Identification of microbial amino acid producers by biosensor-based high-throughput screenings and comparative genome analyses

Inaugural Dissertation

submitted to

the Faculty of Mathematics and Natural Sciences

of the Heinrich-Heine-University Düsseldorf

presented by

Philipp Tobias Baumann

born in Kempten (Allgäu)

Jülich, April 2023

The laboratory work conducted for this thesis in hand has been performed at the Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, from April 2018 until March 2022 under the supervision of Prof. Dr. Michael Bott and Prof. Dr. Jan Marienhagen.

Printed with the permission of

the Faculty of Mathematics and Natural Sciences

of the Heinrich-Heine-University Düsseldorf

Examiner: **Prof. Dr. Michael Bott**

                                     Institute of Bio- and Geosciences, IBG-1: Biotechnology

                                     Forschungszentrum Jülich GmbH

Co-examiner: **Prof. Dr. Jörg Pietruszka**

                                     Institute of Bioorganic Chemistry, IBOC

                                     Heinrich-Heine-University Düsseldorf

                                     at Forschungszentrum Jülich GmbH

Date of oral examination: 10.10.2023

"Don't wish it was easier, wish you were better. Don't wish for less problems, wish for more skills. Don't wish for less challenge, wish for more wisdom."

"How long should you try? Until."

In Gedenken an **Jim Rohn** (1930 – 2009), US-amerikanischer Unternehmer, Autor und Motivationstrainer – meinem größten philosophischen Einfluss während dieser Arbeit.

Results presented in this thesis have been published / are about to be published in:

Della Corte D., van Beek H.L., Syberg F., Schallmey M., Tobola F., Cormann K.U., Schlicker C., Baumann P.T., Krumbach K., Sokolowsky S., Morris C.J., Grünberger A., Hofmann E., Schröder G.F., Marienhagen J. (2020). Engineering and application of a biosensor with focused ligand specificity. *Nature Communications* 11: 4851. doi: 10.1038/s41467-020-18400-0.

My contribution to this publication included conducting the laboratory experiments to validate the LysG-A219L biosensor, pSenHis. This included undirected mutagenesis of biosensor-carrying *C. glutamicum* wild-type strains and subsequent FACS-screenings to isolate strain variants with increased fluorescence. Subsequently, I analyzed these variants with regard to their product formation and investigated induced mutations in central genes of histidine metabolism. I also contributed to the writing and proofreading of the corresponding part of the publication. Own contribution (after consultation with supervisor): approx. 15-20 %.

Baumann P.T., Dal Molin M., Aring H., Krumbach K., Müller M.-F., Vroling B., van Summeren-Wesenhagen P.V., Noack S., Marienhagen J. (2023). Beyond rational—biosensor-guided isolation of 100 independently evolved bacterial strain variants and comparative analysis of their genomes. *BMC Biology* 21: 183. doi: 10.1186/s12915-023-01688-x.

My contribution to this publication included the design and implementation of almost all laboratory experiments. My contribution did not include the development of the genome analysis software (Dr. Michael Dal Molin), the construction of some strain variants (Hannah Aring), the computer-assisted analysis of the induced mutations (Dr. Stephan Noack) and the implementation of the bioreactor experiments (Moritz-Fabian Müller). I also completely wrote and corrected the transcript of the publication. Own contribution (after consultation with supervisor): approx. 85-90 %.

Table of contents

Abstract

The modern life style of our ever growing world population is mainly based on the consumption of fossil resources. These developments are the cause of environmental damage and uncontrolled release of greenhouse gases as main drivers of climate change. Therefore, the implementation of more sustainable biotechnological production processes at industrial scale is needed. Such (mostly) sugar-based bioprocesses of the industrial biotechnology rely on microorganisms as catalysts, which need to be constantly improved to become more resource-efficient. In fact, most microbial production strains used today, operate below their theoretical maximum in terms of product yield. However, increasing the production performance of already highly engineered industrial production strains poses a challenge, since novel beneficial targets for rational engineering are difficult to identify. With the availability of biosensor-based FACS-screening technologies, classical strain development strategies comprising random mutagenesis and screening for an improved production phenotype are coming back into fashion.

In this thesis, the biosensor-based FACS-screening of randomly mutated cells was combined with comparative genome analyses and reverse engineering to reveal novel targets in the genome of an industrial *Corynebacterium glutamicum* production strain for the proteinogenic amino acid L-histidine. Since the starting strain was already highly engineered, novel beneficial mutations were expected to be difficult to identify as they might not be directly linkable to L-histidine biosynthesis. Therefore, 100 independently FACS-isolated improved strain variants were subjected to a comparative genome analyses to look for reoccurring mutations in single genes, certain pathways or modules of the microbial metabolism.

To achieve the independent isolation of 100 improved strain variants, > 600 chemical mutagenesis and > 200 biosensor-based FACS-screenings were performed, which allowed for the isolation of $> 50,000$ variants with increased fluorescence. The characterization of $> 4,500$ variants with regard to biomass formation and L-histidine production, yielded 100 improved strain variants, accumulating 10-80 % more L-histidine in comparison to the starting variant. Comparative analyses of their genomes and reconstruction of point mutations/introduction of gene deletions allowed for the identification of six novel targets with a positive impact on L-histidine accumulation in cultures of *C. glutamicum*. In this context, combination of four genome modifications resulted in an improved L-histidine production strain, which was characterized by a doubled product titer (29 mM) and a doubled product yield (0.13 mol L-histidine mol D-glucose^{-1}) in lab-scale batch-mode bioreactor fermentations.

This approach holds the promise to identify novel genomic targets in already highly engineered production strains in a more systematic manner within a very limited time-frame, and might help to push the performance of industrial bioprocesses quickly towards maximum yields.

Zusammenfassung

Die ständig wachsende Weltbevölkerung und ihr Konsumverhalten, das zu einem großen Teil auf fossilen Ressourcen basiert, verursachen Umweltschäden und treiben die Produktion von Treibhausgasen als Hauptverursacher der Klimaveränderung an. Diese Entwicklung erfordert die Implementierung ressourcenschonender und umweltfreundlicher Produktionsverfahren im industriellen Maßstab. Daher müssen auch die (meist) zuckerbasierten Bioprozesse der industriellen Biotechnologie und ihre mikrobiellen Katalysatoren ständig verbessert werden, um ressourceneffizienter zu werden. Tatsächlich arbeiten die meisten der heute verwendeten mikrobiellen Produktionsstämme hinsichtlich der Produktausbeute unterhalb des theoretischen Maximums. Die Steigerung der Produktionsleistung bereits hochentwickelter industrieller Produktionsstämme stellt jedoch eine Herausforderung dar, da neue genomische Ziele für das rationale Engineering schwer zu identifizieren sind. Mit der Verfügbarkeit biosensorbasierter FACS-Screening-Technologien gewinnen die klassischen Stammentwicklungsverfahren mit Zufallsmutagenese und Screening auf einen verbesserten Produktionsphänotyp wieder an Bedeutung.

In dieser Dissertation wurden Zufallsmutagenesen und biosensorbasierte FACS-Screenings in Kombination mit vergleichenden Genomanalysen und genomischen Rekonstruktionen eingesetzt, um leistungssteigernde Modifikationen im Genom eines industriellen *Corynebacterium glutamicum* Stammes für die Produktion der Aminosäure L-Histidin zu identifizieren. Der Ausgangsstamm war bereits stark genetisch modifiziert, sodass die Identifizierung neuer Modifikationen eine Herausforderung darstellte, weil diese nicht rational mit dem L-Histidin-Biosyntheseweg zusammenhängen müssen. Aus diesem Grunde wurden 100 unabhängige, jeweils FACS-isolierte, verbesserte Stammvarianten einer vergleichenden Genomanalyse unterzogen, um nach einer Häufung von Mutationen in einzelnen Genen, bestimmten Stoffwechselwegen oder Modulen des mikrobiellen Stoffwechsels zu suchen, die letztendlich für die beobachtete Steigerung der L-Histidin-Produktion verantwortlich sind.

Zur unabhängigen Isolierung der 100 Stammvarianten wurden über 600 chemische Mutagenesen und über 200 FACS-Screenings durchgeführt, durch welche über 50.000 Varianten mit erhöhter Fluoreszenz isoliert werden konnten. Im Rahmen einer detaillierten Charakterisierung von über 4.500 Varianten hinsichtlich ihrer Biomasse- und Produktbildung, wurden 100 Varianten mit 10-80 % gesteigerter L-Histidin-Produktion identifiziert. Mithilfe vergleichender Genomanalysen und genetischer Rekonstruktionen wurden sechs bisher unbekannte genomische Modifikationen mit positivem Einfluss auf die L-Histidin-Produktion entdeckt. Durch Kombination der Modifikationen wurde ein L-Histidin-Produktionsstamm generiert, welcher sich in einem Batch-Fermentationsprozess im Labormaßstab durch einen verdoppelten Produkttiter (29 mM L-Histidin) und eine verdoppelte Produktausbeute (0,13 mol L-Histidine mol D-Glukose^{-1}) auszeichnete.

Mit diesem neuen Ansatz können neuartige genetische Modifikationen in bereits hochentwickelten Produktionsstämmen systematischer und im Hochdurchsatzverfahren identifiziert werden. Dies birgt ein hohes Potenzial, um die Entwicklung industrieller Bioprozesse schnell in Richtung einer maximalen Ausbeute voranzutreiben.

Abbreviations

ADP	Adenosine diphosphate
AICAR	1-(5′-phosphoribosyl)-5-amino-4-imidazolecarboxamide
ALE	Adaptive laboratory evolution
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
BHI	Brain-Heart-Infusion medium
Bp	Basepair
CDW	Cell dry weight
CGXII	Defined mineral medium for cultivation of *C. glutamicum*
DMSO	Dimethyl sulfoxide
DNA	Desoxyribonucleic acid
EYFP	Enhanced yellow fluorescent protein
FAAMS	Fast Automated Analysis of Multiple Sequences
FACS	Fluorescence-activated cell sorting
FRET	Förster resonance energy transfer
FSC/FSC-H/FSC-W	Front scatter / front scatter height / front scatter width
fTHF	Formylated tetrahydrofuran
GRAS	Generally recognized as safe
HPLC	High Performance Liquid Chromatography
ITC	Isothermal titration calorimetry
LTTR	LysR-type transcriptional regulator
MNNG	Methylnitronitrosoguanidine
mTHF	Methylated tetrahydrofuran
NADH	Nicotinamide adenine dinucleotide
NCBI	National Center for Biotechnology Information

NGS	Next Generation Sequencing
NRPS	Non-ribosomal peptide synthetase
OD_{600}	Optical density measured at a wavelength of 600 nm
PPP	Pentose phosphate pathway
PRPP	Phosphoribosyl pyrophosphate
PTS	Phosphotransferase system
RNA	Ribonucleic acid
SHMT	Serine hydroxymethyltransferase
SNP	Single nucleotide polymorphism
SSC/SSC-H/SSC-W	Side scatter / side scatter height / side scatter width
TCA	Tricarboxylic acid cycle
UV	Ultraviolet radiation

1 Introduction

The global society is facing increasing challenges caused by effects of a growing world population, increased consumption, and growing industrialization of emerging countries (IPCC, 2018). Hence, increasing demand for fossil resources for energy generation and production of petroleum-derived commodities is causing environmental damage and increased greenhouse gas emissions, which are driving global climate change (Jorgenson *et al.*, 2019). In addition, the availability of arable land for production of food, feed and other agricultural feedstocks becomes limited and could become even more scarce due to global warming and land degradation (OECD, 2011; Prăvălie *et al.*, 2021). Recent geopolitical and pandemic crises exacerbate the global situation with reduced supplies of oil, gas and coal as well as agricultural products and impaired supply chains on the world market. Therefore, the global society needs to develop and implement new strategies to reduce greenhouse gas emissions, environmental damage and to limit global climate change, whereas at the same time global production efficiencies of required commodities need to be increased in a sustainable way.

