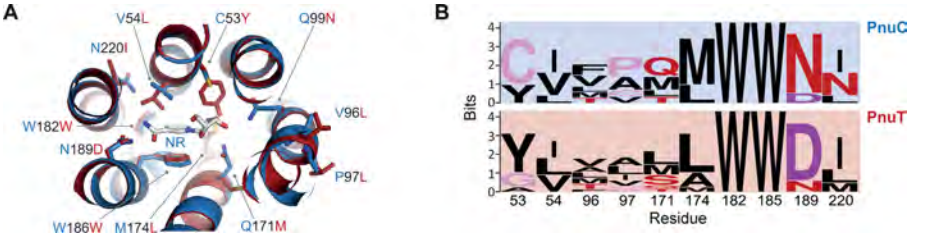


Figure 2. Structural characterization of Pnu transporters. As a foundation for choosing a mutagenesis strategy to change transporter specificity, structural examinations of various Pnu-transporters were carried out. On the basis of the investigations, the amino acid compositions of key residues in the different Pnu-transporters were compared to direct rational mutagenesis and understand the determinants of substrate specificity. (A) The crystal structure of the PnuC nicotinamide transporter substrate-binding site with nicotinamide riboside (NR) from *N. mucosa* (blue) aligned with the homology model of the PnuT thiamine transporter from *B. fragilis* (red). (B) Logo-plots of residues of the PnuC substrate-binding site. The logo-plot is based on a protein alignment of functionally validated PnuCs and PnuTs. The color codes of the amino acids indicate the properties as follows: charged side chains (dark purple); polar uncharged side chains (red); hydrophobic side chains (black); and special cases (light purple). PnuT from *B. fragilis* was chosen as a representative sequence for homology modeling, based on its phylogenetic localization in a clade of validated PnuT transporters.[3]

was based on the sequence of the experimentally validated PnuT transporter from the pathogenic bacteria *B. fragilis*[3] guided by the crystal structure of *N. mucosa* PnuC[10] (see Methods), to enable visual inspection of amino acids of interest. However, a lack of conserved sequence motifs in specific subtypes of the Pnu transporters (Figure 2B) meant that the molecular basis of the differential substrate specificities remained elusive.[11]

In this study, we used random mutagenesis combined with synthetic selections to investigate residues important for the substrate specificity of Pnu transporters. For our analysis, we chose the nicotinamide riboside transporter, PnuC, from *Streptococcus pneumoniae*, which does not exhibit thiamine transport activity *in vivo*.[3] We generated a PnuC mutant library by error-prone PCR and transformed this into an *Escherichia coli* strain harboring a riboswitch-based thiamine selection system which only allows growth of mutant variants that enable

DOI: 10.1021/acssynbio.7b00407
ACS Synth. Biol. 2018, 7, 789−793

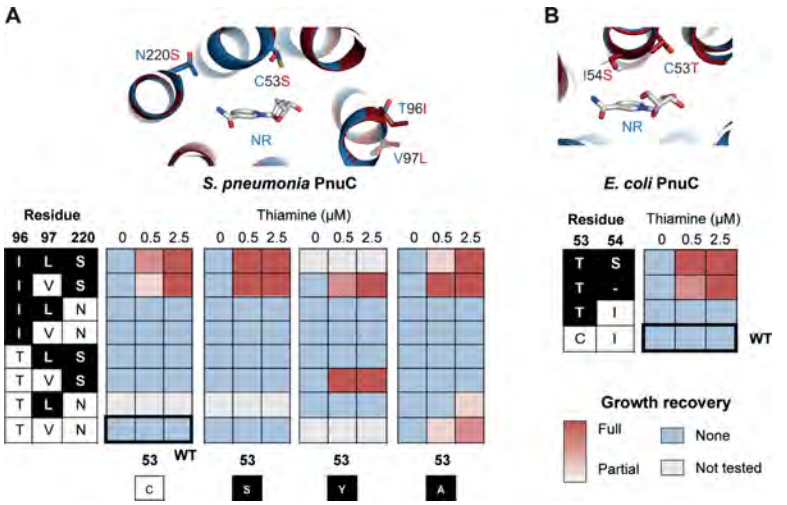


Figure 3. Elucidating the determinants of thiamine substrate specificity by directed evolution. To investigate the sequence determinants of thiamine transport we performed random and site-directed mutagenesis of PnuC nicotinamide riboside transporters from *S. pneumonia* and *E. coli*, respectively. Using thiamine functional selections, the mutant libraries were assayed for variants that enable thiamine uptake when expressed in *E. coli*. (A) Functional analysis of the PnuC mutants from *S. pneumonia* (*N. mucosa* numbering) with mutations at residues 53, 96, 97, and 220. "WT" indicates the wild type PnuC. The mutant proteins were expressed in *E. coli* harboring the selection system, and the thiamine transport of each strain was assayed by spotting 10× serial dilutions of an overnight culture on selective growth medium supplemented with 0.0, 0.5, or 2.5 μM thiamine. The figure above the heatmap represents a comparison of the homology models of an *S. pneumonia* wild type (blue) and a mutant (red) PnuC. (B) *E. coli* PnuC variants with mutations at residues 53 and 54 (*N. mucosa* numbering). The dash (-) indicates a deletion. The figure above the heatmap represents a comparison of the homology models of an *E. coli* wild type (blue) and a mutant (red) PnuC.

thiamine uptake (Figure 1B).[3] This enabled the rapid functional interrogation for the acquired ability to take up thiamine among 6×10^5 PnuC variants (see Methods).

