Technische Universität München

TUM School of Natural Sciences

Analysis of L-cysteine production with recombinant *Escherichia coli*

Kevin Heieck

Vollständiger Abdruck der von der TUM School of Natural Sciences der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Tom Nilges

Prüferin*innen der Dissertation:	1. Prof. Dr. Thomas Brück
	2. Prof. Dr. Michael Groll

Die Dissertation wurde am 24.04.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 03.07.2023 angenommen.

Contents

I Abstract

L-cysteine is an integral chemical building block in pharmaceutical, cosmetic, food and agricultural industries. By convention, L-cysteine production is based on the transformation of keratinous biomass by hydrochloric acid. Today, an alternative process operating on industrial scale involves fermentative production utilising recombinant *E. coli*, in which L-cysteine production is rationalised and facilitated by synthetic plasmid constructs. Yet, metabolic burden on cells and the resulting adaptive evolution are highly constraining factors in industrial biomanufacturing. Thus, yields and process stability are still to be optimised for improved economic viability. In this study, high generation numbers are emulated, which are typically achieved in industrial fermentation processes with *E. coli* harbouring L-cysteine production plasmids. In addition, genetic and phenotypic variations in early and late L-cysteine producing populations are investigated.

Using a comparative experimental design, the *E. coli* K-12 production strain W3110 and the genome-reduced control strain MDS42, which lacks all active insertion sequences, were selected as hosts. Data demonstrates how W3110 populations sacrifice L-cysteine production to acquire growth fitness within 60 generations, while production in MDS42 populations remains robust. The negative effects of transposases of insertion sequence family 3 and 5 on L-cysteine production are reported by combining differential transcriptome analysis with NGS-based deep plasmid sequencing. Moreover, by metabolic clustering of differentially expressed genes, this study bolsters the hypothesis that metabolic stress triggers a rapid propagation of plasmid rearrangements that lead to reduced L-cysteine yields in evolving populations over industrial fermentation intervals.

This study demonstrates the potential of selective deletion of insertion sequence families as a new approach to enhance industrial L-cysteine or even general amino acid production with recombinant *E. coli* hosts.

L-Cystein ist ein wesentlicher chemischer Baustein in der Pharma-, Kosmetik-, Lebensmittel- und Agrarindustrie. Konventionell basiert die L-Cystein-Produktion auf der Umwandlung von keratinhaltiger Biomasse durch Salzsäure. Ein alternatives Verfahren, das heute im industriellen Maßstab eingesetzt wird, ist die fermentative Produktion mit rekombinanten *E. coli*, bei der die L-Cystein-Produktion durch synthetische Plasmidkonstrukte rationalisiert und ermöglicht wird. Der sich auf die Zellen auswirkende metabolische Stress und die daraus resultierende adaptive Evolution sind jedoch äußerst limitierende Faktoren für die industrielle Bioproduktion. Daher müssen die Ausbeute und die Prozessstabilität noch optimiert werden, um die Wirtschaftlichkeit zu verbessern. In dieser Studie werden hohe Generationszahlen emuliert, die typischerweise in industriellen Fermentationsprozessen mit *E. coli* erreicht werden, die Plasmide zur L-Cystein-Produktion tragen. Darüber hinaus werden genetische und phänotypische Variationen in frühen und späten L-Cystein-produzierenden Populationen untersucht.

Unter Verwendung eines vergleichenden Versuchsaufbaus wurden der *E. coli* K-12 Produktionsstamm W3110 und der genomreduzierte Kontrollstamm MDS42, dem alle aktiven Transposons fehlen, als Wirtsstämme ausgewählt. Die Daten zeigen, dass W3110-Populationen die Produktion von L-Cystein opfern, um innerhalb von 60 Generationen Wachstumsfitness zu erlangen, während die Produktion in MDS42-Populationen stabil bleibt. Die negativen Auswirkungen von Transposasen der Insertionssequenz-Familie 3 und 5 auf die L-Cystein-Produktion werden durch die Kombination von differenzieller Transkriptomanalysen mit *NGS*-basierter hochauflösender Plasmidsequenzierung beschrieben. Darüber hinaus stützt diese Studie durch das metabolische *Clustering* differenziell exprimierter Gene die Hypothese, dass metabolischer Stress eine rasche Ausbreitung von Insertionssequenzen in Plasmiden auslöst. Dieses Phänomen führt zwangsläufig aufgrund der sich anpassenden Populationen zu verringerten L-Cystein-Ausbeuten in großtechnischen, industriellen Fermentationen.

Diese Studie unterstreicht das Potenzial der selektiven Deletion von Insertionssequenz-Familien als neuen Ansatz zur Verbesserung der industriellen L-Cystein- oder sogar der allgemeinen Aminosäureproduktion mit rekombinanten *E. coli*.

II Acknowledgements

First, I would like to thank my supervisor Thomas Brück for granting me the opportunity to work under his esteemed supervision and complete my dissertation. Furthermore, I extend my sincerest thanks for providing me with an always-open office and unwavering support.

I would also like to thank my colleagues Felix Melcher, Manfred Ritz, Marion Ringel, Nadim Ahmad, Nathanael Arnold, Nikolaus Stellner, Selina Engelhart-Straub, Sophia Prem and Zora Rerop for their friendship, cooperation, and collaboration. Your diverse expertise, positive attitude, and willingness to help have created a stimulating research environment that has greatly facilitated my research progress. Your insightful feedback, critical evaluation, and constructive criticism have been instrumental in shaping my research work and refining my scientific skills.

I would like to express my deepest gratitude to Anna Moosmann for her unwavering love, understanding, and patience. Your emotional support, motivational words, and caring presence have been the driving force behind my perseverance and resilience in the face of challenges and setbacks.

Finally, I would like to acknowledge the contributions of all the individuals including my family and friends who have helped me in any way during my PhD journey. Your assistance, advice, and encouragement have been essential in making my research work possible.

III List of abbreviations

ALE: Adaptive laboratory evolution

ATP: Adenosine triphosphate

Bp: Base pair

***C. glutamicum*:** *Corynabacterium glutamicum*

Cas: CRISPR-associated

CRISPR: Clustered regularly interspaced short palindromic repeats

DEG: Differentially expressed gene

E. coli: *Escherichia coli*

EDTA: Ethylenediaminetetraacetic acid

EGP: Early generation population

EtOH: Ethanol

GFP: Green fluorescent potein

IPTG: Isopropyl-beta-D-thiogalactopyranoside

IR: Inverted repeat

IS: Insertion sequence

LB: Lysogeny broth

LGP: Late generation population

MDS: Multiple deletion train

MM: Minimal medium

NAD: Nicotinamide

NGS: Next generation sequencing

OAS: O-acetylserine

Orf: Open reading frame

Ori: Origin of replication

OTR: Oxygen transfer rate

R.P.M: Rounds per minute

SIDD: Stress-induced DNA Duplex Destabilisation

SNP: Single-nucleotide polymorphism

TSD: Target site duplication

P. putida: *Pseudomonas putida*

P. aeruginosa: *Pseudomonas aeruginosa*

RAIR: RecA-independent recombination

RBS: Ribosomal binding site

IV List of related articles

This thesis includes the following related articles:

(1) Heieck, K., Arnold, N.D. & Brück, T.B**. Metabolic stress constrains microbial L-cysteine production in Escherichia coli by accelerating transposition through mobile genetic elements.** Microb Cell Fact 22, 10 (2023).

(2) Heieck, K., Engelhart-Straub, S., Pilz, M., Melcher, F., Cavelius, P., Brück, T. (2022). Chapter Agricultural Biocatalysis: **From Waste Stream to Food and Feed Additives.** In (Ed.), Agricultural Biocatalysis (1st ed., Vol. 1, Ser. 1, pp. 130–180). essay, Taylor and Francis Group.

(3) Heieck K, Brück T. **Localization of Insertion Sequences in Plasmids for L-Cysteine Production in *E. coli*.** *Genes*. 2023; 14(7):1317. https://doi.org/10.3390/genes14071317

1 Introduction

1.1 *Escherichia coli* as a whole-cell catalyst for fermentative production of biomolecules

The days where large volumes of animal and plant tissue were needed to chemically extract small amounts of proteins or other biomolecules are almost past. Chemical synthesis processes are usually high yielding, albeit environmentally harmful and often associated with the production of unwanted by-products. In addition, many potentially useful natural products are often too complex to be chemically synthesised or are produced in insufficient quantities. Instead, fermentative production utilizing whole-cell catalysts offer many advantages for production of bulk and fine chemicals: high selectivity (e.g. chiral, positional and functional group-specific), high catalytic efficiencies, multiple step reactions in a single strain with the regeneration of cofactors and environmental friendliness (1, 2). During the early years of fermentation processes, the development of production strains was initially based on classical strain breeding entailing repeated random mutations, each followed by elaborate screening or selection. With the development of molecular genetics, the overproduction of biomolecules accelerated.

Microorganisms used as host systems include bacteria, yeasts, filamentous fungi and unicellular algae. Each of these kingdoms has its own advantages and disadvantages, depending on the biomolecule to be produced (3, 4). If post-translational modifications are required for recombinant protein production for instance, the use of eukaryotic hosts is advisable. In return, *Escherichia coli* is the most prominent host for the production of non-glycosylated recombinant proteins.

Some drawbacks, however, are universally encountered, such as substrate or product inhibition, the presence of metabolic by-products and membranes that act as mass transport barriers. A biocatalyst can be tailored by means of protein engineering and metabolic engineering adapted to overcome these limitations.

Escherichia coli is perhaps the most widely used microbial platform for cell factories because of its many assets (5). The gram-negative, facultative anaerobic, rod shaped bacterium is one of the best known and established organism in laboratory. Cultivated strains (e.g. *E. coli* K-12) have lost their pathogenicity factors and enabled safe handling. With fast growth kinetics and low doubling times *E. coli* can achieve cell densities of 1 x 10^{13} viable cells/ml in optimised culture media (6, 7). Rich complex media can be produced with both available and inexpensive components. In addition, *E. coli* is genetically very accessible through rapid and simple transformation with exogenous DNA as well as the availability of improved genetic tools such as site-specific integration and recombination as well as clustered regularly interspaced short palindromic repeats (CRISPR) with its CRISPR-associated protein (Cas9),(8, 9). Additionally, DNA assembly by means of RecA-independent recombination (RAIR), GoldenGate, Gibson and randomised BioBrick assembly opened the gateway for targeted metabolic engineering approaches by enabling simpler design of synthetic constructs (10-12). Recent advances in *omics* technologies, such as time-resolved proteome-, transcriptome- or metabolome analyses have made it easier to understand complex metabolic pathways on a system-wide level. Thus, the determination and analysis of the metabolome identifies weak points where metabolites accumulate in the metabolism. This indicates how to overcome a limitation of the subsequent synthesis step by gene overexpression. By using transcriptomics, it is possible to test whether the expression patterns of production strains during fermentations correlate with high product yields. If this is the case, the identified genes can be selected for strain optimisation. In addition, genome-wide studies at the sequence level (Genomics) can identify mutants that have better production properties or faster growth. These mutations can subsequently be incorporated into the wild-type genome.

Overall, optimising the flux by a synthetic pathway plays a crucial role for achieving the best volumetric productivity of a bioconversion, which in turn reduce the production costs for the targeted chemicals. Numerous research groups have successfully exploited *E. coli* for the fermentative production of a wide variety of chemicals from alcohols, via fatty acids to amino acids and even complex proteins (13-17).

The engineering of *E. coli* as effective whole-cell biocatalysts demands the introduction of single or multiple enzymes into host cells to provide synthetic pathways for the conversion of feedstock to the desired products (*Fig. 1.1.1*).

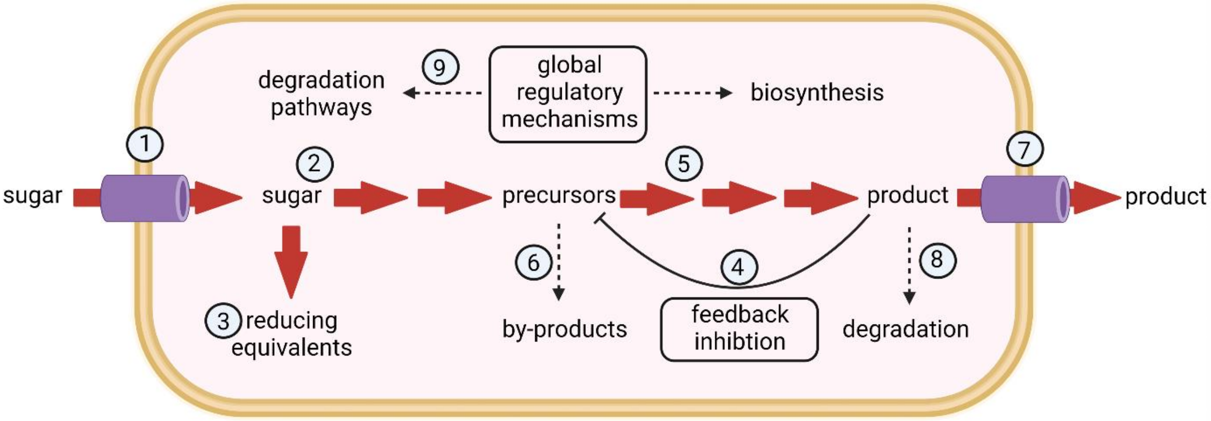


Figure 1.1.1: E. coli as a whole-cell catalyst. *Possible approaches towards enhanced production of chemicals using engineered microorganisms consist of (1) improved sugar intake, (2) supply of precursor, (3) increased formation of reduction equivalents, (4) overcoming feedback inhibition of enzymes, (5) elimination of limiting biosynthetic steps, (6) prevention of by-product formation, (7) improved export, (8) prevention of degradation and (9) optimisation of global regulatory mechanisms for synthesis.*

Particularly important for pathway optimization is the identification of rate-limiting steps. Once the bottleneck genes are identified, its limiting factors can be eliminated by overexpressing those genes, by substituting more efficient enzymes from other species, or by protein engineering. While overexpressing rate-limiting enzymes is usually the easiest to implement, it is often rewarded with high yields (18, 19).

A challenge arises when producing biomolecules that are either toxic themselves if accumulating in large quantities, or whose production is accompanied by the formation of toxic intermediates. The resulting metabolic burden on host cells leads to production decline and suboptimal population growth rates. Therefore, pathway balancing, which involves optimising expression levels through codon usage, promoter and RBS enhancement, should aim at improving the flux towards the product and its export out of cells (20, 21).

Blocking competing pathways prevents the capturing of substrates and intermediates destined to be directed towards the desired biosynthetic pathway (18).

Once a heterologous pathway is integrated into a production host, the pathway will inevitably compete with native metabolism for available precursors. This is especially the case in the production of primary metabolites. For this reason, the efficiency of bioconversion must be improved through an increased supply and faster turnover of precursors (22).

Another major issue that becomes apparent in microbial production with *E. coli* is the provision of cofactors in targeted biosynthetic reactions. The regeneration of cofactors is particularly relevant since they are present in low concentrations in the cell. Nicotinamide (NAD), acetyl-CoA, 2-oxoglutarate and ATP are majorly utilised in glycolysis and the citric acid cycle. Traditionally, cofactor recycling is accomplished via an in-situ regeneration reaction (*Fig. 1.2*).

Most of the reactions take place via oxidoreductases, which depend on NAD. Recycling here is facilitated by a cascading reaction with dehydrogenases, such as the formate-, glucose-, alcohol- or phosphite dehydrogenases (23, 24). Thereby a sacrificial co-substrate gets enzymatically oxidised. Most commonly the formate- and glucose dehydrogenase are exploited for the recycling of NAD(P)H, where formate and glucose are used as co-substrates, respectively (25).

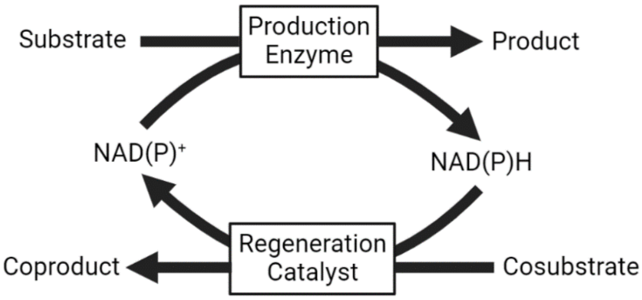


Figure 1.1.2: Engineering cofactor or co-substrate supply. *NAD(P)H recycling systems via coupling with a regeneration step. Figure adapted from Torrelo et al (26).*

When using *E. coli* for the heterologous production of recombinant proteins and biomolecules, the strain selection is as crucial as the synthetic plasmid construct to be used. The selection of an appropriate host strain depends on the protein or biomolecule to be produced. Traditionally, *E. coli* B strains are selected because of their low acetate accumulation, allowing high concentrations of glucose to be consumed as a carbon source. For a first expression screen, the popular strain BL21(DE3) and some derivatives of the K-12 lineage are sufficient. BL21 origins from the B line with various modifications, such as the lack of the Lon protease and the outer membrane protease OmpT, which degrade many foreign proteins (27, 28). Furthermore, plasmid loss is prevented due to mutations in the *hsd*SB gene, resulting in a disruption of DNA methylation and degradation. When the gene of interest (GOI) is placed under the control of the T7 promoter, the BL21(DE3) strain is used because of the insertion of λDE3 prophage into the chromosome, containing the T7 RNA polymerase which can be induced with IPTG.

Other commonly used *E. coli strains* include K-12 lineages such as the Origami^{TM} (Novagen) and Rosetta (DE3) strain. The former has a mutation of the *trxB* gene, which encodes for a thioredoxin reductase. This enhances disulphide formation in the cytoplasm. The Rosetta strain, on the other hand, is used to compensate for codons that are rarely used in *E. coli.* Furthermore, as a result of a *recA* mutation, the K-12 strain HMS174 was found to be beneficial due to increased plasmid stabilities (29). Yet, many strains used for biomolecule production on industrial scale are K-12 "wild types", such as W3110 and MG1655, that have emerged as mutant strains with increased product yields as a result of multiple evolutionary cycles. These are usually stored as cryo- cultures in designated master cell banks and reused for further fermentation processes.

When using synthetic plasmid constructs for production of proteins or other biomolecules, several parameters need to be considered. The origin of replication (ORI) on the plasmid can be used to adjust the copy number. With pMB1 or pACYC ORIs for example 15-60 plasmid copies are possible while a mutated form of pMB1 500-700 copies can be present in a cell, which can allow very high protein yields due to the high number of expression units (30, 31). On the other hand, the high plasmid number can impose a metabolic burden, reducing growth rate and product yield. With the choice of suitable promoters, the expression rate of the downstream proteins can be precisely adjusted. If the target component is a primary metabolite whose synthesis is an integral part of the normal growth process, constitutive promotors, preferably not subject to nutritional repression, are ideal. Stronger expression can be achieved with inducible promoters. The induction can be either chemical, for example through lactose, isopropyl beta-D-thiogalactoside (IPTG) or arabinose, or through physical signals such as temperature or pH (32-36). However, strong expression can result in inclusion bodies (IBs) where proteins accumulate as insoluble aggregates because of incorrect folding and disulphide bond formation. To counteract these effects, it is possible to direct proteins into the oxidative periplasm, to co-express chaperones, to add affinity tags or to perform the induction at lower temperatures. With stationary phase promoters, it is possible to uncouple bacterial growth from product formation. The advantage lies in the auto-inducibility, which saves the costs of expensive inducers and ensures a more stable cell growth and protein production (37).

Affinity tags additionally support keeping the desired protein soluble and active. Besides, tagged proteins are easy to track and extract from the cellular environment of *E. coli* during expression and the subsequent purification. Small peptide tags, such as poly-his, strep-tag II and poly-arg, are less likely tending to interfere with the protein’s structure, therefore keeping biological activity. However, in some cases they may disrupt the tertiary structure of the fused chimeric protein (38-40). Once the crystal structure of the protein of interest is available, it is advisable to check which end is embedded inside the fold and attach the tag to the end accessible to the solvent. Tags can either be positioned on the N-terminal or the C-terminal end with readily available vectors. The latter option is advantageous for recombinant proteins carrying a signal peptide at the N-terminal end for secretion. Adding non-peptide fusion partners on the other hand, such as maltose-binding protein (41), thioredoxin (42), and glutathione S-transferase (43), has the benefit of enhancing solubility (44). If biochemical or structural studies on the recombinant protein are needed, the removal of tags by enzymatic cleavage is crucial. In this process, expression vectors contain a sequence that serves as a recognition site for proteases downstream of the gene coding for the tag (45).

In order to prevent plasmid loss during fermentation and to maintain stable selection pressure, antibiotic resistance markers are integrated. In addition to unstable selection markers, where resistance is based on degradation, such as ampicillin, chloramphenicol or kanamycin, the resistance of tetracycline is based on active efflux, which proved to be very stable during cultivation (46-48). In large-scale biopharmaceutical productions, where antibiotics and the corresponding resistance genes cause high costs and exhibit safety issues regarding inflammatory reactions, the use of antibiotic-free plasmid systems is possible (49). Thereby, auxotroph complementation systems, where an essential gene is deleted from the bacterial genome and added to the plasmid, can be implemented. In consequence, bacteria that are plasmid-free after cell division die (50). Other so-called plasmid addiction systems include toxin / antitoxin-based systems and operator repressor titration systems (50).

Considering the above-mentioned possibilities of designing an expression system, a perfect combination cannot be obtained at the first attempt. Instead, trial and error must be used to screen and subsequently increase the product yield. This time-consuming screening process can be accelerated by micro-scale expression in in 96-well plates. However, promising candidates need to be scaled up in order to develop the good production properties on a larger scale and thus be applicable for industrial use. Special attention must be paid, when designing *E. coli* production systems for therapeutic use. Here, tight compliance and regulatory standards must be met. Quality factors such as the minimisation of *E. coli* host cell contaminants or the maximisation of correctly folded biologically active proteins must be maintained.

If the last mentioned heterologous expression of proteins is a rather recent trend, then the production of primary metabolites with microorganisms has a long tradition. The utilization of microorganisms for the production of fermented beverages and food stuffs has a very long tradition, although the term '*fermentation'* was later introduced by Louis Pasteur. The alcoholic fermentation of yeasts to produce ethanol, as is the case with beer can be dated back to 10,000 years before Christ. Primary metabolites are microbial products that are produced during the exponential growth phase and whose synthesis is an integral part of the regular growth process. These include intermediates or final compounds of anabolic metabolism, which are used as building blocks for essential macromolecules such as amino acids and nucleotides or are converted into coenzymes (i.e. vitamins). Other primary metabolites are not used for the formation of cellular components, but for energy production and substrate use. Hence, they belong to the catabolic metabolism. Product titres are generally very high, with 1-100 g/L depending on the metabolite to be produced (51). The product class of amino acids, for example, qualifies for platform processing, where a few manufacturing units are used for the production of various amino acids. The use of *Escherichia coli*, *Corynebacterium*, *Aspergillium*, *Bacillus* and other microbes for the production of primary metabolites since the 1950s has established this product class as one with a wide range of applications, such as nutritional supplements, livestock farming, flavouring agents and even as building blocks for pharmaceuticals and cosmetics (51).

1.2 Microbial production of amino acids

Amino acids are the most important primary metabolites as building blocks of proteins and thus essential for all organisms. However, humans and animals have only limited synthesis possibilities. Humans cannot synthesise eight of the 20 protein-forming amino acids. These essential amino acids are the three branched-chain amino acids L-leucine, L-isoleucine and L-valine, as well as L-threonine, L-lysine, L-methionine, L-phenylalanine and L-tryptophan. The need for these amino acids must be covered by the consumption of protein sources. In addition to the use of amino acids in the pharmaceutical sector, such as in infusion solutions for chronic inflammatory bowel diseases or in cases of disturbed amino acid balance, such as phenylketonuria, amino acids are used in various sectors. Amino acids are of particular economic importance in the animal feed industry. The addition of amino acids (involving L-lysine, DL-methionine, L-cysteine, L-threonine and L-tryptophan) to feedstuffs constitutes the largest share (56%) of the total amino acid market. Animal feed supplement leads to a balanced amino acid composition that is adapted to the daily requirement. The addition not only reduces the cost of feed raw materials, but also increases the efficiency of feed utilisation and reduces nitrogen excretion, which is harmful to the environment. The annual global demand for feed-grade amino acids is estimated with a value of 2.4 million tons (52).

The history of industrial amino acid production with microorganisms goes back to the 1950s, when Kinoshita et al. discovered *Corynebacterium glutamicum* as a superior amino acid producer (53, 54). Until then, amino acids were obtained exclusively by extraction methods or chemical synthesis with the disadvantage of the subsequent racemate separation. In Japan, extracted seaweed called ‘kelp’ was previously used for seasoning food. It was only in 1908 that Professor Kikunae Ikeda discovered the main ingredient responsible for the “*umami*” flavour, mono-sodium glutamate (MSG). During the first half of the 20th century, MSG was obtained by extraction from wheat soybeans and other plant protein sources via hydrolysis with concentrated acetic acid. Soon after Kinoshita's et al. discovery, Kyowa-Hakko introduced the fermentative production of MSG by *Micrococcus glutamicus*. Later, *M. glutamicus* was renamed *C. glutamicum* (55). Today, nearly 6 million tons of amino acids are produced per year, most of them by utilising coryneform bacteria and *E. coli* in fermentation processes (*Table 1.2*) and the market size is expected to grow by 7.4% each year (56). These numbers are key drivers for the further development of bioprocesses and downstream technologies. Thereby, the organism plays the central role. Microbial amino acid synthesis is preferred because only the L-enantiomer is synthesised, the feedstock containing sugar is renewable and the process is carried out environmentally friendly at low temperatures.

With the exception of glycine, D, L-methionine and aspartic acid, amino acids are now produced microbially. Glycine is achiral and is hence produced chemically. D, L-methionine is also produced chemically. It is used as a racemate in animal feed because animals possess the enzymes D-amino acid oxidase and transaminase, for converting D-methionine into the nutritionally relevant L-methionine. L-aspartate is obtained enzymatically from fumarate in a whole cell process.

Table 1.2: Estimated amount of worldwide annually produced amino acids [t/a] and their production method (57).

Amino acid	Amount [t/a]	Production method	Organism
L-glutamic acid	2.300.000	Fermentative	*C. glutamicum*
L-lysine	1.300.000	Fermentative	*C. glutamicum*, *Brevibacterium* spp.
D,L-methionine	850.000	Chemical	
L-threonine	190.000	Fermentative	*E. coli*, *C. glutamicum*, *Brevibacterium* spp., *Serratia* sp., *Proteus* sp.
Glycine	16.000	Chemical	
L-aspartic acid	14.000	Enzymatic	*E. coli*, *C. glutamicum*
L-phenylalanine	14.000	Fermentative	*E. coli*
L-cysteine	13.000	Chemical /Fermentative	*E. coli*, *C. glutamicum*, *Pseudomonas* sp.
L-tryptophan	4.500	Fermentative	*E. coli*
L-arginine	2.000	Fermentative	*E. coli*, *C. glutamicum*, *Bacillus* sp., *Serratia* sp.
L-alanine	1.500	Fermentative	*Pseudomonas* sp.
L-leucine	1.200	Fermentative	*C. glutamicum*
L-valine	1.000	Fermentative	*C. glutamicum, E. coli, B. subtilis*
L-isoleucine	500	Fermentative	*C. glutamicum*

The classical development of production strains began with random mutations in the genome of the organisms by UV mutagenesis and subsequent screening procedures. After several selection criteria, a strain was available possessing mutations relevant for product formation. No knowledge of the molecular function of gene loci was required. Much effort has been made to isolate wild strains with genetic mutations. This involves identifying auxotrophic or regulatory mutants that accumulate amino acids directly from glucose. Amino acids where production relied on auxotrophic mutants included L-Leucine, L-lysine, L-proline, L-threonine and L-tyrosine. Additionally, most original glutamic acid-producing *C. glutamicum* strains were biotin auxotrophs, in which biotin-deficiency has the property of altering the cell membrane due to suboptimal fatty acid biosynthesis and is thus responsible for increased production. Regulatory mutants that are resistant to feedback inhibition were more widely used for the accumulation of L-arginine, L-histidine, L-isoleucine, L-leucine, L-methionine, L-phenylalanine, L-threonine, L-tryptophan, L-tyrosine and L-valine.

With the advent of modern genetic techniques and metabolic engineering, as discussed in the previous section, the overproduction of amino acids is accomplished by rational design. Thereby intracellular metabolic fluxes are of particular interest. Using the concept of building blocks that arise from carbon sources in microbial metabolism, stoichiometric calculations can be applied to determine flux distributions in amino acid production (58). Typically, substrate uptake rates and secretion rates of ^{13}C labelled metabolites serve as input for the calculations. For *C. glutamicum* eleven central intermediates were considered for central metabolism (Diaminopimelate, glucose 6-phosphate, fructose 6-phosphate, ribose 5-phosphate, erythrose 4-phosphate, glyceraldehyde 3-phosphate, 3-phosphoglycerate, phosphoenolpyruvate, acetyl coenzyme A, oxaloacetate and α-ketoglutarate) (59). The majority of amino acids are formed either directly via glycolysis-derived pyruvate or via intermediates from subsequent steps in the citrate cycle. The formation of the aromatic amino acids is achieved through the shikimate pathway, while glycine, serine and cysteine are built from 3-phospho-glycerate (*Fig. 1.2*).

After strain generation, the culture conditions for each individual strain must be designed to achieve the best microbial performance. For a process to be economically implemented, basic research must be successfully transferred to industrial scale, which often pose a challenge. At the start, a fermentation vessel is filled with culture medium and sterilised. The medium contains a suitable carbon source (e.g. sugar cane molasses, sucrose or starch) as well as the required nitrogen- (e.g. urea or ammonium sulphate), phosphorus- and sulphur sources. In addition, trace elements are supplemented to the medium. Afterwards a seed culture of the production strain, grown in a smaller scale fermenter in advance, is added to the fermentation vessel and cultivated under strict temperature-, pH- and aeration conditions. In order to achieve optimal yields, additional nutrients are added with controlled settings during fermentation – depending on the requirements of the culture. After the amino acid has been secreted from the microorganism, a purification step has to be implemented for extraction from the fermentation broth. Thereby two main processes are applied for downstream processing. After the separation of cells by centrifugation or ultrafiltration, either ion exchange chromatography with subsequent crystallisation or spray drying is performed. Alternatively, the product is simply concentrated and crystallised. While the former method achieves a higher purity, it is more time-consuming and produces more waste water. The latter approach is simpler, but produces less pure products that are suitable for animal feed. Instead, the mother liquor can be supplemented to fertilisers. Industrially, the secretion of amino acids by some bacteria has therefore economic relevance for microbial fermentation, since it simplifies extraction and purification of these metabolites.

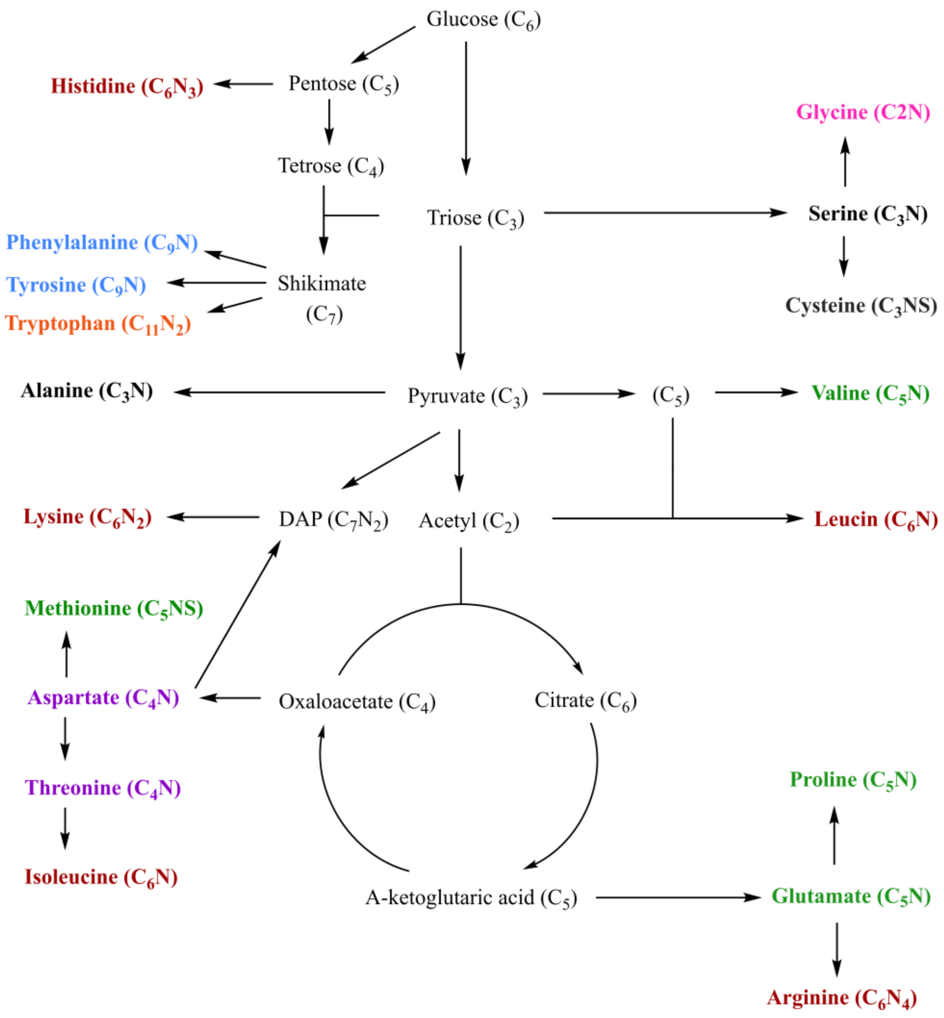


Figure 1.2: Amino acid biosynthesis with glucose as a carbon source. *Chemical formulas are written in brackets. Amino acids are classified and coloured depending on the carbon atoms they contain. C3: black, C4: purple, C5: green, C6: red, C9: blue, C11: orange. Precursors are shown with a thinner font and a grey colour. Arrows indicate different synthesis steps. DAP: diaminopimelate.*

1.3 L-cysteine production

1.3.1 L-cysteine

Large-scale microbial production of L-cysteine has not been studied and developed to a great extent, despite its value in many aspects (*Fig. 1.4*). Since L-cysteine contains a thiol group, it is the only metabolic entry point for reduced sulphur into cellular metabolism in most organisms. Besides the biosynthesis of sulphur-containing substances such as L-methionine, biotin and coenzyme A, it is also required for Fe/s clusters of the catalytic moieties of some enzymes (60). L-cysteine also facilitates disulphide-bond formation in protein folding, assembly and stability. In addition, glutaredoxin (Grx) and thioredoxin (Trx) that utilise L-cysteine-containing peptides (L-glutathione) as a co-factor, are involved in protecting cells under oxidative stresses.