In this context, Industrial Biotechnology can play a central role (OECD, 2011). With the aim of *"transitioning from a petroleum-based economy to a new bio-based economy, renewable biological resources, waste streams and even CO_2 shall be converted by engineered microorganisms into the energy, chemicals, materials, food and medicines that drive the world economy"* (Montaño López, Duran and Avalos, 2021). This is often considered as one of the most promising approaches for pollution prevention, resource conservation, cost reduction and more sustainable production processes (Kiss, Grievink and Rito-Palomares, 2015). Its scope of application includes chemical, food and feed, pharmaceutical, detergent, textile, bioplastics and bioenergy industries. From the economic perspective, the global market of biotechnological products is expected to further increase by 10-15 % annually (IMARC Group, 2020; Mordor Intelligence, 2020).

Microorganisms, however, are very economic with their resources and possess tight regulation on genetic and metabolic levels to ensure maximum growth and avoid waste of carbon or energy amidst ever-changing environmental conditions (Parekh, Vinci and Strobel, 2000; Nielsen and Keasling, 2016). Hence, most biotechnologically desired products are not overproduced by microorganisms naturally, or are not synthesized at industrially relevant levels. As a consequence, microbial production strains have to be genetically engineered to push their metabolism towards overproduction of a certain chemical compound of interest (Bailey, 1991). However, this process is very time-consuming and in light of the high demand for environmentally-friendly biotechnological production processes, more high-throughput

engineering approaches are needed to obtain microbial high-performance production strains faster (Nielsen and Keasling, 2016).

1.1   Microbial strain engineering for industrial use

The first application of microorganisms in winemaking and beer brewing goes back to 7,000 BC – long before their discovery (Demain *et al.*, 2016). In 1857, Louis Pasteur, for the first time postulated fermentation to be a living process of yeast. After a series of other discoveries in the interim, the discovery of penicillin in 1927 by Alexander Fleming led to rapid expansion of the new industrial biotechnology to produce antibiotics as the first chemotherapeutic drugs for use in World War II. In the 1940s, significant advancements were made in fermentation technology for penicillin production, which led to the first industrial fed-batch fermentation process (Demain *et al.*, 2016). Meanwhile, the newly developed "strain improvement" technology made extensive use of random X-ray and UV-radiation mutagenesis (Raper, 1946). Despite huge efforts in clone selection, this technology enabled the isolation of improved variants, even though cellular metabolism and its complex regulation were neither measurable nor controllable and thus, yet poorly understood. Later, chemical mutagens such as *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG) were found highly effective as well (Haerlin, Süssmuth and Lingens, 1970). Hence, combined random mutagenesis, selection and screening procedures were thoroughly applied to improve industrial strains (Rowlands, 1984). Referred to as classical microbial strain engineering, this technology shares a long history of engineering success (Parekh, Vinci and Strobel, 2000). In the 1970s, DNA sequencing and thus the identification of induced mutations became attainable (Sanger, Nicklen and Coulson, 1977). The identification of the individual beneficial mutations, however, remained an unsolved challenge due to the very many mutations induced and yet mostly unknown genotype-phenotype relationships.

The advent of recombinant DNA technology led to the development of targeted metabolic engineering techniques in the 1990s (Bailey, 1991; Stephanopoulos, Aristidou and Nielsen, 1998; Nielsen, 2001). This targeted manipulation of cellular metabolism enabled new insights and the investigation of single engineering targets without the high background of neutral or detrimental mutations acquired from random mutagenesis. For production purposes, besides the introduction of heterologous genes or pathways, metabolic engineering focused on the attenuation of flux-limiting steps by overexpression of respective genes and deregulation of feedback inhibitions in order to redirect carbon flux towards the desired metabolite (Bailey, 1991; Sahm *et al.*, 1995; Bailey *et al.*, 1996; Sahm, Eggeling and de Graaf, 2000). Bailey *et al.* referred to this way of posing the engineering problem as *constructive metabolic engineering* (Bailey *et al.*, 1996). With the broadening of substrate spectrum, elimination of

by-product formation and increase of export activity also other metabolic modules soon became central engineering goals (Sang Yup Lee and E. Terry Papoutsakis, 1999; Nielsen, 2001; Keasling, 2010). For the targeted implementation of genetic alterations in microbial hosts, homologous recombination was (and still is) used as one of the most important methods (Court, Sawitzke and Thomason, 2003; Kirchner and Tauch, 2003). For the identification of promising engineering targets, Metabolic Flux Analysis as well as genome-scale mathematic models for quantification and perturbation of intracellular fluxes were developed and expanded the knowledge on cellular metabolism (Wiechert, 2001, 2002; Toya and Shimizu, 2013). Today, highly data-driven metabolomics and automated machine learning approaches are becoming essential tools in metabolic engineering (Teleki and Takors, 2019; Kim *et al.*, 2020; Lawson *et al.*, 2021). Despite recent advancements and significant academic successes in the field, *constructive metabolic engineering* faces some inherent challenges (Lee and Kim, 2015; Nielsen and Keasling, 2016). Due to the complex interconnections of metabolic, regulatory and signaling networks, *in vivo* effects of rational genetic alterations are still difficult to predict (Kaczmarek and Prather, 2021). Aiding genome-scale mathematical models are improving but may still work inaccurately due to the limited input of global regulatory circuits (Lee and Kim, 2015; Fang, Lloyd and Palsson, 2020). In this context, targeted metabolomics can help to provide quantitative data on metabolite pools but such studies are typically laborious and cost-intensive (Teleki and Takors, 2019; Feith *et al.*, 2020). With increasing degree of engineering, fine-tuning of pathway expression is becoming a predominant issue (Santos and Stephanopoulos, 2008; Hwang, Lee and Lee, 2018; Xu *et al.*, 2021). To test a large number of experimental combinations, several multiplexed genetic engineering techniques were developed (Wang *et al.*, 2009; Cong *et al.*, 2013; Li *et al.*, 2019). Robot-aided automated strain construction platforms, however, are still in early development stages (Tenhaef *et al.*, 2021). Hence, with the methods applied in *constructive metabolic engineering*, only genetic targets with rational connection to the product metabolite or ones included in a genome-scale metabolic model can be considered for perturbation predictions and subsequent strain engineering.

In contrast to rational metabolic engineering, mainly concerned with the precise introduction of a limited number of genetic modifications, classical strain engineering typically yields a large number of unknown neutral or detrimental mutations acquired in the random mutagenesis step. As a result, improved production strains obtained after screening also carry these undesired mutations, potentially affecting growth and robustness (Ohnishi *et al.*, 2002; Ikeda *et al.*, 2009; Warner, Patnaik and Gill, 2009). Thus, to alleviate the drawbacks of each strategy, combined approaches of classical and rational engineering were investigated already in the 1990s (Bailey *et al.*, 1996). At that time, *constructive metabolic engineering*

was hampered by limited knowledge on metabolic pathways and moreover regulatory systems along with infinite genetic perturbation possibilities, such that metabolic consequences differed substantially to desired outcomes in many cases (Bailey, 1991). Therefore, an alternative approach termed *inverse metabolic engineering* was proposed (Bailey *et al.*, 1996). Here, a desired phenotype is identified first (by classical methods), then the causative genetic basis is determined in the second step, and finally the phenotype-conferring genetic modification is transferred to an industrial organism by targeted genome engineering (Bailey *et al.*, 1996; Warner, Patnaik and Gill, 2009). This laid the foundation to later reverse engineering, which promised more efficient industrial production strains with faster growth, greater robustness and hence better reproducibility (Ohnishi *et al.*, 2002; Ikeda *et al.*, 2009; Warner, Patnaik and Gill, 2009).

Many of today's industrially used production strains are still product of random approaches, or a combination of classical and rational strain engineering, with often still unknown genotype-phenotype relationships (Ohnishi *et al.*, 2002; Becker and Wittmann, 2012; Kulis-Horn, Persicke and Kalinowski, 2014; Eggeling and Bott, 2015). More importantly, many of these industrial strains still operate far from theoretical maximum yields (Takors *et al.*, 2007; Blombach *et al.*, 2008; Eggeling and Bott, 2015; Schwentner *et al.*, 2019). Hence, the identification of beneficial mutations for reverse engineering is key to drive industrial strains towards their maximum yields. Comparative genome analysis can aid the identification process by revealing clusters of mutations that are likely to contribute to improved phenotypes. However, this was only applied to a small number of isolated variants, by which beneficial mutations were identified in known key pathways, such as L-lysine, L-arginine and L-citrulline production, or glycerol degradation (Ohnishi *et al.*, 2002; Herring *et al.*, 2006; Ikeda *et al.*, 2009). Recent advancements in high-throughput cultivation systems as well as Next Generation Sequencing have improved speed, throughput and overall quality of the resulting sequence while reducing the overall costs for genetic analyses (Long *et al.*, 2014; Goodwin, McPherson and McCombie, 2016). With this technology, comparative genome analyses of a significantly larger numbers of variants can now be performed. In such a large set of genome sequences, beneficial mutations could be identified with higher accuracy and, more importantly, potentially at unknown genomic locations. Thus, with the combined power of classical approaches, inducing random beneficial mutations, and targeted approaches, transferring beneficial mutations to producer strains, superior industrial production strains could be generated.

1.2 Using biosensors for high-throughput screenings

"*Cells are filled with biosensors, molecular systems that measure the state of the cell and respond by regulating host processes. In much the same way that an engineer would monitor a chemical reactor, the cell uses these sensors to monitor changing intracellular environments and produce consistent behavior despite the variable environment*" (Michener *et al.*, 2012).

The number of enzyme or strain variants that can be generated using recent targeted or random genetic engineering methods greatly exceeds the throughput of standard cultivation and analysis methods, typically requiring chromatographic or mass spectrometric analysis (Bott, 2015). These labor-intensive and repetitive procedures usually limit the screening process of large genetically diverse mutant libraries for improved variants, if the desired product has no intrinsic properties, such as formation of color, fluorescence or conferring of a resistance, all of which can be readily measured (Parekh, Vinci and Strobel, 2000; Schallmey, Frunzke and Eggeling, 2014; Eggeling, Bott and Marienhagen, 2015). To tackle the limitations in screening-throughput, genetically-encoded natural sensing units are used to build biosensors, which translate intracellular signals or product concentrations into a graded, measurable output signal, enabling *in vivo* monitoring of cellular metabolism (Michener *et al.*, 2012; Zhang, Jensen and Keasling, 2015). Their applicability in research or industrial screening setups is mainly determined by the operational range (range of detectable inducer concentrations) and dynamic range (fluorescence intensity between ON and OFF state). In general, three types of genetically encoded biosensors have been developed.