Interestingly, several colonies that are able to import thiamine were selected, and among these, one showed a particularly strong growth phenotype. The sequencing of *pnuC* isolated from this clone revealed four mutated residues, comprising Cys53Ser, Thr96Ile, Val97Leu, and Asn220Ser (*N. mucosa* numbering). Of these, only Cys53 is predicted to directly interact with the substrate by forming a hydrogen bond to the 2′-OH group of the ribose unit of nicotinamide riboside in PnuC from *N. mucosa*.[10] This is likely to be true for *S. pneumoniae* PnuC as well, based on the structural homology modeling (Figure 3A, top). Our comparative analysis of Pnu sequences showed that Cys53 is furthermore highly conserved among the Pnu transporters specific to nicotinamide riboside, while the Pnu transporters specific to thiamine display variability, with the majority of sequences containing a Tyr at position 53 (Figure 2B). Asn220 is predicted to be proximal to the substrate-binding site but is not conserved, whereas Thr96 and Val97 are positioned at the exterior of the protein facing the lipid bilayer for both PnuC and PnuT and are also not conserved (Figure 2A,B). Whereas the mutation of Asn220 and in particular Cys53 was expected to impact substrate binding due to the proximity of these residues to the binding pocket, mutation of the more distant Thr96 and Val97 was highly surprising.

To dissect which of the four residues (Cys53, Thr96, Val97, and Asn220) are required for thiamine acceptance by *S. pneumonia* PnuC, we constructed several mutant transporters with combinations of wild-type and mutant residues. In addition to Ser, Cys53 was mutated to Tyr and Ala, as these are found in Pnu transporters specific to thiamine (Figure 2B). In total, we successfully generated 28 of 32 possible PnuC mutant variants (Figure 3A). The ability of each variant to facilitate thiamine uptake was evaluated by thiamine-dependent uptake assays using the thiamine selection assay strain (Figure 3A).

The thiamine-dependent growth selections showed that the Asn220Ser and surprisingly the Thr96Ile combination is critical for thiamine transport (Figure 3A and Supplementary Figure 1B). When combined with the Asn220Ser and Thr96Ile mutations, both Cys53Ala and Cys53Tyr generated strong thiamine uptake phenotypes, similar to the Cys53Ser combinatorial mutant (Figure 3A and Supplementary Figure 1B). Strikingly, the introduction of Cys53Ala alone enabled thiamine transport. Although the resulting phenotype of this variant is not as strong as that of the combinatorial mutants, it demonstrates that even a single residue mutation is capable of changing the substrate specificity of a Pnu transporter.

To further explore the effect on substrate specificity by changing the conserved Cys53 (Figure 2B), we generated a site saturation mutational library of *E. coli* PnuC at Cys53. *E. coli* PnuC is distantly related to *S. pneumoniae* PnuC and shows less than 20% sequence identity (and less than 30% sequence similarity) on the amino acid level. The resulting PnuC mutant library was transformed into an *E. coli* thiamine selection strain and was assayed for thiamine uptake by thiamine-dependent growth selections as described previously. An analysis of the selected PnuC sequences revealed two unique variants (Supplementary Figure 2). One variant encoded a Cys53Thr substitution with a concomitant Ile54 deletion, and another

DOI: 10.1021/acssynbio.7b00407
ACS Synth. Biol. 2018, 7, 789−793

clone encoded a Cys53Thr and an Ile54Ser substitution. Since mutagenic PCR primers were designed for Cys53 mutagenesis only, the Ile54 mutations were likely introduced as an error during PCR but were identified due to the high-throughput capacity of the selection system. On the basis of the growth assays at selective conditions with and without thiamine, the functionally selected PnuC variants displayed a thiamine uptake performance that was similar to the positive control (*H. pylori* PnuT expressed in *E. coli* (Figure 3B and Supplementary Figure 1A)). No change in the phenotype was observed when Cys53 was mutated alone (Figure 3B and Supplementary Figure 1A).

In summary, we have demonstrated how synthetic biology can effectively be deployed to solve complex biological challenges related to substrate specificity of small molecule transporters. Using synthetic selections our investigations led to the identification of several PnuC variants with an expanded substrate range that now includes thiamine. Whether the ability to import nicotinamide riboside is maintained or is reduced was not analyzed as part of this study, but additional biochemical analyses will likely uncover such properties and further expand our understanding of the sequence-function relationship of the Pnu transporter family. Similarly, further mutational and structural studies could be conducted to mechanistically explain the change in substrate specificity of the selected mutant variants. The majority of previous efforts toward engineering substrate recognition of membrane transporters have achieved changes in specificity to molecules that are structurally related, such as purines[12] and hexose and pentose sugars.[13−17] In contrast, thiamine and nicotinamide riboside are structurally highly diverse molecules with distinct chemical properties. Our finding that even a single residue enables such a change in substrate specificity demonstrates an impressive degree of sequence-function flexibility of the Pnu transporter family. While our functional investigation confirmed residue 53 as being central to substrate specificity as suggested by the structural analysis,[10] our broad analysis illustrates that multiple combinations of residue substitutions even distant to the binding pocket are important to substrate specificity. Such nonintuitive results would have been challenging to obtain without the robust selection system used in this study, highlighting that synthetic selections represent a powerful approach to elucidate complex genotypic−phenotypic relations of membrane transporters. By tailoring synthetic selection systems to new compounds of interest through the replacement of the biosensor module, the approach is expandable to other important compound and transporter families.