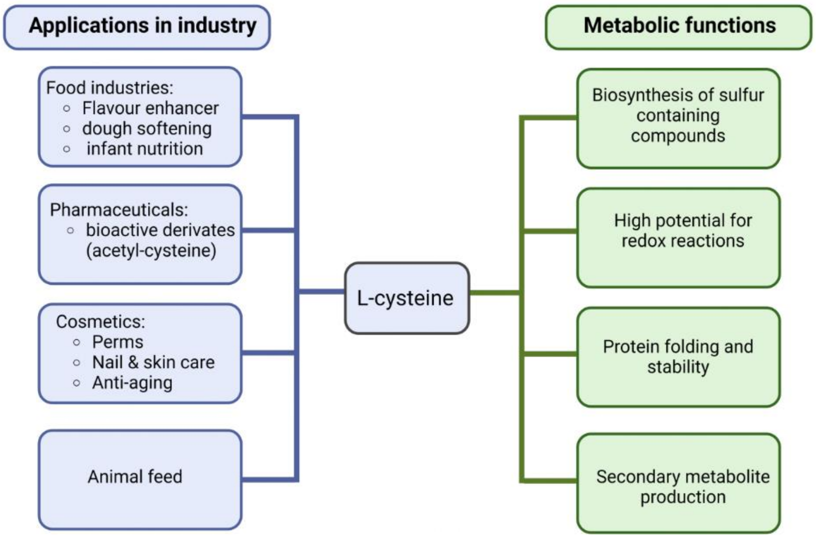


Figure 1.3.1: Metabolic functions and industrial applications of L-cysteine.

Beyond its functions in cellular metabolism, L-cysteine is of considerable industrial importance with applications ranging from pharmaceutical products, cosmetics over food production to feed additives in livestock farming.

In the pharmaceutical field, L-cysteine is used as a precursor for expectorant agents (acetyl cysteine), as an antidote to counteract high toxicity levels of paracetamol, as well as to treat GSH deficiency in genetic defects, metabolic disorders and infections including chronic obstructive pulmonary disease (COPD) (61).

Owing its ability to break disulphide bonds in keratin, L-cysteine is used in cosmetics as a substitute for the strong-smelling and allergenic thioglycolic acid in permed hairstyles. In turn, the acetylated form (N-acetyl-cysteine) is used as a formulation agent for anti-aging and skin care products (62). In addition, L-cystine (L-cysteine dimer) promotes the strength and rigidity of nails and is therefore used in nail care (63).

L-cysteine finds further use in the food industry, especially in bakery. Since its ability to degrade gluten by reducing disulphide bonds and thus changing the dough consistency, L-cysteine is supplemented in various types of flour to ensure proper kneading and increase dough's elasticity (64). In addition, the sulphuric amino acid can be used to create meaty and savory reaction flavours via Maillard reaction when heated with glucose.

The partial requirement of sulphur-containing amino acids in animal feed, can be complemented with L-cystine. Thereby, 50% of the sulphur amino acids (SAAs) demand can be provided in young, fast growing animals and even 80% in older animals (65). Thereby, supplemented L-cystine gets either incorporated in proteins or metabolised and excreted (66). Conversion between L-cysteine and L-cystine is freely reversible.

Its market size is projected to reach a value of 584.2 million US-Dollars by 2027 at a compound annual growth rate (CAGR) of 5.1% during 2022-2027 with an annual global production volume of 14.000 t in 2015 (67). To date, production is facilitated by a combination of different technologies, involving chemical hydrolysis of keratinous biomass such as feathers, pig bristles and human hair, enzymatic conversion and fermentation (68, 69).

1.3.2 Chemical extraction from keratinous biomass

Although chemical hydrolysis is the cheapest and most widely used method, several drawbacks such as low yields, negative environmental impacts and problematic waste disposal remain. In order to obtain 100 kg of a racemic mixture of cysteine from 1,000 kg of raw material, up to 27 tonnes of hydrochloric acid are needed. To mitigate the detrimental environmental impact associated with the hydrochloric waste disposal and the use of animal-derived raw materials, alternative green technologies have rapidly gained significance since their introduction.

1.3.3 Enzymatic bioconversion

Enzymatic bioconversion is facilitated through liquid membrane reactors where precursor substrates are transformed into the desired product with immobilised enzymes or whole cells. Upon extraction of L-cysteine, a chemically synthesised precursor, D, L-2-amino-Δ^2-thiazoline-4-carboxylic acid (D, L-ATC), is converted with enzymes or whole cells of *Pseudomonas spp.* via an intermediate step (*Fig. 1.3.3*). In a first step D-ATC gets racemised to L-ATC, followed by a ring-opening reaction of L-ATC to S-carbamyl-L-cysteine (L-SCC) as intermediate. Lastly, L-SCC is hydrolysed to L-cysteine, which is oxidised to L-cystine under aerobic conditions. The enzymatic bioconversion process, however, is limited by product inhibition of L-cysteine (70, 71).

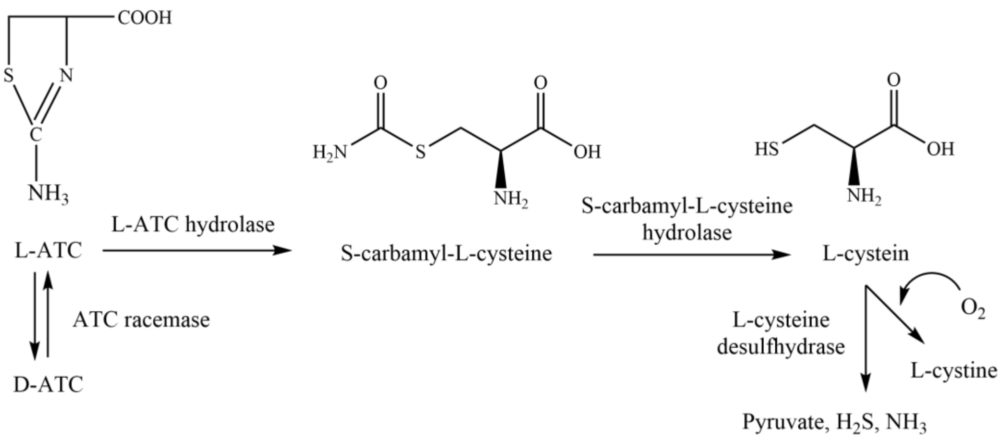


Figure 1.3.3: Enzymatic bioconversion of D, L-2-amino-Δ^2-thiazoline-4-carboxylic acid (D, L-ATC) to L-cysteine via S-carbamyl-L-cysteine (L-SCC) in *Pseudomonas spp.* The conversion is facilitated with L-ATC acid hydrolase and L-SCC amidohydrolase.

Another whole-cell bioconversion process is described with recombinant *E. coli* expressing a tryptophan synthase (72, 73). L-serine proved to be the best substrate among different β-substituted L-alanines and sulphides, which are provided for conversion to L-cysteine. For industrial application, the yields are generally low at approx. 58%, rendering the process uneconomical and inefficient due to high substrate costs.

1.3.4 Fermentative production of L-cysteine

To date, large-scale production of L-cysteine is facilitated by advanced fermentation technologies with predominantly *C. glutamicum* and *E. coli* as hosts. Thereby, the metabolisms were tailored for an optimal flux towards L-cysteine production. Yet the L-cysteine production pathway in both organisms are very similar and hence the strategies for improving the flux towards L-cysteine and its export (*Fig.1.3.4*).

The synthesis of L-cysteine in *Escherichia coli* can be segmented into the following four stages: (I) build-up of 3-phospho-glycerate (3-PG) by glycolysis, (II) capture of 3-PG from glycolysis to form the precursor amino acid L-serine, (III) synthesis of L-cysteine from L-serine, and (IV) formation of L-cysteine from thiosulphate via assimilatory sulphate reduction (74-76).

During the initial step, glucose is imported via the phosphotransferase system (pts) and enters glycolysis. In the course of glycolysis, 3-PG is formed, providing the substrate for the second step of the biosynthesis, the L-serine formation. 3-PG is converted into L-serine via three intermediate steps. First, 3-P-hydroxypyruvate is formed with the help of D-3-phosphoglycerate dehydrogenase, which is further processed to O-phosphoserine by a phosphoserine aminotransferase. Within a phosphatase reaction, the phosphate (Pi) is cleaved off, resulting in L-serine. An acetylation of the precursor amino acid L-serine is carried out by serine acetyltransferase (CysE), which transfers an acetyl residue from acetyl-coenzyme A (Acetyl-CoA) to serine, resulting in O-acetylserine (OAS). This reaction is the rate-limiting step of L-cysteine biosynthesis in *E. coli*. Subsequently, O-acetylserine (thiol) lyase is responsible for the synthesis of L-cysteine in the presence of free sulphide. Alternatively, thiosulphate is imported from the environment by various sulphate transport systems and converted to S-sulphocysteine (SSC) by cysteine synthase B and with OAS as substrate. SSC is converted to L-cysteine and sulphite by an uncharacterised reaction, but thioredoxins and glutaredoxins are postulated to be reduction accelerators of SSC in *E. coli* (77). Final export of L-cysteine from *E. coli* cells to the surrounding medium is realised by efflux transporters, of which EamA and EamB are characterised in more detail (78, 79).

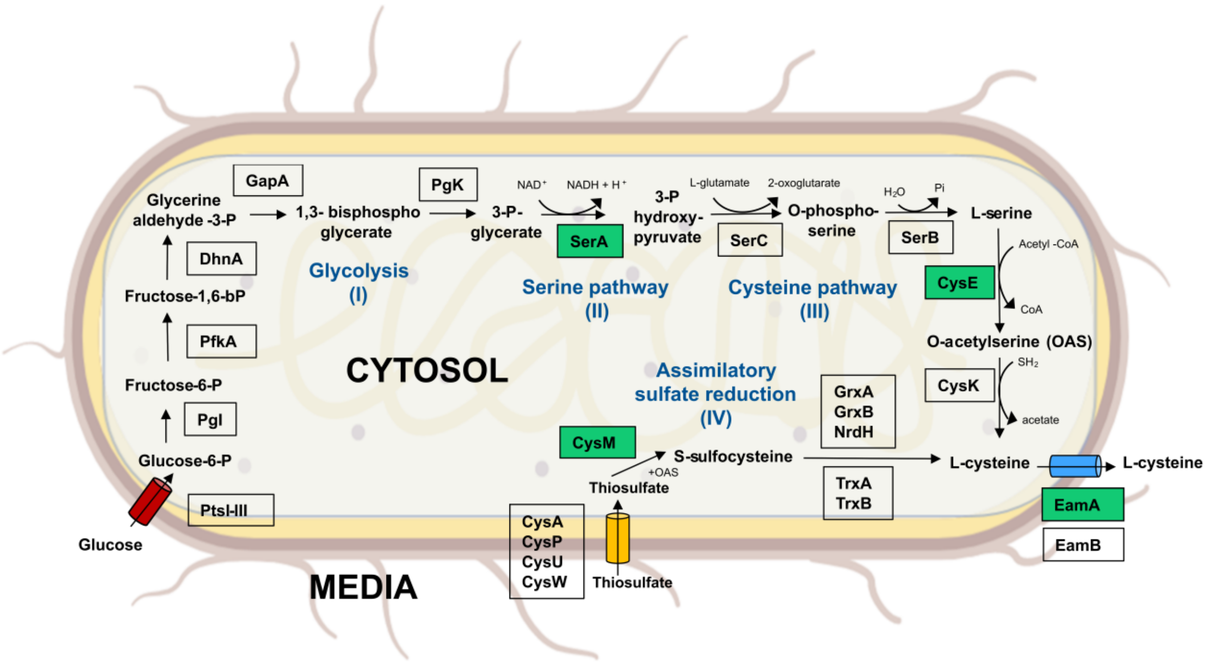


Figure 1.3.4: Production of L-cysteine in Escherichia coli and strategies for metabolic engineering of L-cysteine biosynthesis (80): *Through glycolysis (I, black arrows), the essential intermediate 3-phopsho-glycerate is produced, which is incorporated into L-serine biosynthesis (II, blue arrows). As a precursor amino acid, L-serine is used for the formation of L-cysteine (III, orange arrows). Alternatively, L-cysteine can be synthesised by the assimilatory sulphate reduction pathway (IV, yellow arrows). L-cysteine is transported out of the cell. Corresponding proteins for conversion of substrates are shown in black boxes. Boxes with green background represent potential targets for metabolic engineering and expressed proteins within the synthetic plasmid constructs in this work. PtsI-III Phosphotransferase system I-III, Pgl Glucose-6-phosphate isomerase, PfkA Phosphofructokinase 1, DhnA Fructose-bisphosphate aldolase, GapA Glyceraldehyde-3-phosphate dehydrogenase A, PgK Phosphoglycerate kinase, SerA D-3-phosphoglycerate dehydrogenase, SerC Phosphoserine aminotransferase, SerB Phosphoserine phosphatase, CysE Serine acetyltransferase, CysK Cysteine synthase A, CysA,P,U,W ATP-dependent sulphate/thiosulphate uptake system, CysM Cysteine synthase B, GrxA,B Glutaredoxin 1,2, NrdH Glutaredoxin-like protein, TrxAB Thioredoxin 1,2, EamA,B Cysteine exporter.*

Metabolic engineering targets the overexpression of specific bottleneck genes as illustrated in Figure 1.3.4. In order to drive or accelerate the metabolic flux towards the precursor amino acid L-serine, it is important to capture 3-PG from glycolysis and feed it into the L-serine pathway. For this purpose, a feedback-resistant SerA variant is employed, which is no longer inhibited by high amounts of L-serine and is consequently not autoregulable (81).

The same applies to the enzyme serine acetyltransferase (CysE), which catalyses the reaction of L-serine to the L-cysteine precursor O-acetylserine. Large amounts of L-cysteine do not exert a negative effect on the enzyme itself anymore (82). Similarly, the flux in the direction of glycine is minimised and at the same time driven in the direction of L-cysteine.

In order to facilitate an alternative production route for L-cysteine, the gene encoding for cysteine synthase B (CysM) is overexpressed. This allows fed thiosulphate to be rapidly processed to L-cysteine via assimilatory sulphur reduction.

When overproducing L-Cysteine, high concentrations of L-cysteine accumulate within *E. coli* cells causing inhibitory or even toxic effects (83, 84). Therefore, an export system capable of effectively transporting L-cysteine out of the cell is essential. Overproduction of identified L-cysteine exporters (EamA, EamB) demonstrated parallel export of L-cysteine and OAS.

The regulation of sulphur utilisation and sulphonate-sulphur catabolism via cysteine biosynthesis is negatively autoregulated via the transcription factor CysB (85). In the presence of OAS, binding of CysB tetramers to the DNA target sequence is facilitated, while binding to the repressor site in the *cysB* regulatory region is inhibited. Expression of *cysB* within a synthetic plasmid construct has been demonstrated to increase L-cysteine yields in *E. coli* (86).

The aforementioned approaches ultimately lead to the release of large quantities of L-cysteine into the fermentation broth. This not only evades the inhibitory potential of cysteine, but also facilitates easier extraction and purification in downstream processing steps. Under oxidising conditions, the dimer L-cystine forms and sediments as a white precipitate due to its low solubility in water at neutral pH. Therefore, L-cystine can subsequently be separated from the *E. coli* cells and reduced to the L-cysteine monomer via an electrolysis reaction as a final step.

However, fermentative L-cysteine production can still be improved in terms of yields and production capacities. Large-scale production reveals a phenomenon that generally arises in microbial populations that are exposed to production stress over several generations: Genetic and phenotypic instability.

1.4 Genomic and phenotypic instability of *E. coli* populations during large-scale fermentations

The transition to sustainable bio-based production of chemicals is crucial for green growth. Productivity and yield of engineered organisms frequently deteriorates in large industrial fermentations. In particular, the lack of robustness of synthetic production strains is attributed as the main obstacle to the implementation of large-scale bioprocesses (87, 88); (*Fig. 1.4*).

In continuous fermentations, the appearance of non-producing cells was observed, which was attributed to phenotypic variation of cells within a population (89). This nongenetic variation in microbial cultures is speculated to originate from naturally inherent factors, such as uneven cell division, variations in gene copy numbers, stochastic gene expression and protein activities (90). Furthermore, as a result of insufficient mixing in bioreactors, zones with different environmental conditions can form, which in turn cause different stress to the cells (91). Phenotypic instability was detected in smaller subpopulations temporarily ceasing production properties then resuming to produce at a later time (92). While this effect is not favoured in terms of lower yields and productivity, it also has advantages regarding the rapid adaptation to new conditions and the associated robustness of the fermentation process. These temporally limited phenotypic variations, comprising repetitive cycles of production, re-assimilation of overflow metabolites and relaxation of stress responses, have an impact on overall yield, but are not as detrimental as genetic heterogeneity.

High cellular activity by biosynthetic pathways for economically viable bioprocesses reduces cell fitness and exerts selection pressure towards non-producing cells by adaptive evolution. The irreversible loss of production properties in subpopulations does not only affect the current fermentation, but propagates in continuous fermentation processes where residual volumes of broth is used as inoculum for following fermentation cycles. Thus, populations can quickly establish in which non-producing cells outcompete producer cells and a complete loss of production occurs within a few generations. Similarly, populations that were subjected to a gradual scale-up from a master cell bank to 200 m^3 fed-batch fermenters reach about 60-80 generations, in which non-producers can easily emerge (93).

Several mechanisms are known in microorganisms that lead to fitness enhancement and a concomitant loss of synthetic production properties. The most common involves the loss of plasmids encoding components of the biosynthetic pathway. Several strategies have already been developed to limit plasmid loss, e.g. punishing segregation errors by using plasmid-encoded selection genes as mentioned in section 1.1 and chromosomal integration of pathway genes (94).

Other mechanisms involve error-prone DNA polymerases, such as DNA polymerase IV (Pol IV). Pol IV is controlled by the major stress-response sigma factor RpoS (95). When populations are subjected to non-lethal stress some non-proliferating cells induce DNA polymerase IV as part of the SOS response to DNA damage in order to accumulate mutations and thus alleviate selection pressure. This effect was described as transient hypermutation. Together with DNA polymerase V (Pol V), both facilitate translesion DNA synthesis (TLS) where postreplicative gaps are repaired. Thereby the intrinsic low fidelity of TLS polymerases together with ineffective mismatch repair amplifies genetic instability (96).

Furthermore, micro-organisms contain classes of mobile genetic elements (MGEs) comprised in a mobilome. Such elements have both trans-activity, i.e. they are transferred between different genomes, known as horizontal gene transfer (HGT), and intra-activity, in which genetic regions within a genome are relocalised. These MGEs carry extensive evolutionary potential, both parasitic and beneficial.

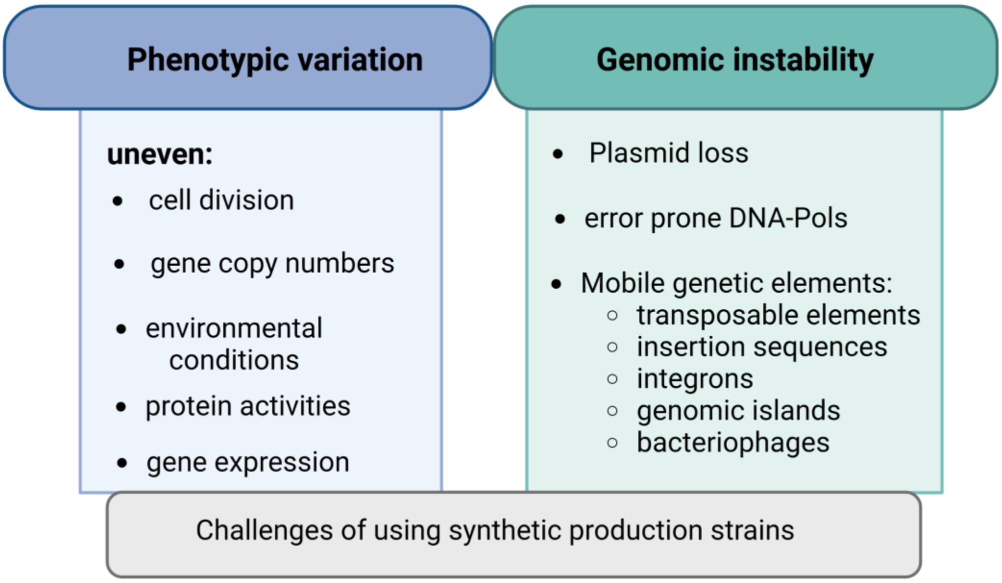


***Figure 1.4.: Challenges of using synthetic production strains in large-scale bioprocesses:** Phenotypic variations are displayed in the blue box (left) and effects comprising genomic instability are shown in the green box (right). Pols: polymerases.*

1.5 Mobile genetic elements (MGEs)

Mobile genetic elements include plasmids, transposable elements, insertion sequences, gene cassettes, integrons, genomic islands and bacteriophages. To date, about 2000 non-core genes from 20 different sequenced genomes have been identified in *Escherichia coli* indicating a strong contribution to the plasticity and variance of genomes (97). This allows microorganisms to adapt to new environmental conditions and stress situations in order to colonise new niches and get rid of burdensome genes.

Plasmids are among the best researched MGEs. These extrachromosomal DNA elements are self-replicating and carry additional information such as antibiotic resistance genes, resistance to heavy metals, virulence genes and other metabolic functions (98). By virtue of specific functions, selected plasmids are used as cloning vectors in recombinant DNA technology, as described in detail in section 1.1. Plasmids can be transferred among bacteria by conjugation, a horizontally genetic transfer process where a donor and recipient get in direct cell-to-cell contact. Indeed, several cases are reported that show how conjugative *E. coli* plasmids promote and spread multidrug-resistant genes (99, 100). In addition to antibiotic resistance genes, plasmids transmit virulence-associated genes. A highly significant outbreak was caused by the hybrid strain enterohemorrhagic *E. coli* (EHEC)-enteroaggregative *E. coli* EAEC O104:H4 in Germany carrying. This strain carries three different plasmids, promoting virulence through fimbriae for adherence, a secretion system, proteases and antibiotic resistance (101).

Involvement in the spread of resistance genes also applies to so-called integrons and gene cassettes. Gene cassettes are free, circular, non-replicating DNA regions that usually consist of only a single gene, often an antibiotic resistance gene, and a specific recombination site. Gene cassettes can be part of an integron, a larger DNA element that promotes site-specific recombination through integrases. Genes located in gene cassettes often lack promoters and are expressed by promoters of the integron (102).

Another class that has the ability to transfer genes among bacteria by transduction are bacteriophages. Bacteriophages are viruses that are able to infiltrate and manipulate the host's biosynthetic machinery, ultimately killing bacterial cells to release a variety of newly formed phages. Phage contamination is a persistent problem in large-scale microbial fermentations, where early process termination result in low-quality and inconsistent product and ultimately in a significant economic loss (103). Phage contamination has been encountered in various industrial fermentation processes including the fermentation of milk, cheese and cucumber by lactic acid bacteria (104-106). This issue is now being tackled through a variety of approaches, such as intensive cleaning and sanitation, changing starter cultures, and introducing plasmid-encoded phage defence mechanisms such as CRISPR-Cas, and antisense mRNA into the starter cultures (107-109).

The most relevant class of MGEs in terms of genetic variation are transposable elements (TEs) or transposons. TEs are DNA fragments that are able to change their position within the genome and between genomes via transposition. TEs are divided into insertion sequence elements (ISEs), composite and non-composite transposons. While the last two carry antibiotic resistance genes in addition to the genetic information of mobility, ISEs are the simpler version that induce large duplications, inversions and genomic rearrangements. With compact lengths ranging from 0.7 to 2.5 kb, they usually possess only a single gene encoding for a transposase and a *cis*-acting site upon which the transposase acts (*Fig. 1.5.1*); (110). *Cis*-sites are short inverted repeats found at both ends of ISE. Transposition can occur in copy-paste or cut-paste mechanisms, whereas a complete IS loss is very rare (111). After insertion, transposons create "footprints" consisting of target site duplications (TSD), i.e. identical sequences at both ends of the transposon.

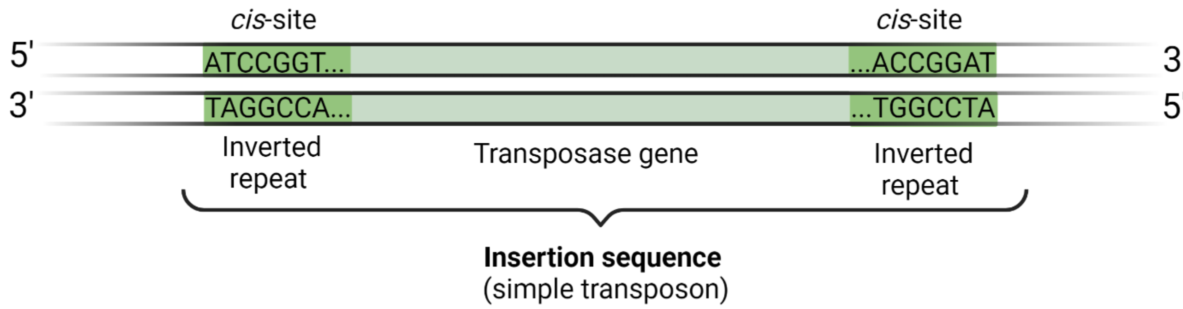


Figure 1.5.1: Schematic structure of a typical Insertion Sequence element (ISE). *Inverted repeats within the cis-sites are highlighted in a darker green while the corresponding transposase gene is displayed in a lighter green.*

Insertion sequences are often classified according to the conserved amino acid motif of their transposases, such as D-D-E, which is the active site of the protein. Within the DDE family, IS3 and IS5 are the most common (*Fig. 1.5.2*).

There is crucial diversity in the types and copy numbers of the different IS families that bacteria carry. Both activity and presence play an important role in genome structure and gene expression, thus affecting fitness. Indeed, ISE have been shown to introduce many mutations that improve fitness under an array of environmental conditions (112-114). Furthermore, ISE cause a critical instability of plasmids harbouring heterologous genes, which ultimately leads to a production decline of the encoded proteins. For example, during the production of a DNA vaccine produced from a plasmid in *E. coli*, the insertion of ISE from the chromosome into the plasmid, reduced productivity drastically (115, 116). Similarly, in other cases, cells with IS-inserted plasmids had a better growth rate than cells with non-IS-inserted plasmids encoding the recombinant target protein, because of sacrificing productivity (93, 117). Hence, a stable host strain without the detrimental effects of ISE is highly desirable for industrial processes.

However, due to the lack of target specificity, designing ISE immune synthetic pathways is an almost impossible task, especially since swapping several bases in the open reading frame will often lead to non-functional proteins with no guarantee that the sequence will no longer be recognised by transposases. Instead, various working groups have managed to generate strains that are completely free of active ISE in elaborate steps. The *E. coli* K-12 derivative MDS42 was created by the group of Blattner et al. and has additional deletions of genes encoding error-prone DNA polymerases and prophages (118). Overall, this minimal genome strain lacks 14.3% of its chromosome compared to its ancestral strain. By deleting ISE in the genome of *Corynebacterium glutamicum*, production could be increased with three models: GFP, poly(3-hydroxybutyrate) and γ-aminobutyrate (119).

By utilising such strains, it has been possible to increase the yields of the amino acid L-threonine and the fatty acid mevalonate in *E. coli* as well as recombinant proteins in *C. glutamicum (93, 119, 120).*

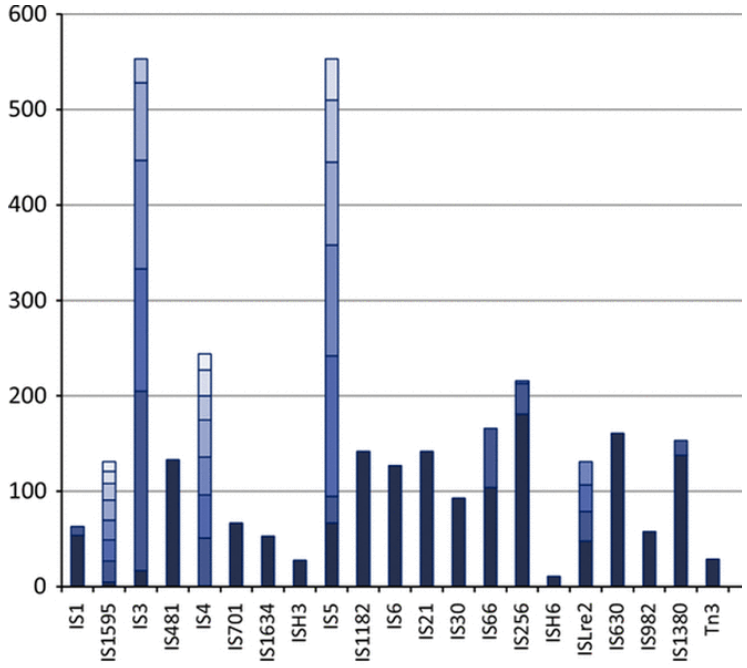


Figure 1.5.2: Distribution of DDE IS families in the ISfinder database. *The histogram shows the number of IS of a given family. Adapted from Siguier et al., (121). Rights for reuse granted by Oxford University Press.*

2 Scope of this work

The aim of this work is to investigate and improve L-cysteine production utilising recombinant *Escherichia coli*. The production dynamics of a population at the beginning and at the end of a cultivation are to be analysed. Thereby, potential differences and drawbacks are uncovered, which should help to increase L-cysteine yields in future rational metabolic engineering approaches.

In the first part, three different synthetic plasmids for L-cysteine production are designed and transformed into two different *E. coli* strains. Afterwards, long-term fermentations are simulated where populations accumulate generation numbers of 55-60, which are often found in large-scale industrial fermentations. During this long-term fermentation simulation phenotypic variation and genetic heterogeneity are analysed. Growth rate and L-cysteine yields are recorded to gain deeper insights in growth fitness and productivity throughout the experiment.

In the second part, comparative transcriptomics between early and late populations is conducted in order to reveal the extent of metabolic stress in *E. coli* cells during L-cysteine production. These results should be constantly assessed in the context of growth rates and L-cysteine yields to identify potential patterns and correlations.

Finally, genetic heterogeneity of the plasmid DNA is investigated using an NGS-based ultra-deep sequencing approach. This enables the identification of single mutations within large plasmid copies in high resolution. Together with the transcriptome data, information on the origin of the mutations should be obtained and an exact localisation within the plasmids should be possible.

Overall, a combined omics-approach should be applied in order to unravel the underlying mechanisms within an industrially relevant L-cysteine production system in *E. coli*.

3 Material & methods

3.1 Chemicals and reagents

Chemicals within this study were obtained from standard sources at the highest purity grade available. Media components and L-cysteine were purchased from Sigma-Aldrich (St. Louis, USA), AppliChem GmbH (Darmstadt, Germany) or Roth chemicals (Karlsruhe, Germany). Ninhydrin reagent was purchased from Merck (Darmstadt, Germany). Enzymes and buffers for polymerase chain reaction (PCR), restriction and ligation were purchased from Thermo Fisher Scientific (Waltham, USA) or New England Biolabs (NEB) (Ipswich, USA). Monarch® PCR & DNA Cleanup Kits and Monarch® DNA Gel Extraction Kits were obtained from New England Biolabs (Ipswich, USA). SV Total RNA Isolation System was purchased from Promega (Madison, USA).

3.2 Media composition and stock solutions

All media described below had their pH adjusted with the appropriate acid or base before autoclaving or sterile filtration. Heat-sensitive components such as amino acids, trace elements, vitamins, sugars, antibiotics were sterile filtrated and added to the final media afterwards. Sterile media and stock solutions were stored at 4°C. Tetracycline stock solution was stored at -20°C.

Table 3.2.1: Composition of Luria-Bertani Broth / Agar (pH 7), (122)

Component	Amount
Yeast extract	5 g/L
NaCl	10 g/L
Tryptone	10 g/L
Agar-Agar	15 g/L
ddH_2O	Up to 1 L

Table 3.2.2: Composition of the trace element solution (pH 4.0)

Component	Amount
H_3BO_4	3.75 g/L
$CoCl_2$ x 6 H_2O	1.55 g/L
$CuSO_4$ x 5 H_2O	0.55 g/L
$MnCl_2$ 4 H_2O	3.55 g/L
$ZnSO_4$ x 7 H_2O	0.65 g/L
Na_2MoO_4 x 2 H_2O	0.33 g/L
ddH_2O	Up to 1 L

Table 3.2.3: Composition of the adapted minimal medium for L-cysteine production (pH 7). *Tetracycline, vitamins, $CaCl_2$ and $MgSO_4$ were added immediately before inoculation.*

Component	**Amount**
Glucose	10 g/L
KH_2PO_4	5 g/L
$(NH_4)_2SO_4$	5 g/L
NaCl	0.5 g/L
Na_3Citrat x 2 H_2O	1 g/L
L-isoleucine	0.9 g/L
D,L-methionine	0.6 g/L
Ammonium thiosulphate	2 g/L
$MgSO_4$ x 7 H_2O	1.2 g/L
$CaCl_2$ x 2 H_2O	0.23 g/L
Thiamine-HCl	18 mg/L
Pryidoxine-HCl	9 mg/L
Tetracyclin (1000x)	15 mg/L
LB-Broth	100 ml/L
Trace element solution	10 ml/L
ddH_2O	Up to 1 L

3.3 Bacterial strains and plasmids

In this work two different strains were used, the laboratory wild-type strain *E. coli* K-12 W3110 retrieved from the industrial cooperation partner Wacker Chemie AG and the minimal genome strain *E. coli* K-12 MDS42 which lacks 14.3% of the parental MG1655 genome including all active insertion sequence elements, cryptic prophages, error-prone DNA polymerases and 699 additional genes without any function in *E. coli*. The MDS42 strain was used as a control strain to determine adaptive evolution under L-cysteine production stress (*Table 3.3.1*).