Förster resonance energy transfer (FRET)-based biosensors typically consist of two fluorescent proteins with overlapping emission and excitation spectra of the FRET donor and acceptor, respectively, fused to a ligand-binding peptide via linker sequences. Upon ligand binding, the peptide undergoes a conformational change such that FRET occurs when donor and acceptor fluorophores get in close proximity below 10 nm distance (Zhang, Jensen and Keasling, 2015). Besides conformational change, other modes of action such as cleavage are employed as well (Hochreiter, Garcia and Schmid, 2015). The alterations between donor and acceptor emission can be measured as FRET biosensor signal and allow for time-resolved monitoring of respective metabolite concentrations or other cellular processes. Although characterized by high temporal resolution and orthogonality, FRET-based biosensors are not used for high-throughput screenings due to their low dynamic range and only nanomolar to micromolar operational range (Schallmey, Frunzke and Eggeling, 2014; Zhang, Jensen and Keasling, 2015). Instead, they are applied for measuring intracellular dynamics of carboxylic acids, amino acids, sugar phosphates, nucleotides as well as ions

and redox status (Vinkenborg *et al.*, 2009; Oku *et al.*, 2013; Mohsin and Ahmad, 2014; San Martín *et al.*, 2014; Tang *et al.*, 2014; Peroza *et al.*, 2015; Yoshida *et al.*, 2019).

RNA-aptamer- or riboswitch-based biosensors use (m)RNA motifs, which undergo structural change upon ligand binding, thereby altering the expression of their downstream reporter gene (Michener *et al.*, 2012; Schallmey, Frunzke and Eggeling, 2014). In addition to regulation of gene transcription or translation, alterations of RNA stability or ribozyme activity can be employed as modes of action as well (Serganov and Nudler, 2013). Thus, riboswitch biosensors can be used to dynamically regulate genes or operons and monitor metabolite concentrations and have been used for high-throughput screenings (Michener and Smolke, 2012; Hwang *et al.*, 2021; Liu *et al.*, 2021). Riboswitches not only provide fast response times, but can also be engineered towards synthetic or *de novo* affinities in a more targeted manner (Zhang, Jensen and Keasling, 2015). Many of those synthetic designs, however, perform poorly *in vivo* (Schallmey, Frunzke and Eggeling, 2014; Berens and Suess, 2015).

Transcriptional biosensors exploit the natural diversity of native regulatory circuits using metabolite-responsive transcription factors and their activating or repressing function on gene transcription (Liu, Evans and Zhang, 2015). Genetically encoded, they typically comprise the transcription factor gene in inverse orientation to the target gene promotor of the transcription factor, which controls the expression of a fluorescent reporter gene (Sonntag *et al.*, 2020). Thereby the respective intracellular metabolite concentration is translated into a graded fluorescence output. Due to the modular design, transcriptional biosensors are straightforward in construction, however, low orthogonality limits their application in other microbial hosts (Zhang, Jensen and Keasling, 2015; Sonntag *et al.*, 2020). Besides metabolite-responsive transcription factors, two component systems and stress-response mechanisms were employed to construct transcriptional biosensors as well (Liu, Evans and Zhang, 2015).

Prior to their application in microbial strain engineering, transcription-based whole-cell biosensors were built for detection of environmental chemicals (Harms, Wells and Van Der Meer, 2006; Fernandez-López *et al.*, 2015). Today, transcriptional biosensors are not only widely used for screening purposes, but also for dynamic modulation of metabolic pathways, identification of optimal production conditions, fine-tuning of pathway expression and investigations of single-cell heterogeneity (Zhang, Carothers and Keasling, 2012; Dahl *et al.*, 2013; Dietrich *et al.*, 2013; Mustafi *et al.*, 2014; Eggeling, Bott and Marienhagen, 2015; Rogers and Church, 2016; Doong, Gupta and Prather, 2018; Koch *et al.*, 2019; Kaczmarek and Prather, 2021).

For high-throughput screening of strain or enzyme libraries, fluorescence-activated cell sorting (FACS) has emerged as valuable tool in addition to screenings using well plates, agar plates or selection-based methods (Eggeling, Bott and Marienhagen, 2015; Kaczmarek and Prather, 2021). Before the first biosensors were employed in screenings, FACS was already used to isolate metabolite-overproducing strain variants of fluorescent products such as the carotenoid astaxanthin (An *et al.*, 1991). Noteworthy, up to 80,000 cells s^{-1} can be analyzed by FACS at the single-cell level. Reliable cell sorting can be realized with a throughput of 10,000 cells s^{-1}. Secondary screens are often applied as repeated FACS or chromatography-based analysis schemes to distinguish true positives from false positives. Hence, FACS is used as an effective high-throughput enrichment strategy (Kaczmarek and Prather, 2021).

The first application of transcriptional biosensors to FACS-based high-throughput screening in order to isolate improved metabolite producers was performed with the pSenLys biosensor (Binder *et al.*, 2012). This biosensor originated from the genomic *lysEG* locus of *Corynebacterium glutamicum,* using the LTTR-type transcriptional activator gene *lysG,* its promoter and target promoter area including a fragment of the native target gene *lysE* (*lysE'*) for its construction (Figure 1). A transcriptional fusion of the *eyfp* reporter gene, coding for enhanced yellow fluorescent protein EYFP, with *lysE'* ensured the expression of *eyfp* under control of LysG by maintaining its native transcription initiation mechanisms (Bellmann *et al.*, 2001; Binder *et al.*, 2012). Consequently, a LysE' polypeptide and EYFP are translated separately. Thereby, the native regulatory circuit of LysG was exploited to drive the expression of *eyfp*. Hence, intracellular concentrations of the three basic amino acids L-lysine, L-arginine and L-histidine can be translated into a graded, fluorescence output signal (Figure 1).
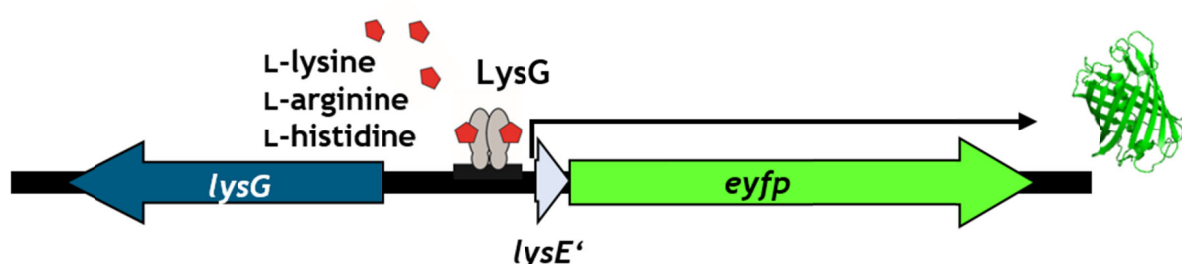
Figure 1: Genetically encoded biosensor unit of pSenLys, adopted from the *Corynebacterium glutamicum lysEG* locus. The homotetrameric transcriptional regulator LysG is responsive to the basic amino acids L-lysine, L-arginine and L-histidine. Its encoding gene *lysG* is divergently transcribed from its target promoter, which controls expression of *lysE* in its native genetic setup. LysE is a basic amino acid exporter which transports L-lysine and L-arginine, but not L-histidine (Bellmann *et al.*, 2001). Upon DNA and ligand binding, LysG undergoes a conformational change by which *lysE* expression is induced. In pSenLys, the gene encoding enhanced Yellow Fluorescent Protein (*eyfp*) is fused to the N-terminal *lysE* fragment (*lysE'*) by transcriptional fusion to maintain native transcription initiation mechanisms of LysG. With this biosensor, the intracellular concentrations of the three basic amino acids can be translated into a graded fluorescent output by *eyfp* expression.

The pSenLys biosensor was used for FACS-based high-throughput screenings of randomly mutated *C. glutamicum* wild type cultures. A majority of isolated single cells with increased fluorescence produced L-lysine (69 %), whereas the *C. glutamicum* wild type excreted no L-lysine. Some variants additionally secreted L-arginine, however, no L-histidine producer was identified in this screening campaign. It was speculated that this was due to the length and tight regulation of the L-histidine biosynthesis pathway (Binder *et al.*, 2012). Targeted sequencing resulted in identification of potential beneficial mutations in known key enzymes of L-lysine biosynthesis. From whole genome sequencing of one isolated strain variant and a subsequent comparative analysis with the wild-type sequence identified an hitherto unknown beneficial mutation in *murE*, coding for UDP-*N*-acetylmuramoylalanyl-D-glutamate-2,6-diaminopimelate ligase – a gene with a rational link to L-lysine biosynthesis.

The application of transcriptional biosensors to improve metabolite production to industrially relevant levels can cause some biosensor-, host-, or screening-specific challenges. Regarding the screening host, biosensor crosstalk can lead to isolation of false positives, if non-producing cells import the respective metabolite and hence can adopt the fluorescent phenotype of overproducing cells (Kaczmarek and Prather, 2021). This was recently addressed by optimization of expression and cultivation conditions (Flachbart, Sokolowsky and Marienhagen, 2019). Since transcriptional regulatory circuits have evolved naturally for physiological metabolite concentrations, their application as biosensors may result in operational ranges not suitable for the production capacity of industrial metabolite producers. In this case, modifications of biosensor units, such as the expression level of the transcriptional regulator, might become essential (Sonntag *et al.*, 2020). In addition, the ligand specificity of transcriptional regulators can limit their application in biosensor-based screenings. In such cases, respective regulators can be engineered towards new specificities or more focused ligand spectra (Della Corte *et al.*, 2020; Flachbart *et al.*, 2021).

The use of randomly mutated (*C. glutamicum)* wild type cultures for high-throughput screenings carries the risk to reveal mainly known beneficial mutations (Binder *et al.*, 2012). To reveal novel beneficial mutations many studies, however, still focused on the analysis of the respective synthesis pathways or connected pathways of the desired metabolites, thereby omitting the potential of a more systematic genome-wide evaluation of the induced mutations (Ohnishi *et al.*, 2002; Ikeda *et al.*, 2009; Binder *et al.*, 2012; Mahr *et al.*, 2015; Zhang *et al.*, 2018; Liu *et al.*, 2021). Working with industrial production strains instead could leverage their genomic modifications to identify unknown genome-wide beneficial mutations that contribute to a further improvement of production performance – beyond rational engineering targets and towards theoretical maximum yields. One of the most industrially

used production hosts in microbial biotechnology, especially for amino acid production, is *Corynebacterium glutamicum*.

1.3 *Corynebacterium glutamicum* as amino acid production host

C. glutamicum was isolated 1956 as part of a screening campaign for naturally L-glutamate-producing microorganisms and was soon used for large-scale fermentative production of this flavor enhancing amino acid (Kinoshita, Udaka and Shimono, 1957). *C. glutamicum* was characterized as a Gram-positive, facultative anaerobic, non-motile, biotin-auxotrophic and non-pathogenic actinobacterium (Eggeling and Bott, 2005; Tatsumi and Massyuki, 2012). Since then, the cellular biology and metabolism of this bacterium have been studied in detail, and numerous techniques have been developed for its genetic manipulation and its cultivation from laboratory to industrial-scale (Eggeling and Bott, 2005; Tatsumi and Massyuki, 2012). The publication of its genome sequence in 2003 further boosted strain engineering and underlined importance of *C. glutamicum* for industrial biotechnology (Ikeda and Nakagawa, 2003; Kalinowski *et al.*, 2003).

As a member of the *Corynebacterianeae*, *C. glutamicum* possesses a characteristic cell wall structure. In addition to the cytoplasmic membrane, lipomannan, lipoarabinomannan, peptidoglycan, arabinogalactan, and mycolic acid layers form the outer cell wall (Eggeling and Sahm, 2001; Houssin *et al.*, 2020). Mycolic acids derived from fatty acid synthase pathway are important for the efflux properties of the cell, since alterations in mycolic acid content have been shown to influence secretion of amino acids such as L-glutamate or L-lysine (Eggeling and Sahm, 2001; Gebhardt *et al.*, 2007a).