METHODS

Materials and General Considerations. Plasmid DNA and PCR products were purified using the QIAprep Spin Miniprep Kit and QIAquick PCR Purification Kit, respectively (Qiagen). Gel extractions were performed using NucleoSpin Gel and PCR clean-up protocol (Macherey−Nagel). Thiamine−HCl, thiamine monophosphate−HCl, thiamine pyrophosphate−HCl, and all antibiotics used herein were purchased from Sigma-Aldrich. Synthetic oligonucleotides were purchased from Integrated DNA Technologies. *Escherichia coli* DH10B was used for all the experiments. For all the incubations in liquid media, the cells were grown with shaking at 250 rpm at 37 °C. The strains were stored at −80 °C in a 15% v/v glycerol solution. The plasmid manipulations were performed using USER cloning.[18,19] The sequences of all the vectors constructed were verified by DNA sequencing (Beckman Coulter). The cells were cultured in Luria Broth (LB), or modified rich MOPS medium without thiamine (mrMOPS) (Supporting Information Note 2).

Structural Comparisons of the PnuC and PnuT Proteins. The PnuC and PnuT homology models were computed using the HHPred prediction server[20] with the default settings. The *Neisseria mucosa* PnuC crystal structure (PDB entry 4qtn)[10] was used as a template for generating all the homology models, and the structures were visualized and analyzed using PyMOL v1.7.[21] To identify the conserved residues and motifs, a multiple sequence alignment of 26 functionally verified PnuTs[3] in addition to PnuT from *Helicobacter pylori*, as well as the five functionally validated PnuCs (*E. coli, Haemophilus influenzae, N. mucosa, Salmonella enterica, Streptococcus pneumoniae*), was computed using CLC Main Workbench 6. Ten positions, of which eight are predicted to interact with nicotinamide riboside in *N. mucosa* PnuC,[10] were selected and presented as a sequence logo plot generated using WebLogo.[22] For the logo plot visualization, representative sequences from 11 phylogenetically different groups of the 26 validated PnuT sequences[3] were used in order avoid biasing the logo plot.

Error-Prone PCR Mutagenesis. Random mutagenesis of *S. pneumoniae pnuC* by error-prone PCR was performed with the GeneMorphII Random Mutagenesis Kit (Agilent Technologies, No. 200550) using the primers oGEN292/oGEN291. pGEN64 encoding *S. pneumoniae pnuC* was used as the template for the backbone amplification with the PCR primers oGEN308/oGEN320. The generated *pnuC* mutational PCR fragment library was assembled with the pGEN64 backbone by Gibson Assembly following the manufacturer's instructions. The ligation mix was desalinated using 0.025 μM CSWP membrane filters (Merck Millipore Ltd.) and was subsequently transformed into electrocompetent *E. coli* DH10B cells already harboring the pGEN37 thiamine selection plasmid. The resulting cell libraries were washed twice in mrMOPS and assayed for mutated *S. pneumoniae* PnuC variants capable of transporting thiamine through plating on mrMOPS agar plates with 0 to 5 μM thiamine. EcGEN186 (encoding the *S. pneumoniae* wild type *pnuC*) was used as a negative control.

Construction of *S. pneumoniae* PnuC Combinatorial Variants by Multiple Site-Directed Mutagenesis. To explore the combinatorial effects of the Cys53Ser, Thr96Ile, Val97Leu, and Asn220Ser mutations identified in the isolated *S. pneumoniae* PnuC variant capable of transporting thiamine, we used the USER cloning strategy for multiple-site directed mutagenesis.[19] Using the AMUSER Web server,[19] a set of three semidegenerate primer sets were designed for the combinatorial introduction of Ala/Cys/Ser/Tyr at position 53 (primers oGEN306/308), Thr/Ile at position 96, Val/Leu at position 97 (primers oGEN323/324), and Asn/Ser at position 220 (primers oGEN309/310). Following PCR amplification, the three fragments were assembled to generate a library that covered all 32 combinatorial variants. Following ligation, the library was transformed into electrocompetent *E. coli* DH10B harboring the pGEN37 selection plasmid. We sequenced 264 randomly selected transformants and identified 28 unique combinatorial PnuC mutants. The capability of the individual PnuC mutants to transport thiamine was assayed by spot assays as described above under "Characterization of specific Pnu transporters" using 0.0, 0.5, and 2.5 μM thiamine in the agar plates.