Table 3.3.1: Name, genotype and sources of E. coli strains used in this work

Strain	**Genotype**	**Sources**	**Literature**
E. coli K-12 W3110	F- λ- rph-1 INV(rrnD, rrnE)	Wacker Chemie AG	*(123)*
E. coli K-12 MDS42	fhuACDB(del) endA(del) + deletion of 699 additional genes, including all IS elements and cryptic prophages	Scarab Genomics ®	(124)

All plasmids used in this study are pACYC184-derived vectors with a p15A origin of replication. As selection marker, only tetracycline resistance is present. The origin chloramphenicol resistance got deleted. The original plasmid pCYS possesses the bottleneck genes *serA*, *cysE* and *ydeD* for L-cysteine biosynthesis and secretion with the corresponding genomic promotor regions. CysE and SerA are feedback insensitive variants as described in references (81, 82). In pCYS_i, these bottleneck genes were arranged in a coherent operon with GAPDH as a constitutive promoter. Moreover, non-coding backbone sequences were trimmed to potentially reduce metabolic burden on production cells. An additional gene *cysM* under the control of the stationary phase promoter *fic1*, coding for cysteine synthase B, was included in pCYS_m (*Table 3.3.2*).

Table 3.3.2: Name, features and sources of plasmids used in this work

Plasmid	Features	Sources	Literature
pCYS	p_{GAPDH}:*ydeD*, $p_{serAp1,2}$:*serA317*, p_{cysE}: *cysE-XIV*, tet^R, p15A	Wacker Chemie AG	(82),(81),(78),(125)
pCYS_i	p_{GAPDH}:*cysE-XIV-ydeD-serA317*, tet^R p15A	This study	
pCYS_m	p_{GAPDH}:*ydeD*, $p_{serAp1,2}$:*serA317*, p_{fic1}:*cysM*, p_{cysE}: *cysE-XIV*, tet^R, p15A	This study	

3.4 Cloning and plasmid construction

For pCYS_i ribosomal binding sites (RBS) and translational rates were balanced according to calculations of the SalisLab webtool box (126). The plasmid pCYS_i was created by ligation of four amplified and digested amplicons in an equimolar ratio in a 40 µl reaction. The fragments contained the metabolic pathway genes (*cysE-XIV* and *serA317*), the L-cysteine exporter gene (*ydeD*) of pCYS and the empty vector as a backbone. The digestions were conducted with SacI, XhoI, BamHI, PacI, and NcoI. For amplification of different fragments, specific PCR primer were used (*Table 3.4*).

PCYS_m was generated by introducing the stationary phase promoter *fic1* and the *cysM* gene into the backbone of pCYS via Gibson cloning standard procedure. PCRs were performed with specifically designed primers (*Table 3.4*).

Standard protocols were employed for polymerase chain reaction (PCR), DNA restriction and ligation. PCR products were checked using a 1% (w/v) agarose gel in TAE buffer and got purified subsequently with the Monarch® PCR & DNA Cleanup Kit (New England Biolabs). For introduction of the plasmids into the strains, standard heat shock transformations with chemical competent cells were performed and finally plated on LB agar supplemented with tetracycline (127).

Afterwards, colony PCRs were conducted for a minimum of 10 clones for each construct. Clones showing a positive colony PCR result were subsequently grown in liquid LB broth supplemented with tetracycline for plasmid extraction. The GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) was used for plasmid purification. All obtained clones were checked and sequenced by Eurofins Genomics (Ebersberg, Germany). DNA concentrations in all cloning steps were checked with a NanoPhotometer® (Implen, Munich, Germany). Clones with positive sequencing results were deposited as cryo stocks in 50% glycerol at -80°C. Wherever applicable, manufacturer's protocols were followed for all procedures.

Table 3.4: Oligonucleotides used for the assembly of plasmids pCYS_i and pCYS_m. *The primers were synthesised by Eurofins Genomics (Ebersberg, Germany)*

Primer name	**DNA sequence (5'->3')**	**Associated assembled plasmid**
Fragment 1_fw	CCGGAGCTCCCGCTTGACGCTGCGTAAGGTTTTTGTAATTTTACAGGCTACCTAGCACTTCGGTTTTATTTTAGGAGAACTTTAATGTCGTGTGAAGAACTGGAA	pCYS_i
Fragment 1_rev	TCATCACCTCGAGTTACGTATTAATCCATTGATGGCTTTCGCTGTCTGG	
Fragment 2_fw	ATACTCGAGGGAAAAAGATGAAATTCAGAGGCG	
Fragment 2_rev	AATGGATCCGGCTTATTAACTTCCCACC	
Fragment 3_fw	TGGGGATCCGCTTATGTTAAGTACAGTCACACTACATGCAAATGATCAAAGGC	
Fragment 3_rev	GGAGCCTTAATTAAGGCGTCAGATCATTTCACAATGGT	
Fragment 4_fw	AGTAGTTTAATTAAGAGTCCTGGCTAACCCACAAGAAGGTTTCAAATGGCAAAGGTATCGCTG	
Fragment 4_rev	ATATCGCCATGGCTGGAGTACTTAGTCAGAATACTT	
Pfic_fw	CCTGAGGCTGCAGCTGCCGTAATGATTT	pCYS_m
Pfic_rev	TTCTAATGTACTCATATGTTGATGCCTCCCTGAACGT	
cysM_fw	CAACATATGAGTACATTAGAACAAACAATAGGCAATACGCCTCTGGTGA	
cysM_rev	GGGAGAGCCTGAGCAAACTGGCCTCAGGTTTAAAAGATAAAAAACGCCCGGCGGCAACCGAGCGTTCTTAAGCCGC	

3.5 Simulated long-term fermentation

This method for simulating long-term fermentation was adapted from Rugbjerg et al. (93) (*Figure 3.5*). After transformation, fresh colonies were inoculated into 250 ml baffled shaking flasks containing 25 ml of the adapted minimal medium (*Table 3.2.3*). The *E. coli* W3110 and MDS42 strains got cultivated at 32°C and horizontal shaking of 150 r.p.m (New Brunswick Innova 44). Each strain containing one of three plasmids got cultivated in triplicates. In order to keep the cultures in an exponential state, serial transfer of the culture was performed every 5 h for the MDS42 strains and every 10 hours for the W3110 strains. Growth rates of MDS42 strains were higher, which resulted in more frequent passaging. After each time point, cultures were inoculated into 25 ml fresh medium (initial OD_{600} = 0.05) and incubated under the same conditions for another 5 h or 10 h. These cycles were repeated until the cultures had accumulated approximately 60-65 generations. At each passage, sample's OD_{600} and accumulated generations were determined, while 1 ml got snap-frozen with 1 ml glycerol (50%) and stored at -80°C immediately. Afterwards, the sampled cryo-cultures were re-cultivated for

72 hours with an initial OD_{600} of 0.01 in 25 ml fresh medium under the above conditions. Growth rates were recorded every 30 min. L-cysteine yields were determined after 72 h. RNA extraction was carried out at OD_{600} values of 0.6-0.8. Extractions of plasmid DNA was performed from the same cultures.

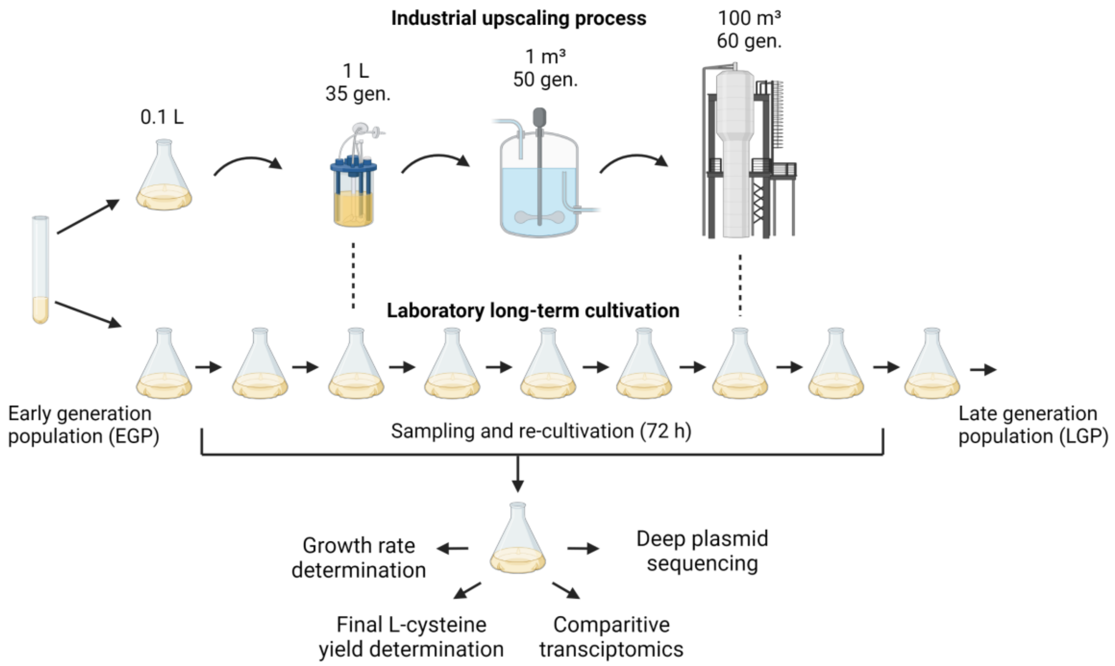

Industrial upscaling process
0.1 L
1 L
35 gen.
1 m³
50 gen.
100 m³
60 gen.
Laboratory long-term cultivation
Early generation population (EGP)
Sampling and re-cultivation (72 h)
Late generation population (LGP)
Growth rate determination
Deep plasmid sequencing
Final L-cysteine yield determination
Comparitive transciptomics

***Figure 3.5: Laboratory long-term fermentation simulation of L-cysteine producing E. coli cells (80)*: *Adapted from Rugbjerg et al., 2018. One "wild-type" E. coli MG1655 strain (W3110) and a reduced genome strain (MDS42) were cultivated harbouring one out of three plasmid constructs). By serially transferring samples every 10 or 5 h into fresh medium, respectively, we imitate generation numbers found in large-scale industrial bioreactors. Thereby the cultures were kept in the exponential phase throughout the experiment. Sample collection was continued until desired generation numbers (> 60) were achieved. Subsequently, all cryo-samples were re-cultivated for 72 h during which growth rate- and final L-cysteine yields were monitored. RNA extractions for comparative transcriptomics and plasmid extractions for deep plasmid sequencing were performed on the early and late generation population (EGP, LGP) samples for each strain and plasmid construct, respectively. Created with BioRender.com*

3.6 Determination of L-cysteine yields by spectrophotometry

Each cryo-culture sample was used to inoculate 25 ml medium, which was cultivated at 32°C at 150 r.p.m for 72 h (New Brunswick Innova 44). 1 ml culture was centrifuged at 15.000 x g for 1 min afterwards to proceed with the L-cysteine yield determination. While the supernatant was directly transferred to a new 1.5 ml reaction tube, the pellet was incubated with 1 ml of mixture 1 (*Table 3.6.1*) and shaken for 2 h in a Thermomix (Eppendorf) with 1.000 rpm. At 70°C. Then, samples were centrifuged at 14.000 g for 5 min and the supernatants transferred to a new 1.5 ml reaction tube. In order to reduce the disulphide bonds, 100 µl of both reaction mixtures (supernatant and pellet) were incubated with 385 µl of Tris-HCl (pH 8) and 30 µl of a freshly prepared 0.1 solution of dithiothreitol (DTT) in 2.0 ml reaction tubes. Afterwards, 500 µl ninhydrin reagent (*Table 3.6.2*) were added to the mixtures and heated for 10 min at 100°C. Upon expiration of the 10 minutes, the mixture changed to different shades of pink, dependent on the L-cysteine concentration. To stabilise the pink product, 1 ml of EtOH (95%) was added and cooled on ice. Finally, the absorbance could be measured at 560 nm (NanoPhotometer, Implen, Germany). A sample with minimal medium subjected to the same protocol was used as blank. Furthermore, ninhydrin assays of W3110 and MDS42 cultivated in minimal medium with transformed empty vectors were performed as negative controls. These samples showed no absorption at all. The A560 absorbance as a function of different L-cysteine concentrations was plotted as a calibration curve (*Fig. 3.6*).

Table 3.6.1: Composition of mixture 1. *(Storage at room temperature)*

Compound	Volume [ml]
Phosphoric acid [85%]	15.25
Sulphuric acid [96%]	1.53
$H_2O_{dd.}$	83.22

Table 3.5.2: Composition of ninhydrin reagent. (*Storage at 4°C for up to 2 weeks)*

Compound	Amount
Ninhydrin	1.25 g
Hydrochloride acid	20 ml
Acetic acid	80 ml

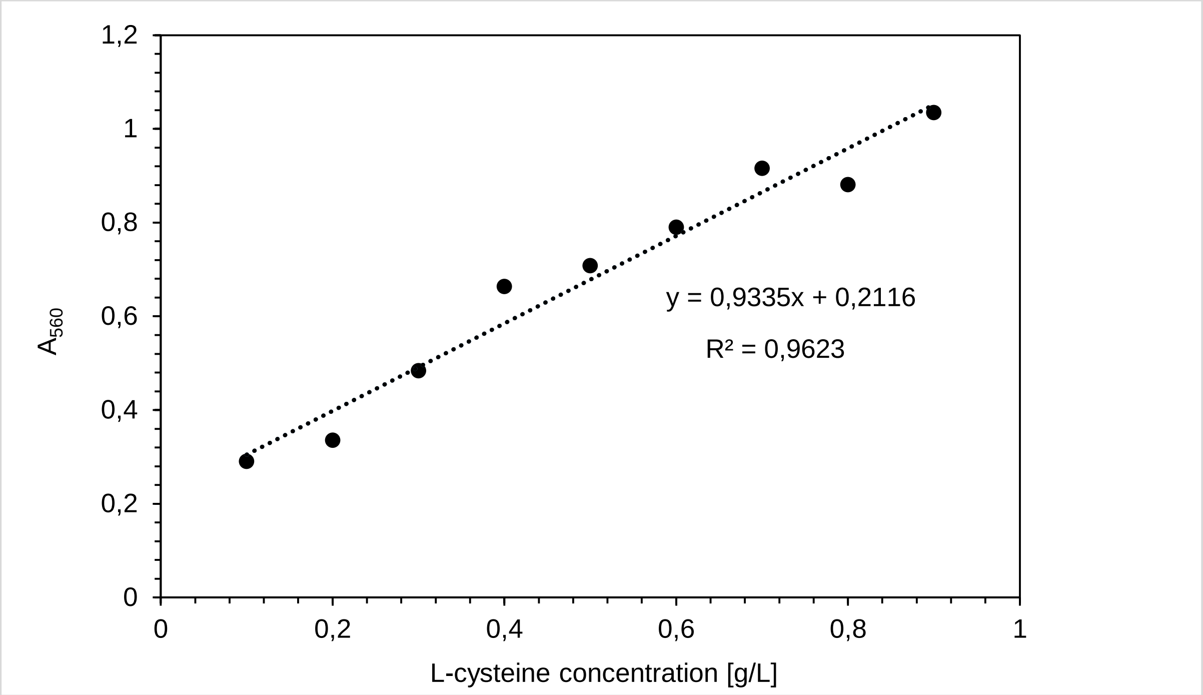


***Figure 3.6: Calibration curve of the A560 absorbance as a function of different L-cysteine concentrations performed with L-cysteine (98.5%) from Sigma**. Concentrations from 0.1 g/L to 0.9 g/L were measured. The function y = 0.9335x + 0.2116 of the according trend line was used to calculate L-cysteine concentrations.*

3.7 Measurement of population growth rates

Population growth rates were calculated from cultures grown for L-cysteine productivity with hourly measured OD_{600} values. For this, the following formula was utilised:

$$r = \frac{N(t)^{\frac{1}{t}}}{N(0)} - 1$$

where: N(t): cell number at time t, N (0): cell number at time 0, r: growth rate and t: time passed. Population growth rates were normalised based on growth rates of the corresponding non-producing strains harbouring empty vectors. The cell number counts were calculated with Agilent's online tool "*E. coli* Cell Culture Concentration from OD_{600} Calculator" (https://www.agilent.com/store/biocalculators/calcODBacterial.jsp, accessed 02/14/2023).

3.8 RNA sequencing and analysis

One sample per early and late generation population, plasmid and strain that have undergone re-cultivations for L-cysteine yield and growth rate determination was chosen for RNA extraction using a standard RNA extraction protocol (Promega SV total RNA isolation system). In total, 12 samples were processed (EGPs and LGPs of MDS42 and W3110 populations with integrated pCYS, pCYS_i and pCYS_m). Ribosomal RNA depletion, transcriptome library construction and sequencing of the corresponding samples were performed by Eurofins Genomics. Sequencing was performed using the Illumina NovaSeq 6000 technology with a 150 bp paired-end reading. The number of raw reads was generated using feature counts (version 1.5.1), (128). Only reads that overlapped coding sequences (CDS) features were counted. All reads that mapped to features with the same meta-feature attribute were summed. A read count was only considered for reads with unique mapping positions and mapping quality scores of at least 10. Supplementary alignments were disregarded. As opposed to counting paired-end reads twice, they were only counted once, i.e. as a single fragment. Reads that mapped to multiple features were assigned to the feature with the largest number of overlapping bases. Normalization was performed with a Trimmed-Mean of M-Values (TMM) algorithm using the edgeR package (version 3.16.25), (129, 130). Mapping of reads to the genome reference sequences of *E. coli* K-12 W3110 and *E. coli* K12 MDS42 was conducted using Burrows-Wheeler-Maximal-Exact-Match (BWA-MEM), (version 0.7.12-r1039).

In order to identify features that are significantly differentially expressed, a small threshold (<<0.01) was applied to the false discovery rate (FDR) values, which is the p-value adjusted for multiple testing using Benjamini-Hochberg procedure. Fold changes were calculated by dividing values of the later generation population (LGP) by values of the early generation population (EGP).

Metabolic clustering of differentially expressed genes (DEGs) was performed with the EcoCyc E. coli database regarding all features with p-values < 0.05. Mapping and expression profiling statistics are shown in Supplementary Tables 3-4.

3.9 Deep sequencing of plasmids from early and late generation population

Plasmid extraction was performed from the same cultures described in the previous section. Thus, a total of 12 samples were processed for plasmid heterogeneity analysis. Plasmid DNA extraction was performed using a standard kit (ThermoFisher Scientific GeneJET Plasmid Miniprep Kit). To investigate potential plasmid loss during the long-term cultivations, we recorded pDNA contents in early and late generation populations. The subsequent library preparation and deep sequencing was conducted by Eurofins Genomics with the Illumina Novaseq 6000 S4 technology with paired-end reads of 150 bp. A per base coverage depth of at least 140.000 x could be obtained from all sequences plasmid samples (*Supplementary figures 1-3*). Adapter trimming, quality filtering and per-read pruning were performed to obtain only high quality bases. Afterwards, reads were mapped to the corresponding reference plasmid sequence, followed by a single nucleotide variant calling. Reads that could not be mapped to the plasmid reference sequence, were then mapped against an insertion sequence elements database (131) with the Burrows-Wheeler Aligner. Reads with positive hits, were then blasted against each individual insertion sequence family with NCBI's megablast algorithm (132) and counted afterwards. Only highly similar sequences were regarded (*Fig. 3.9)*.

3.10 Localisation of Insertion sequence entry sites

The software tool Genome Artist (**Ar**tificial **T**ransposon **I**nsertion **S**ite **T**racker) was used to localise so-called target site duplications (TSD), i.e. reads that have duplicated plasmid sequences indicating IS insertion (133). Those reads that could be at least partially mapped to Insertion sequence elements were specified as queries. Alignments were performed against the plasmids sequences pCYS, pCYS_i and pCYS_m. Only TSDs with highest alignment scores were selected. The exact insertion sites were allocated to the genes directly affected by the transposon insertion and resolved to the nucleotide level.

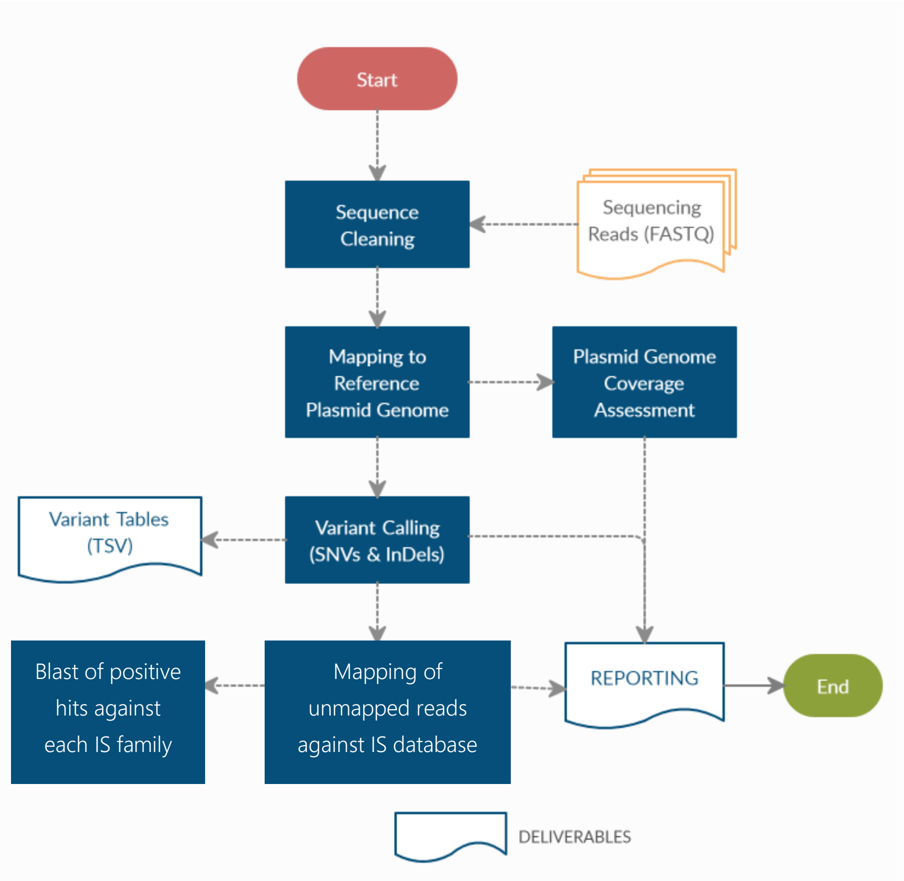

Start
Sequence Cleaning
Sequencing Reads (FASTQ)
Mapping to Reference Plasmid Genome
Plasmid Genome Coverage Assessment
Variant Tables (TSV)
Variant Calling (SNVs & InDels)
Blast of positive hits against each IS family
Mapping of unmapped reads against IS database
REPORTING
End
DELIVERABLES

Figure 3.9: Pipeline for sequencing and bioinformatic analysis of plasmid DNA isolated from early and late generation of E. coli W3110 and MDS42 populations*. Sequencing was conducted with the technology of Illumina Novaseq 6000 S4 technology with paired-end reads of 150 bp. Reference genomes were plasmid sequences of pCYS, pCYS_i and pCYS_m. Unmapped reads were mapped against an Insertion sequence database ISfinder and later blasted against each Insertion sequence family with NCBI's megablast algorithm.*

4 Results

4.1 Long-term fermentation simulation

In order to investigate phenotypic and genetic diversity in L-cysteine producing *E. coli* populations over relevant generation numbers accumulating on an industrial scale, a gradual scale-up growth process was established (*Fig. 3.4*). Thereby the *E. coli* K-12 strains W3110 and MDS42 harbouring different L-cysteine production plasmids were serially transferred into fresh culture medium every 10 and 5 hours, respectively (*Supplementary Table 1*). Baffled shaking flasks and high r.p.m. ensured proper O_2 supply for the cells. Hence, an exponential growth phase was guaranteed at all times throughout the experiment. Constant antibiotic selection with tetracycline was applied in an effort to prevent potential plasmid loss. At each transfer step, optical density was measured and the generation counts recorded until the cells had finally progressed through more than 60 cell division cycles (*Fig. 4.1*). Furthermore, growing *E. coli* populations were immediately cryopreserved to subsequently investigate time-dependent phenotypic and genetic variability.

MDS42 populations reached 60 generations at about the same time with different integrated plasmids while the time after W3110 populations reached 60 generations was different, depending on the plasmid. Populations with integrated pCYS required the most time with approximately 150 hours. Overall, results demonstrated that the MDS42 strain reached the threshold much faster (65 hours) than the W3110 strain (115-150 hours), indicating much higher growth rates of the MDS42 strains.

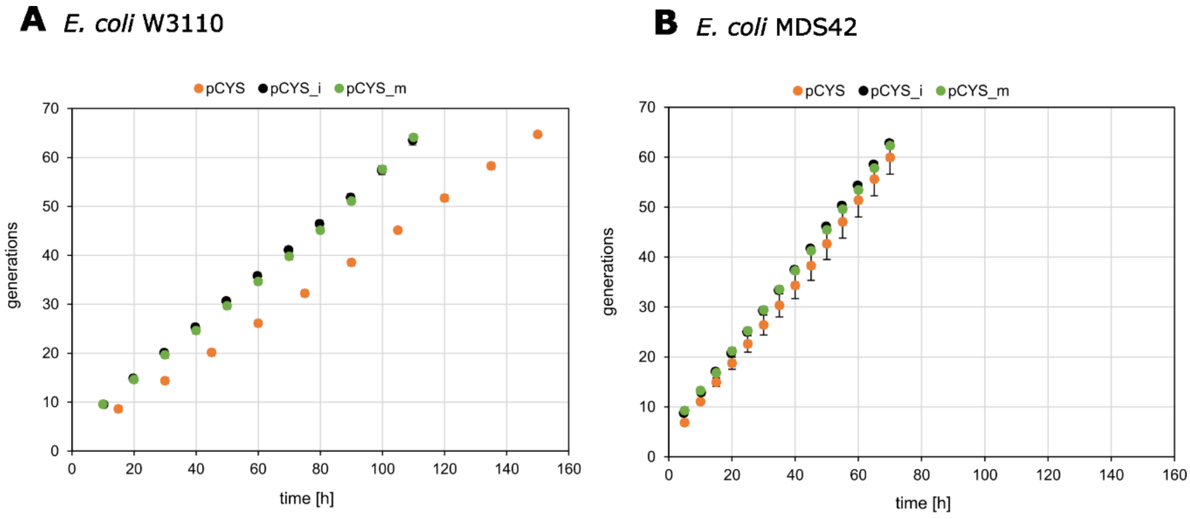


Figure 4.1: *Average number of generations after each passage of the simulated fermentation of E. coli W3110 (A) and MDS42 (B) with integrated pCYS, pCYS_i and pCYS_m. Each strain was cultivated in biological triplicates.*

4.2 Stability of L-cysteine-producing phenotypes

To investigate phenotypic variability and the stability of L-cysteine producing phenotypes, the generated cryo-cultures, as described in 4.1, were re-cultivated for a further 72 h in a new batch.

Using three different plasmid constructs designed for L-cysteine production, titres and yields were analysed in the *E. coli* W3110 and MDS42 populations during long-term cultivation (*Fig. 4.2A*).

All three plasmids carried bottleneck genes for the synthesis of L-cysteine, i.e. *cysE* and *serA*, which code for serine acetyltransferase (SATase) and phosphoglycerate dehydrogenase (PHGDH), respectively. In particular, *serA* encodes a feedback-insensitive variant of PHGDH that cannot be inhibited by the presence of L-serine (81). The same applies to *cysE* through the presence of L-cysteine (82). For cell viability, it is vital to diminish high concentrations of L-cysteine because of its toxicity. Therefore, the gene *eamA* is included in all plasmids, with the corresponding protein being an L-cysteine/O-acetylserine exporter. The plasmid pCYS_m carries an additional gene *cysM*, coding for cysteine synthase B, under control of the stationary phase promoter *fic1*. Thus, the delayed L-cysteine production in the stationary phase was intended to be decoupled from cell growth. While the genes in pCYS are under the control of native genomic promoters, the genes in pCYS_i were arranged in a single operon under the control of the constitutive promoter GAPDH. Moreover, non-coding backbone sequences were trimmed to potentially reduce the metabolic load on the production cells.

During the long-term cultivation, final L-cysteine yields of each sampled population were determined. Although an increase in relative L-cysteine yields of modified W3110 strains was observed initially, reaching a maximum, the product concentration at 60 accumulated generations only amounted to 15-35% of the maximum (*Fig. 4.2.1B*). The W3110 strain with integrated pCYS_i plasmid consistently produced high amounts L-cysteine up to approx. 60 generations, at which point there was a decline. In contrast, the other plasmids showed a gradual decrease in L-cysteine yields from generation 15-20 onwards. Transformed MDS42 strains showed no decrease in relative L-cysteine yields as generations accumulated. Instead, L-cysteine yields remained stable at 75-100% (*Fig. 4.2.1D*). Similar findings could be noted regarding the overall L-cysteine production in W3110 and MDS42 populations (*Fig. 4.2.2*). However, MDS42 populations exhibited up to 50% lower overall yields of L-cysteine compared to W3110 populations.

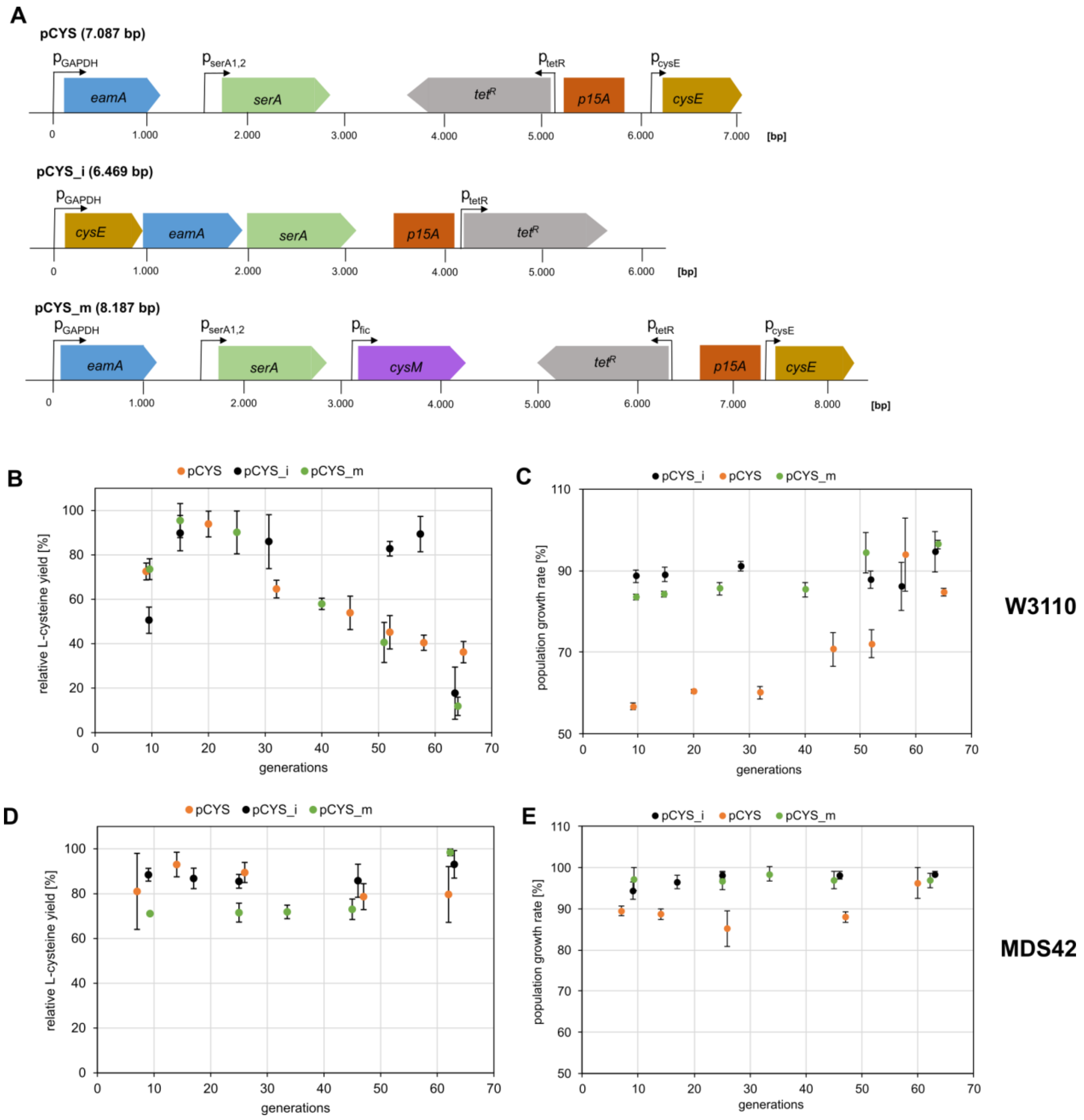


Fig. 4.2.1: Stability of L-cysteine producing phenotypes (80). A*: Plasmids used for transformation in E. coli W3110 and MDS42 to study the stability of L-cysteine production. Boxed arrows indicate genes located on the plasmid, whereas bended arrows display corresponding promoters.* ***B-E****: Plots of relative L-cysteine yields in % (B + D) and the population growth rates (C + E) as a function of the number of accumulated generations. Cultivation and subsequent L-cysteine yield measurements were carried out in biological triplicates of W3110 and MDS42 with the three different plasmids shown in A. L-cysteine yields were normalised based on the OD_{600} and the highest yield of the corresponding biological replicate. Population growth rates were normalised based on growth rates of the corresponding non-producing strains harbouring empty vectors (Supplementary Table 2).*

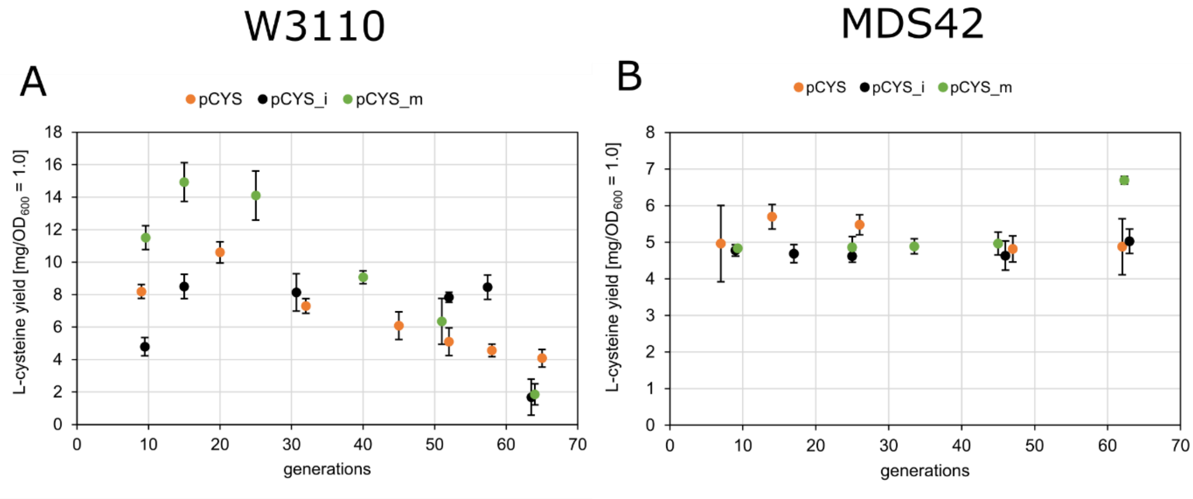


Fig. 4.2.2: Plots of total L-cysteine yields in mg/OD$_{600}$ = 1.0 (80)*. Cultivation and subsequent L-cysteine yields determination was carried out in biological triplicates of W3110 (A) and MDS42 (B) with the three different plasmids pCYS, pCYS_i and pCYS_m.*

To assess the strain viabilities, growth rates of the sampled populations were computed. The growth rates of W3110 showed variability depending on the transformed plasmid and the generation number (*Fig. 4.2.1C*). No considerable alterations were observed in the growth rate of W3110 with the integrated pCYS_i plasmid, whereas a substantial shift in growth fitness was detected in the case of W3110 harbouring pCYS and pCYS_m. By comparing initial and final growth rates, an increase of 13% for pCYS_m and even 27% for pCYS was observed. No noticeable effects on growth fitness were recorded in MDS42 populations.