C. glutamicum offers unique advantages over other industrial production hosts such as its comparatively fast growth to high cell densities and its remarkable robustness against physical forces and oxygen and substrate supply oscillations in large-scale fermentations (Tatsumi and Massyuki, 2012; Grünberger *et al.*, 2013; Buchholz *et al.*, 2014; Käß *et al.*, 2014). In addition, products derived from *C. glutamicum* are generally recognized as safe (GRAS), since no endotoxins are produced by this organism (Taguchi *et al.*, 2015). Existing export systems and lack of degradation pathways for some proteinogenic amino acids render *C. glutamicum* a valuable host for microbial strain engineering and application in industrial amino acid production (Eggeling and Sahm, 2003; Tatsumi and Massyuki, 2012; Kulis-Horn, Persicke and Kalinowski, 2014; Eggeling and Bott, 2015).

L-glutamate and L-lysine represent main examples of industrial amino acid production, which accounted for 6 million tons produced in 2015 – almost exclusively by *C. glutamicum* (Eggeling and Bott, 2015; Lee and Wendisch, 2017). In 2021, the global amino acid market

reached a volume of 10.3 million tons and is expected to continue growing by 4.7 % annually (IMARC Group, 2021). In addition to L-glutamate and L-lysine, other amino acids such as L-threonine, L-phenylalanine, L-tryptophan, L-valine, L-alanine, L-cysteine, L-glutamine, L-isoleucine, L-leucine, L-methionine, L-proline, L-serine, L-tyrosine, L-arginine or L-histidine can be produced by classically and/or rationally evolved *C. glutamicum* strains (Becker and Wittmann, 2012; Tatsumi and Massyuki, 2012). Moreover, D-amino acids, vitamins, organic acids, nucleic acids, alcoholic bulk chemicals, pharmaceuticals and others can be produced by engineered *C. glutamicum* lab strains in commercially relevant amounts (Becker and Wittmann, 2012; Tatsumi and Massyuki, 2012; Lee and Kim, 2015).

The difficulty of engineering a specific metabolite production depends, among others, on the respective pathway, its regulations and interconnection to other pathways. In addition, suitable high-throughput screening systems for improvement by classical methods may not be readily available. As a result, many industrial production strains still produce far below their theoretical maximum yields (Takors *et al.*, 2007; Blombach *et al.*, 2008; Eggeling and Bott, 2015). Microbial L-histidine production represents a prominent example (Schwentner *et al.*, 2019).

1.4 L-histidine production - a sophisticated strain engineering project

L-histidine is one of the 20 standard proteinogenic amino acids and has aromatic properties due to its imidazole side chain, which confers its unique property of acting as both, proton donor and proton acceptor under physiological conditions (Figure 2) (Kulis-Horn, Persicke and Kalinowski, 2014). Hence, it is catalytically involved in many enzymatic reactions (Rebek, 1990; Polgár, 2005; Liao *et al.*, 2013). Beyond its use as feed and food supplement, it is used as antioxidant and anti-inflammatory agent in the pharmaceutical industry (Wade and Tucker, 1998; Waagboø *et al.*, 2010; Hasegawa *et al.*, 2012; Feng *et al.*, 2013). Its supply is realized by protein hydrolysis or microbial production in a range of million tons per year (Fior Markets, 2020). The global L-histidine market is expected to grow from 2021 to 2026 by 6 % annually, up to a market size of about 300 million USD (MarketWatch, 2021).
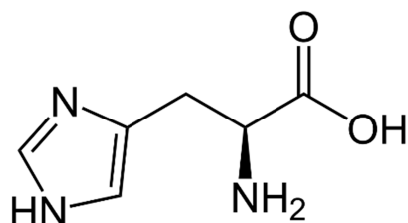

Figure 2: structural formula of L-histidine

L-histidine biosynthesis is an ancient metabolic pathway, which most likely evolved before the diversification of bacteria, archaea and eukaryotes (Fondi *et al.*, 2009). It has been

extensively studied in *Salmonella typhimurium* and *Escherichia coli* and served as a model pathway for a wide range of biochemical studies on operon structure, gene expression and gene regulation (Alifano *et al.*, 1996). The biosynthesis is an unbranched metabolic pathway involving ten enzymatic reactions, which begins with the condensation of phosphoribosyl pyrophosphate (PRPP) and ATP as part of the structural backbone (Figure 3).
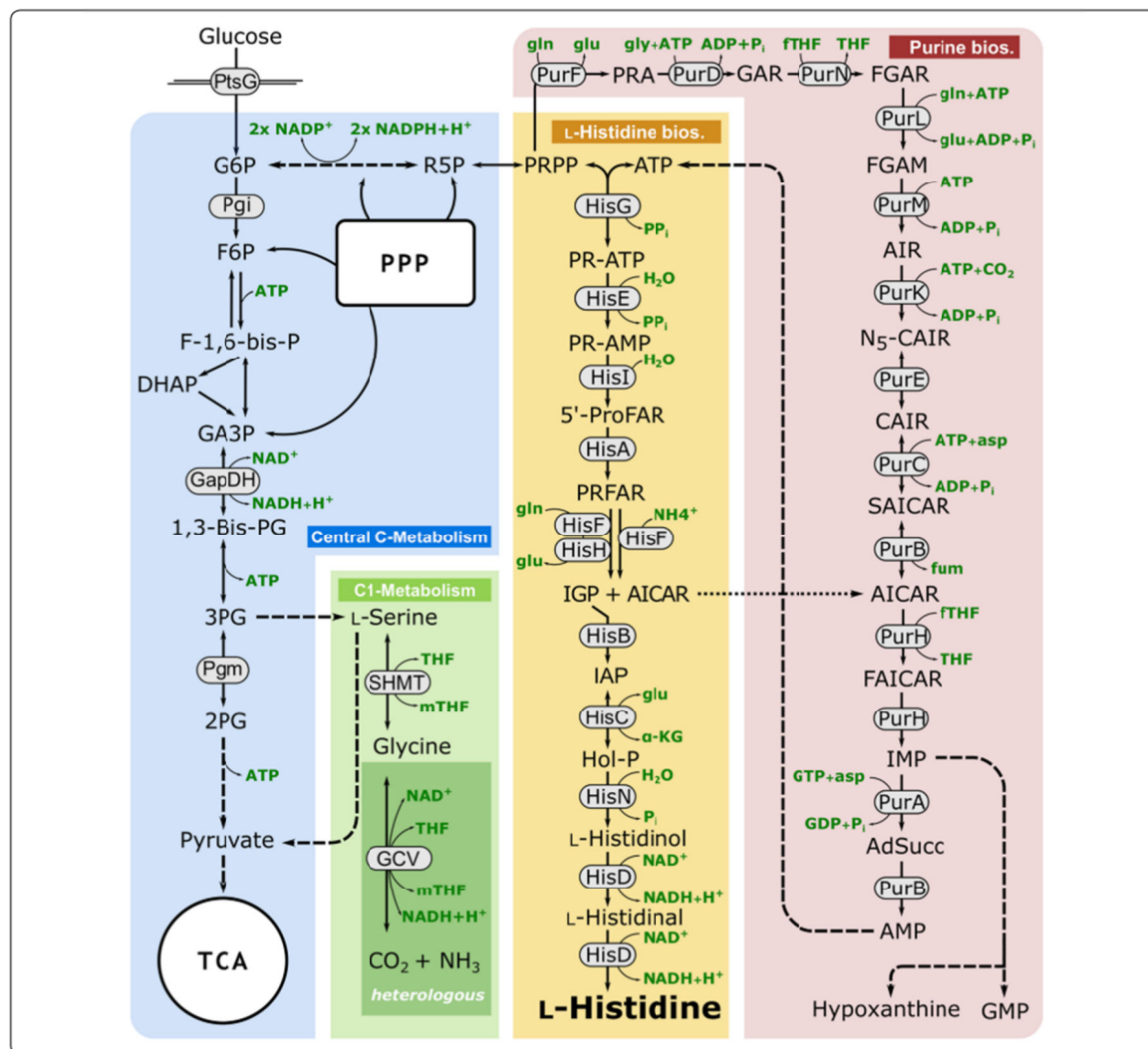
Figure 3: Scheme of L-histidine biosynthesis and connected pathways in *C. glutamicum* (Schwentner *et al.*, 2019)**.** Starting from D-glucose, relevant pathways for L-histidine biosynthesis are glycolysis, pentose phosphate pathway (PPP), C_1-metabolism and *de novo* purine biosynthesis. L-histidine and purine biosynthesis share the same precursor phosphoribosyl pyrophosphate (PRPP) and are strongly connected via ATP and AICAR (1-(5′-phosphoribosyl)-5-amino-4-imidazolecarboxamide). To metabolize by-product AICAR towards ATP, C_1 donors are provided by the endogenous serine hydroxymethyltransferase (SHMT) reaction. ATP represents a structural part of the L-histidine backbone.

Notably, PRPP is a shared precursor with other biosynthetic pathways of purines, pyrimidines, tryptophan and nicotinamide dinucleotides (Kulis-Horn, Persicke and Kalinowski, 2014). HisG, the first enzyme of the pathway, is subject to allosteric feedback inhibition by the end-product L-histidine and is thus the rate-limiting reaction of the pathway. HisG is

furthermore regulated by competitive inhibition by adenosine mono- and diphosphates (AMP, ADP) as well as their structural analog 1-(5′-phosphoribosyl)-5-amino-4-imidazolecarboxamide (AICAR) (Malykh *et al.*, 2018). AICAR is formed as by-product in the fifth step of the metabolic pathway and directed into the purine biosynthesis pathway to replenish nucleotides such as the ATP precursor. For conversion of AICAR, C_1 donors (mTHF, fTHF) are provided by the endogenous hydroxymethyltransferase (SHMT) reaction from L-serine (Schwentner *et al.*, 2019).

In *C. glutamicum* ATCC 13032, nine enzymes catalyze ten enzymatic reactions of the L-histidine pathway with HisD (histidinol dehydrogenase) serving as bifunctional enzyme (for oxidation of both L-histidinol to L-histidinal and finally L-histidine). The transcriptional organization is composed of four operons *hisD–hisC–hisB–cg2302– cg2301*, *hisH–hisA–impA–hisF–hisI–cg2294*, *cg0911–hisN*, and *hisE–hisG* (Kulis-Horn, Persicke and Kalinowski, 2014).

Due to the high ATP demand of 9.4 mol_{ATP} mol_{His}^{-1}, L-histidine biosynthesis is strictly regulated (Kulis-Horn, Persicke and Kalinowski, 2014). In addition to feedback regulation at the enzymatic level, a T-box mediated attenuation mechanism has been proposed for transcriptional regulation of the *hisDCB–cg2302–cg2301 (-hisHA–impA–hisFI–cg2294)* operon and a riboswitch mechanism has been discussed for translational control of *hisD* expression (Jung *et al.*, 2010; Kulis-Horn, Persicke and Kalinowski, 2014). Furthermore, the expression of the biosynthetic L-histidine genes is positively regulated by stringent response (Brockmann-Gretza and Kalinowski, 2006).