DOI: 10.1021/acssynbio.7b00407
ACS Synth. Biol. 2018, 7, 789−793

Site-Saturation Mutagenesis. Site-saturation mutagenesis of Cys53 (*N. mucosa* numbering) of *E. coli* and *S. pneumoniae* PnuC was performed using the USER cloning strategy for site-directed mutagenesis.[19] Mutagenic primers containing NNN degenerate nucleotides at the Cys53 codon were designed using the AMUSER-Web server.[19] pGEN62, encoding *E. coli pnuC*, and pGEN64, encoding *S. pneumoniae pnuC*, were used as the templates for whole plasmid amplifications with the PCR primers oGEN321/oGEN322 and oGEN308/oGEN320 for the two plasmids, respectively. Following USER assembly and ligation,[19] the ligation mixes were desalinated using 0.025 μm CSWP membrane filters (Merck Millipore Ltd.) and were subsequently transformed into electrocompetent *E. coli* DH10B cells already harboring the pGEN37 selection plasmid. The resulting cell libraries were grown, stored, and assayed for that mutated PnuC variants capable of transporting thiamine as performed for the metagenomic libraries described above. EcGEN184 (encoding the *E. coli* wild-type *pnuC*) and EcGEN186 (encoding the *S. pneumoniae* wild-type *pnuC*) were used as negative controls for all the selections. The capability of the individual PnuC mutants to transport thiamine was assayed by the spot assays as described above under "Characterization of specific Pnu transporters," using 0.0, 0.5, and 2.5 μM thiamine in the agar plates.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.7b00407.

Additional figures and notes as described in the text (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: msom@biosustain.dtu.dk.

ORCID

Morten O. A. Sommer: 0000-0003-4005-5674

Author Contributions

#A.P.B. and H.J.G. contributed equally. A.P.B. designed and performed experiments and wrote the paper. H.J.G. designed and performed experiments, supervised and wrote the paper. M.S. supervised and wrote the paper.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge Eric van der Helm for help with structural analysis and homology modeling. The work was supported by Innovation Fund Denmark (Industrial Ph.D. program, No. 5016-00135B), the Novo Nordisk Foundation, and the European Union Seventh Framework Programme (FP7-KBBE-2013-7-single-stage) under Grant Agreement No. 613745, Promys.

■ REFERENCES

(1) Lacapère, J.-J., Pebay-Peyroula, E., Neumann, J.-M., and Etchebest, C. (2007) Determining membrane protein structures: still a challenge! *Trends Biochem. Sci. 32*, 259−70.
(2) Nugent, T., and Jones, D. T. (2012) Membrane protein structural bioinformatics. *J. Struct. Biol. 179*, 327−337.
(3) Genee, H. J., Bali, A. P., Petersen, S. D., Siedler, S., Bonde, M. T., Gronenberg, L. S., Kristensen, M., Harrison, S. J., and Sommer, M. O. A. (2016) Functional mining of transporters using synthetic selections. *Nat. Chem. Biol. 12*, 1015−1022.
(4) Vogl, C., Grill, S., Schilling, O., Stülke, J., Mack, M., and Stolz, J. (2007) Characterization of riboflavin (vitamin B2) transport proteins from Bacillus subtilis and Corynebacterium glutamicum. *J. Bacteriol. 189*, 7367−75.
(5) Kemmer, G., Reilly, T. J., Schmidt-brauns, J., Zlotnik, G. W., Green, B. A., Fiske, M. J., Herbert, M., Kraiß, A., Smith, A., and Reidl, J. (2001) NadN and e (P4) Are Essential for Utilization of NAD and Nicotinamide Mononucleotide but Not Nicotinamide Riboside in Haemophilus influenzae. *J. Bacteriol. 183*, 3974−3981.
(6) Jaehme, M., Singh, R., Garaeva, A. A., Duurkens, R. H., and Slotboom, D. J. (2018) PnuT uses a facilitated diffusion mechanism for thiamine uptake. *J. Gen. Physiol. 150*, 1−10.
(7) Herbert, M., Sauer, E., Smethurst, G., Krai, A., Hilpert, A., and Reidl, J. (2003) Nicotinamide Ribosyl Uptake Mutants in Haemophilus influenzae. *Infection and Immunity 71*, 5398−5401.
(8) Hemberger, S., Pedrolli, D. B., Stolz, J., Vogl, C., Lehmann, M., and Mack, M. (2011) RibM from Streptomyces davawensis is a riboflavin/roseoflavin transporter and may be useful for the optimization of riboflavin production strains. *BMC Biotechnol. 11*, 119.
(9) Jaehme, M., and Slotboom, D. J. (2015) Structure, function, evolution, and application of bacterial Pnu-type vitamin transporters. *Biol. Chem. 396*, 955−966.
(10) Jaehme, M., Guskov, A., and Slotboom, D. J. (2014) Crystal structure of the vitamin B3 transporter PnuC, a full-length SWEET homolog. *Nat. Struct. Mol. Biol. 21*, 1013−1015.
(11) Jaehme, M., and Slotboom, D. J. (2015) Structure, function, evolution, and application of bacterial Pnu-type vitamin transporters. *Biol. Chem. 396*, 955−966.
(12) Amillis, S., Koukaki, M., and Diallinas, G. (2001) Substitution F569S converts UapA, a specific uric acid-xanthine transporter, into a broad specificity transporter for purine-related solutes. *J. Mol. Biol. 313*, 765−774.
(13) Young, E. M., Tong, A., Bui, H., Spofford, C., and Alper, H. S. (2014) Rewiring yeast sugar transporter preference through modifying a conserved protein motif. *Proc. Natl. Acad. Sci. U. S. A. 111*, 131−6.
(14) Farwick, A., Bruder, S., Schadeweg, V., Oreb, M., and Boles, E. (2014) Engineering of yeast hexose transporters to transport D-xylose without inhibition by D-glucose. *Proc. Natl. Acad. Sci. U. S. A. 111*, 5159−64.
(15) Nijland, J. G., Shin, H., de Jong, R. M., de Waal, P. P., Klaassen, P., and Driessen, A. (2014) Engineering of an endogenous hexose transporter into a specific D-xylose transporter facilitates glucose-xylose co-consumption in Saccharomyces cerevisiae. *Biotechnol. Biofuels 7*, 168.
(16) Reinders, A., Sun, Y., Karvonen, K. L., and Ward, J. M. (2012) Identification of amino acids important for substrate specificity in sucrose transporters using gene shuffling. *J. Biol. Chem. 287*, 30296−30304.
(17) Reider Apel, A., Ouellet, M., Szmidt-Middleton, H., Keasling, J. D., and Mukhopadhyay, A. (2016) Evolved hexose transporter enhances xylose uptake and glucose/xylose co-utilization in Saccharomyces cerevisiae. *Sci. Rep. 6*, 19512.
(18) Bitinaite, J., and Nichols, N. M. (2009) DNA cloning and engineering by uracil excision. *Curr. Protoc. Mol. Biol. Chapter 3*, Unit 3.21.10.1002/0471142727.mb0321s86.
(19) Genee, H. J., Bonde, M. T., Bagger, F. O., Jespersen, J. B., Sommer, M. O. a, Wernersson, R., and Olsen, L. R. (2015) Software-supported USER cloning strategies for site-directed mutagenesis and DNA assembly. *ACS Synth. Biol. 4*, 342.
(20) Söding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res. 33*, 244−248.
(21) Delano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Sci. LLC, San Carlos, CA.
(22) Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: A sequence logo generator. *Genome Res. 14*, 1188−1190.