4.3 Transcriptome analysis of early and late *E. coli* W3110 and MDS42 populations

To investigate the molecular and cellular mechanisms responsible for the L-cysteine production decline, transcriptomes of W3110 and MDS42 populations subjected to long-term cultivation were examined. Thereby, RNA from early and late generation populations (EGP, LGP) was extracted and sequenced to create transcriptome libraries (*Fig. 4.3.1A*). Transcriptomes from EGPs were then compared with the corresponding LGPs in a principal component analysis (*Fig. 4.3.1B*). Using this approach, one can assess global relationships between samples. Indeed, the EGPs and LGPs of the MDS42 strains were located in very close proximity, indicating minimal disparity between the transcriptome datasets. Large deviations could be observed between early and late W3110 populations regardless of the integrated plasmid. However, the variance of 15% in the EGPs and LGPs of W3110 strains carrying the pCYS_i plasmid was lower than of early and late W3110 populations harbouring either pCYS or pCYS_m, which exhibited a variance of 30%.

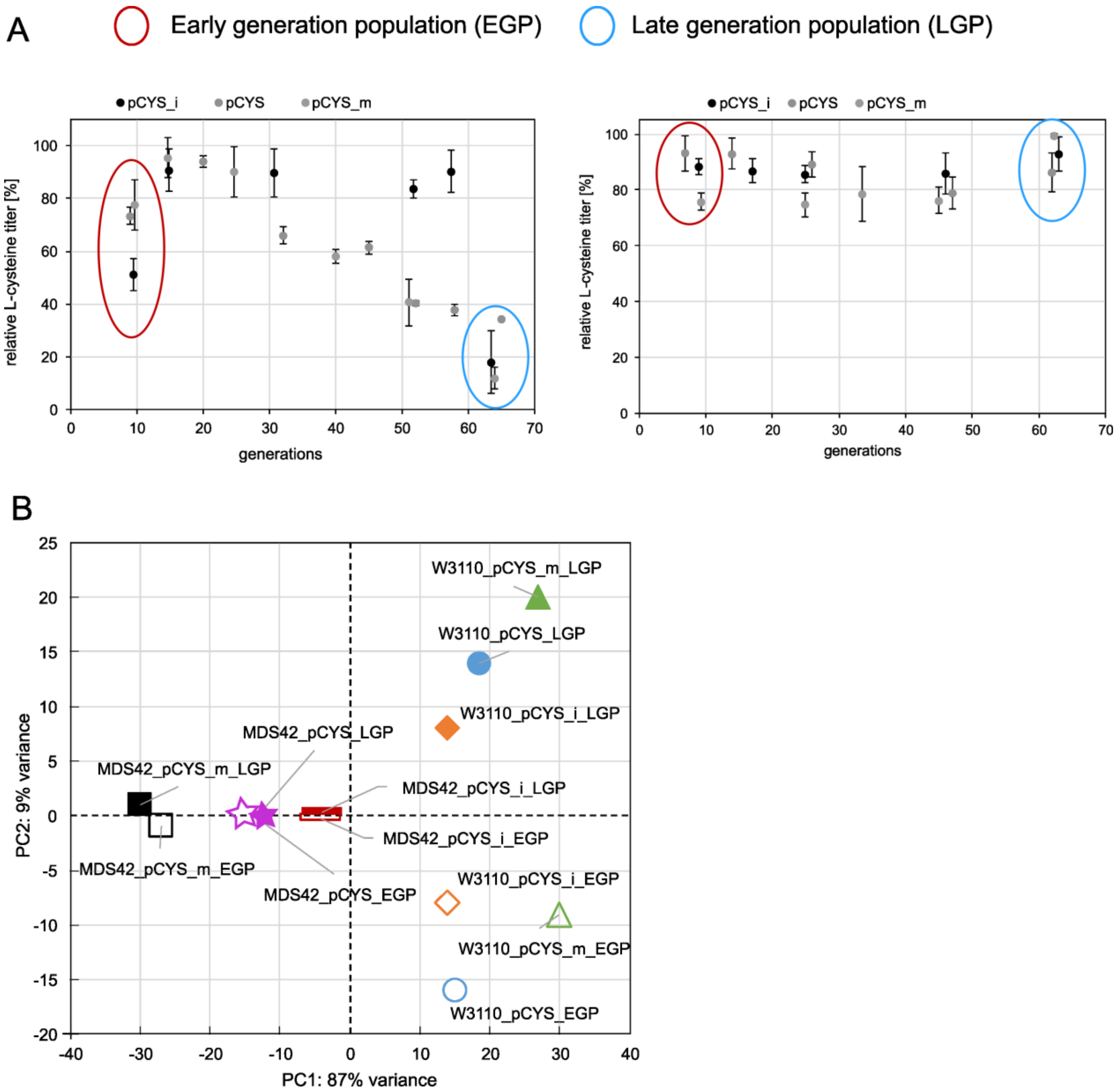


Figure 4.3.1: Comparative transcriptome analysis of sampled populations during long-term cultivation (80). *RNA was extracted from early and late generation population (EGP = red, LGP = blue) of E. coli MDS42 (right) and W3110 (left) harbouring the different L-cysteine production plasmids pCYS, pCYS_i and pCYS_m (**A**). Extracted RNA was sequenced and the resulting transcriptome libraries subsequently compared. (**B**): Principal component analysis (PCA) of all samples subjected to transcriptome analysis. Early generation populations are marked with unfilled forms, while forms of the late generation populations are filled. EGPs and LGPs correspond to populations with 7-10 and 60-65 evolved generations, respectively. Related samples are labelled with the same colour. Related samples in close proximity indicate minimal differences between them whilst samples placed more distantly suggest bigger variations.*

In order to obtain a more profound understanding of the variations in expression levels between EGPs and LGPs, an examination of differentially expressed genes (DEGs) with a p-value cut-off of 0.05 was conducted (Supplementary tables 5-10). Once the DEGs were clustered based on their metabolic function, a plot was created illustrating the number of DEGs within a cluster as a function of the mean fold change of the DEGs within the same cluster (*Fig. 4.3.2*). Fold changes were calculated by dividing values of the later generation population by values of the early generation population. Genes with unknown function as well as clusters mapped with only one gene were disregarded (single gene cluster) were disregarded here, but can be found in the appendix (*Supplementary tables 5-10*).

Overall, a lower number of different metabolic clusters was observed in MDS42 populations compared to W3110 populations. In early generation populations, general stress features such as acid resistance, hyperosmotic stress, acetyl-CoA metabolism and carbon assimilation were frequently up-regulated. Furthermore, genes for nitrate assimilation, degradation of pyrimidines and L-arginine were highly abundant in EGPs, indicating nitrogen deficiency during the phase when L-cysteine production was highest. However, one feature in EGPs most prominent with 17-20 genes within the cluster was assigned to sulphur- and L-cysteine starvation (*Table 4.3.1*). Genes related to the taurine and aliphatic sulphonate operon, as well as transport proteins for sulphur and thiosulphate, were identified within the cluster.

In late generation populations cluster were upregulated including biofilm formation, iron starvation, anaerobic respiration and metal transporters (e.g. copper, silver, nickel). Two clusters were particularly interesting regarding L-cysteine metabolism and genetic instability of *E. coli* cells. With a 3.5 to 4.5-fold upregulation, the cluster of an operon for L-cysteine detoxification could be identified in W3110 populations (*Table 4.3.1*). The operon comprises two specific genes *cyuA* and *cyuP*, the former encodes a putative L-cysteine desulphidase and the latter encodes a transporter for D-serine. Moreover, features involving transposases of insertion sequence families 3 and 66, mismatch repair genes and cellular response genes to DNA damage were upregulated in late generation populations.

Next, it was crucial to determine whether genetic modifications or errors responsible for the decline in L-cysteine production and the resulting increase in fitness were located in the genome or on plasmids.

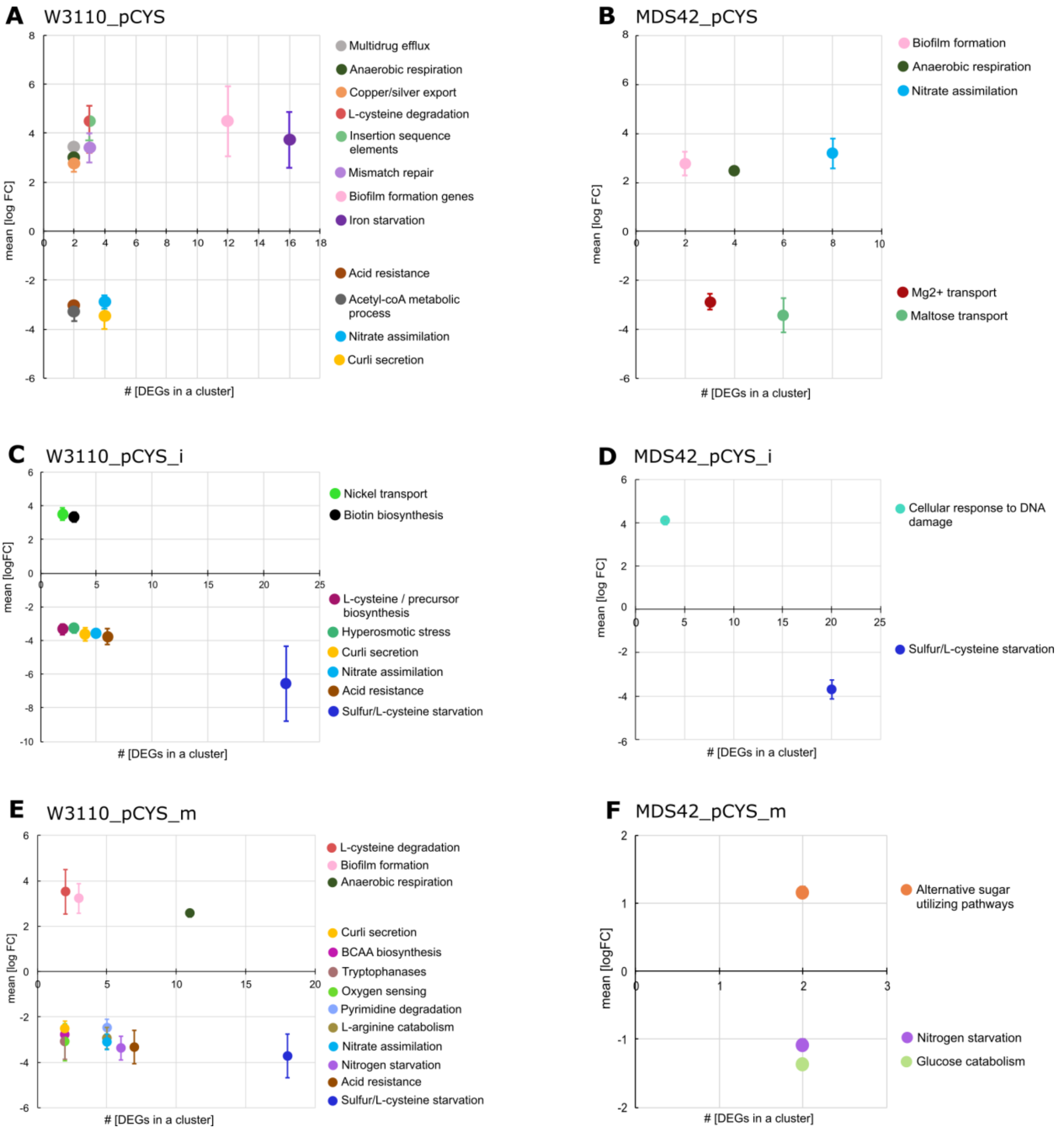


Fig. 4.3.2: Clustering of differentially expressed genes (DEGs) based on the metabolic function in E. coli (80)*. For this, logarithmic fold change (logFC) medians of DEGs with the same metabolic function were plotted against the number of DEGs within this group. Fold changes were calculated by dividing values of the later generation population (LGP) by values of the early generation population (EGP). LGPs correspond to populations with 60–65 evolved generations, while EGPs correspond to populations with 7–10 evolved generations. E. coli strains W3110 (A, C, E) and MDS42 (B, D, F) transformed with one of three plasmids engineered for L-cysteine production (pCYS: A, B; pCYS_i: C, D; pCYS_m: E, F) were analysed. A p-value cut-off < 0.05 was selected. Genes with unknown function as well as clusters mapped with only one gene were disregarded here, but can be found in Additional file 1: Tables S6-S11.* ***A****: 81 genes in total (57 genes within 12 cluster, 20 genes with unknown function and 4 single gene clusters).* ***B****: 34 genes in total (34 genes within 5 cluster, 6 genes with unknown function and 2 single gene clusters).* ***C****: 65 genes in total (47 genes within 8 cluster, 10 genes with unknown function and 8 single gene clusters).* ***D****: 23 genes within 2 cluster.* ***E****: 105 genes in total (70 genes within 12 cluster, 17 genes with unknown function and 18 single gene clusters.* ***F****: 14 genes in total (6 genes within 3 cluster, 4 genes with unknown function and 4 single gene clusters)*

Table 4.3.1: List of sulphur and L-cysteine metabolism related operons detected as clusters in transcriptome analysis (80).

DEGs of sulphur-/ L-cysteine metabolism related operons	**Metabolic function**	**Literature**
cyuAP	L-cysteine desulphidase, L-cysteine utilization permease	(134, 135)
aslB	Putative anaerobic sulphatase maturation enzyme	(136)
tauABCD	Taurine utilization proteins	(137)
ssuEADCB	Aliphatic sulphonates utilization proteins	(138)
cysPUWA, sbp	Sulphate/thiosulphate transport proteins	(139, 140)
cysDNC	Sulphate activation proteins	(141)
cysJIH	Sulphite reductase proteins	(142)
tsuAB	Thiosulphate transport proteins	(143)

Table 4.3.2: List of genes related to genetic instability detected in transcriptome analysis (80).

DEGs belonging to genetic instability	**Metabolic function**	**Literature**
insJK	IS3 family transposase	(144)
yjgZ	Putative IS66 family transposase	InterPro
ybcN, rusA, iprA	DNA repair	(145), (146), (147)

4.4 Source of genetic error modes

A simple attempt for investigating whether mutant plasmids are accountable for the decline in L-cysteine production was to extract plasmids from EGPs and LGPs and subsequently introduce them into fresh strains to test L-cysteine production and growth rates (*Fig. 4.4*).

It became evident that plasmids from LGPs in fresh W3110 strains had substantially lower L-cysteine yield than EGPs. For pCYS and pCYS_i the yields were 50% lower and for pCYS_m even 83%. At the same time, W3110 strains with integrated LGP plasmids exhibited higher growth rates than those with EGP plasmids.

Neither L-cysteine yields nor growth rated of MDS42 varied with integrated EGP- and LGP plasmids.

As a consequence of this experiment, there was a clear indication that the subsequent investigations would focus on the plasmids rather than the genomes of W3110 and MDS42 populations. Based on observations regarding the upregulation of genes indicating genetic instability in combination with the loss of L-cysteine production in evolving populations and the concomitant increase in fitness, an approach for the examination of plasmid sequences and quantities had to be established.

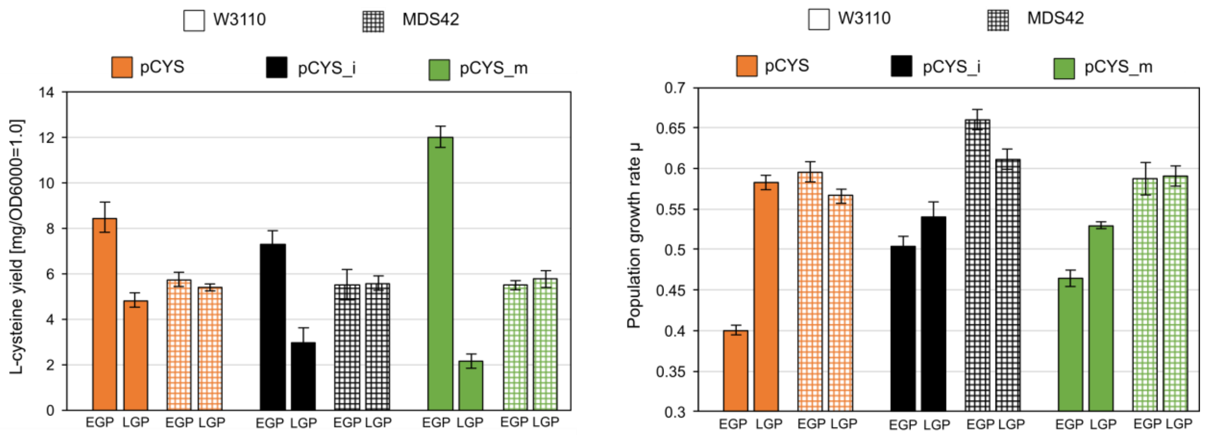


Figure 4.4: Determination of L-cysteine yields and population growth rates of fresh strains transformed with extracted plasmids from EGPs and LGPs of W3110 and MDS42. *Plasmids from early and late generation populations (EGP, LGP) were transformed into fresh E. coli MDS42 and W3110 strains and tested for L-cysteine yields and population growth rates. Filled = W3110, gridded = MDS42. Different plasmids are color-coded (pCYS: orange, pCYS_i: black, pCYS_m: green).*

4.5 Plasmid deep sequencing and examination of early and late *E. coli* W3110 and MDS42 population

Initially, an investigation should be conducted to determine if the reduction of L-cysteine productivity and enhancement of growth fitness in evolving populations was attributed to plasmid losses. Therefore, contents of extracted plasmids per µg cells of EGs and LGPs where compared (*Table 4.5.1*). No relevant differences were observed in plasmid contents between early and late population of *E. coli* W3110 and MDS42. As a result, the decrease in L-cysteine production could not be attributed to a reduction in plasmid quantities in LGPs.

Finally, the plasmid sequences should be checked for any sort of mutations. The difficulty in identifying mutations in isolated plasmids arose from the insufficient sequencing depth in classical sanger sequencing to resolve low abundant mutations within a huge amount of plasmid sequences from an entire population. Given that PCR-based approaches coupled with gel electrophoresis can solely provide qualitative results pertaining to the presence of mutations, an approach for quantitative evaluation of mutated sequences was sought.

Therefore, a deep-sequencing approach was performed using Illumina Novaseq 6000 technology. With 150 bp, paired-end read sequencing, average per base coverage depths of > 140.000 x were achieved (*Supplementary figures 1-3*).

After alignment of reads to the different plasmid sequences, a single nucleotide polymorphism analysis (SNP) was performed. Thereby no SNP could be detected within any open reading frame of the L-cysteine pathway genes. Instead, negligible SNPs were found in the backbones (*Table 4.5.2*).

However, when examining the alignment statistics, the number of unmapped reads in LGPs of W3110 almost doubled compared to EGPs (*Fig. 4.5A*). In a next step, those unmapped reads were aligned against an insertion sequence database of *E. coli (*ISFinder), resulting in higher coverages of aligned reads to IS in LGPs. Although reads that could be mapped to insertion sequence elements (ISE) were also present in plasmids of MDS42 populations, their number remained unchanged. In addition, the read count was 10 times smaller than for plasmids from W3110 populations.

Since the presence of ISE fractions in isolated plasmid sequences had been demonstrated, it was crucial to identify and assign them precisely. For this task, NCBI’s megablast algorithm became beneficial. This algorithm has been designed with the focus on identifying long alignments between very similar sequences, making it an ideal tool for discovering an exact match to an ISE query sequence. Additionally, reads mapped to ISAs1 family were very abundant.

The distribution of the various IS families exhibited high similarity across all three L-cysteine production plasmids. This observation was the same when comparing EGPs to LGPs. (*Fig. 4.5B*). Among the plasmids obtained from W3110 populations, reads that could be mapped to the IS3 and IS5 family were identified most frequently. The plasmids derived from MDS42 populations primarily contained reads that could be aligned to sequences of the IS200 and IS110 families.

Table 4.5.1: Values of calculated plasmid DNA contents extracted from early and late generation populations (EGP, LGP) (80). *Plasmid DNA was extracted from 10 ml cultures and eluted in 50 µl each. Cell dry weights were extrapolated with the factor of 0.33 g/L/OD$_{600= 1.0}$ for E. coli K-12 MG1655 cells according to Sauer et al (148).*

Plas-mid	**pCYS_m**				**pCYS_i**				**pCYS**			
Strain	**W3110**		**MDS42**		**W3110**		**MDS42**		**W3110**		**MDS42**	
Population	**EGP**	**LGP**	**EGP**	**LGP**	**EGP**	**LGP**	**EGP**	**LGP**	**EGP**	**LGP**	**EGP**	**LGP**
OD$_{600}$	4.3	4.8	4.6	4.7	4.4	4.6	4.8	4.4	4.5	4.3	4	4.4
Cell Dry mass [µg]	1.42E +04	1.58E +04	1.52E +04	1.55E +04	1.45E +04	1.52E +04	1.58E +04	1.45E +04	1.49E +04	1.42E +04	1.32E +04	1.45E +04
pDNA [ng/µl]	**96**	**112**	**52**	**38**	**103**	**123**	**44**	**47**	**67**	**75**	**51**	**62**

pDNA mass [µg]	4.82	5.63	2.62	1.9	5.15	6.18	2.23	2.36	3.4	3.8	2.6	3.13
µg pDNA/ µg cells	3.40E -04	3.56E -04	1.73E -04	1.23E -04	3.55E -04	4.08E -04	1.40E -04	1.63E -04	2.29E -04	2.65E -04	1.94E -04	2.16E -04

Table 4.5.2: Single-nucleotide polymorphism (SNP) variant table (80). *The SNP calling was done using VarSCan2 (149). Allele frequency cut-off used for variant calling was 1%. For each sample, the following variant summary is provided: Strain, plasmid, population. Additionally, POS: Position at which the variant was observed, BB/ORF: Location affected by the mutation (BB: backbone, ORF: open reading frame), REF: Reference base, ALT: Alternative base, Allele Freq: Variant allele frequency in percentage, Alt Depth: Depth of variant-supporting bases, total depth: Depth of variant-supporting bases and reference-supporting bases. Mutations which showed allele-frequencies >95% were assumed to be originally present in the plasmids.*

Strain	Plasmid	Population	POS	BB /ORF	REF	ALT	Allele-Freq	Alt_Depth	Total_Depth
MDS42	pCYS	EGP	3057	BB	A	C	55.14	3425	206153
		LGP	3057	BB	A	C	54.19	4131	248335
	pCYS_i	EGP	-	-	-	-	-	-	-
		LGP	-	-	-	-	-	-	-
	pCYS_m	EGP	1253	BB	G	GT	95.65	140272	134362
			1397	BB	A	T	1.07	1162	108962
			2523	BB	C	A	3.4	3818	112358
			2637	BB	G	A	99.93	151469	151574
			2625	BB	A	T	99.87	148937	149131
		LGP	2523	BB	C	A	3.38	4360	129097
			2637	BB	G	A	99.98	179578	179647
			2625	BB	A	T	99.89	170523	170718
			2645	BB	A	C	99.98	179578	179647
			1253	BB	G	GT	95.83	157198	152298
W3110	pCYS	EGP	3057	BB	A	C	54.09	3040	171653
		LGP	3057	BB	A	C	54.66	2446	140778
	pCYS_i	EGP	-	-	-	-	-	-	-
		LGP	-	-	-	-	-	-	-
	pCYS_m	EGP	1253	BB	G	GT	95.3	105873	93750
			2523	BB	C	A	3.89	3147	80897
			2625	BB	A	T	99.91	108989	109092
			2637	BB	G	A	99.95	169794	169866
			2645	BB	A	C	99.96	113604	113669
		LGP	1397	BB	A	T	1.11	1308	117871
			2523	BB	C	A	3.65	4455	122235
			2645	BB	A	C	99.97	169794	169866
			2625	BB	A	T	99.9	161681	161857
			1253	BB	G	GT	95.5	153242	144568

Despite using an experimental approach to identify various IS subgroups/ families within plasmids computationally, an accurate localization of the insertion sites has not been achieved. Indeed, it was intriguing to ascertain the existence of particular sequence motifs within plasmids that enable transpositions and the degree of dissimilarity in such motifs across detected IS families.

A

Plasmid	pCYS				pCYS_i				pCYS_m			
Strain	W3110		MDS42		W3110		MDS42		W3110		MDS42	
Population	EGP	LGP	EGP	LGP	EGP	LGP	EGP	LGP	EGP	LGP	EGP	LGP
Unmapped reads to plasmid sequence [%]	4.4	8.2	2.7	2.5	3.3	6	2.5	2.4	3.8	7.1	2	2.1
Reads mapped to IS [%]	0.13	0.24	0.01	0.01	0.11	0.18	0.01	0.01	0.12	0.23	0.01	0.01
Reads mapped to IS [total]	3803	5852	271	249	2361	5253	261	245	2101	6070	180	196

B

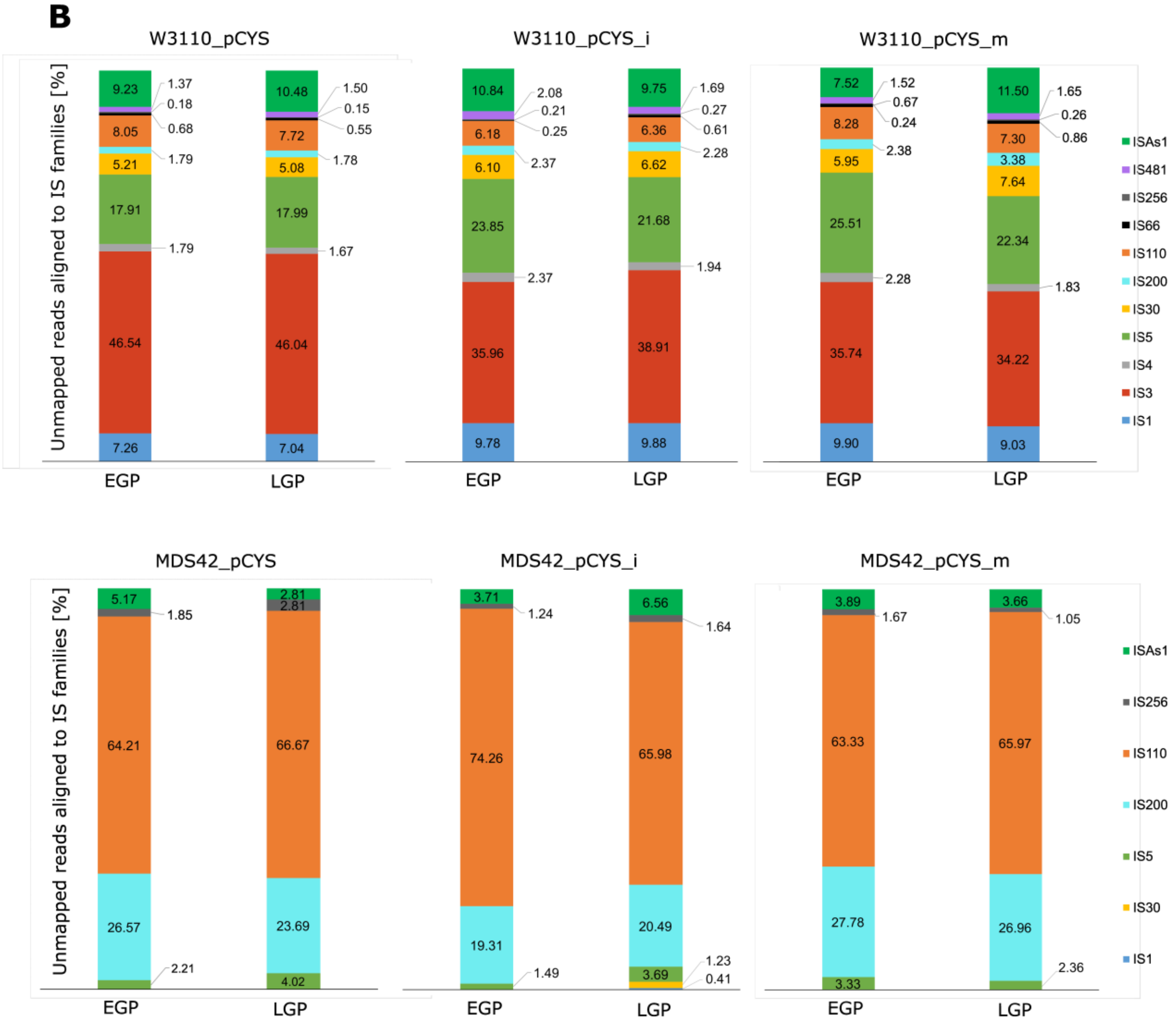


***Figure 4.5: Mapping of sequenced reads against insertion sequence (IS) families (80). A**: Table showing the percentages of reads that did not map to the plasmid sequences as well as the percentages of those unmapped reads mapped to IS. Plasmids pCYS, pCYS_i and pCYS_m extracted from early and late generation populations of W3110 and MDS42 got deep sequenced (Methods). **B**: Proportions of unmapped reads aligned to the different IS families in percent. The caption is arranged in the same order as the stacked bars. Sequencing was conducted with Iluumina Novaseq 6000, 150 bp paried-end reads.*

4.6 Localisation of IS entry sites within plasmids of late W3110 populations

Employing an Illumina sequencing approach and utilizing short 150 base pair reads, it proved to be a daunting task to identify junction reads that encompassed both the plasmid sequence and the insertion sequence. Indeed, the presence of lengthy and repetitive sequences, such as those located in the inverted repeat regions within IS, made it almost impractical to pinpoint the insertion sites using conventional mapping algorithms.

Instead, a strategy involving the appearance of target site duplications (TSD) subsequent to the transposition of IS was employed. Those TSDs arise from staggered double-strand breaks at the target site, which were utilised to determine the IS entry sites in this study.

Due to the predominant identification of reads mapping to families 3 and 5 of insertion sequence elements, as seen in the previous section, the tracking for insertion sites was confined to these families.

The programme used to locate the TSDs was Genome ARTIST (**Ar**tificial **T**ransposon **I**nsertion **S**ite **T**racker) (133). It is readily determined where exactly the transposon was inserted, the gene that has been impacted by the insertion, as well as the neighbouring genes in close proximity to the insertion site (*Fig. 4.6*). Furthermore, the frequency of specific target duplication sites identified at a particular sequence locus were quantified.

In general, transpositions of IS3 and IS5 were distributed over the entire plasmid sequences. Target site duplications were detected within the plasmid backbone, genes of the pathway cassette, promoter sequences as well as within the propagation region of the p15A origin of replication. Transposition of the insertion sequence families 3 and 5 seemed rather random and not dependent on specific target sequences.

However, although the occurrence of TSDs was rather scattered across plasmid sequences, there was an accumulation of TSDs with the specific sequence motif "*ATAAAGCG*" in pCYS_m, totalling more than 100 copies. This sequence was detected in two locations within the pCYS_m plasmid concurrently, specifically within the open reading frame of both *eamA* and *cysM (Fig. 4.6C).* Thus, it was not possible to precisely determine the location of IS5 insertion in this particular case.

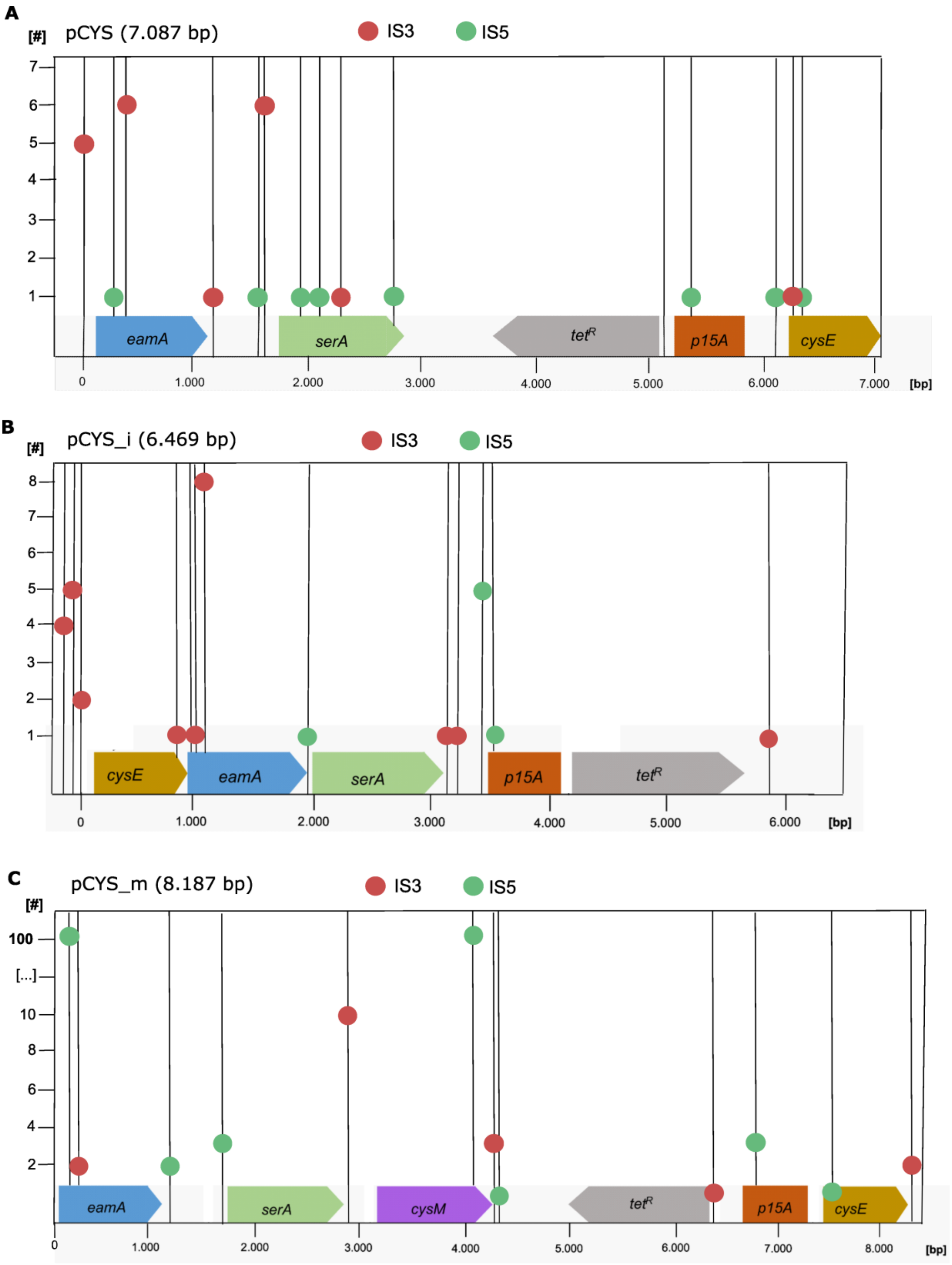


Figure 4.6: Localisation of IS entry sites within plasmids from late W3110 populations: *Insertion sequence entry sites of families 3 (red dots) and 5 (green dots) were determined and quantified based on target site duplications (TSD) originating from transposition events in pCYS (A), pCYS_i (B) and pCYS_m (C). Computational analysis was conducted with the Genome ARTIST (Artificial Transposon Insertion Site Tracker) software (133)*

5 Discussion

5.1 Rational design of L-cysteine overproduction

Designing a strategy for the targeted overproduction of L-cysteine in microorganisms has presented a substantial obstacle. Despite the availability of biotechnological and bioinformatic tools, it remains a compelling task to establish a cellular system that is economically viable.