L-histidine import systems of *C. glutamicum* have been postulated but not identified yet, even though L-histidine uptake could be measured in histidine-auxotrophic mutants (Kulis-Horn, Persicke and Kalinowski, 2014). L-histidine-containing dipeptides are known to be readily taken up and have been shown to increase intracellular L-histidine concentrations (Erdmann, Weil and Krämer, 1993; Bellmann *et al.*, 2001). Since L-histidine degradation systems are absent in *C. glutamicum*, its export is crucial for intracellular amino acid homeostasis (Eggeling and Sahm, 2003; Kulis-Horn, Persicke and Kalinowski, 2014). LysE, however, exporting L-lysine and L-arginine, represents no active L-histidine export system in *C. glutamicum,* even though L-histidine acts as a co-inducer of *lysE* expression (Bellmann *et al.*, 2001). Hence, no such system has been identified for L-histidine export so far (Kulis-Horn, Persicke and Kalinowski, 2014).

In the development of L-histidine-producing *C. glutamicum* strains, initial attempts focused on deregulation of feedback inhibition of HisG using L-histidine analogs (Araki and Nakayama, 1974; Araki, Shimojo and Nakayama, 1974). Later, feedback deregulation of HisG was

pursued by site-directed mutagenesis and found the C-terminal domain involved (Zhang *et al.*, 2012). Site-directed mutagenesis combined with FACS-screening of obtained mutant libraries also revealed several *hisG* mutations inducing L-histidine production (Schendzielorz, Dippong, Grünberger, *et al.*, 2014).

Rational engineering studies focused on deregulation of expression control of the *hisD* promotor and overexpression of *hisE* and *hisG* (Cheng *et al.*, 2013a). Deletion of the C-terminal HisG domain and implementation of the S143F substitution in the catalytic HisG domain combined with *hisEG* overexpression also resulted in significant L-histidine production (Kulis-Horn, Persicke and Kalinowski, 2015). Schwentner *et al* (2019) identified unphysiologically high ATP regeneration from AICAR as necessary for improved L-histidine production and therefore overexpressed *purA*, *purB* and a heterologous glycine cleavage system for sufficient C_1-supply of the PurH-catalyzed reaction (Schwentner *et al.*, 2019). In addition, the start codon of the *pgi*-gene for the glucose-6-phosphate isomerase was attenuated to increase pentose phosphate pathway flux, which finally resulted in a L-histidine producer accumulating 0.093 mol L-histidine per mol D-glucose (Schwentner *et al.*, 2019). A pool influx kinetics approach using ^{13}C labeling dynamics supported the necessity of *purA*-, *purB*- and *purH*- overexpression as well as formyl recycling for a significantly improved L-histidine production (Feith *et al.*, 2020).

In *E. coli,* studies on L-histidine production also demonstrated the necessity of AICAR conversion towards ATP by overexpression of *purA* and *purH* (Malykh *et al.*, 2018). Wu and co-workers additionally expressed a gene for a heterologous NADH-dependent glutamate dehydrogenase from *Bacillus subtilis* (Wu *et al.*, 2020b). Their final *E. coli* strain produced 66.5 g L-histidine L^{-1} in a fed-batch process from D-glucose with a productivity of 1.5 g L^{-1} h^{-1} and a product yield of 0.23 g L-histidine g substrate^{-1}, representing the best published L-histidine production strain to date. In comparison, the apparently most efficient L-histidine production strain of *C. glutamicum* to date, however, is based on a feedback deregulated HisG variant and was published in 1994. Mizukami and Co-workers achieved an L-histidine titer of 23 g L^{-1} (148 mM L-histidine) in a fed-batch process with a volumetric productivity of 0.2 g L^{-1} h^{-1} and an estimated product yield of 0.15-0.20 g g substrate^{-1} (Mizukami *et al.*, 1994). Schwentner *et al* (2019) calculated the maximum theoretical L-histidine yield for *C. glutamicum* to be 0.44 g L-histidine g D-glucose^{-1} at a maximum growth rate (μ_{max}) of 0.1 (Schwentner *et al.*, 2019). This indicates room for improvement beyond the rationally engineered variants. With random genome mutagenesis, existing genetic alterations of L-histidine production strains could be leveraged for the identification of unknown beneficial mutations in non-intuitive off-site targets within the *C. glutamicum* genome.

2 Aims of this thesis

The main goal of this thesis was the identification of novel beneficial mutations and engineering targets in the genome of an already highly engineered *C. glutamicum* strain. Therefore, an available industrial L-histidine production strain was to be improved by whole cell directed evolution. An ultra-high-throughput approach, combining multiplexed random mutagenesis, biosensor-based FACS-screening, whole genome sequencing, automated comparative genome analysis and reverse engineering (at lower throughput) was developed. This enabled a strongly data-driven process for the identification of novel mutational hot spots contributing to improved L-histidine production.

Prior to this work, the LysG-based pSenLys biosensor was engineered towards insensitivity to L-lysine. The resulting LysG-A219L-based biosensor (pSenHis) was characterized with regard to ligand specificity by several methods including dipeptide supplementation experiments (Figure 4).
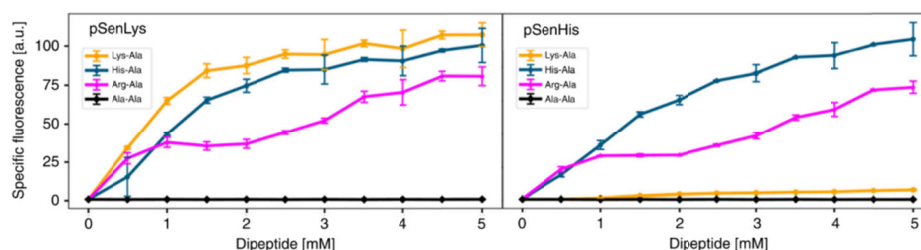
Figure 4: Fluorescence response of pSenLys- and pSenHis-harboring *C. glutamicum* variants to the presence of various dipeptides at different concentrations during microtiter plate cultivations.

In this thesis, pSenHis-based FACS-screenings of mutated *C. glutamicum* wild type cultures were performed to verify the biosensor's applicability for the reliable isolation of L-histidine overproducers. The isolated variants were subsequently analyzed for acquired mutations in the L-histidine biosynthesis pathway by targeted DNA sequencing. Then, a highly engineered industrial L-histidine production strain was subjected to 100 independent chemical and random mutagenesis experiments, pSenHis-based FACS-screening, and subsequent characterization of isolated strain variants to identify 100 independent strain variants with improved L-histidine production capabilities. By using whole genome sequencing of the 100 variants and comparative- and combinatorial analysis of their genomes, hitherto unknown beneficial mutations with high probability of contribution to improved L-histidine production were identified in the *C. glutamicum* genome. Reverse engineering was performed to confirm (and evaluate) the individual effects of identified mutations on L-histidine production and biomass formation. Further strain engineering then focused on the identification of beneficial combinations. Finally, the best variants were characterized in laboratory-scale bioreactor fermentations.

3 Material & Methods

3.1 Bacterial strains

Bacterial strain	Relevant characteristics	Source/ Reference
E. coli strains		
DH5α	F– Φ80*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *recA1 endA1 hsdR17* (rK–, mK+) *phoA supE44 λ– thi*-1 *gyrA96 relA1*	Invitrogen (Karlsruhe, Germany)
C. glutamicum strains		
wild type	*C. glutamicum* ATCC 13032 wild type strain, biotin-auxotroph	
Δ*lysEG*	*C. glutamicum* ATCC 13032 Δ*lysEG*	Binder *et al.*, 2012
CgHis1 (starting strain)	*C. glutamicum* ATCC 13032 derivative Δ*iolR* P$_{trc}$-*hisEG(S143F/ΔC)* P$_{trc}$-*hisDCB* P*tuf*-*hisHAFI* P$_{H36}$-*hisN* P$_{tuf}$-*fbp* pHisOP1	Senseup GmbH (Jülich, Germany)
CgHis2 (screening and reference strain)	*C. glutamicum* CgHis1 Δ*lysEG* pSenHis[*hisEG*]	this study
CgHis2 derivatives	**Isolates from FACS-screening:** strain nomenclature X1-X2-X3-X4 according to X1, # mutagenesis; X2, MNNG condition (i.e. 50 = 5.0 mg mL^{-1}); X3, column number in mutagenesis plate; X4, clone number after single cell sorting	
1-50-81-2	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
1-50-44-2	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
1-50-39-12	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
1-50-38-2	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
1-50-1-8	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
1-50-15-6	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
6-50-398-14	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-60-9-11	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-60-6-4	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-60-5-9	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-60-12-1	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-60-11-2	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-60-10-21	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study

7-55-8-5	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-55-11-16	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-55-9-22	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-55-5-16	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-55-4-6	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-55-10-21	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-50-8-15	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-50-6-18	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-50-4-15	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-50-11-21	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-50-10-8	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-45-7-14	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-45-6-2	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-45-5-5	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-45-4-16	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-45-11-24	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-45-10-9	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-40-9-14	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-40-8-23	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-40-7-8	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-40-6-4	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-40-5-24	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-40-4-11	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-40-12-7	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-40-11-5	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-40-10-2	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-35-9-3	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-35-7-14	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-35-6-8	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study

7-35-5-20	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-35-4-1	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-35-12-20	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-35-11-20	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-30-9-20	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-30-8-11	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-30-7-3	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-30-6-15	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-30-5-7	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-30-10-1	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-25-9-18	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-25-6-10	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-25-5-21	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
7-25-4-3	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
10-30-9-12	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
10-30-12-9	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
10-30-11-11	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
10-20-7-24	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
10-20-5-16	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
10-20-4-20	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
10-20-10-20	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
10-10-6-13	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
10-10-4-2	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
11-25-5-6	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
11-15-9-18	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
11-15-4-6	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
11-10-7-14	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
11-10-6-22	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
11-10-5-2	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study

11-10-11-7	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-20-7-18	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-20-4-16	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-20-15-17	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-20-14-11	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-20-12-3	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-15-9-18	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-15-8-5	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-15-7-9	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-15-6-10	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-15-5-4	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-15-4-15	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-15-21-16	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-15-20-1	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-15-19-24	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-15-18-8	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-15-16-9	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-15-15-6	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-15-12-5	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-15-10-7	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-10-9-5	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-10-8-6	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-10-7-2	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-10-6-23	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-10-5-6	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-10-4-1	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-10-12-5	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-10-11-11	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study
12-10-10-16	mutagenized *C. glutamicum* CgHis2 derivative, isolated in pSenHis-based FACS-screening	this study