Directed evolution of membrane transport using synthetic selections

Anne P. Bali#[1,2], Hans J. Genee#[2], Morten O. A. Sommer[1]*

 # Denotes equal contribution

Affiliations

[1] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hørsholm, Denmark. [2] Biosyntia ApS, Fruebjergvej 3, DK-2100, Østerbro, Denmark

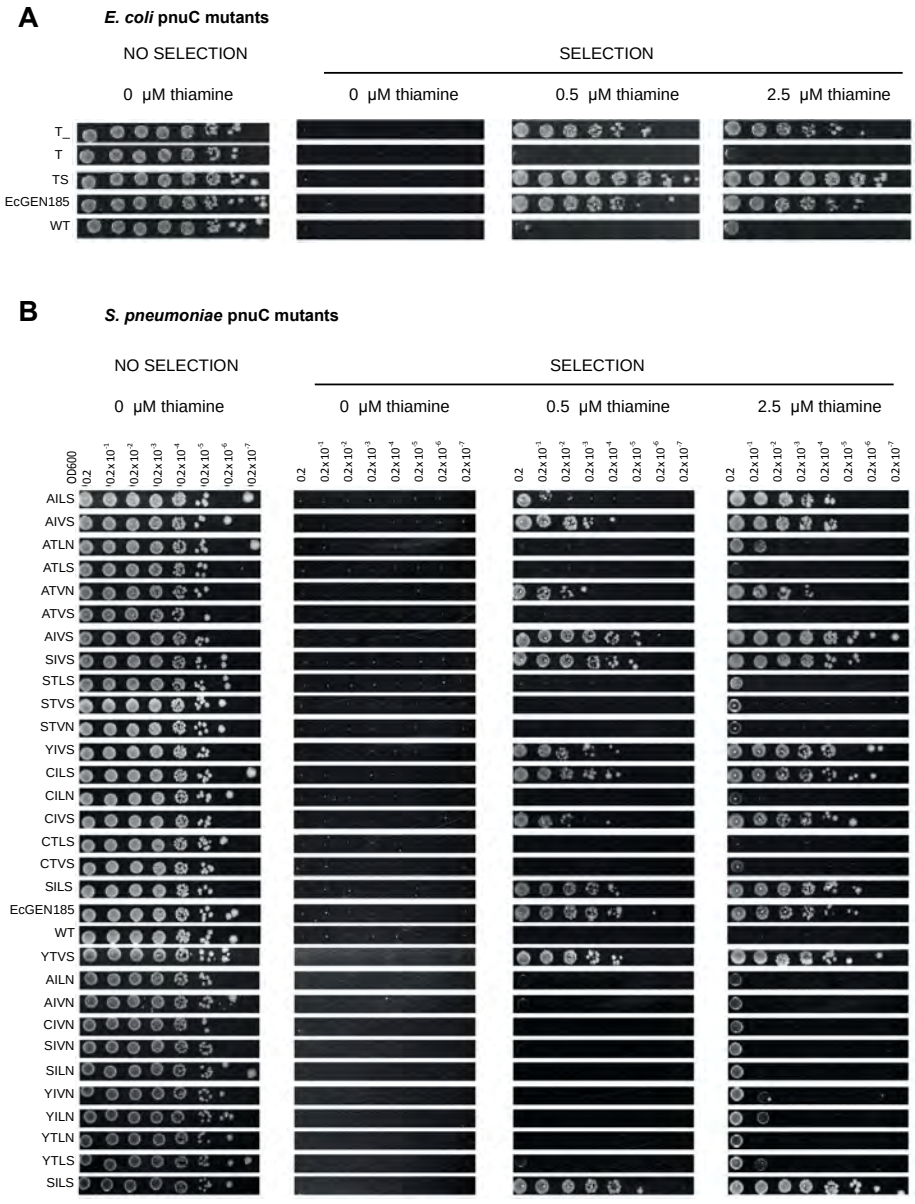
Corresponding author

*Morten O. A. Sommer: msom@bio.dtu.dk

Supplementary Information

Directed evolution of membrane transport using synthetic selections

Anne P. Bali, Hans J. Genee, Morten O. A. Sommer



Supplementary Figure 1: Functional analysis of PnuC mutants for thiamine transport. (A) Functional analysis of *E. coli* PnuC variants with mutations at residue 53 and 54 (*N. mucosa* numbering). "T_" denotes a Cys53Thr and a concomitant Ile54 deletion. "T" denotes a Cys53Thr substitution. "TS" denotes a Cys53Thr and a concomitant Ile54Ser substitution. "WT" denotes the *E. coli* wild-type PnuC protein. EcGEN185 encoded the *H. pylori* PnuT and was included as a positive control. Mutant proteins were expressed in *E. coli* harboring the selection system and thiamine transport of each stain was assayed by spotting 10x serial dilutions of an overnight culture on selective growth medium supplemented with 0.0 µM, 0.5 µM or 2.5 µM thiamine. The plates were incubated at incubated at 37°C for 48 h. **(B)** Functional analysis of PnuC mutants from *S. pneumonia* with mutations at position C53A/S/Y, T96I, V97L, and N220S (*N. mucosa* numbering). E.g. the label "SILS" is PnuC variant with C53S, T96I, V97L, and N220S substitutions. "WT" denotes the *E. coli* wild-type PnuC protein. EcGEN185 encodes the *H. pylori* PnuT and was included as a positive control.

Supplementary Table 1: Primers used in this study.

Primer Name	Sequence (5'-3')
oGEN291	ACTTGCUCGACCTTCCGATDYAAAGATTAHAYMGATAATCCCTGTCAAGCT
oGEN292	AGCAAGUAATTATCTTTTTGGCTTGATTAACBYAGTTATTTACCTTATTTTGGCCCT
oGEN306	AGGGATTAUCKCTGTAATCTTTGTATCGGAAGGTCG
oGEN308	ATAATCCCUGTCAAGCTACAAATCATC
oGEN309	ATCTCATUARCAGTCTAGTTGGTTGGTAT
oGEN310	AATGAGAUAAATTAGATATTTCCCTTGAATTT
oGEN320	AGGGATTAUCNNNGTAATCTTTGTATCGGAA
oGEN321	AGGCCAAUNNNCAGCAACCCGGCGATCGT
oGEN322	ATTGGCCUTGCCAGTCTGGAGAAGATCAG
oGEN323	AAGTAAAGUGTCGTCAGCACCTCACCA
oGEN324	ACTTTACTUCAYASTCATGCAGCCAATTGGAC

Supplementary Table 1: Primers used in this study.

Primer Name	Sequence (5'-3')
oGEN291	ACTTGCUCGACCTTCCGATDYAAAGATTAHAYMGATAATCCCTGTCAAGCT
oGEN292	AGCAAGUAATTATCTTTTTGGCTTGATTAACBYAGTTATTTACCTTATTTTGGCCCT
oGEN306	AGGGATTAUCKCTGTAATCTTTGTATCGGAAGGTCG
oGEN308	ATAATCCCUGTCAAGCTACAAATCATC
oGEN309	ATCTCATUARCAGTCTAGTTGGTTGGTAT
oGEN310	AATGAGAUAAATTAGATATTTCCCTTGAATTT
oGEN320	AGGGATTAUCNNNGTAATCTTTGTATCGGAA
oGEN321	AGGCCAAUNNNCAGCAACCCGGCGATCGT
oGEN322	ATTGGCCUTGCCAGTCTGGAGAAGATCAG
oGEN323	AAGTAAAGUGTCGTCAGCACCTCACCA
oGEN324	ACTTTACTUCAYASTCATGCAGCCAATTGGAC