Thiol-containing cysteine is not only a proteogenic amino acid, but is also crucial for the synthesis of essential metabolites that are involved in various biological processes. Firstly, L-cysteine provides sulphur for the biosynthesis of iron-sulphur (FeS) clusters and coenzyme A (CoA), which are vital for metabolic enzymes in both anabolic and catabolic pathways (60, 150). Secondly, it is a constituent of the Cys-Gly-Pro-Cys motif of thioredoxins that are involved in regulating protein dithiol/disulphide balance across all kingdoms. Hence maintaining L-cysteine and sulphur homeostasis is essential and must be carefully considered when designing an overproduction system.

This study achieved higher yields of L-cysteine by overexpressing feedback-insensitive variants of D-3-phoshphoglycerate dehydrogenase (*serA*) and serine acetyltransferase (*cysE*). In both proteins, mutations were introduced to the allosteric center, which prevented L-serine and L-cysteine from inhibiting the enzymes. The inhibition of SerA through L-serine is intelligible considering that 3-phosphoglycerate is an intermediate in the glycolytic pathway. Thereby L–serine binding has shown a cooperative behaviour, where mutants exhibited a higher sensitivity to allosteric inhibition, indicating a separation between the chemical process that lead to allosteric binding cooperativity and inhibition of the active site (151). The inhibitory mechanism observed in SerA affected Vmax, implying that L-serine primarily influences the reaction rate, rather than the binding affinity of the active site (152, 153). SerA requires the cofactor NAD^+ for the conversion of D-3-phosphoglycerate (PGA) to D-3-phosphohydroxypyruvate (PHP), which is reduced to NADH during the reaction.

The serine acetyltransferase (CysE) enzyme variant that was intentionally overproduced in this study was not an endogenous protein of *E. coli* but rather adapted from the model plant *Arabidopsis thaliana*. There are two serine acetyltransferase enzymes present in this organism, one of which is subject to allosteric inhibition by L-cysteine, while the other is not affected by this mechanism. In the latter, a segment of the N-terminus responsible for L-cysteine binding is truncated, resulting in a feedback-insensitive variant (154). This information was utilised to modify the native serine acetyltransferase in *E. coli* by removing the N-terminal segment. This approach eliminated the requirement for codon optimization, which would have been necessary if the SATase variant from *A. thaliana* was used instead. The truncated version yielded good results in terms of L-cysteine production. In the process of converting L-serine to L-cysteine, the enzyme serine acetyltransferase is activated by forming a cysteine synthase complex (CSC) with the downstream enzyme O-acetylserine sulphhydrylase-A (CysK) (*Fig 5.1.1*). In this process, CysE catalyses an acyl-moiety transfer from the co-factor acetyl-coA to L-serine, when L-serine is present in larger quantities. Normally, L-cysteine induces a conformational change that reduces the affinity for acetyl-coA. Furthermore, in the presence of HS^-, CysK triggers an allosteric modification in the active site of CysE to mitigate substrate and feedback inhibition in the context of CSC when present. Thus, the two mechanisms employed by the CSC serve as regulators, that enable cells to modulate the biosynthesis of L-cysteine in accordance with the prevailing growth conditions, including the internal sulphur content of the cell.

Applying the CSC concept to the protein variants employed in this study, it is inferred that while feedback inhibition of CysE by L-cysteine was absent, the production of L-cysteine remained contingent upon the levels of bisulphide present within the cell.

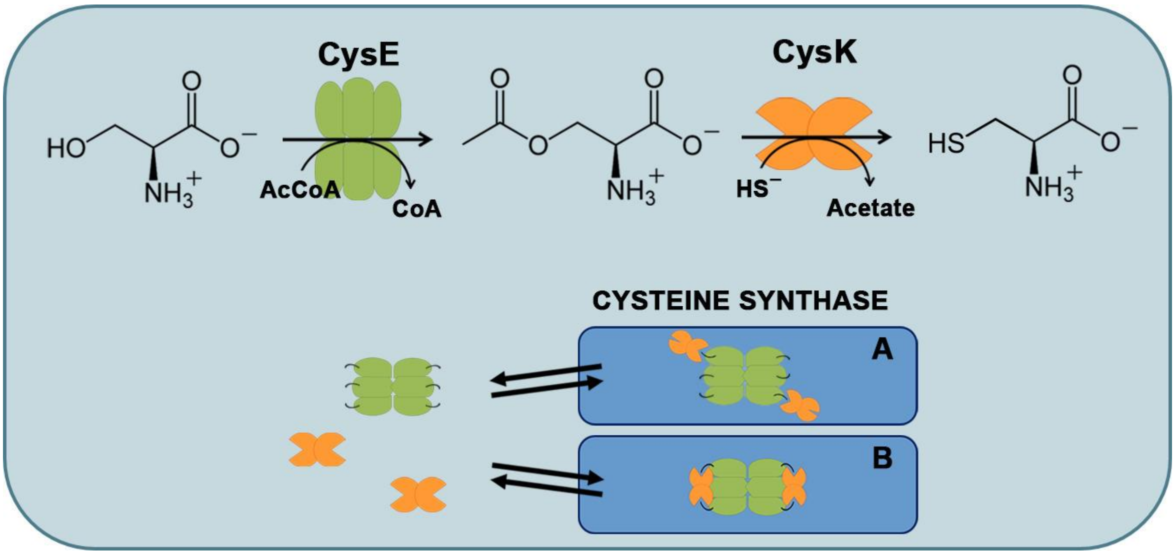


Figure 5.1.1: Cysteine synthase complex (CSC) and cysteine biosynthesis: *CysE and CysK are responsible for catalysing the last two steps of cysteine biosynthesis- The sulphate reduction pathway, which involes multiple steps (not depicted) produces bisulphide (HS-) as a product. CysK and CysE collaborate to form the bienzymatic CSC, but the exact three-dimensional structure of the complex is not yet known. Based on prior functional analysis, two potential models have been suggested for the complex. In model A, the C-terminus of CysE occupies only one active site of each CysK dimer, while in model B, both active sites are involved. Figure taken from Roberto Benini et al., (155). Reuse permission granted by John Wiley and Sons.*

A functional export system was required to prevent the toxicity of elevated levels of intracellular L-cysteine in *E. coli* cells, as cysteine can reduce ferric iron and stimulate the Fenton reaction.

Thereby, highly reactive hydroxyl radicals (OH^-) are generated, which are direct effectors for DNA damage (156). In addition, L-cysteine triggers sulphide production and suppresses a crucial enzyme in the isoleucine biosynthetic pathway by competing with threonine for the active site (83). Sulphides inhibit cytochrome *bo* oxidase, generating H_2O_2 by oxidation and mobilizing ferrous iron, which results in the effect described above (157).

In this study, EamA (*ydeD*) was selected as an effective export protein, given its superior binding affinity for L-cysteine and rapid transportation rates compared to the alternate O-acetylserine/cysteine exporter EamB (*yfik*). The *ydeD* gene was discovered in an industrially utilised *E. coli* strain, that overproduces L-cysteine (158). The gene product has six predicted transmembrane helices and belongs to the PecM family of export proteins. So far, no crystal structure could be obtained; nevertheless, very good structural predictions are available (*Fig. 5.1.2*). Both, overexpression of *ydeD* and *yfik* results in elevated excretion of L-cysteine and its precursor, O-acteylserine (OAS). Since the precise transport mechanism could not been elucidated yet, it is possible for both molecules to be transported via either uniport or symport mechanisms. EamA and EamB exhibit different effects on the temporal dynamics of OAS and L-cysteine export. In an EamA overproducing system, L-cysteine export is enhanced during the stationary phase, whereas in an EamB overproducing system, both OAS and L-cysteine are exported during the exponential phase (79).

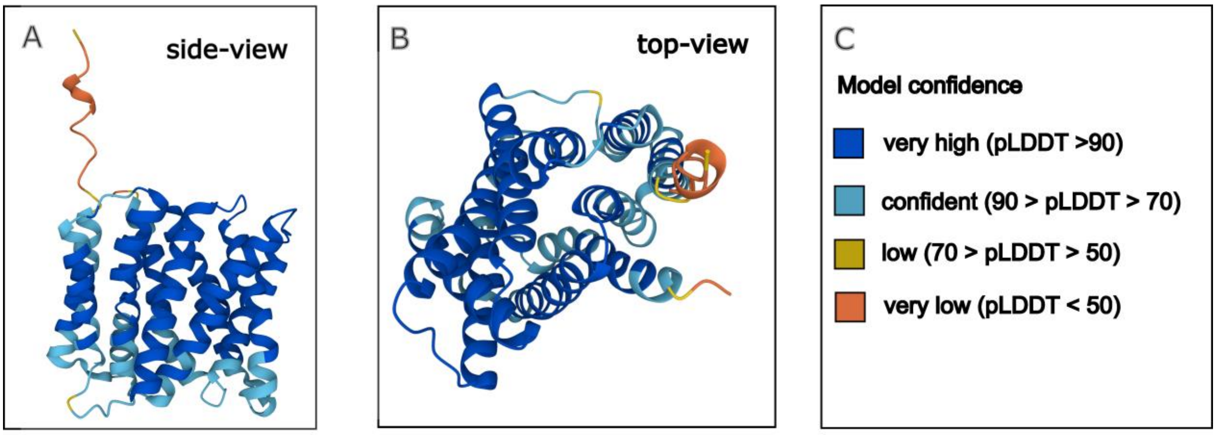


Figure 5.1.2: 3D structure prediction of EamA (ydeD) from Escherichia coli: *Structure prediction of EamA was conducted with AlphaFold and shows the side-view (A), top-view (B) and model confidence (C). AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructed in isolation.*

Exporting OAS comes with both advantages and drawbacks. Due to its harmful effects on cellular systems, it is crucial to regulate its levels, especially when sulphur is scarce. A deficiency in sulphur results in an accumulation of OAS, causing the aforementioned cysteine synthase complex to dissociate and cease L-cysteine production. Furthermore, the feedback-insensitive serine acetyltransferase exacerbates this effect by inhibiting L-cysteine's counteractive properties. However, from the perspective of maximising yield, every exported molecule of OAS results in the loss of one molecule L-cysteine. Ideally, a balanced production system that converts OAS effectively into L-cysteine, along with a transporter that only exports L-cysteine out of the cell, would be optimal. In turn, it is crucial to maintain sufficient intracellular levels of OAS as an activator for the transcription of genes involved in L-cysteine biosynthesis via the transcription factor CysB.

In an attempt to enhance the production of L-cysteine, *cysM* was introduced into the pathway construct of pCYS which encodes cysteine synthase B. CysM is an isozyme of CysK that facilitated the conversion of OAS and sulphide to L-cysteine. Unlike CysK, CysM does not enter into a complex with serine acetyltransferase. In addition, the broader substrate specificity of CysM enables it to effectively metabolise thiosulphate into S-sulphocysteine (SSC), thus making it a promising contender for augmenting L-cysteine synthesis (159). Throughout the cultivations, thiosulphate was fed to utilise the assimilatory sulphate reduction pathway as an alternative target for L-cysteine production. This approach also aimed at negating a potential sulphur limitation, as SSC, unlike L-cysteine, is not toxic to cells and is expected to play a role in the L-cysteine pool. Compared to sulphate, thiosulphate was the preferred substrate in assimilation as it can be transformed into L-cysteine in fewer steps without requiring ATP and necessitating less NADPH as a reduction equivalent (2x ATP and 4x NADPH compared to 2 x NAPDH).

Thiosulphate was imported into *E. coli* via an ATP-dependent sulphate-thiosulphate permease complex consisting of five subunits: Sbp, CysP, CysU, CysW and CysA while CysP was initiating the assimilation (160). As integral membrane proteins, CysU and CysW facilitated the transport of thiosulphate across the inner membrane to CysA, the ATP-binding subunit, which actively transported thiosulphate into the cytosol. Once inside the cytosol, CysM converted thiosulphate into SSC together with OAS.

Theoretically, *E. coli* has multiple options to transform SSC into L-cysteine, including several thioredoxins (Trxs), glutaredoxins (Grxs) and Grx-like proteins. Conversion was experimentally demonstrated with mainly Grx1 and NrdH (161). The two corresponding genes would be good targets for future overexpression in the systems used in this study.

Constitutive promoters were exclusively utilised for the overproduction of L-cysteine, since inducible systems are unprofitable and impractical for industrial large-scale fermentations. They allow continues and stable expression of target genes throughout the fermentation process, unlike inducible systems that require the addition of expensive inducer molecules to trigger gene expression. It also reduces the risk of low yields due to incomplete induction and the likelihood of genetic instability and plasmid loss. Synthetic inducible promoters often elicit expression in the uninduced state (leakiness) and low differences between the induced and uninduced states. Strong expression of L-cysteine pathway genes is rapidly accompanied by a depletion of cofactors, such as NADP(H) as reduction equivalent, and acetyl-CoA. Fine-tuning is therefore difficult and cannot realistically emulate cell demands. Intrinsic, native promoter systems on the other hand possess the characteristic of being repressed by metabolic control, ceasing production until environmental conditions have improved. However, constitutive promoters contribute to cell viability in presence of metabolic burden.

For these reasons, the L-cysteine production genes were cloned under the control of their respective native genomic promoter sequences. In the case of *ydeD*, the native promoter was replaced with the more efficient GAPDH promoter. A cloning strategy was utilised for pCYS_i wherein the genes were assembled into a coherent operon under the regulation of the GAPDH promoter. Furthermore, in the case of pCYS_m, the expression of cysteine synthase B was placed under the control of the constitutive stationary phase promoter fic1. This promoter normally controls the expression of the Fic protein (**f**ilamentation **i**nduced by **c**AMP) in the stationary phase through hybridizing with the RNA polymerase sigma factor RpoS. By utilizing this method, L-cysteine production through assimilatory sulphate reduction was expected to be shifted towards the stationary phase, consequently decoupling L-cysteine synthesis from cell proliferation and providing a more stable production system.

5.2 Simulated long-term cultivations

The process of fermentation simulation and long-term cultivations plays a critical role in studying the evolutionary process of populations during large-scale fermentations. The upscaling process from a master cell bank to a 200 m³ fed-batch bioreactor results in approximately 60-80 cell generations of *E. coli*, with population sizes of around 10^{20} cells. Such population sizes and timescales enable both the emergence and selection of non-producing cells, which can outcompete producer cells with reaching high cell densities in the final fermentation population.

The simulated long-term cultivation approach used in this study is a commonly employed method to study genetic and phenotypic changes that may occur during the evolutionary process. Serial transfer successfully emulated approximately 60 generations within 11-15 passages and 70-150 hours while maintaining populations in the exponential growth phase. Although no marker was used to record the exponential phase, the growth rates were established in a prior experiment to determine the optimal time for over-inoculation. This was a crucial step in avoiding experimental bias that could be caused by nutrient depletion in evolving populations.

It is important to note that long-term cultivation in shake flasks does not precisely represent the exact conditions in large bioreactors. Agitation and aeration conditions in shake flasks are often different from those in bioreactors, leading to disparities in shear stress, oxygen transfer rates as well as mass and heat transfer, which can affect microbial physiology and metabolism (162). In shake flasks, oxygen transfer occurs through two liquid surfaces: the bulk liquid surface and the liquid film on the wetted surface of the flask wall. The oxygen transfer rate (OTR) in shake flasks is influenced by various factors, such as flask size and shape, agitation speed, filling volume and ambient conditions. In bioreactors, the OTR depends on the impeller design, dimensions, sparger type, agitation speed, gas flow rate and gas concentration. All these parameters can be controlled and monitored in real-time with multiple online sensors. This enables oxygen transfer at virtually any location within the bioreactor.

Despite the aforementioned differences between shake flask- and bioreactor cultivation, the fundamental issue remains the same: Cells are driven by evolutionary pressure to eliminate factors that impair growth fitness. The underlying genetic and regulatory mechanisms used by *E. coli* to evade metabolic burden are poorly understood in developing populations and were successfully elucidated in this study using simulated long-term cultivation.

5.3 *E. coli* K-12 MDS42 as a negative control for genetic instability

In this study, the minimal genome strain *E. coli* K-12 MDS42 was used in addition to the classical *E. coli* K-12 W3110 strain. This was intended to serve as a negative control with limited genetic adaptability. In total, 14.3% of the genome of the parental MG1655 strain is deleted. This includes all active transposases and IS elements, error-prone DNA polymerases and 699 additional genes.

During L-cysteine production, the MDS42 strain exhibited remarkable stability, with no significant changes in L-cysteine titres or cellular growth rates observed throughout the cultivation. This stability was corroborated by transcriptome data, which indicated that the strain had lower adaptability, as evidenced by consistent expression levels of global transcripts in both early and late populations. The absence of active transposases was also confirmed when sequencing isolated plasmids from MDS42 populations, as there was no evidence of transposition of insertion sequences during the long-term cultivations.

It is important to note, that the MDS42 strain should not be compared to the W3110 strain beyond its utility as a negative control for genetic instability. Despite both strains being K-12 strains, they exhibit distinct metabolic characteristics. Due to the deletion of 699 genes in the case of MDS42, the strain possesses a lower replication load and a different carbon utilisation profile. Thus, the strain can grow better on minimal medium, as it uses the glyoxylate shunt to bypass the steps associated with CO_2 loss in the tricarboxylic acid cycle, which enables the cells to utilise substrates that enter central carbon metabolism at the level of acetyl-coA. This could be the reason why MDS42 displays an altered fatty acid composition (163). Furthermore, a reduced ability to catabolise certain amino acids was observed (120, 164). In addition, MDS42 has been found to have a slower growth rate under certain conditions, which may reflect differences in its regulatory networks or nutrient requirements. In fact, the analysis of "pathway holes" - referring to pathway reactions without corresponding enzymes identified in the genome - between the two strains revealed notable differences (Biocyc database, accessed on 03/08/2023). Specifically, the MDS42 strain exhibited a higher count of 125 (11.1%) such pathway holes, in contrast to the W3110 strain, which had a lower count of 46 (5.7%).

These metabolic distinctions highlight the importance of using caution when comparing MDS42 to other *E. coli* strains, even those within the K-12 lineage.

5.4 Instability of the L-cysteine producing phenotype

Before discussing the underlying causes of the collapse in L-cysteine production, the extent of the collapse should be illustrated and put into context with cellular fitness in *Escherichia coli*.

During long-term cultivation of W3110 strains carrying plasmids for L-cysteine production, a significant reduction in L-cysteine yields, ranging from 65-85% overall, was observed within 60 generations. In the initial 20 generations, yields increased, conceivably due to cellular adaption to the burdening environmental conditions. However, thereafter, the yields steadily declined, while populations with integrated pCYS_i plasmid maintained stable yields for further generations. It should be mentioned that the measured L-cysteine yields cannot be directly correlated with the L-cysteine production properties. The observed differences in yields could potentially reflect the degradation of L-cysteine, which would not necessarily affect production itself.

As an indicator of cellular fitness and metabolic burden, growth rates were also measured, which were inversely proportional to L-cysteine yields in W3110 populations. Others studies employed the implementation of GFP cassettes as an integrated marker protein to determine the “capacity” of the cells or used ^{13}C-tracer to analyse metabolic fluxes in differently induced cultures during recombinant protein production (165, 166).

Based on the observation of increasing growth rates, a correlation could be established between cellular fitness and the decreasing L-cysteine yields. Specifically, it was evident, that as W3110 populations gained fitness during cultivation, it came at the cost of L-cysteine production. Alternatively, considering the possibility of L-cysteine degradation mentioned earlier, populations exhibiting such degradation might have gained fitness advantage. The prolonged stability of W3110 populations harbouring the pCYS_i plasmid may be attributed to the plasmid’s reduced burden on cells, resulting from the deletion of non-coding backbone regions and the incorporation of L-cysteine pathway genes into a stringent operon. Nevertheless, these measures could not prevent a collapse of the L-cysteine yields towards the end of the cultivation.

Notably, the MDS42 populations exhibited 50% lower total L-cysteine yields compared to W3110 populations. This suggests that the lower L-cysteine concentrations and production capacities may have been less of a burden on the cell, resulting in stable growth rates and L-cysteine levels throughout the long-term cultivation. Such a result was surprising, as the MDS42 strain exhibited superior production of amino acids, such as L-threonine, compared to the wild-type *E .coli* strain MG1655 (120).

Based on the aforementioned findings, phenotypic variation, which could emerge in subpopulations over time, was also considered as a contributing factor. The characteristic trait of repetitive cycles of production, re-assimilation of overflow metabolites and the relaxation of stress responses would have been affecting both relative and absolute levels of L-cysteine as well as growth fitness. In fact, a rapid adaptation of populations to L-cysteine production was observed. L-cysteine yields increased during the first 20 cell division cycles, suggesting that populations were increasingly able to maintain production for a period of time in L-cysteine-rich environments. Subsequently, adaptations occurred that favoured cell growth at the expense of L-cysteine production or L-cysteine as such. However, no cyclic fluctuations of L-cysteine yields and growth rates were detected. Perhaps the duration of long-term cultivations with 60 accumulated generations was too short to observe this effect.

Besides cyclical fluctuations, phenotypic heterogeneity is also described by the emergence of a phenotype by adapting to changing conditions either responsively or stochastically (167). The responsive switching mechanism involves a cellular response to environmental cues, resulting in a change in phenotype to optimise temporal fitness. Conversely, stochastic switching is a long-term strategy that entails the expression of suboptimal phenotypes with lower fitness that may be better suited for future environmental conditions (*Fig. 5.4*). This particular case of phenotypic heterogeneity is known as “bet-hedging”. Numerous strategies for bet-hedging have been identified and reported, including the induction of sporulation and biofilm formation in *Bacillus subtilis*, utilization of respiratory mechanisms in denitrifiers during anoxic conditions, and different utilization of amino acids and carbohydrates in *Lactococci* and *E. coli* (168-171).

Without further extensive research, it is difficult to ascribe the observed effects during the long-term cultivations to a particular switching strategy. Nevertheless, it is plausible that the fundamental prerequisite for phenotypic switching, which involves communication through quorum sensing, occurred within the *E. coli* populations. During this process, small diffusible molecules known as autoinducers are produced and sensed from cell-to-cell. These autoinducers initiate cooperate behaviour within a microbial community and are concentration-dependent, indicating the number of cells in a population (i.e., ‘quorum’). As elaborated in the following section, certain populations exhibited increased gene expression associated with biofilm formation, which may suggest the occurrence of prior quorum sensing and subsequent phenotype switching.

Despite previous considerations, when examining the results of the plasmid transformation experiment, it was evident that the collapse of L-cysteine production, accompanied by an enhanced growth fitness, could not be solely attributed to phenotypic heterogeneity. Instead, the same effect was observed in freshly cultured strains with transformed plasmids extracted from early and late populations, which indicated the presence of a genetic influence alongside to phenotypic diversity. It was later determined that the contributing genetic factor was the presence of defective plasmids.

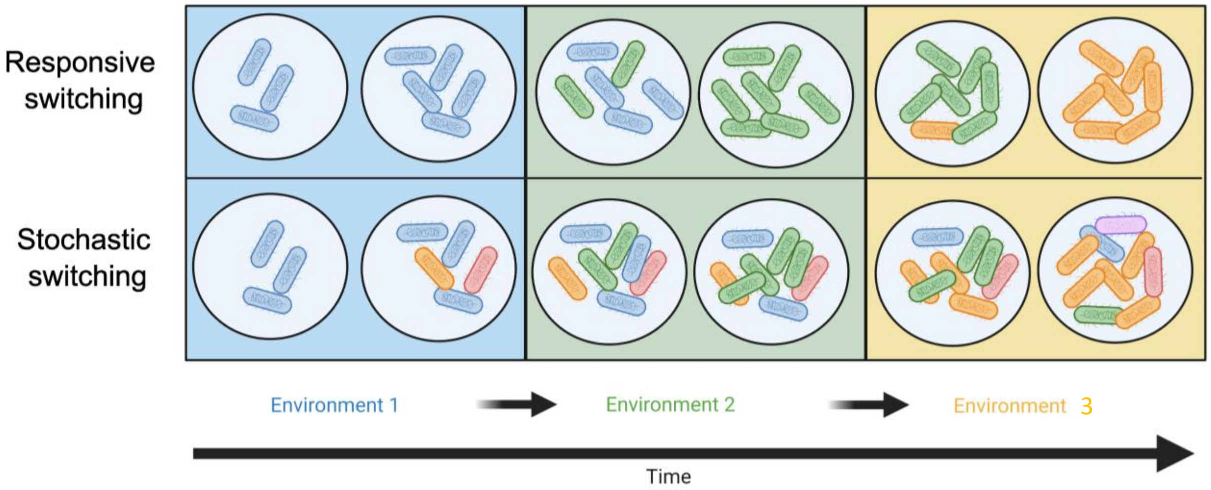


Figure 5.4: Phenotypic variation in microbial populations. *Microbial populations adapt to environmental changes by responsive switching- (upper panel) and stochastic switching mechanisms (lower panel). A Schematic model of the switching strategies is displayed, in which the fittest cells matches the colour of the surrounding environment. Cells with responsive switching strategies change their phenotype after sensing environmental cues to improve fitness temporally. Populations that employ random stochastic switching, express a variation of phenotypes of reduced fitness that may be better suited for future environments. (Figure adapted from Kussell et al. (172), (Created with Biorender).*

5.5 Metabolic stress during the production of L-cysteine

In order to gain deeper insights into the cellular states during L-cysteine production, global transcripts of early and late generation population were examined and compared. Based on the principal component analysis results, it is possible to infer which components can be associated with each axis in a biological context. The x-axis, which exhibits 87% variance, is likely to correspond to the different strains and plasmids, while the y-axis, which showed a variance of 9%, is potentially reflecting the values and numbers of differentially expressed genes. Furthermore, the principal component analysis confirmed the assumption that W3110 populations underwent large metabolic alterations during the long-term cultivation, since transcript levels differed largely. At the same time, *E. coli* MDS42 populations demonstrated minimal differences, indicating its limited ability to adapt to environmental changes, as previously discussed in section 5.3.

Metabolic clustering of the differentially expressed genes revealed an overall resource-draining L-cysteine production process in many aspects. In a phase where L-cysteine production was highest, W3110 populations were confronted with limitations related to both the supply of acetyl-coA and a deficiency in nitrogen. This was reflected in the upregulation of clusters involved in the degradation of pyrimidine-, L-arginine-, and L-tryptophan as well as nitrate assimilation and acetyl-coA metabolic process. The lack of pyrimidines and amino acids due to degradation to obtain scarce nitrogen, can trigger a stringent response. Thereby, the enzyme RelA, a GDP/GTP pyrophosphokinase registers deacetylated tRNA in ribosomes as a result of unloaded tRNAs with amino acids. Consequently, the alarmone and nucleotide messenger guanosine tetra- and pentaphosphate ((p)ppGpp) is synthesised, stalling translation (173, 174). This alarmone regulated the transcription of approximately one-third of the cell’s total genes, redirecting resources from cell growth and division towards amino acid synthesis. Thus, the survival of cells should be ensured until the nutrient environment has improved. Although a comparison of *relA* transcripts in early and late W3110 populations did not exhibit significant differential expression, it is possible that the time of RNA extraction was already too late to detect increased *relA* expression. As a consequence, the previous section's claim of the early W3110 population's adaptation to L-cysteine production can be elaborated by proposing that the stringent response potentially impeded cellular growth.

During the phase of intense L-cysteine production, the deficiency of sulphur and L-cysteine resulted in severe metabolic stress. The cluster of S-/ cysteine starvation included 17-23 features with an average logFC of -5, rendering it the most prominent cluster.

Among the cluster, genes of the *tauABCD* operon were the most abundant. This operon is expressed during growth under sulphur starvation conditions and encodes an ABC-type uptake system for taurine (2-aminoethanesulphonate) and a dichlorphenoxyacetic acid dioxygenase which releases sulphite from taurine (137). Regulation is mediated by CysB, as in the case of the expression of the sulphate-binding protein Sbp but also by CbI, a CysB-like regulator. It was shown that CysB occupies several binding sites, whereas CbI occupied only one site upstream of the transcription start. CysB is a LysR-type (**Lys**ine **R**egulator) transcriptional activator and highly conserved among gram-negative bacteria. Unlike animals and plants, bacteria of different phyla are able to metabolise taurine either as carbon, sulphur, nitrogen or energy source (175). As the first step in taurine utilization, sulphoacetaldehyde is usually formed by oxidation or transamination. For instance, in *Pseudomonas aeruginosa* taurine is metabolised together with pyruvate as an energy source via transamination, similarly in *Achromobacter* spp. together with α- ketoglutarate (176, 177). *E. coli*, on the other hand, can only metabolise taurine as a source of sulphur under aerobic conditions.

Furthermore, the cluster contained genes of the *ssuEADCB* operon, which, in addition to *tauABCD*, plays a crucial role in the mobilisation of sulphur obtained from aliphatic sulphonates (138). In comparison to the *tauABCD* system, the *ssuEADCB* system exhibits a broader substrate spectrum beyond taurine. Expression of *ssu* genes is not directly dependent on the presence of CysB but rather on CbI. The proteins encoded by *ssuA*, *ssuB* and *ssuC* facilitate the transportation of sulphur through an ABC type transport system, while *ssuD* and *ssuE* code for a flavin mononucleotide ($FMNH_2$)-dependent monooxygenase and an NAPDH-dependent FMN reductase, respectively.

Additionally, genes that encode proteins involved in the assimilation of sulphate and thiosulphate were upregulated (*cysPUWA*, *sbp*, *cysDNC*, *cysJIH*). The reaction cascade of assimilated sulphate progresses through activation with ATP and GTP hydrolysis via Adenosine 5'-phosphosulphate (APS) to 3′-Phosphoadenosine-5′-phosphosulphate (PAPS). This is followed by a reductive part in which PAPS is degraded to sulphite via thioredoxin and then reduced to sulphide via sulphite reductases and NADPH. This sulphide is finally converted with OAS via CysK to L-cysteine (178).

A recent study detected a new thiosulphate uptake system in *E. coli*, which is regulated by the CysB regulon and undergoes post-transcriptional mRNA degradation mediated by the RNA 5´ Pyrophosphohydrolase RppH (179). The genes encoding TsuAB are organised in an operon and were identified as upregulated features in the sulphur-/ L-cysteine starvation cluster.

Generally, taurine, aliphatic sulphonates, sulphates, and thiosulphates traverse the outer membrane via porins and enter the periplasm. Subsequently, the binding protein specific to each sulphur-containing compound captures them and the corresponding transport protein complexes facilitate their transport into the cytoplasm. For a more comprehensive understanding, a schematic illustrating all the genes involved in sulphur assimilation that were up-regulated in early *E. coli* W3110 populations was created (*Fig. 5.5*).

Moreover, it is interesting to note that the predicted amino acid sequences of Sbp and TauA binding proteins lack cysteine residues and very few methionine residues. In addition, TauB, TauC and CysK posess a solitary cysteine residue. This sulphur-saving principle in proteins expressed specially during sulphate limitation is observed in other bacteria as well (180). Recently, a working group reconstructed L-cysteine biosynthesis by using engineered cysteine-free enzymes. Specifically, they substituted alternate amino acids in CysE and CysM for cysteine and methionine, with both enzymes retaining their activity (181). This approach could advance the optimisation and fine-tuning of L-cysteine production, reducing its vulnerability to potential sulphur deficiencies.