CgHis2 derivatives	reverse engineered *C. glutamicum* CgHis2 strains harboring point mutations in specific hotspot genes	
cps-G987D	*C. glutamicum* CgHis2 derivative harboring point mutation in *cps* leading to amino acid substitution G987D	this study
emb -G477E	*C. glutamicum* CgHis2 derivative harboring point mutation in *emb* leading to amino acid substitution G477E	this study
emb-T529I/T539I	*C. glutamicum* CgHis2 derivative harboring point mutation in *emb* leading to amino acid substitutions T529I/T539I	this study
fasA-A2702T	*C. glutamicum* CgHis2 derivative harboring point mutation in *fasA* leading to amino acid substitution A2702T	this study
fasA-P783S	*C. glutamicum* CgHis2 derivative harboring point mutation in *fasA* leading to amino acid substitution P783S	this study
fasB-G1921E	*C. glutamicum* CgHis2 derivative harboring point mutation in *fasB* leading to amino acid substitution G1921E	this study
fasB-G2762D	*C. glutamicum* CgHis2 derivative harboring point mutation in *fasB* leading to amino acid substitution G2762D	this study
gltB-G1106D	*C. glutamicum* CgHis2 derivative harboring point mutation in *gltB* leading to amino acid substitution G1106D	this study
gltB-P988S	*C. glutamicum* CgHis2 derivative harboring point mutation in *gltB* leading to amino acid substitution P988S	this study
iolD-S481F	*C. glutamicum* CgHis2 derivative harboring point mutation in *iolD* leading to amino acid substitution S481F	this study
mrpA-L42F	*C. glutamicum* CgHis2 derivative harboring point mutation in *mrpA* leading to amino acid substitution L42F	this study
NCgl0552-G432D	*C. glutamicum* CgHis2 derivative harboring point mutation in *NCgl0552* leading to amino acid substitution G432D	this study
NCgl0552-P823S	*C. glutamicum* CgHis2 derivative harboring point mutation in *NCgl0552* leading to amino acid substitution P823S	this study
NCgl0705-S1847N	*C. glutamicum* CgHis2 derivative harboring point mutation in *NCgl0705* leading to amino acid substitution S1847N	this study
NCgl2789-S265N	*C. glutamicum* CgHis2 derivative harboring point mutation in *NCgl2789* leading to amino acid substitution S265N	this study
NCgl2859-S372F	*C. glutamicum* CgHis2 derivative harboring point mutation in *NCgl2859* leading to amino acid substitution S372F	this study
NCgl2959-D1453N	*C. glutamicum* CgHis2 derivative harboring point mutation in *NCgl2959* leading to amino acid substitution D1453N	this study
NCgl2959-G870D	*C. glutamicum* CgHis2 derivative harboring point mutation in *NCgl2959* leading to amino acid substitution G870D	this study
NCgl2964-E512K	*C. glutamicum* CgHis2 derivative harboring point mutation in *NCgl2964* leading to amino acid substitution E512K	this study
NCgl2964-P863S	*C. glutamicum* CgHis2 derivative harboring point mutation in *NCgl2964* leading to amino acid substitution P863S	this study
NCgl2981-D735G	*C. glutamicum* CgHis2 derivative harboring point mutation in *NCgl2981* leading to amino acid substitution D735G	this study
pks-A1525V	*C. glutamicum* CgHis2 derivative harboring point mutation in *pks* leading to amino acid substitution A1525V	this study
pks-D1186N	*C. glutamicum* CgHis2 derivative harboring point mutation in *pks* leading to amino acid substitution D1186N	this study
putA-P217S	*C. glutamicum* CgHis2 derivative harboring point mutation in *putA* leading to amino acid substitution P217S	this study
pyc-A764V	*C. glutamicum* CgHis2 derivative harboring point mutation in *pyc* leading to amino acid substitution A764V	this study
pyk-T357I	*C. glutamicum* CgHis2 derivative harboring point mutation in *pyk* leading to amino acid substitution T357I	this study
ulaA-V219I	*C. glutamicum* CgHis2 derivative harboring point mutation in *ulaA* leading to amino acid substitution V219I	this study
xylB-G55R	*C. glutamicum* CgHis2 derivative harboring point mutation in *xylB* leading to amino acid substitution G55R	this study

CgHis2 derivatives	*C. glutamicum* CgHis2 strains harboring in-frame deletions in specific hotspot genes	
Δ*cps*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*emb*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*fasB*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*ggtB*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*gltB*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*NCgl0552*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*NCgl0705*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*NCgl1737*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*NCgl2959*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*NCgl2964*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*NCgl2981*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*pknB*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*pks*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*putA*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*pyk1*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
Δ*pyk2*	*C. glutamicum* CgHis2 derivative harboring respective in-frame deletion	this study
CgHis2 derivatives	**Combinatorial *C. glutamicum* CgHis2 strains**	
NCgl0705-S1847N *cps*-G987D	*C. glutamicum* CgHis2 derivative harboring *NCgl0705*-S1847N *cps*-G987D	this study
NCgl2981-D735G *mrpA*-L42F	*C. glutamicum* CgHis2 derivative harboring *NCgl2981*-D735G *mrpA*-L42F	this study
Δ*cps NCgl2981*-D735G	*C. glutamicum* CgHis2 derivative harboring Δ*cps NCgl2981*-D735G	this study
Δ*cps pks*-D1186N	*C. glutamicum* CgHis2 derivative harboring Δ*cps pks*-D1186N	this study
Δ*cps* Δ*pyk1*	*C. glutamicum* CgHis2 derivative harboring Δ*cps* Δ*pyk1*	this study
Δ*cps* Δ*pyk1 NCgl2981*-D735G	*C. glutamicum* CgHis2 derivative harboring Δ*cps* Δ*pyk1 NCgl2981*-D735G	this study
Δ*fasB NCgl2981*-D735G	*C. glutamicum* CgHis2 derivative harboring Δ*fasB NCgl2981*-D735G	this study
Δ*fasB pks*-D1186N	*C. glutamicum* CgHis2 derivative harboring Δ*fasB pks*-D1186N	this study
Δ*fasB* Δ*cps*	*C. glutamicum* CgHis2 derivative harboring Δ*fasB* Δ*cps*	this study
Δ*fasB* Δ*cps NCgl2981*-D735G	*C. glutamicum* CgHis2 derivative harboring Δ*fasB* Δ*cps NCgl2981*-D735G	this study
Δ*fasB* Δ*cps pks*-D1186N	*C. glutamicum* CgHis2 derivative harboring Δ*fasB* Δ*cps pks*-D1186N	this study
Δ*fasB* Δ*cps* Δ*pyk1*	*C. glutamicum* CgHis2 derivative harboring Δ*fasB* Δ*cps* Δ*pyk1*	this study

ΔfasB Δcps Δpyk1 NCgl2981-D735G	C. glutamicum CgHis2 derivative harboring ΔfasB Δcps Δpyk1 NCgl2981-D735G	this study
ΔfasB Δcps Δpyk1 pks-D1186N	C. glutamicum CgHis2 derivative harboring ΔfasB Δcps Δpyk1 pks-D1186N	this study
ΔfasB Δcps Δpyk1 pks-D1186N NCgl2981-D735G	C. glutamicum CgHis2 derivative harboring ΔfasB Δcps Δpyk1 pks-D1186N NCgl2981-D735G	this study
ΔfasB Δpyk1	C. glutamicum CgHis2 derivative harboring ΔfasB Δpyk1	this study
ΔfasB Δpyk1 NCgl2981-D735G	C. glutamicum CgHis2 derivative harboring ΔfasB Δpyk1 NCgl2981-D735G	this study
ΔfasB Δpyk1 NCgl2981-D735G pks-D1186N	C. glutamicum CgHis2 derivative harboring ΔfasB Δpyk1 NCgl2981-D735G pks-D1186N	this study
ΔfasB Δpyk1 pks-D1186N	C. glutamicum CgHis2 derivative harboring ΔfasB Δpyk1 pks-D1186N	this study
Δpyk1 NCgl2981-D735G	C. glutamicum CgHis2 derivative harboring Δpyk1 NCgl2981-D735G	this study
Δpyk1 pks-D1186N	C. glutamicum CgHis2 derivative harboring Δpyk1 pks-D1186N	this study
Δpyk1Δpyk2	C. glutamicum CgHis2 derivative harboring Δpyk1Δpyk2	this study

3.2 Plasmids

Plasmid	Relevant characteristics	Source/ Reference
pHisOP1	pJC1-based hisEG-overexpression plasmid with feedback-resistant hisG(ΔC/S143F) variant, kanamycin resistance	SenseUp GmbH
pSenHis	pJC1-based biosensor plasmid, kanamycin resistance, LysG-A219L based biosensor module, eYFP as fluorescence reporter gene	Della Corte et al., 2020
pSenHis[hisEG]	pSenHis biosensor module subcloned onto pJC1-based hisEG-overexpression plasmid pHisOP1	this study
pk19mobsacB	mobilizable suizide vector for double homologous recombination in C. glutamicum, kanamycin resistance, levansucrase gene sacB	Schäfer et al., 1994
pk19mobsacB derivatives		
pk19-cps-G987D	pk19mobsacB derivative for introduction of cps-G987D	this study
pk19-emb-G477E	pk19mobsacB derivative for introduction of emb -G477E	this study
pk19-emb-T529I/T539I	pk19mobsacB derivative for introduction of emb-T529I/T539I	this study
pk19-fasA-A2702T	pk19mobsacB derivative for introduction of fasA-A2702T	this study
pk19-fasA-P783S	pk19mobsacB derivative for introduction of fasA-P783S	this study
pk19-fasB-G1921E	pk19mobsacB derivative for introduction of fasB-G1921E	this study
pk19-fasB-G2762D	pk19mobsacB derivative for introduction of fasB-G2762D	this study
pk19-gltB-G1106D	pk19mobsacB derivative for introduction of gltB-G1106D	this study
pk19-gltB-P988S	pk19mobsacB derivative for introduction of gltB-P988S	this study
pk19-iolD-S481F	pk19mobsacB derivative for introduction of iolD-S481F	this study
pk19-mrpA-L42F	pk19mobsacB derivative for introduction of mrpA-L42F	this study
pk19-NCgl0552-G432D	pk19mobsacB derivative for introduction of NCgl0552-G432D	this study
pk19-NCgl0552-P823S	pk19mobsacB derivative for introduction of NCgl0552-P823S	this study

pk19-*NCgl0705*-S1847N	pk19*mobsacB* derivative for introduction of *NCgl0705*-S1847N	this study
pk19-*NCgl2789*-S265N	pk19*mobsacB* derivative for introduction of *NCgl2789*-S265N	this study
pk19-*NCgl2859*-S372F	pk19*mobsacB* derivative for introduction of *NCgl2859*-S372F	this study
pk19-*NCgl2959*-D1453N	pk19*mobsacB* derivative for introduction of *NCgl2959*-D1453N	this study
pk19-*NCgl2959*-G870D	pk19*mobsacB* derivative for introduction of *NCgl2959*-G870D	this study
pk19-*NCgl2964*-E512K	pk19*mobsacB* derivative for introduction of *NCgl2964*-E512K	this study
pk19-*NCgl2964*-P863S	pk19*mobsacB* derivative for introduction of *NCgl2964*-P863S	this study
pk19-*NCgl2981*-D735G	pk19*mobsacB* derivative for introduction of *NCgl2981*-D735G	this study
pk19-*pks*-A1525V	pk19*mobsacB* derivative for introduction of *pks*-A1525V	this study
pk19-*pks*-D1186N	pk19*mobsacB* derivative for introduction of *pks*-D1186N	this study
pk19-*putA*-P217S	pk19*mobsacB* derivative for introduction of *putA*-P217S	this study
pk19-*pyc*-A764V	pk19*mobsacB* derivative for introduction of *pyc*-A764V	this study
pk19-*pyk*-T357I	pk19*mobsacB* derivative for introduction of *pyk*-T357I	this study
pk19-*ulaA*-V219I	pk19*mobsacB* derivative for introduction of *ulaA*-V219I	this study
pk19-*xylB*-G55R	pk19*mobsacB* derivative for introduction of *xylB*-G55R	this study
pk19-Δ*lysEG*	pk19*mobsacB* derivative for in-frame deletion of *lysEG*	(Binder *et al.*, 2012)
pk19-Δ*cps*	pk19*mobsacB* derivative for in-frame deletion of *cps*	this study
pk19-Δ*emb*	pk19*mobsacB* derivative for in-frame deletion of *emb*	this study
pk19-Δ*fasB*	pk19*mobsacB* derivative for in-frame deletion of *fasB*	(Radmacher *et al.*, 2005)
pk19-Δ*ggtB*	pk19*mobsacB* derivative for in-frame deletion of *ggtB*	this study
pk19-Δ*gltB*	pk19*mobsacB* derivative for in-frame deletion of *gltB*	this study
pk19-Δ*NCgl0552*	pk19*mobsacB* derivative for in-frame deletion of *NCgl0552*	this study
pk19-Δ*NCgl0705*	pk19*mobsacB* derivative for in-frame deletion of *NCgl0705*	this study
pk19-Δ*NCgl1737*	pk19*mobsacB* derivative for in-frame deletion of *NCgl1737*	this study
pk19-Δ*NCgl2959*	pk19*mobsacB* derivative for in-frame deletion of *NCgl2959*	this study
pk19-Δ*NCgl2964*	pk19*mobsacB* derivative for in-frame deletion of *NCgl2964*	this study
pk19-Δ*NCgl2981*	pk19*mobsacB* derivative for in-frame deletion of *NCgl2981*	this study
pk19-Δ*pknB*	pk19*mobsacB* derivative for in-frame deletion of *pknB*	this study
pk19-Δ*pks*	pk19*mobsacB* derivative for in-frame deletion of *pks*	this study
pk19-Δ*putA*	pk19*mobsacB* derivative for in-frame deletion of *putA*	this study
pk19-Δ*pyk1*	pk19*mobsacB* derivative for in-frame deletion of *pyk1*	this study
pk19-Δ*pyk2*	pk19*mobsacB* derivative for in-frame deletion of *pyk2*	this study