43 **Supplementary Table 2: Strains used and constructed in this study.**

Strain ID	Description	Description of plasmid
EcGEN136	EcGEN46 + pZE21	pZE21, cloning vector [1]
EcGEN184	EcGEN46 + pGEN62	pGEN62: pZE21 with *E. coli* DH10B *pnuC* [2]
EcGEN185	EcGEN46 + pGEN63	pGEN63: pZE21 with *Helicobacter pylori* 26695 *pnuT* [2]
EcGEN186	EcGEN46 + pGEN64	pGEN64: pZE21 with *Streptococcus pneumoniae* CDC0288-04 *pnuC* [2]
EcGEN243	EcGEN46 + pGEN86	pGEN62 with *E.coli* PuC Cys53Thr and Ile54deletion
EcGEN244	EcGEN46 + pGEN87	pGEN62 with *E.coli* PuC Cys53Thr and Ile54Ser
EcGEN245	EcGEN46 + pGEN88	pGEN62 with *E.coli* PuC Cys53Thr
EcGEN246	EcGEN46 + pGEN89	pGEN64 with *S. pneumoniae* PnuC with T96I, V97L, N220S
EcGEN247	EcGEN46 + pGEN90	pGEN64 with *S. pneumoniae* PnuC with T96I, N220S
EcGEN248	EcGEN46 + pGEN91	pGEN64 with *S. pneumoniae* PnuC with T96I, V97L
EcGEN249	EcGEN46 + pGEN92	pGEN64 with *S. pneumoniae* PnuC with T96I
EcGEN250	EcGEN46 + pGEN93	pGEN64 with *S. pneumoniae* PnuC with V97L, N220S
EcGEN251	EcGEN46 + pGEN94	pGEN64 with *S. pneumoniae* PnuC with N220S
EcGEN252	EcGEN46 + pGEN95	pGEN64 with *S. pneumoniae* PnuC with V97L
EcGEN253	EcGEN46 + pGEN96	pGEN64 with *S. pneumoniae* PnuC with C53S, T96I, V97L, N220S
EcGEN254	EcGEN46 + pGEN97	pGEN64 with *S. pneumoniae* PnuC with C53S, T96I, N220S
EcGEN255	EcGEN46 + pGEN98	pGEN64 with *S. pneumoniae* PnuC with C53S, T96I, V97L
EcGEN256	EcGEN46 + pGEN99	pGEN64 with *S. pneumoniae* PnuC with C53S, T96I
EcGEN257	EcGEN46 + pGEN100	pGEN64 with *S. pneumoniae* PnuC with C53S, V97L, N220S
EcGEN258	EcGEN46 + pGEN101	pGEN64 with *S. pneumoniae* PnuC with C53S, N220S
EcGEN259	EcGEN46 + pGEN102	pGEN64 with *S. pneumoniae* PnuC with C53S, V97L
EcGEN260	EcGEN46 + pGEN103	pGEN64 with *S. pneumoniae* PnuC with C53S
EcGEN261	EcGEN46 + pGEN104	pGEN64 with *S. pneumoniae* PnuC with C53A, T96I, V97L, N220S
EcGEN262	EcGEN46 + pGEN105	pGEN64 with *S. pneumoniae* PnuC with C53A, T96I, N220S
EcGEN263	EcGEN46 + pGEN106	pGEN64 with *S. pneumoniae* PnuC with C53A, T96I, V97L
EcGEN264	EcGEN46 + pGEN107	pGEN64 with *S. pneumoniae* PnuC with C53A, T96I
EcGEN265	EcGEN46 + pGEN108	pGEN64 with *S. pneumoniae* PnuC with C53A, V97L, N220S
EcGEN266	EcGEN46 + pGEN109	pGEN64 with *S. pneumoniae* PnuC with C53A, N220S
EcGEN267	EcGEN46 + pGEN110	pGEN64 with *S. pneumoniae* PnuC with C53A, V97L
EcGEN268	EcGEN46 + pGEN111	pGEN64 with *S. pneumoniae* PnuC with C53A
EcGEN269	EcGEN46 + pGEN112	pGEN64 with *S. pneumoniae* PnuC with C53S, T96I, V97L, N220S

EcGEN270	EcGEN46 + pGEN113	pGEN64 with *S. pneumoniae* PnuC with C53S, T96I, N220S
EcGEN271	EcGEN46 + pGEN114	pGEN64 with *S. pneumoniae* PnuC with C53S, T96I, V97L
EcGEN272	EcGEN46 + pGEN115	pGEN64 with *S. pneumoniae* PnuC with C53S, T96I
EcGEN273	EcGEN46 + pGEN116	pGEN64 with *S. pneumoniae* PnuC with C53S, V97L, N220S
EcGEN274	EcGEN46 + pGEN117	pGEN64 with *S. pneumoniae* PnuC with C53S, N220S
EcGEN275	EcGEN46 + pGEN118	pGEN64 with *S. pneumoniae* PnuC with C53S, V97L
EcGEN276	EcGEN46 + pGEN119	pGEN64 with *S. pneumoniae* PnuC with C53S

Supplementary Note 1: Modified rich MOPS medium

Modified rich MOPS medium (mrMOPS) as outlined in Table N1 was used for all thiamine functional selections. K_2HPO_4 solution was prepared by autoclaving 23 g of dibasic K_2HPO_4 in 1 L de-ionized water. Glucose solution was prepared by dissolving and autoclaving 100 g anhydrous dextrose in 500 mL of de-ionized water. A 500 X vitamin solution was prepared by mixing individual sterile filtered vitamins as described in Table N2 and aliquots were stored at -20° C. KOH solution was prepared by dissolving and autoclaving of 1.247 grams of KOH (90% Reagent Grade, Sigma-Aldrich) in 1 L of de-ionized water. Agar plates were prepared by adding melted agar solution (13 g of Bacto agar in 500 mL of water).