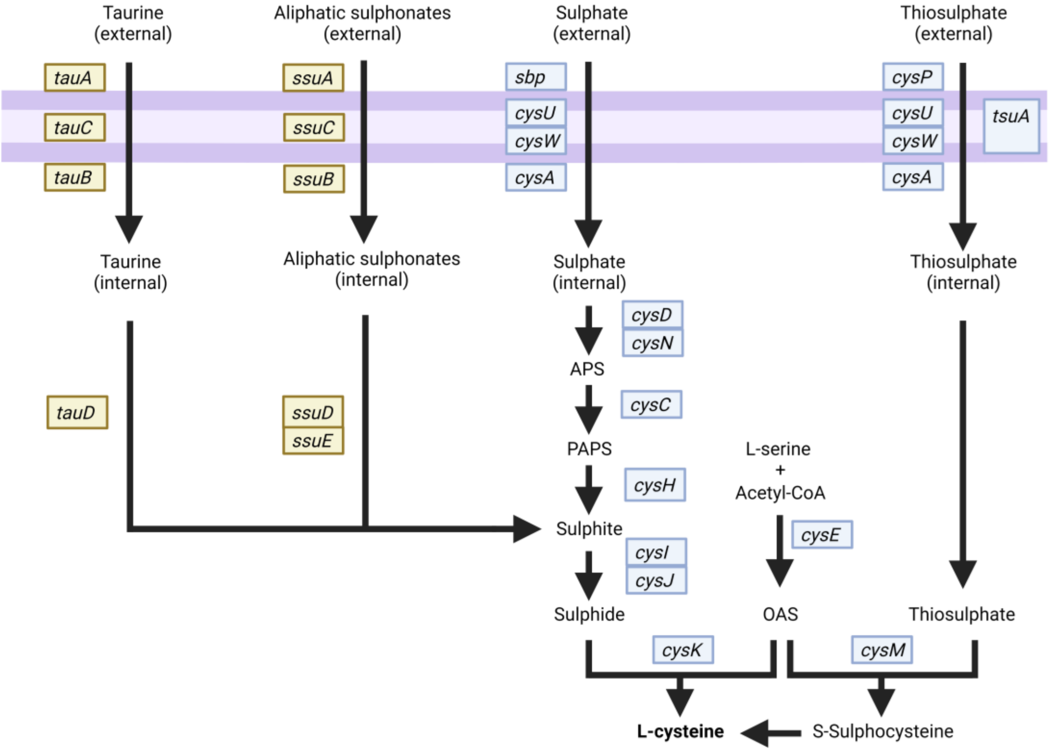


Figure 5.5: Sulphate/ thiosulphate and sulphonate-sulphur assimilatory pathway systems in Escherichia coli: *Genes belonging to the sulphate/ thiosulphate metabolic system and regulated by the transcriptional regulator CysB are highlighted in light blue boxes, while genes related to the sulphonate-sulphur utilization and requiring Cbl (as a direct activator) as well as CysB (as an activator of the cbi gene) are displayed in yellow boxes. APS: Adenosine 5'-phosphosulphate, PAPS: 3'-Phosphoadenosine-5'-phosphosulphate. Created with Biorender.*

In summary, when sulphate was deficient, *E. coli* increased the synthesis of proteins involved either in the biosynthesis of cysteine from sulphate or in the utilisation of alternative sulphur sources such as taurine or aliphatic sulphonates. These findings indicate a higher uptake and metabolism of thiosulphate during periods of increased L-cysteine production. The flux of sulphur towards L-cysteine and its subsequent export was so significant that the amount of thiosulphate supplied was inadequate to fulfil the cellular metabolic demands. The upregulation of the gene *yciW*, resulting in increased tolerance to external L-cysteine, was another indicator showing that extracellular levels of L-cysteine were very high. The corresponding protein is speculated to be involved in the degradation of intracellular L-cysteine (182). Additionally, as a result of the sulphur deficiency, O-acetylserine most likely accumulated in the cytoplasm of cells, triggering the dissociation of the cysteine synthesis complex.

If the concentration of L-cysteine within *E. coli* cells is insufficient, the direct uptake of L-cysteine and/or L-cystine from the environment would be advantageous. However, *E. coli* lacks a specific transporter for the import of L-cysteine and relies on ATP-driven transporters (TcyJLNP) to assimilate the disulphide form L-cystine (183). The surprising absence of a L-cysteine importer is most likely due to the assimilation of hydrogen sulphide in anoxic and reducing environments, such as the natural habitat of *E. coli*, the colon. *E. coli* can readily obtain sulphur from hydrogen sulphide, which can passively diffuse across the membrane and can be captured by O-acetylserine sulphhydrylase CysK to build L-cysteine. Unlike other sulphur sources, the assimilation of sulphides does not require any reducing equivalents.

Under oxic environmental conditions, as they prevail during fermentations, the import of L-cystine would provide a significant energetic benefit for the production of L-cysteine. While the assimilation of sulphate and subsequent formation of L-cysteine requires 32 ATP equivalents, the import of L-cystine and its reduction requires only about 2 ATP molecules. Despite the potential benefits of importing L-cystine, it poses a considerable risk. In *E. coli*, L-cystine is imported at levels that exceed the actual sulphur requirement by up to 50 times (184). Once inside the cell, it is rapidly converted to L-cysteine with the aid of glutathione, in order to mitigate the damaging impact of disulphide stress on proteins. While this solves one problem, an oversupply of L-cysteine creates new challenges for cellular metabolism, leading to the issues mentioned above.

Indeed, metabolic clustering of DEGs in late W3110 populations showed evidence of L-cysteine intoxication. Expression of the *cyuPA* operon, which is regulated by the DNA-binding transcriptional regulator DecR, was approximately five times higher in later populations than in early stage populations. The *cyuP* gene within the operon encodes a transport protein for D/L-serine from the hydroxyl/ amino acid permease (HAAAP) family (185). The *cyuA* gene, on the other hand, encodes a putative L-cysteine desulphidase that catabolises L-cysteine under anaerobic conditions (135). The strong flux towards L-cysteine and its export presumably led to a strong increase in intracellular L-cysteine and a deficiency in free sulphur, prompting the cells to counteract further L-cysteine accumulation by exporting the precursor amino acid L-serine out of the cell before it could be captured by CysE. Furthermore, the degradation of L-cysteine through CyuA would have provided the cell with urgently required sulphur. The study shed light on an additional detrimental effect of high intracellular L-cysteine levels, which was previously mentioned in section 5.1 reflecting an average 4-fold up-regulation of 16 genes that were indicative of iron starvation. It is highly plausible, that L-cysteine bound to ferric iron, leading to its reduction to ferrous iron, which acted as an electron donor for the Fenton reaction and triggered an iron starvation signal. This, in turn, caused ferrous iron to convert hydrogen peroxide to hydroxyl radicals, which directly resulted in DNA damage. It is noteworthy, that the levels of L-cysteine were very low at the time of upregulated L-cysteine detoxification and degradation genes. This may be due to a time discrepancy between the snapshot of intracellular transcripts and their lifespan. While L-cysteine may have already been degraded, mRNA levels could have still been high.

Another potential scenario, which will be discussed later and only briefly mentioned here, is a malfunctioning exporter in late generation populations, which would have resulted in L-cysteine accumulation within cells. With the degradation of intracellular L-cysteine by desulphidases, the cell could have gained access to sulphur.

Finally, high intracellular levels of L-cysteine and L-cystine can both be harmful to the cell. These findings highlight the multifaceted role of L-cysteine in cellular processes and its impact on cellular homeostasis. This cellular stress emanating from sulphur and L-cysteine manifested itself in the emergence of genetic adaptation through activation of mobile genetic elements, such as insertion sequence elements and their transposases. Within the cluster of sulphur starvation stress, *cysD* was detected, which belongs to a network of genes which facilitate stress-induced mutagenesis (SIM) in *E. coli* (186). Thus, upregulated genes encoding IS3 and IS66 family transposases were identified in late W3110 populations.

These observations suggested, that *E. coli* may have overcome L-cysteine production stress genetically by disabling the synthetic plasmid constructs. This hypothesis was supported by the transformation experiment, which revealed lower L-cysteine yields in fresh strains carrying plasmids from later populations. The extent to which potential plasmid mutations occurred and how they affect the overall concept of microbial production is discussed in the following section.

5.6 Genetic heterogeneity in L-cysteine producing populations

Apart from the potential adverse impact of defective plasmids on L-cysteine production, the possibility of plasmid loss in evolving populations was initially ruled out. A complete loss of plasmids was suspected as rather unlikely, as *E. coli* would have to acquire a future escape route beforehand by incorporating antibiotic resistance through genome integration in order to be able to withstand the constant selection pressure from tetracycline.

From an evolutionary perspective, it is more likely that individual plasmid genes, which impose an enormous burden on production cells, are disrupted by mutations to provide a fitness advantage. To study multiple plasmids within populations, a sequencing approach with high resolution was employed, capable of detecting low-frequency mutations.

Initially, it was unexpected that no single nucleotide polymorphism (SNP) mutations were observed in production genes, which would represent the simplest form of inactivation of burdensome genes. However, upon review of published literature, it became evident that SNP-induced mutations are a rather rare occurrence with a basal spontaneous mutation rate of $5.4x10^{- 10}$/bp/generation (89, 93, 187, 188). This means that, on average, 1-2 x 10^{-3} mutations occur per generation per genome with a very low chance of affecting genes actually required for L-cysteine production (189). In fact, SNPs were exclusively detected in the plasmid backbone, without any discernible impact on the L-cysteine production or cellular fitness. Mutations with a 99% allele frequency were regarded as pre-existing.

The identification of insertion sequences within plasmids by alignment against an IS database with the Burrows-Wheeler-Aligner (BWA) was inspired by examination of mapping statistics, which indicated an accumulation of unmapped reads in plasmids from late generation populations of *E. coli* W3110. Initially, the discovery of IS on plasmids from MDS42 populations was surprising, given the assumption, that all IS had been removed from the genome and that transposition was no longer possible. Yet, after thorough investigation, it was determined that the sequences had pre-existed on the plasmid. The quantity of reads attributed to IS remained constant, and the identified IS were classified as "silent", reflecting inactive transposases. IS200, for example, achieves tight repression of transposase synthesis through a combination of mechanisms, including inefficient transcription, repression of translation via a stem-loop mRNA structure, and protection from impeding transcription by a terminator (190). IS200 and IS110 elements are therefore classified as ancestral elements of bacterial genomes which remain constant in natural populations. As a result, reads obtained from plasmids derived from MDS42 populations can be considered as a baseline or negative control, representing sequences that were already present on the plasmids. Thus, any additional reads that emerge are likely the outcome of transposition of genomic IS elements into plasmids.

Indeed, plasmids originating from the W3110 populations exhibited a significantly higher abundance of insertion sequences compared to those from the MDS42 populations. Furthermore, a diverse array of IS families were detected in the reads of the latter. The homogenous distribution of various IS families within plasmids obtained from both early and late populations, suggests that the transposition propagated in similar frequencies. Since, the complete sequence is always replicated during plasmid segregation, it was not surprising. Once transpositions have taken place, they are expected to become established over time, as long as transpositions confer a selection advantage.

The predominance of IS3 and IS5 reads is consistent with the high abundance of these two IS families within the genome of *E. coli* K-12 MG1655, with six and ten copies, respectively. However, several family subgroups with similar sequence identities and transposition mechanisms have been identified to be present in more than 267 bacterial species distributed over 145 genera. Thereby IS3 is one of the most coherent, widely distributed IS families (191). This strong abundance of IS3 in plasmid reads is corroborated by the increased *insJK* expression, which is a transposase belonging to the IS3 family, observed in late generation populations of W3110. IS3 family members exhibit a size ranging from 1200 to 1550 bp and display well conserved inverted terminal repeats of 16-40 base pairs. Upon integration, they create target site duplications of about 3-5 base pairs. In addition, IS3 generally has two consecutive partially overlapping, shifted reading frames, *orfA* and *orfB*, in the -1 and 0 reading phase, respectively (*Fig. 5.6*).

The OrfA upstream element and the actual transposase OrfAB are both synthesised, with the latter being a fusion protein that undergoes activation through a programmed translational frameshifting mechanism (192). The OrfA protein possesses a distinct helix-turn-helix (HTH) motif that is thought to facilitate binding to the terminal inverted repeats of the OrfAB transposases in a sequence-specific manner (193). Nevertheless, except for the IS3 family members IS911 and IS150, no insertion specificity could be detected. This is supported by the failed attempt to localise discernible sequence patterns within IS3 entry sites in this study. Instead, the unpredictable insertions render it impossible to anticipate and avoid potential susceptible sites during a plasmid design process.

At the C-terminus, downstream of the HTH motif, a conserved leucine zipper (LZ) motif can be found which is associated with protein multimerisation (194). The OrfAB fusion protein possesses an additional DDE motif, which is also found in retroviral integrases and which catalysis the actual transposase reaction (195, 196). The transposition process involves a copy-out-paste-in mechanism, where the original site is retained and a double-strand circular DNA intermediate is utilised (197). Numerous other members of the IS3 family are organised the same way (198, 199). As such, this study revealed an upregulation of both elements of the IS150 transposase (*insJK*). The rate of frameshifting varies among different elements with approximately 50% for IS150 (200).

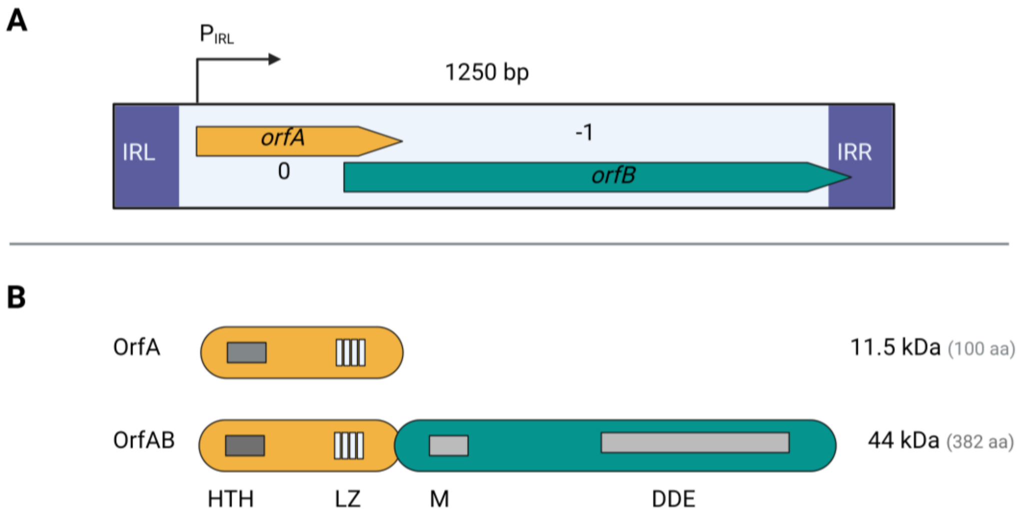


Figure 5.6: Organisation of IS3 transposases: A: *Overall structure of the 1250 bp IS3 transposase complex. Highlighted in purple are the left and right terminal repeats (IRL, IRR). The operon is organised in two open reading frames, orfA (yellow) and orfB (green) with reading frames 0 and -1 and is under control of the P$_{IRL}$ promoter. A frameshifting signal is necessary to create OrfAB and is controlled in the overlapping region of orfA and orfB.* ***B:*** *Function map of OrfA and OrfAB. DNA binding is facilitated by a helix-turn-helix motif (HTH) and a leucine zipper motif (LZ). OrfAB is generated by a programmed translational frameshifting, where a multimerisation reaction (M) fuses both components. The DDE motif is important for the catalytic copy-out mechanism. Figure adapted from Chandler et al.,(191) and with BioRender.*

Unlike IS3, the IS5 family, consisting of approximately 550 members, displays significantly greater diversity in both sequence motifs- and lengths (201). The majority of members are organised in a solitary open reading frame that encodes for the transposase. About 20% exhibit programmed transcriptional realignment frameshifting instead of translational frameshifting, as observed in IS3 members (198). Neither the transposition mechanism nor specific target site sequences are known yet. In this study, insertions were identified – similar to IS3 – rather random at different sites in plasmids. Yet, the pCYS_m plasmid exhibited a high frequency of target site duplications with the sequence motif “*ATAAAGCG*”, which was evidenced by the occurrence of over 100 reads. Despite a meticulous comparison between this motif and the IS5 sequence, no similarities were discerned, even at the crucial terminal inverted repeats, which are typically accountable for target site recognition in other IS families. Perhaps certain DNA structure features outside the potential target site are involved in recognition, which should be addressed in future studies. Upon checking the target site duplications of IS5 identified in previous studies, it became apparent, that they consistently were tetranucleotides, in contrast to the eight base pair repeats identified in this work. The insertion of IS5 can be traced back to two sites within the pCYS_m plasmid through the duplicated target sequence (TSD). Specifically, within the open reading frame of the L-cysteine exporter EamA, as well as within the ORF of Cysteine Synthase B (*cysM*). In the case of an impaired exporter, populations harbouring the pCYS_m plasmid would have accumulated L-cysteine within cells. Cells, which previously had to cope with sulphur deficiency, could have restored sulphur balance by catabolising larger quantities of L-cysteine. This would explain the elevated expression levels of desulphidases and the increased growth rates. In the other case, a defective L-cysteine synthase would have relieved the cells equally in the sulphur balance by directing the sulphur flow less towards L-cysteine and more towards other metabolic processes.

Overall, target site duplications and insertion sequences were detected in both the plasmid backbone and the open reading frames of production genes, along with their promoters, in W3110. Starvation stress could have triggered an activation of insertion sequences that may have led to their propagation and deactivation of L-cysteine production genes. One of the most well-documented instances of IS activation due to environmental factors is observed in the *glpFK*/*Crp* system in *E. coli*. This system demonstrates, that the integration of an IS5 transposase into the *glpFK* promoter region can activate the utilisation of glycerol in conditions of starvation (202, 203). In another case, the operon responsible for ß-glucoside utilisation in *E. coli*, *bglGFB*, which is typically dormant in wild-type cells, is stimulated by the insertion of an IS upstream of the operon (204). Humayun et al., speculate a theoretical framework by which insertion events are facilitated at particular sites and how stress conditions are correlated with increased insertion at a specific locus. Specifically they show that a specific DNA structure known as superhelical stress-induced duplex destabilisation (SIDD) is connected with IS5 insertions in the previously mentioned cases of *glpFK* and *bglGFB* (205). SIDD was employed as a bioinformatic model that assesses the likelihood of denaturation of a specific DNA sequence and generates an energy profile (206). Base pairs exhibiting reduced energy levels are deemed less stable. Stress conditions negatively influence the linking number, a mathematical parameter, that characterises the magnitude of DNA twisting (207). In this way, regions with low twisting properties, the duplex destabilisation sites, serve as potential hotspots for IS insertions. Perhaps these destabilised DNA structures outside the TSD also had an influence on the insertion events discovered in this study.

Nevertheless, the small percentage of reads that could be aligned to insertion sequences (0.1-0.3%), coupled with the identification of only a few TSDs, raises the question, to what extent defective plasmids may have contributed to the decline in L-cysteine yields, as well as the increase in growth rates. Owing to the limited length of Illumina sequencing reads (150 bp) and the stringent filter criteria employed during the alignment of these reads with the IS database, a considerable portion of reads may have gone undetected. Therefore, it is plausible, that only a fraction of the actual IS sequences present within the plasmids were identified. By utilising long read sequencer like PacBio or Oxford Nanopore, a more comprehensive understanding of the plasmid sequences could have been attained. This would have allowed both for the precise identification and localisation of junction reads of IS and plasmid sequences, instead of relying on the TSD footprint approach. However, long read sequencer typically have lower coverages, making it near impossible to cover all plasmids within an entire population.

6. Conclusion & Outlook

The present study investigated the phenotypic and genetic alterations of recombinant *Escherichia coli* populations during L-cysteine production. The metabolic stress encountered during L-cysteine production, particularly the perturbation of tightly regulated sulphur balance within *E. coli* cells, brought to light the challenges that need to be addressed during a fermentation process. Insertion sequence transpositions contributed to the genetic instability of *E. coli* strains, leading to the disruption of plasmid genes critical for L-cysteine overproduction. Within 60 generations, as achieved in large-scale fermentations, *E. coli* sacrificed up to 80% production capacities in order to gain growth fitness.

Using time resolved plasmid deep sequencing and a transcriptomics approach, insertion sequence families 3 and 5 were identified as a root cause for the stress induced disruption of plasmid constructs, tailored for L-cysteine production. Moreover, it was demonstrated how the minimal genome *E. coli* strain MDS42, with the majority of IS and 699 genes being deleted, is superior in process stability compared to the established W3110 production host. However, while MDS42 showed higher growth rates and stability compared to W3110, its overall L-cysteine space-time yield was lower.

Therefore, a more targeted approach would be the selective deletion of insertion sequence families and their corresponding transposases, such as IS3 and 5, in genomes of industrially relevant production strains, like W3110. Our study postulates, that this new approach could not only optimise the *E. coli* based industrial L-cysteine production, but may be generally applicable to other recombinant amino acid production systems. Furthermore, a trade-off between increased plasmid stability over the entire production phase while maintaining evolutionary adaptability of the cell system linked to high cell density fermentations at maximal growth rates would be provided.

This approach is supported by the fact, that biotechnology industries commonly store production strains as frozen aliquots without extensive prior sequence screening. However, even though the population may seem stable for a certain number of cell divisions, there is a possibility, that a few cells in the starting seed may contain genetic mutations. These mutations can negatively affect the production process and lead to reduced yields or even complete failure of the production run, particularly in continuous fed-batch fermentations. Therefore, it is crucial to screen the cell bank aliquots before starting the production process, especially for rare pre-existing mutations, that may disrupt the production process. To ensure a comprehensive analysis, deep-sequencing techniques should be employed for accurate identification and characterisation of any genetic mutations present in the cell bank aliquots.

7. Appendix

Supplementary table 1: *Average number of generations after each passage of the simulated fermentation of W3110 and MDS42 with integrated pCYS, pCYS_i and pCYS_m. Each strain was cultivated in biological triplicates.*

W3110						
	pCYS		**pCYS_i**		**pCYS_m**	
Time [h]	Accumulated generations	STD	Accumulated generations	STD	Accumulated generations	STD
Preculture (0)	2.67	0.08	4.50	0.14	3.83	0.07
10	9.52	0.14	9.53	0.05	9.57	0.003
20	14.37	0.21	14.84	0.13	14.61	0.05
30	20.16	0.21	20.13	0.24	19.65	0.05
40	26.13	0.25	25.33	0.32	24.63	0.03
50	32.23	0.31	30.66	0.43	29.68	0.05
60	38.53	0.33	35.79	0.47	34.65	0.02
70	45.12	0.35	41.11	0.58	39.77	0.04
80	51.71	0.37	46.42	0.70	45.13	0.21
90	58.25	0.43	51.82	0.77	51.07	0.62
100	64.69	0.42	57.40	0.90	57.48	0.65
110	-	-	63.52	0.97	64.10	0.62
MDS42						
Preculture (0)	2.70	0.30	5.21	0.08	5.64	0.14
5	6.85	0.42	8.75	0	9.26	0.01
10	11.05	0.42	12.88	0.04	13.29	0.15
15	14.93	0.86	17.09	0.07	16.84	0.41
20	18.74	1.22	20.75	0.09	21.19	0.32
25	22.60	1.62	25.04	0.12	25.24	0.39
30	26.40	1.98	29.26	0.11	29.41	0.39
35	30.35	2.33	33.37	0.10	33.54	0.37
40	34.32	2.64	37.48	0.11	37.27	0.34
45	38.29	2.93	41.73	0.08	41.29	0.42
50	42.66	3.15	46.14	0.14	45.45	0.41
55	47.05	3.28	50.34	0.16	49.59	0.43
60	51.39	3.36	54.35	0.26	53.42	0.40
65	55.63	3.38	58.56	0.20	57.83	0.31
70	60.63	3.37	62.81	0.19	62.31	0.21

Supplementary Table 2. *Average growth rates of samples taken after each passage of the simulated long-term cultivation of W3110 and MDS42 with integrated pCYS, pCYS_i and pCYS_m. Each strain was cultivated in biological triplicates. Growth rates were calculated according to the formula found in the manuscripts' methods section of "Measurement of population growth rates".*

W3110								
	pCYS		**pCYS_i**		**pCYS_m**		**Empty vector**	
Generation [#]	Growth rate µ	STD	Growth rate µ	STD	Growth rate µ	STD	Growth rate µ	STD
10	0.42	0.006	0.51	0.009	0.48	0.004	0.79	0.023
15	-	-	0.51	0.014	0.48	0.004	-	-
20	0.45	0.003	-	-	-	-	0.77	0.045
25	-	-	-	-	0.49	0.010	-	-
30	-	-	0.52	0.007	-	-	0.79	0.060
32	0.45	0.012	-	-	-	-	-	-
40	-	-	-	-	0.49	0.010	0.81	0.041
45	0.54	0.012	-	-	-	-	-	-
51	-	-	0.5	0.013	0.54	0.028	0.80	0.063
52	0.54	0.025	-	-	-	-	-	-
59	0.70	0.067	0.49	0.033	-	-	-	-

63	0.63	0.007	0.54	0.029	0.55	0.006	0.72	0.030
				MDS42				
7	0.60	0.030	-	-	-	-	-	-
9	-	-	0.78	0.018	0.79	0.023	0.68	0.045
14	0.58	0.010	-	-	-	-	-	-
17	-	-	0.80	0.013	-	-	0.67	0.030
25	-	-	0.81	0.008	0.79	0.017	-	-
26	0.58	0.020	-	-	-	-	0.67	0.051
32	-	-	-	-	0.80	0.015	-	-
45	-	-	0.81	0.088	0.79	0.017	0.69	0.042
47	0.58	0.008	-	-	-	-	-	-
60	0.57	0.041	-	-	-	-	0.66	0.063
62	-	-	0.81	0.004	0.79	0.015	-	-

Supplementary table 3: Mapping statistics overview. *For each sample, the following statistics are provided: Reads mapped: the total number of reads mapped to the reference genome. Unique: number of uniquely mapped reads, i.e. read can only be mapped to one reference locus. Reference covered: reference bases covered by at least one read. Mean read coverage: average read coverage of the reference sequence. HGP: high generation population, LGP: low generation population.*

Sample	Reads mapped [Mio]	Unique [Mio]	Reference covered [Mb]
W3110_pCYS_HGP	15.74 (98.6%)	15.43 (96.6%)	4.35 (93.7%)
W3110_pCYS_LGP	17.48 (97.5%)	17.16 (95.7%)	4.26 (91.8%)
MDS42_pCYS_HGP	12.52 (98.2%)	12.35 (96.9%)	3.54 (76.3%)
MDS42_pCYS_LGP	21.15 (98.1%)	20.73 (96.1%)	2.63 (78.2%)
W3110_pCYS_i_HGP	63.00 (98.0%)	62.31 (96.9%)	4.61 (99.2%)
W3110_pCYS_i_LGP	38.42 (97.1%)	37.97 (96.0%)	4.56 (98.3%)
MDS42_pCYS_i_HGP	52.81 (98.3%)	52.30 (97.3%)	3.93 (84.6%)
MDS42_pCYS_i_LGP	32.11 (98.5%)	31.88 (97.8%)	3.89 (83.7%)
W3110_pCYS_m_HGP	13.74 (98.5%)	13.62 (97.7%)	4.59 (99.1%)
W3110_pCYS_m_LGP	10.93 (97.3%)	10.84 (96.5%)	4.35 (97.4%)
MDS42_pCYS_m_HGP	12.75 (98.4%)	12.64 (97.6%)	3.89 (84%)
MDS42_pCYS_m_LGP	13.14 (97.9%)	13.05 (97.3%)	3.76 (83.3%)

Supplementary table 4: Expression profiling statistics. *For each sample, the following statistics are provided: Effective library size: The total number of reads mapped to reference features. Normalised library size: The total number of reads mapped to reference features normalised by the associated normalization factor, which can be derived by dividing the normalised library size by the effective library size. No feature: The number of reads mapping to the reference sequence that could not be assigned to any annotated feature, i.e. mapping positions and feature positions do not overlap. Filtered: The number of reads that were filtered due to insufficient mapping quality or ambiguous mapping location. These reads were ignored for read counting.*

Sample	Effective Library Size	Normalised Library Size	No Feature	Filtered
W3110_pCYS_HGP	14,252,994	14,476,778.3	1,165,261	538,202
W3110_pCYS_LGP	15,767,859	17,145,605.9	1,379,155	776,008
MDS42_pCYS_HGP	10,911,912	11,159,581.3	1,433,356	390,282
MDS42_pCYS_LGP	18,559,438	17,419,716.3	2,160,469	832,562
W3110_pCYS_i_HGP	28,130,536	27,532,366.2	2,882,373	1,119,315
W3110_pCYS_i_LGP	16,502,506	18,445,755.2	2,402,562	869,322
MDS42_pCYS_i_HGP	23,361,202	21,780,445.7	2,574,658	915,004
MDS42_pCYS_i_LGP	14,731,777	13,971,444.8	981,102	571,985
W3110_pCYS_m_HGP	18,597,401	18,489,582	1,065,231	481,982
W3110_pCYS_m_LGP	18,351,569	18,154,799	995,471	571,451
MDS42_pCYS_m_HGP	17,581,965	17,281,118	1,921,667	581,949

MDS42_pCYS_m_LGP	16,411,682	15,984,381	1,763,955	739,814

Supplementary table 5: Table showing accession numbers, gene names, features, logarithmic fold changes (logFC) and p-values of all differentially expressed genes (DEGs) of W3310_pCYS with p-values <0.05. *LogFC and logCPM were calculated by dividing values of the later generation population (LGP) by values of the early generation population (EGP). *: Genes were excluded because they did not fall within the FC range of the metabolic cluster.*

Accession	Gene	Feature	LogFC	P-Value
b4205	*ytfA*	biofilm formation	6.31	0.0004
b3110	*cyuP*	l-cysteine degradation	5.12	0.001
b1505	*ydeT*	biofilm formation	5.07	0.001
b1556	*essQ*	unknown function	5.12	0.002
b1504	*YdeS*	biofilm formation	5.02	0.003
b4293	*FecI*	iron starvation	4.28	0.003
b1503	*ydeR*	biofilm formation	4.95	0.003
b4002	*ZraP*	biofilm formation	4.25	0.004
b1560	*YdfU*	unknown function	4.23	0.004
b4470	*CyuA*	l-cysteine degradation	4.11	0.004
b4277	*yjgZ*	insertion sequence element	5.05	0.004
b1038	*csgF*	curli secretion	-4.15	0.004
b3558	*insK*	insertion sequence element	4.05	0.005
b0585	*fes*	iron starvation	4.04	0.005
b4292	*fecR*	iron starvation	4.00	0.005
b0584	*fepA*	iron starvation	3.98	0.005
b0590	*fepD*	iron starvation	3.99	0.005
b4567	*yjjZ*	unknown function	4.20	0.005
b0691	*ybfG*	biofilm formation	7.66	0.006
b1309	*ycjM*	alternative sugar utilizing pathway	4.36	0.007
b4011	*yjaA*	biofilm formation	3.87	0.007
b1502	*ydeQ*	biofilm formation	4.05	0.007
b2229	*yfaT*	unknown function	4.40	0.007
b0587	*fepE*	iron starvation	7.53	0.007
b0236	*prfH*	unknown function	4.30	0.008
b2298	*yfcC*	unknown function	3.65	0.010
b0589	*fepG*	iron starvation	3.61	0.010
b1554	*rrrQ*	unknown function	3.73	0.011
b0986	*gfC*	unknown function	3.72	0.011
b1039	*csgE*	curli secretion	-3.53	0.012
b1600	*mdtJ*	multidrug efflux	3.55	0.013
b0583	*entD*	iron starvation	3.46	0.013
b2221	*atoD*	acetyl-coA metabolic process	-3.56	0.014
b3557	*insJ*	Insertion sequence element	3.34	0.016
b1553	*rzpQ*	unknown function	3.55	0.017
b1496	*yddA*	iron starvation	3.81	0.017
b1037	*csgG*	curli secretion	-3.25	0.018
b1599	*mdtI*	multidrug efflux	3.32	0.019
b1466	*narW*	nitrate assimilation	-3.27	0.019
b3800	*aslB*	L-cysteine degradation	3.20	0.020
b0985	*gfcC*	biofilm formation	3.29	0.020
b2155	*cirA*	iron starvation	3.18	0.020
b0591	*entS*	iron starvation	3.15	0.022
b3158	*ubiU*	ubiquinone biosynthesis	3.15	0.022
b0894	*dmsA*	anaerobic respiration	3.07	0.024
b0549	*ybcO*	unknown function	3.67	0.024
b1492	*gadC*	acid resistance	-3.05	0.025
b3708	*tnaA*	tryptophanases	-3.06	0.025
b3060	*ttdR*	transcriptional regulator	3.15	0.025
b0575	*cusA*	copper/silver export	3.04	0.025
b2343	*yfcZ*	unknown function	3.02	0.026
b1493	*gadB*	acid resistance	-3.00	0.028
b1040	*csgD*	curli secretion	-2.98	0.028
b4183	*yjkK*	unknown function	3.21	0.029

b2339	*yfcV*	biofilm formation	3.48	0.029
b2997	*hybO*	anaerobic respiration	2.94	0.030
b2534	*yfhR*	unknown function	3.12	0.032
b0547	*ybcN*	mismatch repair	3.17	0.032
b0593	*entC*	iron starvation	2.90	0.032
b0984	*gfcD*	unknown function	2.92	0.033
b3408	*feoA*	iron starvation	2.89	0.033
b4511	*ybdZ*	iron starvation	3.07	0.033
b2222	*atoA*	acetyl-coA metabolic process	-3.00	0.034
b1409	*ynbB*	biofilm formation	3.10	0.035
b1674	*ydhY*	iron starvation	3.02	0.036
b1468	*narZ*	nitrate assimilation	-2.81	0.037
b3654	*xanP*	xanthine transport	2.81	0.037
b0913	*ycaI*	unknown function	2.80	0.038
b0836	*bssR*	biofilm formation	2.79	0.038
b1469	*narU*	nitrate assimilation	-2.79	0.039
b2172	*yeiQ*	unknown function	2.78	0.039
b3020	*ygiS*	unknown function	2.77	0.039
b1121	*ycfZ*	unknown function	4.17	0.040
b0550	*rusA*	mismatch repair	2.98	0.042
b1161	*ycgX*	unknown function	3.42	0.043
b4367	*fhuF*	iron starvation	2.72	0.043
b0375	*iprA*	mismatch repair	4.08	0.046
b0574	*cusB*	copper/silver export	2.68	0.046
b1541	*ydfZ*	unknown function	2.71	0.046
b1467	*narY*	nitrate reductase	-2.67	0.047
b4314	*fimA**	biofilm formation	-2.63	0.049

Supplementary table 6: Table showing accession numbers, gene names, features, logarithmic fold changes (logFC) and p-values of all differentially expressed genes (DEGs) of MDS42_pCYS with p-values <0.05. *LogFC and logCPM were calculated by dividing values of the later generation population (LGP) by values of the early generation population (EGP).*