3.3   Construction of plasmids and strains

Recombinant DNA work was performed in *E. coli* DH5α, which was grown in LB medium (Bertani, 1951) at 37 °C, supplemented with 50 µg mL^{-1} kanamycin if appropriate. Plasmids were constructed following standard protocols of molecular cloning such as PCR, DNA restriction and ligation and Gibson cloning (Gibson *et al.*, 2009; Sambrook and Russel, 2012). Transformation of *C. glutamicum* was performed following the standard electroporation protocol (Eggeling and Bott, 2005). Double homologous recombination was employed for introduction of point mutations and in-frame deletions using the pK19mobsacB vector (Schäfer *et al.*, 1994). Oligonucleotide synthesis and Sanger sequencing was performed at Eurofins SE (Ebersberg, Germany). Oligonucleotides used in this study are listed in the Appendix (Chapter 8.8).

3.4   Cultivation of *C. glutamicum* strains

In general, all cultivations of *C. glutamicum* were performed at 30 °C in undefined Brain-Heart-Infusion (BHI) medium (Difco Laboratories, Detroit, MI, USA) or in defined CGXII medium supplemented with 2 % (w/v) D-glucose as source of carbon and energy (Keilhauer, Eggeling and Sahm, 1993). Where appropriate, 15 µg mL^{-1} kanamycin was added. Biomass formation was measured by the optical density at 600 nm (OD$_{600}$).

Initially, cultivations were performed using the BioLector system (m2p-labs GmbH, Baesweiler, Germany). 48-well microtiter plates (Flowerplates, m2p-labs GmbH, Baesweiler, Germany) were applied as cultivation vessels. Formation of biomass and culture fluorescence was monitored over a 48 h time course (30 °C, 900 rpm, 75 % humidity and 3 mm throw). Biomass formation was thereby determined as backscatter light at 620 nm. For measurement of culture fluorescence, EYFP emission was determined at 532 nm upon excitation at 510 nm. The inoculation of the main culture (800 µL CGXII, 2 % D-glucose) was followed by a seedtrain consisting of a stationary first pre-culture (650 µL BHI) and a second stationary pre-culture (600 µL CGXII, 2 % D-glucose) as described above. Dipeptides for supplementation were purchased from Bachem (Bubendorf, Switzerland).

High-throughput characterization of isolated *C. glutamicum* CgHis2 variants was performed following a two-step cultivation workflow. Both steps were performed in 48-well microtiter plates (Flowerplates, m2p-labs GmbH, Baesweiler, Germany) in a Multitron Pro HT Incubator (InforsAG, Bottmingen, Switzerland) at 30 °C, 900 rpm, 75 % humidity and 3 mm throw over a cultivation time course of 48 h. For sample preparation of main cultures, a Hamilton robotic liquid handling platform (Hamilton, Reno, NV, USA) was employed to achieve high-throughput and high accuracy. Biomass formation was measured as the final OD$_{600}$ in a

Synergy Mx microplate reader (BioTek, Winooski, VT, USA). The first cultivation step was designed to pre-characterize isolated *C. glutamicum* CgHis2 variants in high-throughput. Therefore, main cultures (800 µL CGXII, 2 % D-glucose) were inoculated to an OD_{600} of 1.0 from a pre-culture (650 µL BHI), which had been grown for 22 h. Where appropriate, 15 µg mL^{-1} kanamycin was added. For high-throughput, the second pre-culture was omitted and no technical replicates were cultivated. L-histidine titers of *C. glutamicum* CgHis2 variants, as determined by HPLC (see 3.5), which exceeded the 90 % confidence interval of the control strains (as determined by four biological and three technical replicates each) were considered for re-cultivation if they exceeded an additional threshold of 1 mM increase in L-histidine titer.

Re-cultivation in the second cultivation step included technical triplicates of selected variants and a second pre-culture (600 µL CGXII, 2 % D-glucose) in the applied seedtrain to achieve high accuracy for characterization with regard to biomass and product formation. To be considered for Next Generation Sequencing, L-histidine titers of *C. glutamicum* CgHis2 variants had to exceed the 90 % confidence interval of control strains (as before) plus a threshold of at least 10 % increase in L-histidine titer.

Standard cultivation conditions for characterization of reverse engineered *C. glutamicum* CgHis2 strains were accordingly defined: A first pre-culture (650 µL BHI) was grown for 22 h before used to inoculate the second pre-culture (600 µL CGXII, 2 % D-glucose). The second pre-culture was also grown for 22 h before used for inoculation of the main culture (800 µL CGXII, 2 % D-glucose) to an OD_{600} of 1.0, using the Hamilton robotic platform (Hamilton, Reno, NV, USA). Where appropriate, 15 µg mL^{-1} kanamycin was added. All cultivation steps were performed in 48-well microtiter plates (Flowerplates, m2p-labs GmbH, Baesweiler, Germany) in a Multitron Pro HT Incubator (InforsAG, Bottmingen, Switzerland) at 30 °C, 900 rpm, 75 % humidity and 3 mm throw. The main culture was cultivated for 48 h before harvest.

3.5 Amino acid quantification

First, culture supernatants were diluted 1:100 using a Hamilton robotic liquid handling platform (Hamilton, Reno, NV, USA). Then, High-performance liquid chromatography (HPLC) was performed on an uHPLC 1290 Infinity system (Agilent Technologies, Santa Clara, CA, USA), which was equipped with a Zorbax Eclipse AAA C18 3.5 micron 4.6 x 75 mm and a fluorescence detector. As mobile phase, a gradient of 0.01 M Na-borate buffer pH 8.2 with increasing concentrations of methanol was applied. Quantification of amino acids was achieved by their o-phthaldialdehyde derivatives via pre-column-derivatization. Fluorescence detection of the derivatives was performed at an excitation wavelength of 230 nm and an emission wavelength of 450 nm.

3.6 Whole genome mutagenesis

MNNG is a highly effective chemical mutagen, methylating nucleobases at their oxygen-residues, thereby inducing base mispairing and hence introduction of random point mutations with strong mutational bias towards GC→AT transitions (Ohta, Watanabe-Akanuma and Yamagata, 2000; B Singer and Kusmierek, 2003; Ohnishi *et al.*, 2008). A MNNG mutagenesis protocol was modified for application of multiplexed mutagenesis using 96-well plates. To achieve a desired culture mortality of 50-90 %, a MNNG dilution series was applied as follows:

A *C. glutamicum* CgHis2 pre-culture (50 mL BHI, 91 g L^{-1} Sorbitol, 1 % D-glucose and 15 µg mL^{-1} kanamycin in 500 mL baffled shake flask) was grown over night. From this pre-culture, the main culture (50 mL BHI, 91 g L^{-1} Sorbitol, 1 % D-glucose and 15 µg mL^{-1} kanamycin in 500 mL baffled shake flask) was inoculated to an OD_{600} of 1.5. Over 2-3 h the culture was grown in a rotary shaker (Infors, Bottmingen, Switzerland) at 30 °C and 120 rpm until an OD_{600} of 4.4 was reached. Then, the culture was immediately distributed to a 96-well plate (2 mL vessels, Eppendorf, Hamburg, Germany) in 400 µL aliquots. For mutagenesis, a MNNG dilution series (6.0; 5.5; 5.0; 4.5; 4.0; 3.0; 2.0; 1.0 mg mL^{-1} MNNG in DMSO) was prepared. 20 µL solution were applied to each well with decreasing MNNG concentration from row A to H. As controls, Columns 1-3 were treated with DMSO only. Mutagenesis was performed by incubation for 15 minutes in a rotary shaker (Infors, Bottmingen, Switzerland) at 30 °C and 120 rpm, including transportation time. Mutagenesis was stopped by centrifugation for 7 minutes at 20 °C and 4000 rpm (Heraeus Multifuge X3R Centrifuge, ThermoFisher Scientific Inc., Waltham, MA, USA) and subsequent discarding of the supernatant. Then, cultures were washed twice using NaCl (0.9 %, 1.6 mL per well) and re-cultivated for 3 h (400 µL BHI, 91 g L^{-1} Sorbitol, 1 % D-glucose and 15 µg mL^{-1} kanamycin per well). For evaluation of culture mortality, representative cultures of respective MNNG conditions were applied to single cell sorting of the main population using FACS. Before FACS-screening for improved *C. glutamicum* CgHis2 variants, glycerol was added to the mutagenized cultures (80 %, 400 µL per well) upon which they were stored at -80 °C.

3.7 pSenHis-based FACS-screening

Prior to FACS-screening, the biomass from 400 µL of mutagenized *C. glutamicum* CgHis2 cultures was used to inoculate the respective cultures considered for FACS-screening (800 µL CGXII, 2 % D-glucose, 15 µg mL^{-1} kanamycin). Controls were processed accordingly. The culture was grown for 5-7 h in 48-well microtiter plates (Flowerplates, m2p-labs GmbH, Baesweiler, Germany) in a Multitron Pro HT Incubator (Infors AG, Bottmingen, Switzerland), at 30 °C, 900 rpm, 75 % humidity and 3 mm throw.

FACS-screening was performed using a BD FACSAria Fusion Flow Cytometer (BD Biosciences, Franklin Lakes, NJ, USA). Therefore, cultures were diluted (CGXII base) in order to reach 10,000 events s^{-1} at a flow rate of 1.0, when using a 70 µm nozzle at a sheath pressure of 70 psi. For excitation of EYFP, a blue solid laser (488 nm) was employed. Single cell parameters were recorded as forward scatter (FSC, as small angle scatter), side scatter (SSC, as orthogonal scatter) and fluorescence from the 488 nm laser. Therefore, the system was equipped with a 502 nm long-pass and 530/30 band-pass filter combination. Electronic noise and cell debris were reduced by electronic gating of FSC-H against SSC-H. Subsequent gating of SSC-H against SSC-W and FSC-H against FSC-W ensured doublet discrimination. Based on this gating scheme, fluorescence and FSC characteristics were analyzed. An electronic gate comprising the top 5 % of fluorescent cells (with respect to the FSC distribution) was used for single cell sorting. From each mutagenized *C. glutamicum* CgHis2 culture, 240 events were sorted on agar plates (BHI, 91 g L^{-1} Sorbitol, 15 µg mL^{-1} kanamycin). Control cultures were analyzed and sorted accordingly, when appropriate.