Table N1. Modified MOPS medium (mrMOPS).

Ingredient	Volume (mL) for 1000 mL
10 X Modified MOPS (M2101, Teknova)	100
0.132 M K2HPO4	10
10X ACGU (M2103, Teknova)	100
500 X Vitamin solution without thiamine (see below)	2
20% CAS amino acids, vitamin assay (228830, BD)	20
20% D-glucose	50
Other desired supplements	
Sterile de-ionized water	up to 1000 mL

Table N2. 500 x vitamin solution.

Vitamin	Mass	Volume	
calcium pantothenate (C8731, Sigma-Aldrich)	0.238 g	25 mL	dissolve in de-ionized water
p-aminobenzoic acid (A9878, Sigma-Aldrich)	0.069 g	25 mL	dissolve in 0.02 M KOH
p-hydroxybenzoic acid (240141, Sigma-Aldrich)	0.069 g	25 mL	dissolve in 0.02 M KOH
2,3-dihydroxybenzoic acid (126209, Sigma-Aldrich)	0.077 g	25 mL	dissolve in 0.02 M KOH

References

(1) Lutz, R., and Bujard, H. (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res. 25*, 1203–10.

(2) Genee, H. J., Bali, A. P., Petersen, S. D., Siedler, S., Bonde, M. T., Gronenberg, L. S., Kristensen, M., Harrison, S. J., and Sommer, M. O. A. (2016) Functional mining of transporters using synthetic selections. *Nat. Chem. Biol. 12*, 1015–1022.

CHAPTER V
CONCLUSIONS AND PERSPECTIVES

Sustainable production is essential to continuously meeting the demands of our modern lifestyle, increasing consumption patterns and growing population. We cannot keep on pumping the planet for resources and expect these to be replenished at an increasing speed. It is well known that burning petroleum in the car leads to enhanced CO_2 levels and increase global warming. Production based on fossil fuels also contributes to rising CO_2 levels and chemical manufacturing (of e.g. vitamins) is an energy intensive process. As a consequence, extreme weather conditions have grown more frequent; the numbers of wildfires increase, hurricanes are getting more severe and the global minimum and maximum temperatures becomes more extreme. The research conducted in this thesis hopefully provide work that can become pieces of the solution to replace unsustainable, chemical production of a variety of products as well as giving inspiration to what is possible to achieve with microbial cell factories.

The overall aim of this PhD thesis has been to develop sustainable cell factories for production of small molecules, particularly vitamins, along with tools and approaches that assists the development. This thesis contributes to engineering of vitamin-producing microbes directly by describing modifications that increase vitamin production and indirectly through development of different relevant selection systems and finally, more broadly, by describing different transporter proteins of vitamins and their derivatives.

Different selection systems based on auxotrophy, protein toxicity and riboswitch mediated expression, are developed and/or applied in **Chapter II, III** and **IV** for specific purposes. Here they are used to either improve production titers of cell factories, obtain insight into specific protein families important for metabolic engineering or identify enzymes, improving understanding and reducing cost of microbial production. The general strategy behind these approaches are well-known within the field,

but these new applications of classic selection approaches are likely to be beneficial for future engineering efforts developing bio-based production of other compounds.

Improved cell factories for thiamine, lipoic acid and biotin are presented in **Chapter II**. The identified engineering strategy is expected to also benefit biological production of other related compounds with FeS-cluster enzymes in the biosynthesis pathway. Even though the genomic mutations in IscR responsible for improved production were identified with respect to the bottleneck step of biotin production, BioB, we show that the positive effect can be extended to *de novo* production of both lipoic acid and thiamine. Additional work is needed to take each of the cell factories to commercial scale, but the results in this thesis provides a significant step in the right direction and emphasizes the likelihood of finding synergistic effects that may help accelerating the development. In **Chapter III** and **IV** identification and engineering of transporters with affinity towards the industrial relevant molecules biotin sulfoxide and thiamine are carried out. **Chapter III** exemplifies how functional mining of metagenomic libraries can help to identify unknown transporters as well as expanding the knowledge of already identified proteins with experimental, phenotypic data. Increased understanding of another transporter family, PnuC, is the focus of **Chapter IV**, where directed evolution is used to unveil new information of residues participating in substrate recognition. Transporters are likely to be important tools for improving future cell factories. Furthermore, the knowledge gained in these studies could feed into other research areas such as drug development. As both biotin (reduced biotin sulfoxide) and thiamine pyrophosphate (phosphorylated thiamine) are essential nutrients, blocking transport in pathogens relying on salvage pathways could serve as novel antibiotic targets.

Continued development of the results presented in this thesis are ongoing in Biosyntia ApS. Hopefully these efforts will allow for bio-based manufacturing of vitamins and other small molecule in a not too distant future.