Accession	**Gene**	**Feature**	**LogFC**	**P-Value**
b4035	*malK*	maltose transport	-4.54	0.0005
b4036	*lamB*	maltose transport	-4.03	0.0013
b1224	*narG*	nitrate assimilation	3.86	0.0019
b1225	*narH*	nitrate assimilation	3.78	0.0022
b1223	*narK*	nitrate assimilation	3.74	0.0025
b3060	*ttdR*	transcriptional regulator	3.56	0.0043
b1226	*narJ*	nitrate assimilation	3.40	0.0052
b4034	*malE*	maltose transport	-3.34	0.0064
b1227	*narI*	nitrate assimilation	3.30	0.0065
b4037	*malM*	maltose transport	-3.00	0.0127
b4242	*mgtA*	Mg2+ transport	-2.96	0.0131
b4032	*malG*	maltose transport	-3.01	0.0133
b1750	*zdjX*	unknown function	3.00	0.0145
b1436	*yncJ*	unknown function	-3.87	0.0193
b0894	*dmsA*	anaerobic respiration	2.65	0.0243
b4702	*mgtL*	Mg2+ transport	-3.13	0.0254
b4033	*malF*	maltose transport	-2.70	0.0259
b3366	*nirD*	nitrate assimilation	2.59	0.0274
b2111	*yehD*	biofilm formation	3.12	0.0279
b2298	*yfcC*	unknown function	2.56	0.0303
b2243	*glpC*	anaerobic respiration	2.53	0.0307
b3367	*nirC*	nitrate assimilation	2.50	0.0329
b0836	*bssR*	biofilm formation	2.44	0.0368
b3158	*ubiU*	ubiquinone biosynthesis	2.45	0.0375
b3960	*argH*	L-arginine biosynthesis	2.43	0.0375
b1826	*mgrB*	Mg2+ transport	-2.51	0.0395
b1748	*astC*	L-arginine degradation	-2.37	0.0426

b3365	*nirB*	nitrate assimilation	2.36	0.0429
b2242	*glpB*	anaerobic respiration	2.36	0.0430
b1182	*hlyE*	hemolysin	2.67	0.0435
b0895	*dmsB*	anaerobic respiration	2.35	0.0438
b3508	*yhiD*	unknown function	2.52	0.0448
b1608	*rstA*	unknown function	-2.34	0.0451
b1751	*ydjy*	unknown function	2.30	0.0498

Supplementary table 7: Table showing accession numbers, gene names, features, logarithmic fold changes (logFC) and p-values of all differentially expressed genes (DEGs) of W3110_pCYS_i with p-values <0.05. *LogFC and logCPM were calculated by dividing values of the later generation population (LGP) by values of the early generation population (EGP).*

Accession	**Gene**	**Feature**	**LogFC**	**P-Value**
b0365	*tauA*	sulphur/L-cysteine starvation	-10.82	4.37E-07
b0366	*tauB*	sulphur/L-cysteine starvation	-10.57	6.44E-07
b0367	*tauC*	sulphur/L-cysteine starvation	-9.97	1.60E-06
b0368	*tauD*	sulphur/L-cysteine starvation	-8.97	7.24E-06
b0937	*ssuE*	sulphur/L-cysteine starvation	-8.61	2.22E-05
b3917	*sbP*	sulphur/L-cysteine starvation	-7.98	3.27E-05
b0935	*ssUD*	sulphur/L-cysteine starvation	-7.91	3.94E-05
b0936	*ssuA*	sulphur/L-cysteine starvation	-7.36	9.94E-05
b0934	*ssuC*	sulphur/L-cysteine starvation	-6.63	2.81E-04
b2752	*cysD*	sulphur/L-cysteine starvation	-6.56	2.92E-04
b2750	*cysC*	sulphur/L-cysteine starvation	-6.54	3.01E-04
b2751	*cysN*	sulphur/L-cysteine starvation	-6.24	4.68E-04
b2422	*cysA*	sulphur/L-cysteine starvation	-5.51	1.43E-03
b0933	*ssUB*	sulphur/L-cysteine starvation	-5.37	1.81E-03
b2763	*cysI*	sulphur/L-cysteine starvation	-5.16	2.40E-03
b2424	*cysU*	sulphur/L-cysteine starvation	-5.16	2.47E-03
b2423	*cysW*	sulphur/L-cysteine starvation	-5.07	2.80E-03
b4721	*ytiD*	unknown function	-5.99	3.16E-03
b2762	*cysH*	sulphur/L-cysteine starvation	-4.95	3.29E-03
b2764	*cysJ*	sulphur/L-cysteine starvation	-4.70	4.77E-03
b1492	*gadC*	acid resistance	-4.42	7.18E-03
b1493	*gadB*	acid resistance	-4.15	1.07E-02
b4518	*ymdF*	unknown function	-4.09	1.17E-02
b3517	*gadA*	acid resistance	-4.01	1.29E-02
b2425	*cysP*	sulphur/L-cysteine starvation	-3.98	1.38E-02
b1038	*csgF*	curli secretion	-3.98	1.38E-02
b1039	*csgE*	curli secretion	-3.87	1.61E-02
b1467	*narY*	nitrate assimilation	-3.77	1.84E-02
b0897	*ysaC*	unknown function	-3.73	1.94E-02
b3477	*nikB*	nickel transport	3.75	1.97E-02
b1489	*dosP*	oxygen-sensing	-3.72	1.98E-02
b3491	*yhiM*	unknown function	-3.72	1.99E-02
b2379	*alaC*	L-alanine biosynthesis	-3.71	2.00E-02
b3555	*yiaG*	unknown function	-3.69	2.04E-02
b0753	*ybgS*	unknown function	-3.64	2.19E-02
b0775	*bioB*	biotin biosynthesis	3.67	2.23E-02
b1466	*narW*	nitrate assimilation	-3.63	2.32E-02
b2414	*cysK*	L-cysteine /precursor biosynthesis	-3.58	2.37E-02
b1287	*yciW*	hyperosmotic stress	-3.58	2.40E-02
b3073	*patA*	nitrogen limitation indicator	-3.56	2.45E-02
b1465	*narV*	nitrate assimilation	-3.57	2.48E-02
b1040	*csgD*	curli secretion	-3.55	2.51E-02
b2427	*murR*	muramic acid regulator	-3.56	2.51E-02
b4187	*aidB*	cellular response to DNA damage	-3.51	2.64E-02
b1468	*narZ*	nitrate assimilation	-3.48	2.74E-02
b1469	*narU*	nitrate assimilation	-3.45	2.90E-02
b2241	*glpA*	anaerobic respiration	3.43	3.01E-02
b1259	*yciG*	acid resistance	-3.44	3.15E-02

b3510	*hdeA*	acid resistance	-3.36	3.24E-02
b0553	*nmpC*	unknown function	3.31	3.53E-02
b2749	*ygbE*	unknown function	-3.29	3.60E-02
b3511	*hdeD*	acid resistance	-3.25	3.77E-02
b3478	*nikZ*	nickel transport	3.23	4.01E-02
b0774	*bioA*	biotin biosynthesis	3.21	4.18E-02
b2013	*tsuA*	sulphur/L-cysteine starvation	-3.17	4.21E-02
b1732	*katE*	hyperosmotic stress	-3.16	4.25E-02
b2012	*ysuB*	sulphur/L-cysteine starvation	-3.16	4.26E-02
b0776	*bioF*	biotin biosynthesis	3.16	4.44E-02
b4376	*osmY*	hyperosmotic stress	-3.12	4.49E-02
b1137	*ymfD*	unknown function	3.40	4.52E-02
b2465	*tktB*	L-cysteine /precursor biosynthesis	-3.11	4.53E-02
b4568	*ytjA*	unknown function	-3.10	4.58E-02
b1037	*csgG*	curli secretion	-3.10	4.60E-02
b3514	*mdtF*	multidrug efflux	-3.08	4.75E-02
b3661	*nlpA*	methionine transport	-3.05	4.89E-02

Supplementary table 8: Table showing accession numbers, gene names, features, logarithmic fold changes (logFC) and p-values of all differentially expressed genes (DEGs) of MDS42_pCYS_i with p-values <0.05. *LogFC and logCPM were calculated by dividing values of the later generation population (LGP) by values of the early generation population (EGP).*

Accession	**Gene**	**Feature**	**LogFC**	**P-Value**
b0936	*ssuA*	sulphur-/L-Cysteine starvation	-4.48	0.0068
b0937	*ssuE*	sulphur-/L-Cysteine starvation	-4.43	0.0075
b3603	*lldP*	cellular response to DNA damage	4.35	0.0081
b3917	*sbp*	sulphur-/L-Cysteine starvation	-4.33	0.0082
b0934	*ssuC*	sulphur-/L-Cysteine starvation	-4.20	0.0101
b0935	*ssuD*	sulphur-/L-Cysteine starvation	-4.06	0.0122
b2752	*cysD*	sulphur-/L-Cysteine starvation	-4.03	0.0128
b3604	*lldR*	cellular response to DNA damage	4.03	0.0130
b3605	*lldD*	cellular response to DNA damage	3.99	0.0134
b2751	*cysN*	sulphur-/L-Cysteine starvation	-3.78	0.0180
b2764	*cysJ*	sulphur-/L-Cysteine starvation	-3.77	0.0184
b2763	*cysI*	sulphur-/L-Cysteine starvation	-3.64	0.0219
b2750	*cysC*	sulphur-/L-Cysteine starvation	-3.62	0.0226
b2425	*cysP*	sulphur-/L-Cysteine starvation	-3.56	0.0250
b2013	*tsuA*	sulphur-/L-Cysteine starvation	-3.55	0.0253
b2012	*tsuB*	sulphur-/L-Cysteine starvation	-3.52	0.0264
b2422	*cysA*	sulphur-/L-Cysteine starvation	-3.43	0.0294
b2762	*cysH*	sulphur-/L-Cysteine starvation	-3.42	0.0300
b0933	*ssuB*	sulphur-/L-Cysteine starvation	-3.37	0.0321
b2424	*cysU*	sulphur-/L-Cysteine starvation	-3.27	0.0372
b2423	*cysW*	sulphur-/L-Cysteine starvation	-3.22	0.0397
b1729	*tcyP*	sulphur-/L-Cysteine starvation	-3.20	0.0401
b1287	*yciW*	sulphur-/L-Cysteine starvation	-3.12	0.0447

Supplementary table 9: Table showing accession numbers, gene names, features, logarithmic fold changes (logFC) and p-values of all differentially expressed genes (DEGs) of W3110_pCYS_m with p-values <0.05. *LogFC and logCPM were calculated by dividing values of the later generation population (LGP) by values of the early generation population (EGP).*

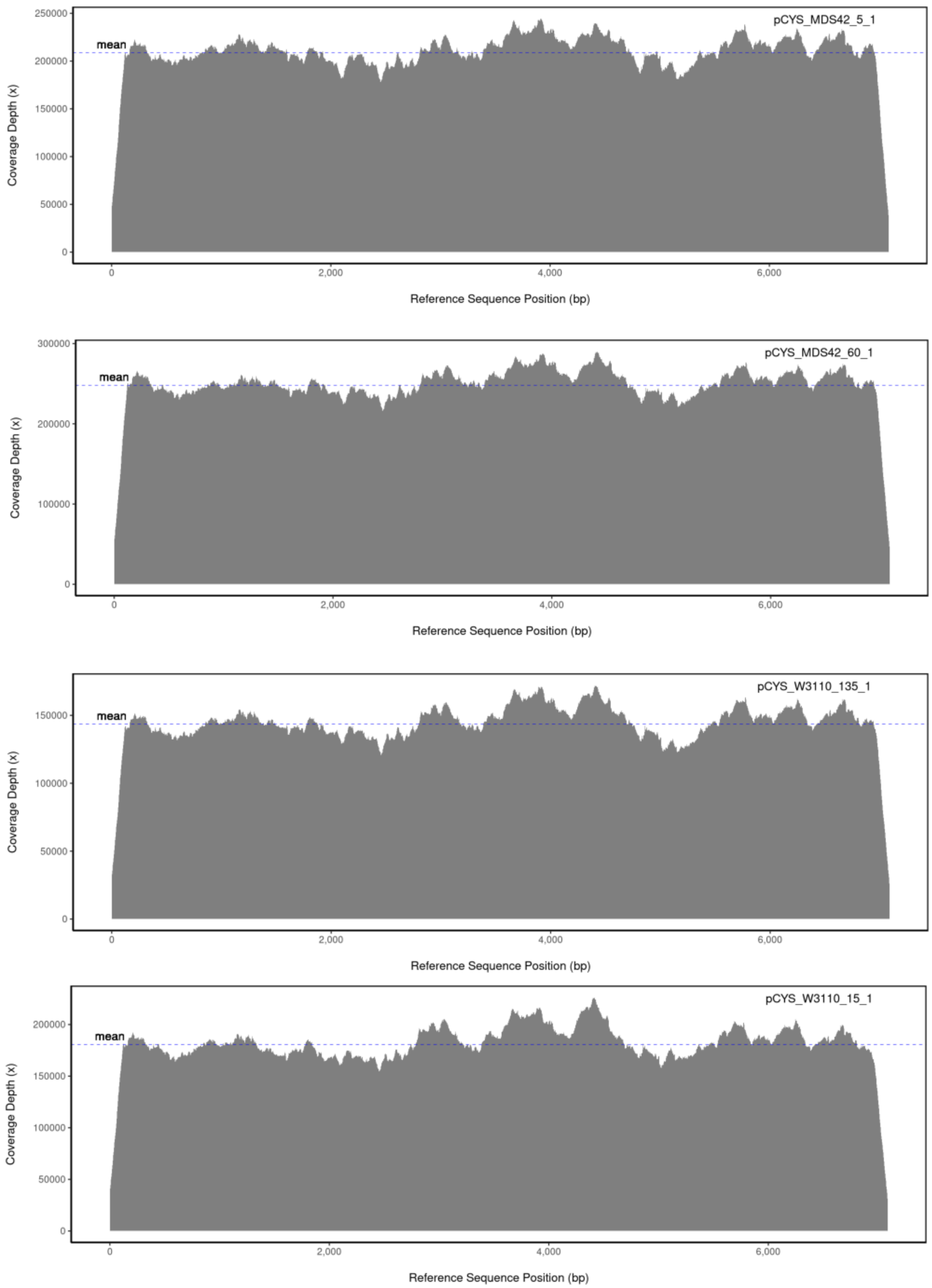
Accession	**Gene**	**Feature**	**logFC**	**P-Value**
b0366	*tauB*	sulphur-/ L-cysteine starvation	-5.82	2.98E-05
b0367	*tauC*	sulphur-/ L-cysteine starvation	-5.08	0.000155955
b0365	*tauA*	sulphur-/ L-cysteine starvation	-4.90	0.000213135
b0936	*ssuA*	sulphur-/ L-cysteine starvation	-4.78	0.000363555

b0935	*ssuD*	sulphur-/ L-cysteine starvation	-4.42	0.00064772
b1493	*gadB*	acid resistance	-4.39	0.000650756
b4470	*cyuA*	l-Cysteine degradation	4.20	0.00099003
b1492	*gadC*	acid resistance	-4.20	0.001002402
b0937	*ssuE*	sulphur-/ L-cysteine starvation	-4.30	0.001055832
b0368	*tauD*	sulphur-/ L-cysteine starvation	-4.13	0.00123138
b0450	*glnK*	nitrogen starvation	-3.95	0.001776567
b3707	*tnaC*	transcriptional attenuation	-4.18	0.002304927
b0934	*ssuC*	sulphur-/ L-cysteine starvation	-3.81	0.002799372
b0451	*amtB*	nitrogen starvation	-3.67	0.003164359
b1489	*dosP*	oxygen sensing	-3.69	0.003226504
b3708	*tnaA*	tryptophanases	-3.63	0.003575459
b2750	*cysC*	sulphur-/ L-cysteine starvation	-3.59	0.003912051
b2298	*yfcC*	unknown function	3.59	0.004078273
b1488	*ddpX*	nitrogen starvation	-3.62	0.004573246
b1748	*astC*	L-arginine catabolysis	-3.52	0.004670165
b1988	*nac*	nitrogen starvation	-3.43	0.005431418
b3917	*sbp*	sulphur-/ L-cysteine starvation	-3.43	0.005444333
b1467	*narY*	nitrate assimilation	-3.43	0.005601404
b3511	*hdeD*	acid resistance	-3.361	0.006279399
b1466	*narW*	nitrate assimilation	-3.39	0.006610653
b1747	*astA*	L-arginine catabolysis	-3.32	0.006992027
b1012	*rutA*	pyrimidin degradation	-3.42	0.007184279
b1182	*hlyE*	hemolysis	3.51	0.00775016
b3517	*gadA*	acid resistance	-3.22	0.008425411
b2751	*cysN*	sulphur-/ L-cysteine starvation	-3.21	0.008634876
b1468	*narZ*	nitrate assimilation	-3.15	0.009868409
b4314	*fimA*	cell adhesion	-3.01	0.013196354
b3671	*ilvB*	BCAA biosynthesis	-2.99	0.013343519
b2763	*cysI*	sulphur-/ L-cysteine starvation	-2.98	0.013714836
b3509	*hdeB*	acid resistance	-2.94	0.014698199
b3268	*yhdW*	unknown function	-2.94	0.015223895
b1038	*csgF*	curli secretion	-2.98	0.015299404
b1487	*ddpA*	nitrogen starvation	-2.92	0.01666648
b1541	*ydfZ*	unknown function	2.90	0.016843664
b1465	*narV*	nitrate assimilation	-2.89	0.017708658
b3514	*mdtF*	multidrug efflux	-2.85	0.018047176
b3110	*cyuP*	l-Cysteine degradation	2.82	0.018927267
b1987	*cbl*	sulphur-/ L-cysteine starvation	-2.82	0.019316605
b1746	*astD*	L-arginine catabolysis	-2.82	0.019391912
b2111	*yehD*	biofilm formation	3.26	0.019492368
b3510	*hdeA*	acid resistance	-2.79	0.020018914
b2422	*cysA*	sulphur-/ L-cysteine starvation	-2.77	0.020788403
b0992	*yccM*	unknown function	2.84	0.02087771
b2752	*cysD*	sulphur-/ L-cysteine starvation	-2.72	0.02271106
b4072	*nrfC*	anaerobic respiration	2.93	0.022806289
b2343	*yfcZ*	unknown function	2.71	0.023304408
b4070	*nrfA*	anaerobic respiration	2.74	0.023744641
b2423	*cysW*	sulphur-/ L-cysteine starvation	-2.71	0.023758522
b0933	*ssuB*	sulphur-/ L-cysteine starvation	-2.72	0.024362651
b0896	*dmsC*	anaerobic respiration	2.66	0.026118345
b2110	*yehC*	biofilm formation	3.86	0.026778193
b4116	*adiY*	transcriptional activation	2.61	0.028356012
b1745	*astB*	L-arginine catabolysis	-2.62	0.028460359
b1224	*narG*	nitrate assimilation	-2.59	0.029222593
b3158	*ubiU*	ubiquinone biosynthesis	2.61	0.029286652
b2310	*argT*	nitrogen starvation	-2.58	0.030323738
b2764	*cysJ*	sulphur-/ L-cysteine starvation	-2.57	0.030819373
b3670	*ilvN*	BCAA biosynthesis	-2.56	0.031168241
b1443	*ydcV*	unknown function	-2.59	0.031564249
b3269	*ydhX*	unknown function	-2.59	0.032143511
b1674	*ydhY*	unknown function	2.64	0.032660893
b0895	*dmsB*	anaerobic respiration	2.53	0.033062582
b2203	*napB*	anaerobic respiration	2.59	0.03328397
b1490	*dosC*	oxygen sensing	-2.53	0.033441933

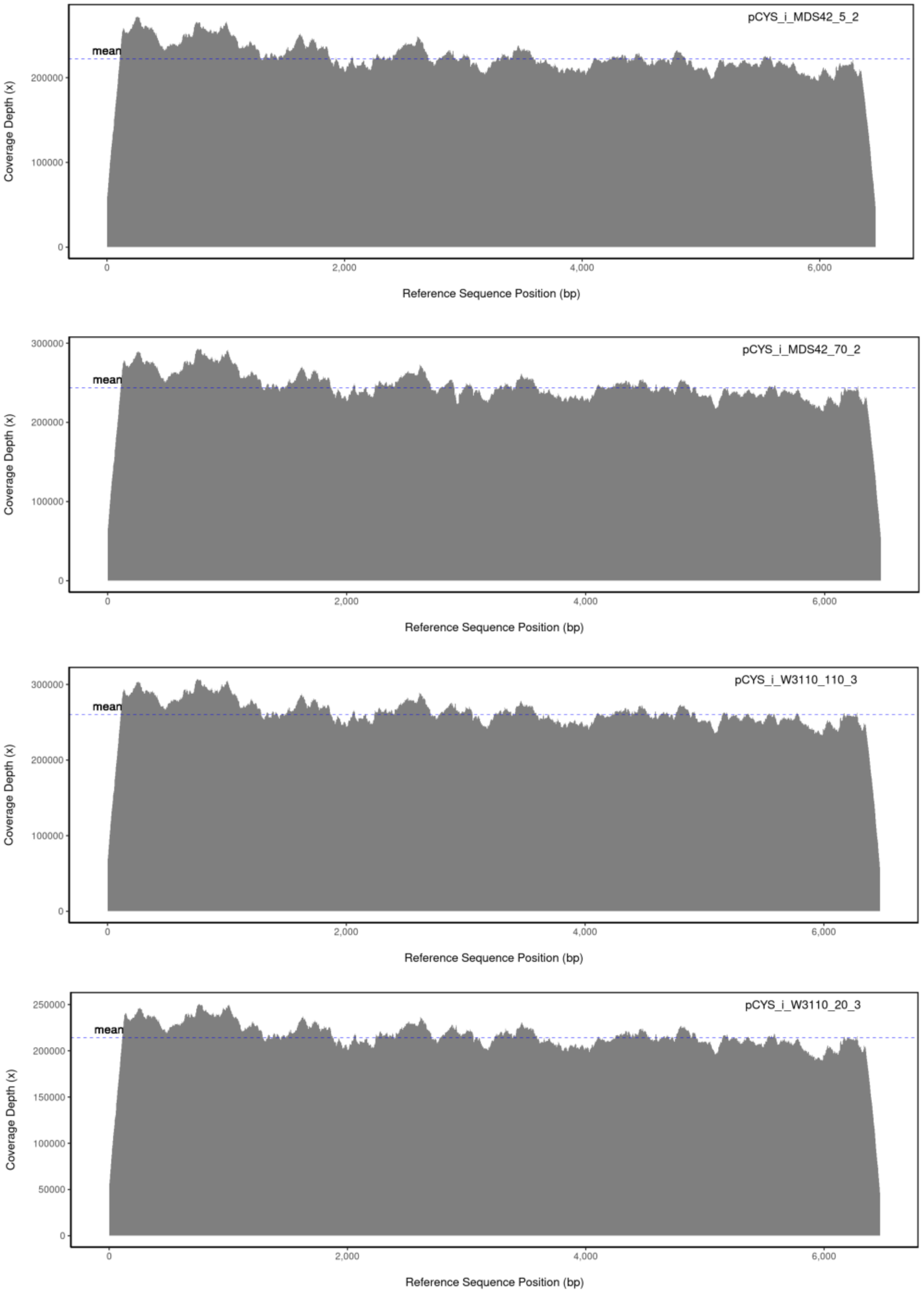
b4245	*pyrB*	pyrimidin degradation	-2.52	0.033568073
b3369	*yhfL*	unknown function	3.10	0.033702782
b1011	*rutB*	pyrimidin degradation	-2.69	0.034889092
b2345	*yfdF*	unknown function	3.18	0.03551058
b2208	*napF*	anaerobic respiration	2.53	0.035939355
b0991	*ymcE*	unknown function	3.06	0.036027242
b3073	*patA*	l-lysin degradation	-2.45	0.03821933
b1731	*cedA*	cell division regulation	2.68	0.038563357
b2292	*yfbS*	unknown function	2.45	0.038713109
b2996	*hybA*	anaerobic respiration	2.45	0.038806123
b0990	*cspG*	cold shock response	2.50	0.03895797
b0894	*dmsA*	anaerobic respiration	2.44	0.039041251
b1287	*yciW*	hyperosmotic stress	-2.44	0.039274106
b2676	*nrdF*	pyrimidin degradation	-2.49	0.03953287
b1675	*fumA*	anaerobic respiration	2.54	0.039793075
b1025	*dgcT*	biofilm formation	2.54	0.039812429
b1441	*ydcT*	unknown function	-2.47	0.040116258
b3709	*tnaB*	tryptophanases	-2.60	0.040244207
b2201	*ccmA*	cytochrome c maturation	2.44	0.040712435
b2727	*hypB*	protein maturation	2.42	0.041180646
b4244	*pyrI*	pyrimidin degradation	-2.40	0.041669821
b0621	*dcuC*	succinat efflux	2.42	0.041722113
b2202	*napC*	anaerobic respiration	2.42	0.041909772
b0036	*caiD*	L-carnitine degradation	2.51	0.043274337
b2997	*hybO*	anaerobic respiration	2.38	0.043938516
b4606	*ypfM*	unknown function	2.49	0.044010761
b1014	*putA*	L-proline degradation	-2.37	0.044714409
b2749	*ygbE*	unknown function	-2.39	0.045600936
b3508	*yhiD*	unknown function	-2.37	0.046274455
b3513	*gadW*	acid resistance	-2.35	0.046389608
b4713	*agrB*	small regulatory RNA	2.89	0.047659431
b0618	*citC*	citrate lyase	2.78	0.047944746
b1039	*csgE*	curli secretion	-2.38	0.047956706
b4517	*gnsA*	unknown function	2.43	0.048711147
b1744	*astE*	L-arginine catabolysis	-2.36	0.049168527
b2150	*mglB*	chemotaxis	-2.34	0.049424165

Supplementary table 10: Table showing accession numbers, gene names, features, logarithmic fold changes (logFC) and p-values of all differentially expressed genes (DEGs) of MDS42_pCYS_m with p-values <0.05. *LogFC and logCPM were calculated by dividing values of the later generation population (LGP) by values of the early generation population (EGP).*

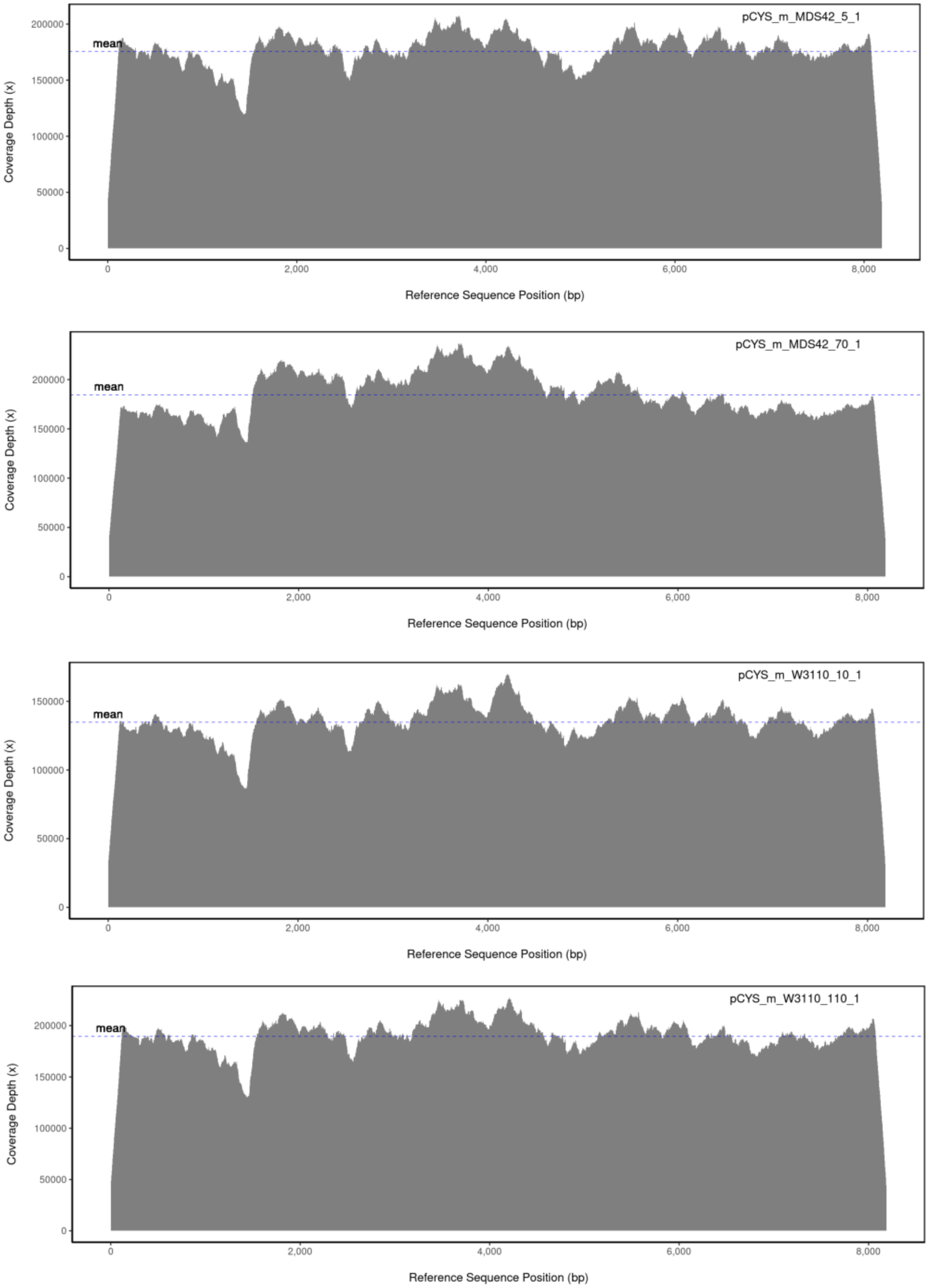
Accession	Gene	Feature	logFC	P-Value
gene-b4668	*ibsB*	unknown function	-2.17	0.022
gene-b3643	*rph*	rRNA 3'-end and tRNA processing	-4.98	0.027
gene-b1539	*ydfG*	uracil catabolism	3.04	0.028
gene-b1137	*ymfD*	unknown function	2.14	0.028
gene-b2848	*yqeJ*	unknown function	1.37	0.038
gene-b0150	*yceK*	iron starvation	4.59	0.039
gene-b2724	*hycB*	glucose catabolism	-1.33	0.039
gene-b2481	*hyfA*	alternative sugar utilizing pathways	1.22	0.043
gene-b3900	*frvA*	glucose catabolism	-1.40	0.044
gene-b4268	*idnK*	alternative sugar utilizing pathways	1.12	0.047
gene-b0532	*sfmD*	pilus organization	-1.13	0.047
gene-b0040	*caiT*	nitrogen starvation	-1.13	0.048
gene-b4181	*yjfI*	unknown function	1.24	0.048
gene-b4254	*argI*	nitrogen starvation	-1.02	0.049



Supplementary figure 1: Per base coverage depths (x) of sequenced pCYS extracted from early and late generation populations (EGPs and LGPs) of E. coli W3110 and MDS42. *Sequencing was conducted with Illumina Novaseq paired end 2x150bp.*



Supplementary figure 2: Per base coverage depths (x) of sequenced pCYS_i extracted from early and late generation populations (EGPs and LGPs) of E. coli W3110 and MDS42. *Sequencing was conducted with Illumina Novaseq paired end 2x150bp.*



Supplementary figure 3: Per base coverage depths (x) of sequenced pCYS_m extracted from early and late generation populations (EGPs and LGPs) of E. coli W3110 and MDS42. *Sequencing was conducted with Illumina Novaseq paired end 2x150b.*

8 Bibliography

1. Straathof AJ. Transformation of biomass into commodity chemicals using enzymes or cells. Chem Rev. 2014;114(3):1871-908.

2. Wachtmeister J, Rother D. Recent advances in whole cell biocatalysis techniques bridging from investigative to industrial scale. Curr Opin Biotechnol. 2016;42:169-77.

3. Demain AL, Vaishnav P. Production of recombinant proteins by microbes and higher organisms. Biotechnol Adv. 2009;27(3):297-306.

4. Adrio J-L. Recombinant organisms for production of industrial products. Bioengineered Bugs. 2010:116-31.

5. Terpe K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol. 2006;72(2):211-22.

6. Lee SY. High cell-density culture of Escherichia coli. Trends in Biotechnology. 1996;14(3):98-105.

7. Shiloach J, Fass R. Growing *E. coli* to high cell density - a historical perspective on method development. Biotechnol Adv. 2005;23(5):345-57.

8. Pope B. High efficiency 5 min transformation of Escherichia coli. Nucleic Acid research. 1996;24:536-7.

9. Zhao D, Yuan S, Xiong B, Sun H, Ye L, Li J, et al. Development of a fast and easy method for Escherichia coli genome editing with CRISPR/Cas9. Microb Cell Fact. 2016;15(1):205.

10. Smanski MJ, Bhatia S, Zhao D, Park Y, L BAW, Giannoukos G, et al. Functional optimization of gene clusters by combinatorial design and assembly. Nat Biotechnol. 2014;32(12):1241-9.

11. Watson JF, Garcia-Nafria J. In vivo DNA assembly using common laboratory bacteria: A re-emerging tool to simplify molecular cloning. J Biol Chem. 2019;294(42):15271-81.

12. Rokke G, Korvald E, Pahr J, Oyas O, Lale R. BioBrick assembly standards and techniques and associated software tools. Methods Mol Biol. 2014;1116:1-24.

13. Cannazza P, Rabuffetti M, Donzella S, De Vitis V, Contente ML, de Oliveira M, et al. Whole cells of recombinant CYP153A6-E. coli as biocatalyst for regioselective hydroxylation of monoterpenes. AMB Express. 2022;12(1):48.

14. Oh YH, Kang KH, Kwon MJ, Choi JW, Joo JC, Lee SH, et al. Development of engineered Escherichia coli whole-cell biocatalysts for high-level conversion of L-lysine into cadaverine. J Ind Microbiol Biotechnol. 2015;42(11):1481-91.

15. Shin J, Yu J, Park M, Kim C, Kim H, Park Y, et al. Endocytosing Escherichia coli as a Whole-Cell Biocatalyst of Fatty Acids. ACS Synth Biol. 2019;8(5):1055-66.

16. Yang SY, Han YH, Park YL, Park JY, No SY, Jeong D, et al. Production of L-Theanine Using Escherichia coli Whole-Cell Overexpressing gamma-Glutamylmethylamide Synthetase with Bakers Yeast. J Microbiol Biotechnol. 2020;30(5):785-92.

17. Wilms B. Development of an *Escherichia coli* whole cell biocatalyst for the production of L-amino aicds. Journal of Biotechnology. 2001;86:19-30.

18. Lin BX, Zhang ZJ, Liu WF, Dong ZY, Tao Y. Enhanced production of N-acetyl-D-neuraminic acid by multi-approach whole-cell biocatalyst. Appl Microbiol Biotechnol. 2013;97(11):4775-84.

19. Chen F, Tao Y, Jin C, Xu Y, Lin BX. Enhanced production of polysialic acid by metabolic engineering of Escherichia coli. Appl Microbiol Biotechnol. 2015;99(6):2603-11.

20. Anthony JR, Anthony LC, Nowroozi F, Kwon G, Newman JD, Keasling JD. Optimization of the mevalonate-based isoprenoid biosynthetic pathway in Escherichia coli for production of the anti-malarial drug precursor amorpha-4,11-diene. Metab Eng. 2009;11(1):13-9.

21. Pitera DJ, Paddon CJ, Newman JD, Keasling JD. Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. Metab Eng. 2007;9(2):193-207.

22. Li C, Ying LQ, Zhang SS, Chen N, Liu WF, Tao Y. Modification of targets related to the Entner-Doudoroff/pentose phosphate pathway route for methyl-D-erythritol 4-phosphate-dependent carotenoid biosynthesis in Escherichia coli. Microb Cell Fact. 2015;14:117.

23. Lorenz E, Klatte S, Wendisch VF. Reductive amination by recombinant Escherichia coli: whole cell biotransformation of 2-keto-3-methylvalerate to L-isoleucine. J Biotechnol. 2013;168(3):289-94.

24. Hummel W, Groger H. Strategies for regeneration of nicotinamide coenzymes emphasizing self-sufficient closed-loop recycling systems. J Biotechnol. 2014;191:22-31.

25. Kratzer R, Woodley JM, Nidetzky B. Rules for biocatalyst and reaction engineering to implement effective, NAD(P)H-dependent, whole cell bioreductions. Biotechnol Adv. 2015;33(8):1641-52.

26. Torrelo G, Hanefeld U, Hollmann F. Biocatalysis. Catalysis Letters. 2014;145(1):309-45.

27. Studier FW. Use of bacteriophage t7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol 1986;189:113-30.

28. Gottesmann S. Proteases and their targets in *Escherichia coli*. Annu Rev Genet 1996;30:465-506.

29. Marisch K. Evaluation of three industrial *Escherichia coli strains* in fed-batch cultivation during high-level SOD protein production. Microbial Cell Factories. 2013;12:11.

30. Minton NP. Improved plasmid vectors for the isolation of translational lac gene fusions. Gene. 1984;1105:269-73.

31. Bolivar F. Construction and characterization of new cloning vehicles, a multipurpose cloning system. Gene. 1977;2:95-113.

32. Wanner BL. Regulation of *lac* Operon Expression: Reappraisal of the theory of catabolite repression.pdf>. Journal of Bacteriology. 1978;136:947-54.

33. Boer HAd. The tac promoter: a functional hybrid derived from the trp and lac promoters. Biochemistry. 1983;0:21-5.

34. Guzman L-M. Tight regulation, Modulation, and High-level Expression by Vectors containing the arabinose P_{BAD} promoter. Journal of Bacteriology. 1995;177:4121-30.

35. Menart V, Jevsevar S, Vilar M, Trobis A, Pavko A. Constitutive versus thermoinducible expression of heterologous proteins in Escherichia coli based on strong PR,PL promoters from phage lambda. Biotechnol Bioeng. 2003;83(2):181-90.

36. Valdez-Cruz NA. Production of recombinant proteins in *E. coli* by the heat inducible expression system based on the phage lambda pL and / or pR promoters. Microb Cell Fact. 2010;9.

37. Jaishankar J, Srivastava P. Strong synthetic stationary phase promoter-based gene expression system for Escherichia coli. Plasmid. 2020;109:102491.

38. Khan F, Legler PM, Mease RM, Duncan EH, Bergmann-Leitner ES, Angov E. Histidine affinity tags affect MSP1(42) structural stability and immunodominance in mice. Biotechnol J. 2012;7(1):133-47.

39. Klose J, Wendt N, Kubald S, Krause E, Fechner K, Beyermann M, et al. Hexa-histidin tag position influences disulphide structure but not binding behavior of in vitro folded N-terminal domain of rat corticotropin-releasing factor receptor type 2a. Protein Sci. 2004;13(9):2470-5.

40. Bucher MH, Evdokimov AG, Waugh DS. Differential effects of short affinity tags on the crystallization of *Pyrococcus furiusos* maltodextrin-binding protein. Acta Crystallogrpahica D. 2001;58:392-7.

41. Kapust RB. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused.pdf>. Protein Sci. 1999;8:1668-74.

42. LaVallie ER. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. Nature Biotechnology. 1993;11:187-93.

43. Smith DB. Single-step purifaction of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene. 1988;67:31-40.

44. Hammarstrom M, Hellgren N, van Den Berg S, Berglund H, Hard T. Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Sci. 2002;11(2):313-21.

45. Waugh DS. An overview of enzymatic reagents for the removal of affinity tags. Protein Expr Purif. 2011;80(2):283-93.

46. Korpimäki T, Kurittu J, Karp M. Surprisingly fast disappearance of β-lactam selection pressure in cultivation as detected with novel biosensing approaches. Journal of Microbiological Methods. 2003;53(1):37-42.

47. Shaw WV. Chloramphenicol acetyltransferase: enzymology and molecular biology. CRC Crit Rev Biochem. 1983;14(1):1-46.

48. Roberts MC. Tetracycline resistance determinants: mechanisms of action, regulation of expression, genetic mobility, and distribution. FEMS Microbiological Reviews. 1996;19:1-24.

49. Yew NS, Zhao H, Wu IH, Song A, Tousignant JD, Przybylska M, et al. Reduced inflammatory response to plasmid DNA vectors by elimination and inhibition of immunostimulatory CpG motifs. Mol Ther. 2000;1(3):255-62.

50. Kroll J, Klinter S, Schneider C, Voss I, Steinbuchel A. Plasmid addiction systems: perspectives and applications in biotechnology. Microb Biotechnol. 2010;3(6):634-57.

51. Hermann T. Industrial production of amino acids by coryneform bacteria. Journal of Biotechnology. 2003;104(1-3):155-72.

52. Nampoothiri KM. Bioenergy Research: Advances and Applications. In: Gupta VG, Tuohy, M., Kubicek, C. P., Saddler J. & Xu F., editor.2014. p. 337-52.

53. Kinoshita S. Studies on the amino acid fermentation; Production of L-glutamic acid by various microorganisms. J Gen Appl Microbiol. 1957;3:193-205.

54. Udaka S. Screening method for microorganisms accumulating metabolites and its use in the isolation of *micrococcus glutamicus*. J Bacteriol Res 1959;79:754-5.

55. Kumagai H. Microbial production of Amino acids in Japan. Advances in Biochemical Engineering/Biotechnology. 692000. p. 71-85.

56. research Gv. Amino acids market size, share & trends analysis report. 2022. Report No.: 978-1-68038-453-6.

57. Hermann Sahm, Garabed Antranikian, Klaus-Peter Stahmann, Takors R. Amino acids. Industrial Microbiology: Springer Spektrum; 2009. p. 110.

58. F.C. Neidhardt JLI, M. Schaechter. Physiology of the bacterial cell. Biochemical Education1990. p. 133-73.

59. Marx A, de Graaf AA, Wiechert W, Eggeling L, Sahm H. Determination of the fluxes in the central metabolism ofCorynebacterium glutamicum by nuclear magnetic resonance spectroscopy combined with metabolite balancing. Biotechnology and Bioengineering. 1996;49(2):111-29.

60. Cicchillo RM, Baker MA, Schnitzer EJ, Newman EB, Krebs C, Booker SJ. Escherichia coli L-serine deaminase requires a [4Fe-4S] cluster in catalysis. J Biol Chem. 2004;279(31):32418-25.

61. Atkuri KR, Mantovani JJ, Herzenberg LA, Herzenberg LA. N-Acetylcysteine--a safe antidote for cysteine/glutathione deficiency. Curr Opin Pharmacol. 2007;7(4):355-9.

62. Ott DM, inventorPersonal care and medical products incorporating bound organosulphur groups. United States of America2006.

63. Iorizzo M, Piraccini BM, Tosti A. Nail cosmetics in nail disorders. J Cosmet Dermatol. 2007;6(1):53-8.

64. AG WC. Fermopure - plant-based L-cystine and L-cysteine2010 12.12.2022.

65. Buttery PJ. Amino acids in farm animal nutrition. In: D'Mello JPF, editor. Amino acid metabolism in farm animals: an overview: CAB International 1994. p. 1-10.

66. Scientific Opinion on the safety and efficacy of L-cystine for all animal species. EFSA Journal. 2013;11(4).

67. IndustryArc. Cysteine market - industry analysis, market size, share, trends, application analysis, growth and forecast 2022-2027. 2022. Report No.: CMR10433.

68. Ismail NI. Production of Cysteine: Approaches, Challenges and Potential Solution. International Journal of Biotechnology for Wellness Industries. 2014;3:95-101.

69. Renneberg R. High-grade cysteine no longer has to be extraxted from hair. In: Demain AL, editor. Biotechnology for Beginners. 2: Elsevier; 2017. p. 132.

70. Hee RO. Analysis of the reaction steps in the bioconversion of D,L-ATC to L-cysteine. Journal of Microbiology and Biotechnology. 1991;1(1):50-3.

71. Hee RO. Continuous L-cysteine production using immobilized cell reactors and product extractors. Process Biochemistry. 1996;32(3):201-9.

72. Xu L, Wu F, Li T, Zhang X, Chen Q, Jiang B, et al. Ultrasound-assisted l-cysteine whole-cell bioconversion by recombinant Escherichia coli with tryptophan synthase. Green Processing and Synthesis. 2021;10(1):842-50.

73. Ishiwata K-I. Enzymatic production of L-cysteine with tryptophan synthase of *Escherichia coli*. Journal of fermentation and bioengineering. 1989;67:169-72.

74. Wada M, Takagi H. Metabolic pathways and biotechnological production of L-cysteine. Appl Microbiol Biotechnol. 2006;73(1):48-54.

75. Dhillon GS. Microbial process for L-cysteine production. Enzyme Microb Technol. 1986;9:277-80.

76. Wirtz M, Droux M. Synthesis of the sulphur amino acids: cysteine and methionine. Photosynth Res. 2005;86(3):345-62.

77. Nakatani T. Enhancement of thioredoxin/glutaredoxin-mediated L-cysteine synthesis from S-sulphocysteine increases L-cysteine production in Escherichia coli. Microbial Cell Factories. 2012;11(62).

78. Daßler T. Identification of a major facilitator protein from Escherichia coli involved in efflux of metabolites of the cysteine pathway. Molecular Microbiology. 2000;36(5):1101-12.

79. Franke I, Resch A, Dassler T, Maier T, Bock A. YfiK from Escherichia coli promotes export of O-acetylserine and cysteine. J Bacteriol. 2003;185(4):1161-6.

80. Heieck K, Arnold ND, Bruck TB. Metabolic stress constrains microbial L-cysteine production in Escherichia coli by accelerating transposition through mobile genetic elements. Microb Cell Fact. 2023;22(1):10.

81. Bell JK, Pease PJ, Bell JE, Grant GA, Banaszak LJ. De-regulation of D-3-phosphoglycerate dehydrogenase by domain removal. Eur J Biochem. 2002;269(17):4176-84.

82. Leinfelder W HP, inventor; Wacker Chemie AG, assignee. Process for preparing O-Acetylserine, L-cysteine and L-cysteine-related products2001.

83. Harris CL. Cysteine and Growth Inhibition of Escherichia coli: Threonine Deaminase as the Target Enzyme. Journal of Bacteriology. 1981;145:1031-5.

84. Sorenson MA. Cysteine, Even in Low Concentrations, Induces Transient Amino Acid Starvation in Escherichia coli. Journal of Bacteriology. 1991;173:5244-6.

85. Jagura-Burdzy G. Use of Gene Fusions to Study Expression of cysB, the Regulatory Gene of the Cysteine Regulon. Journal of Bacteriology. 1981;147:744-51.

86. Liu H, Wang Y, Hou Y, Li Z. Fitness of Chassis Cells and Metabolic Pathways for l-Cysteine Overproduction in Escherichia coli. J Agric Food Chem. 2020;68(50):14928-37.

87. Nielsen J, Keasling JD. Engineering Cellular Metabolism. Cell. 2016;164(6):1185-97.

88. Borkowski O, Ceroni F, Stan GB, Ellis T. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. Curr Opin Microbiol. 2016;33:123-30.

89. Xiao Y, Bowen CH, Liu D, Zhang F. Exploiting nongenetic cell-to-cell variation for enhanced biosynthesis. Nat Chem Biol. 2016;12(5):339-44.

90. Lidstrom ME, Konopka MC. The role of physiological heterogeneity in microbial population behavior. Nat Chem Biol. 2010;6(10):705-12.

91. Lara AR. Living with heterogeneities in bioreactors. Molecular Biotechnology 2006;34:355-81.

92. Mustafi N, Grunberger A, Mahr R, Helfrich S, Noh K, Blombach B, et al. Application of a genetically encoded biosensor for live cell imaging of L-valine production in pyruvate dehydrogenase complex-deficient Corynebacterium glutamicum strains. PLoS One. 2014;9(1):e85731.

93. Rugbjerg P, Myling-Petersen N, Porse A, Sarup-Lytzen K, Sommer MOA. Diverse genetic error modes constrain large-scale bio-based production. Nat Commun. 2018;9(1):787.

94. Tyo KE, Ajikumar PK, Stephanopoulos G. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. Nat Biotechnol. 2009;27(8):760-5.

95. Layton JC, Foster PL. Error-prone DNA polymerase IV is controlled by the stress-response sigma factor, RpoS, in Escherichia coli. Mol Microbiol. 2003;50(2):549-61.

96. Fujii S. A Comprehensive View of Translesion Synthesis in Escherichia coli. Microbiology and Molecular Biology Reviews. 2020;84(3):1-29.

97. Rankin DJ, Rocha EP, Brown SP. What traits are carried on mobile genetic elements, and why? Heredity (Edinb). 2011;106(1):1-10.

98. Fernandez-Alarcon C, Singer RS, Johnson TJ. Comparative genomics of multidrug resistance-encoding IncA/C plasmids from commensal and pathogenic Escherichia coli from multiple animal sources. PLoS One. 2011;6(8):e23415.

99. Zhang WJ, Wang XM, Dai L, Hua X, Dong Z, Schwarz S, et al. Novel conjugative plasmid from Escherichia coli of swine origin that coharbors the multiresistance gene cfr and the extended-spectrum-beta-lactamase gene blaCTX-M-14b. Antimicrob Agents Chemother. 2015;59(2):1337-40.

100. Moran RA, Holt KE, Hall RM. pCERC3 from a commensal ST95 Escherichia coli: A ColV virulence-multiresistance plasmid carrying a sul3-associated class 1 integron. Plasmid. 2016;84-85:11-9.

101. Brzuszkiewicz E, Thurmer A, Schuldes J, Leimbach A, Liesegang H, Meyer FD, et al. Genome sequence analyses of two isolates from the recent Escherichia coli outbreak in

Germany reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic Escherichia coli (EAHEC). Arch Microbiol. 2011;193(12):883-91.

102. Fluit AC. Class 1 Integrons, Gene Cassettes, Mobility and Epidemiology. Eur J Clin Microbiol Infect Dis. 1999;18:761-70.

103. Los M. Minimization and Prevention of Phage Infections in Bioprocesses. Methods in Molecular Biology. 834: Springer; 2011. p. 305-15.

104. Garneau JE. Bacteriophages of lactic acid bacteria and their impact on milk fermentations. Microbial Cell Factories. 2011;10:1-10.

105. Murphy J, Royer B, Mahony J, Hoyles L, Heller K, Neve H, et al. Biodiversity of lactococcal bacteriophages isolated from 3 Gouda-type cheese-producing plants. J Dairy Sci. 2013;96(8):4945-57.

106. Lu Z, Perez-Diaz IM, Hayes JS, Breidt F. Bacteriophage ecology in a commercial cucumber fermentation. Appl Environ Microbiol. 2012;78(24):8571-8.

107. Los M. Bacteriophage contamination: Is there a simple method to reduce its deleterious effects in laboratory cultures and biotechnological factories. Journal of Applied Genetics 2004;45:111-20.

108. Liu L, Zhao D, Ye L, Zhan T, Xiong B, Hu M, et al. A programmable CRISPR/Cas9-based phage defense system for Escherichia coli BL21(DE3). Microb Cell Fact. 2020;19(1):136.

109. Walker SA. An Explosive Antisense RNA strategy for Inhibition of a Lactococcal Bacteriophage. Applied and Environmental Microbiology. 2000;66:310-9.

110. Hickman AB, Dyda F. Mechanisms of DNA Transposition. Microbiol Spectr. 2015;3(2):MDNA3-0034-2014.

111. Lee H, Doak TG, Popodi E, Foster PL, Tang H. Insertion sequence-caused large-scale rearrangements in the genome of Escherichia coli. Nucleic Acids Res. 2016;44(15):7109-19.

112. de Visser JA, Akkermans AD, Hoekstra RF, de Vos WM. Insertion-sequence-mediated mutations isolated during adaptation to growth and starvation in Lactococcus lactis. Genetics. 2004;168(3):1145-57.

113. Wiser MJ. Long-Term Dynamics of Adaptation in Asexual Populations. Science. 2013;342(6164):1364-7.

114. Cooper VS, Schneider D, Blot M, Lenski RE. Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of Escherichia coli B. J Bacteriol. 2001;183(9):2834-41.

115. van der Heijden I, Gomez-Eerland R, van den Berg JH, Oosterhuis K, Schumacher TN, Haanen JB, et al. Transposon leads to contamination of clinical pDNA vaccine. Vaccine. 2013;31(32):3274-80.

116. Prather KL, Edmonds MC, Herod JW. Identification and characterization of IS1 transposition in plasmid amplification mutants of E. coli clones producing DNA vaccines. Appl Microbiol Biotechnol. 2006;73(4):815-26.

117. Park MK, Lee SH, Yang KS, Jung SC, Lee JH, Kim SC. Enhancing recombinant protein production with an Escherichia coli host strain lacking insertion sequences. Appl Microbiol Biotechnol. 2014;98(15):6701-13.

118. Umenhoffer K, Feher T, Baliko G, Ayaydin F, Posfai J, Blattner FR, et al. Reduced evolvability of Escherichia coli MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. Microb Cell Fact. 2010;9:38.

119. Choi JW, Yim SS, Kim MJ, Jeong KJ. Enhanced production of recombinant proteins with Corynebacterium glutamicum by deletion of insertion sequences (IS elements). Microb Cell Fact. 2015;14:207.

120. Lee JH, Sung BH, Kim MS, Blattner FR, Yoon BH, Kim JH, et al. Metabolic engineering of a reduced-genome strain of Escherichia coli for L-threonine production. Microb Cell Fact. 2009;8:2.

121. Siguier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev. 2014;38(5):865-91.

122. Protocols CSH. LB (Luria-Bertani) liquid medium2006.

123. Bachmann BJ. Pedigrees of Some Mutant Strains of Escherichia coli K-12. Bacteriological Reviews. 1972;36:525-57.

124. Pósfai Gr. Emergent Properties of Reduced-Genome Escherichia coli. Science. 2006;312(5776):1044-6.

125. Winterhalter C, Leinfelder W, inventors; Consortium für elektrchemische Industrie GmbH, assignee. Microorganisms and processes for the fermentative production of L-cysteine, l-cystin, N-acetylserine or thiazolidin derivatives. Germany1998.

126. Cetnar DP, Salis HM. Systematic Quantification of Sequence and Structural Determinants Controlling mRNA stability in Bacterial Operons. ACS Synth Biol. 2021;10(2):318-32.

127. Inoue H. High efficiency transformation of Escherichia coli with plasmids. Gene. 1990;96:23-8.

128. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.

129. Robsinson MD. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology. 2010;11:1-9.

130. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40.

131. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 2006;34(Database issue):D32-6.

132. Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schaffer AA. Database indexing for production MegaBLAST searches. Bioinformatics. 2008;24(16):1757-64.

133. Ecovoiu AA, Ghionoiu IC, Ciuca AM, Ratiu AC. Genome ARTIST: a robust, high-accuracy aligner tool for mapping transposon insertions and self-insertions. Mob DNA. 2016;7:3.

134. Loddeke M, Schneider B, Oguri T, Mehta I, Xuan Z, Reitzer L. Anaerobic Cysteine Degradation and Potential Metabolic Coordination in Salmonella enterica and Escherichia coli. J Bacteriol. 2017;199(16).

135. Nonaka G, Takumi K. Cysteine degradation gene yhaM, encoding cysteine desulphidase, serves as a genetic engineering target to improve cysteine production in Escherichia coli. AMB Express. 2017;7(1):90.

136. Berteau O, Guillot A, Benjdia A, Rabot S. A new type of bacterial sulphatase reveals a novel maturation pathway in prokaryotes. J Biol Chem. 2006;281(32):22464-70.

137. Ploeg JRvd. Identification of a Sulphate Starvation-Regulated Genes in *Escherichia coli*: a Gene Cluster Involved in the Utilization of Taurine as a Sulphur Source. Journal of Bacteriology. 1996;178:5438-46.

138. van Der Ploeg JR, Iwanicka-Nowicka R, Bykowski T, Hryniewicz MM, Leisinger T. The Escherichia coli ssuEADCB gene cluster is required for the utilization of sulphur from aliphatic sulphonates and is regulated by the transcriptional activator Cbl. J Biol Chem. 1999;274(41):29358-65.

139. Sirko A. Sulphate and thiosulphate transport in Escherichia coli K-12 nucleotide sequence and expression of the cysTWAM gene cluster. Journal of Bacteriology. 1990;172:3351-7.

140. Kertesz MA. Proteins induced by sulphate limitation in Escherichia coli, Pseudomonas putida, or Staphylococcus aureus. Journal of Bacteriology. 1993;175:1187-90.

141. Leyh TS, Vogt TF, Suo Y. The DNA sequence of the sulphate activation locus from Escherichia coli K-12. Journal of Biological Chemistry. 1992;267(15):10405-10.

142. Ostrowski J, Wu JY, Rueger DC, Miller BE, Siegel LM, Kredich NM. Characterization of the cysJIH Regions of Salmonella typhimurium and Escherichia coli B. Journal of Biological Chemistry. 1989;264(26):15726-37.

143. Tanaka Y. Crystal structure of a YeeE/YedE family protein engaged in thiosulphate uptake. Science Advanses. 2020;6(35):1-10.

144. Haas M, Rak B. Escherichia coli insertion sequence IS150: transposition via circular and linear intermediates. J Bacteriol. 2002;184(21):5833-41.

145. Chen CS, Korobkova E, Chen H, Zhu J, Jian X, Tao SC, et al. A proteome chip approach reveals new DNA damage recognition activities in Escherichia coli. Nat Methods. 2008;5(1):69-74.

146. Kumar A, Beloglazova N, Bundalovic-Torma C, Phanse S, Deineko V, Gagarinova A, et al. Conditional Epistatic Interaction Maps Reveal Global Functional Rewiring of Genome Integrity Pathways in Escherichia coli. Cell Rep. 2016;14(3):648-61.

147. Posfai G, Plunkett G, 3rd, Feher T, Frisch D, Keil GM, Umenhoffer K, et al. Emergent properties of reduced-genome Escherichia coli. Science. 2006;312(5776):1044-6.

148. Sauer U. Metabolic flux ratio analysis of genetic and environmental modulations of Escherichia coli central carbon metabolism. AMS for Microbiol 1999;181(21):6679-88.

149. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 2009;25(17):2283-5.

150. Johnson DC, Dean DR, Smith AD, Johnson MK. Structure, function, and formation of biological iron-sulphur clusters. Annu Rev Biochem. 2005;74:247-81.

151. Grant GA, Hu Z, Xu XL. Specific interactions at the regulatory domain-substrate binding domain interface influence the cooperativity of inhibition and effector binding in Escherichia coli D-3-phosphoglycerate dehydrogenase. J Biol Chem. 2001;276(2):1078-83.

152. Grant GA. The Contribution of Adjacent Subunits to the Active Sites of D-3-Phosphoglycerate Dehydrogenase. The Journal of Biological Chemistry. 1999;274(9):5357-61.

153. Grant GA. A model for the regulation of D-3-phosphoglycerate dehydrogenase, a V_{max}-type allosteric enzyme. Protein Science. 1996;5:34-41.

154. Inoue K, Noji M, Saito K. Determination of the sites required for the allosteric inhibition of serine acetyltransferase by L-cysteine in plants. Eur J Biochem. 1999;266(1):220-7.

155. Benoni R, De Bei O, Paredi G, Hayes CS, Franko N, Mozzarelli A, et al. Modulation of Escherichia coli serine acetyltransferase catalytic activity in the cysteine synthase complex. FEBS Lett. 2017;591(9):1212-24.

156. Park S, Imlay JA. High levels of intracellular cysteine promote oxidative DNA damage by driving the fenton reaction. J Bacteriol. 2003;185(6):1942-50.

157. Korshunov S, Imlay KRC, Imlay JA. Cystine import is a valuable but risky process whose hazards Escherichia coli minimizes by inducing a cysteine exporter. Mol Microbiol. 2020;113(1):22-39.

158. Dassler T, Maier T, Winterhalter C, Bock A. Identification of a major facilitator protein from Escherichia coli involved in efflux of metabolites of the cysteine pathway. Molecular Microbiology. 2000;36(5):1101-12.

159. Sirko AE. Identification of the *Escherichia coli cysM* Gene Encoding O-Acetylserine Sulphhydylase B by Cloning with Mini-Mu-*lac* Containing Plasmid Replicon. Microbiology 1987;133(10):2719-25.

160. Hryniewicz M. Sulphate and Thiosulphate Transport in Escherichia coli K-12: Identification of a Gene Encoding a Novel Protein Involved in Thiosulphate Binding. Journal of Bacteriology. 1990;172:3358-66.

161. Nakatani T. Enhancement of thioredoxin/glutaredoxin-mediated L-cysteine synthesis from S-sulphocysteine increases L-cysteine production in Escherichia coli. Microbial Cell Factories. 2012;11.

162. Giese H, Klöckner W, Peña C, Galindo E, Lotter S, Wetzel K, et al. Effective shear rates in shake flasks. Chemical Engineering Science. 2014;118:102-13.

163. Horinouchi T, Sakai A, Kotani H, Tanabe K, Furusawa C. Improvement of isopropanol tolerance of Escherichia coli using adaptive laboratory evolution and omics technologies. J Biotechnol. 2017;255:47-56.

164. Maeda T, Horinouchi T, Sakata N, Sakai A, Furusawa C. High-throughput identification of the sensitivities of an Escherichia coli DeltarecA mutant strain to various chemical compounds. J Antibiot (Tokyo). 2019;72(7):566-73.

165. Ceroni F, Algar R, Stan GB, Ellis T. Quantifying cellular capacity identifies gene expression designs with reduced burden. Nat Methods. 2015;12(5):415-8.

166. Heyland J, Blank LM, Schmid A. Quantification of metabolic limitations during recombinant protein production in Escherichia coli. J Biotechnol. 2011;155(2):178-84.

167. Morawska LP, Hernandez-Valdes JA, Kuipers OP. Diversity of bet-hedging strategies in microbial communities-Recent cases and insights. WIREs Mech Dis. 2022;14(2):e1544.

168. Siebring J, Elema MJ, Drubi Vega F, Kovacs AT, Haccou P, Kuipers OP. Repeated triggering of sporulation in Bacillus subtilis selects against a protein that affects the timing of cell division. ISME J. 2014;8(1):77-87.

169. Bergaust L, Shapleigh J, Frostegard A, Bakken L. Transcription and activities of NOx reductases in Agrobacterium tumefaciens: the influence of nitrate, nitrite and oxygen availability. Environ Microbiol. 2008;10(11):3070-81.

170. Hernandez-Valdes JA, van Gestel J, Kuipers OP. A riboswitch gives rise to multi-generational phenotypic heterogeneity in an auxotrophic bacterium. Nat Commun. 2020;11(1):1203.

171. Kremling A, Geiselmann J, Ropers D, de Jong H. Understanding carbon catabolite repression in Escherichia coli using quantitative models. Trends Microbiol. 2015;23(2):99-109.

172. Kussell E. Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments. Science. 2005;309(5743):2075-8.

173. Haseltine WA. Synthesis of Guanosin Tetra- and Pentaphosphate Requires the Presence of a Codon-Specific, Uncharged Transfer Ribonucleic Acid in the Acceptor Site of Ribosomes. Proc Nat Acad Sci 1973;70:1564-8.

174. Agirrezabala X, Fernandez IS, Kelley AC, Carton DG, Ramakrishnan V, Valle M. The ribosome triggers the stringent response by RelA via a highly distorted tRNA. EMBO Rep. 2013;14(9):811-6.

175. Huxtable RJ. Physiological Actions of Taurine. Physiological Reviews. 1992;72:1-63.

176. Shiamoto G. Catabolism of Taurine in Pseudomonas aeruginosa. Biochimica et Biophysica Acta. 1979;569:287-92.

177. Toyama S. Occurrence of taurine: a-ketoglutarateaminotransferase in bacterial extracts. J Bacteriol. 1972;109:533-8.

178. Leyh TS, Taylor JC, Markham GD. The sulphate activation locus of Escherichia coli K12: cloning, genetic, and enzymatic characterization. Journal of Biological Chemistry. 1988;263(5):2409-16.

179. Tanaka Y. Crystal structure of a YeeE/yedE family protein engaged in thiosulphate uptake. Structural Biology. 2020;6(35):1-10.

180. Mazel D. Adaptive eradication of methionine and cysteine from cyanobacterial light-harvesting proteins. Nature. 1989;341:245-8.

181. Fujishima K, Wang KM, Palmer JA, Abe N, Nakahigashi K, Endy D, et al. Reconstruction of cysteine biosynthesis using engineered cysteine-free enzymes. Sci Rep. 2018;8(1):1776.
182. Kawano Y, Ohtsu I, Takumi K, Tamakoshi A, Nonaka G, Funahashi E, et al. Enhancement of L-cysteine production by disruption of yciW in Escherichia coli. J Biosci Bioeng. 2015;119(2):176-9.

183. Zhou Y, Imlay JA. Escherichia coli K-12 Lacks a High-Affinity Assimilatory Cysteine Importer. mBio. 2020;11(3).

184. Downs DM. Balancing cost and benefit: How E. coli cleverly averts disulphide stress caused by cystine. Mol Microbiol. 2020;113(1):1-3.

185. Connolly JP, Gabrielsen M, Goldstone RJ, Grinter R, Wang D, Cogdell RJ, et al. A Highly Conserved Bacterial D-Serine Uptake System Links Host Metabolism and Virulence. PLoS Pathog. 2016;12(1):e1005359.

186. Al Mamun AA, Lombardo MJ, Shee C, Lisewski AM, Gonzalez C, Lin D, et al. Identity and function of a large gene network underlying mutagenic repair of DNA breaks. Science. 2012;338(6112):1344-8.

187. Porse A, Gumpert H, Kubicek-Sutherland JZ, Karami N, Adlerberth I, Wold AE, et al. Genome Dynamics of Escherichia coli during Antibiotic Treatment: Transfer, Loss, and Persistence of Genetic Elements In situ of the Infant Gut. Front Cell Infect Microbiol. 2017;7:126.

188. Drake JW. Rates of Spontaneous Mutation. Genetics. 1998;148:1667-86.

189. Foster PL, Lee H, Popodi E, Townes JP, Tang H. Determinants of spontaneous mutation in the bacterium Escherichia coli as revealed by whole-genome sequencing. Proc Natl Acad Sci U S A. 2015;112(44):E5990-9.

190. Beuzón CR. IS200: an old and still bacterial transposon. International Microbiology 2004;7:3-12.

191. Chandler M, Fayet O, Rousseau P, Ton Hoang B, Duval-Valentin G. Copy-out-Paste-in Transposition of IS911: A Major Transposition Pathway. Microbiol Spectr. 2015;3(4).

192. Chandler M, Fayet O. Translational frameshifting in the control of transposition in bacteria. Mol Microbiol. 1993;7(4):497-503.

193. Rousseau P, Gueguen E, Duval-Valentin G, Chandler M. The helix-turn-helix motif of bacterial insertion sequence IS911 transposase is required for DNA binding. Nucleic Acids Res. 2004;32(4):1335-44.

194. Haren L, Normand C, Polard P, Alazard R, Chandler M. IS911 transposition is regulated by protein-protein interactions via a leucine zipper motif. J Mol Biol. 2000;296(3):757-68.

195. Khan E. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. Nucleic Acid research. 1991;19:851-60.

196. Haren L. Integrating DNA: Transposases and Retroviral Integrases. Annual Review of Microbiology. 1999;53:245-81.

197. Sekine Y. Linearization and Transposition of Circular Molecules of Insertion Sequence IS3. J Mol Biol. 1999;294:21-34.

198. Sharma V, Firth AE, Antonov I, Fayet O, Atkins JF, Borodovsky M, et al. A pilot study of bacterial genes with disrupted ORFs reveals a surprising profusion of protein sequence recoding mediated by ribosomal frameshifting and transcriptional realignment. Mol Biol Evol. 2011;28(11):3195-211.

199. Sharma V, Prere MF, Canal I, Firth AE, Atkins JF, Baranov PV, et al. Analysis of tetra- and hepta-nucleotides motifs promoting -1 ribosomal frameshifting in Escherichia coli. Nucleic Acids Res. 2014;42(11):7210-25.

200. Vögele K. High-level ribosomal frameshifting directs the synthesis of IS150 gene products. Nucleic Acid research. 1991;19:4377-85.

201. Mahillon J. Insertion Sequences. Microbiology and Molecular Biology Reviews. 1998;62(3):725-74.

202. Zhang Z, Saier MH, Jr. A mechanism of transposon-mediated directed mutation. Mol Microbiol. 2009;74(1):29-43.

203. Zhang Z, Saier MH, Jr. A novel mechanism of transposon-mediated gene activation. PLoS Genet. 2009;5(10):e1000689.

204. Schnetz K. IS5: A mobile enhancer of transcription in Escherichia coli Proc Natl Acad Sci U S A. 1992;89:1244-8.

205. Humayun MZ, Zhang Z, Butcher AM, Moshayedi A, Saier MH, Jr. Hopping into a hot seat: Role of DNA structural features on IS5-mediated gene activation and inactivation under stress. PLoS One. 2017;12(6):e0180156.

206. Fye RM. Exact method for numerically analyzing a model of local denaturation in superhelically stressed DNA. Physical Review E. 1999;59(3).

207. Sheridan SD, Benham CJ, Hatfield GW. Activation of gene expression by a novel DNA structural transmission mechanism that requires supercoiling-induced DNA duplex destabilization in an upstream activating sequence. J Biol Chem. 1998;273(33):21298-308.