After incubation on agar plates at 30 °C for two days, 22 randomly picked *C. glutamicum* CgHis2 strain variants were transferred to liquid medium (600 µL BHI, 91 g L^{-1} Sorbitol, 1 % D-glucose, 15 µg mL^{-1} kanamycin) from each mutagenized culture. Cultivation was performed for 24 h in a Multitron Pro HT Incubator (Infors AG, Bottmingen, Switzerland) at 30 °C, 900 rpm, 75 % humidity and 3 mm throw in 96-well microtiter plates. After addition of glycerol (85 %, 400 µL per well), cultures were stored at -80 °C until strain characterization, as described above.

3.8 Whole genome sequencing

First, chromosomal DNA was isolated from *C. glutamicum* according to established protocols (Eikmanns *et al.*, 1994). Then, whole genome sequencing was performed at Eurofins Genomics Germany GmbH (Ebersberg, Germany) on an Illumina High-Seq in paired-end mode using 150 bp read length.

3.9 Automated comparative genome analysis

For automated comparative genome analysis of the high number of sequenced *C. glutamicum* CgHis2 strain variants, Dr. Michael Dal Molin developed a software tool called Fast Automated Analysis of Multiple Sequences (FAAMS). Included in the automated workflow was the download of reference data for *C. glutamicum* ATCC 13032 from NCBI genome download tool (https://github.com/kblin/ncbi-genome-download). At first, pre-processing of the sequence data (fastq files) including trimming and filtering of bad reads was performed with fastp (Chen *et al.*, 2018). Subsequently the alignment of the sequence

data to the *C. glutamicum* ATCC 13032 wild type reference was performed by the Burrows-Wheeler Aligner using the BWA-MEM algorithm (Li and Durbin, 2009). The Picard Toolkit (Broad Institute, 2019) and Samtools were applied to determine statistical data for quality control (Li *et al.*, 2009). Samtools was also applied for pileup of the sequence data before identification of mutational variants (Single nucleotide polymorphisms, deletions and insertions) by VarScan (Koboldt *et al.*, 2009, 2012; Li *et al.*, 2009). The sequences of the *C. glutamicum* CgHis2 strain variants were compared to the *C. glutamicum* CgHis2 reference. Differences were annotated on gene level by resulting codon change and induced amino acid substitution, whereas mutations in a hypothetical 200 bp upstream area of a gene were labeled as potential promoter mutations. Statistical analysis of sequence reads and alignments was performed by the R software. Finally, visualization of the data output was performed by a customized shiny dashboard in R (Bunn and Korpela, 2019).

3.10 Combinatorial sequence analysis

FAAMS delivered hotspot genes ranked by the number of mutations, which clustered in specific *C. glutamicum* genes. Hence, genes with high numbers of mutations (single nucleotide polymorphisms) were transferred into a spreadsheet. All *C. glutamicum* CgHis2 strain variants were listed with their specific L-histidine titers and final OD_{600}, which were revealed from strain characterizations and the total number of acquired mutations from random genome mutagenesis. Then, nonsynonymous mutations of the *C. glutamicum* CgHis2 strain variants, which applied to the hotspot genes, were transferred to a spreadsheet. The number of strain variants, which harbored nonsynonymous mutations in a specific gene, was called hotspot significance and used to compare the relevance of a gene for its potential contribution to improved L-histidine production. In the end, a threshold of 10 % was set for a gene to be considered a significant hotspot gene, which thereby qualified for reverse engineering. To select specific mutations in hotspot genes for reverse engineering, three more criteria were defined: Hotspot abundance, variant performance and network effects (Chapter 4.9).

3.11 Bioreactor fermentations

All bioreactor fermentations were conducted by Moritz-Fabian Müller from AG Noack, Forschungszentrum Jülich GmbH. The seedtrain comprised two pre-cultures before inoculation of the bioreactor. The first pre-culture (10 mL BHI in 100 mL shaking flask with baffles) was inoculated from a fresh agar plate. The culture was incubated for 18-38 h at 30 °C and 250 rpm on a rotary shaker (Infors, Bottmingen, Switzerland). 5 mL of the first pre-culture were used to inoculate the second pre-culture (100 mL CGXII, 2 % D-glucose in a 1 L shake flask), which was cultivated in accordance to the first pre-culture for 18-28 h. The

stationary cultures were centrifuged and washed using 0.9 % NaCl to a final volume of 5 mL. An appropriately diluted aliquot was transferred to the bioreactor to inoculate the main culture (988 mL CGXII, 2 % D-glucose, 2 mL antifoam, 10 mL inoculum) to an OD_{600} of 0.5. Kanamycin was added to all cultivations appropriately.

The bioreactor fermentation was performed on a dasgip system (Eppendorf SE, Hamburg Germany) with four simultaneously operated reactors. Regulation of pH was performed both-sided with 5 M H_2PO_4 and 25 % NH_4OH. Dissolved oxygen was fixed via a cascade to a minimum of 30 % by increasing agitation from 400 to 1200 rpm. Hence, the airflow increase from 6 to 40 standard liter h^{-1} when needed. The fermentation was stopped when the optical density had surpassed its peak and agitation had returned to 400 rpm, which meant the carbon source was depleted. During the fermentation, a series of 6 mL samples were taken from the cultures for determination of D-glucose and L-histidine concentrations as well as OD_{600} and cell dry weight.

Glucose measurements were performed on an Agilent 1260 Infinity II HPLC System (Santa Clara, CA, USA) equipped with a 300x8mm Organic Acid Column (Chromatographie Service GmbH, Langerwehe, Germany) with a up to four-fold dilution. Cell dry weight was determined from 1.8 mL samples, washed in 0.9 % NaCl and dried at 80 °C for at least one day with subsequent incubation in a desiccator for one day. The weight was determined according to pre-weighed tubes on a precision scale.

4 Results

Prior to this work, no L-histidine-producing variants could be isolated in any of the conducted pSenLys-based FACS-screenings, even though L-histidine acts as inducer of transcriptional regulator LysG, which is the key component of pSenLys (Bellmann *et al.*, 2001; Binder *et al.*, 2012). To achieve the isolation of L-histidine-producing variants, LysG was engineered towards insensitivity against L-lysine by semi-rational protein engineering using FACS. As a result of this directed evolution campaign, the LysG-A219L variant could be isolated, which did not accept L-lysine as ligand anymore. As a consequence, the resulting LysG-A219L-based biosensor (pSenHis) showed an unaltered fluorescence response to L-histidine-containing dipeptides, whereas the supplementation of L-lysine-containing dipeptides did not result in a significant biosensor response. Cultivations of pSenHis carrying cells in microfluidic chips as well as isothermal titration calorimetry performed with LysG-A219L and molecular dynamics simulations of the engineered regulator confirmed these results. However, the applicability of the pSenHis biosensor for the isolation of L-histidine-producing variants in real FACS-based screening campaigns had not been demonstrated.

4.1 Biosensor-based FACS-screening of a *C. glutamicum* wild type culture for L-histidine production

The functionality of pSenHis in biosensor-based FACS-screenings for the isolation of L-histidine producing strain variants was demonstrated using mutated *C. glutamicum* wild type cells. For this purpose, pSenHis-harboring *C. glutamicum* ATTC13032 wild type cultures were randomly mutagenized by incubation with *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG). MNNG is a highly effective chemical mutagen, methylating nucleobases at their oxygen-residues, thereby inducing base mispairing and hence introduction of random point mutations with strong mutational bias towards GC→AT transitions (Ohta, Watanabe-Akanuma and Yamagata, 2000; B Singer and Kusmierek, 2003; Ohnishi *et al.*, 2008). The resulting mutant library was subjected to pSenHis-based high-throughput screening in order to isolate L-histidine-producing strain variants using FACS. From 450 isolated and characterized variants, 217 were identified with L-histidine titers > 0.1 mM. Of these variants, the best 25 accumulated L-histidine in the supernatant at concentrations ranging from 0.4 – 0.7 mM. *C. glutamicum* wild type cultures served as control where no L-histidine could be detected in the supernatant. None of the variants produced significant amounts of L-lysine or L-arginine, thus the applicability of pSenHis for the identification of L-histidine-producing variants could be successfully verified.

4.2 Analysis of induced mutations in L-histidine biosynthetic pathway genes

The L-histidine biosynthetic pathway genes represent intuitive engineering targets for L-histidine production (Mizukami *et al.*, 1994; Zhang *et al.*, 2012; Schendzielorz, Dippong, Grunberger, *et al.*, 2014; Kulis-Horn, Persicke and Kalinowski, 2015; Schwentner *et al.*, 2019; Wu *et al.*, 2020). Hence, it was speculated, that the strain variants isolated in pSenHis-based FACS-screening, which were evolved from the *C. glutamicum* wild type, most likely acquired beneficial mutations for L-histidine production in the L-histidine biosynthetic pathway genes.

To confirm this hypothesis, targeted DNA sequencing was performed to investigate the biosynthetic genes of the L-histidine pathway (*hisA*, *hisB*, *hisC*, *hisD*, *hisE*, *hisF*, *hisG*, *hisH*, *hisI*, and *hisN*) for acquired mutations in the 25 best producers. All strain variants were found to carry mutations in *hisG*, which encodes for the ATP phosphoribosyltransferase. This enzyme catalyzes the first committed step of L-histidine biosynthesis and is subject to non-competitive feedback inhibition by the pathway end-product L-histidine (Kulis-Horn, Persicke and Kalinowski, 2014). Notably, most of the identified *hisG* mutations of isolated strain variants were already known from previous studies (Table 1) (Zhang *et al.*, 2012; Schendzielorz, Dippong, Grunberger, *et al.*, 2014; Kulis-Horn, Persicke and Kalinowski, 2015). Only few mutations were found in other L-histidine biosynthetic genes, which caused amino acid substitutions in HisB (A105T), HisD (T107A) and HisH (G76A, S12F, R16H). Thus, the major relevance of feedback regulated HisG (catalyzing the dedicated first step of L-histidine biosynthesis) for the overproduction of this amino acid could be confirmed.

Table 1: Amino acid substitutions in HisG of strain variants isolated from pSenHis-based FACS-screening of a mutagenized *C. glutamicum* wild type culture and L-histidine titer produced by respective strain variant.

Amino acid substitutions in HisG	L-histidine titer of respective variant [mM]
Wild type	0
S143F	0.67 ± 0.21
S143F + G233D	0.43 ± 0.1
D213N	0.68 ± 0.19
G230S	0.57 ± 0.09
G230D	0.57 ± 0.23
S232P	0.46 ± 0.08
G233D	0.62 ± 0.14
T235M	0.47 ± 0.17

Mutations acquired elsewhere in the *C. glutamicum* genome may have contributed to the increased L-histidine production as well, but were not analyzed in the context of these

experiments, which were conducted to demonstrate the suitability of pSenHis for the isolation of L-histidine-producing strain variants. However, when the pSenHis-based FACS screening strategy is supposed to be used to identify novel genetic targets contributing to L-histidine production, a *C. glutamicum* strain, which already harbors beneficial mutations in *hisG* or other L-histidine pathway genes, should be used as starting variant.

4.3 Engineering an industrial L-histidine production strain for pSenHis-based FACS-screening

FACS-based high-throughput screening can be used as an effective enrichment strategy, by which the effort of chromatographic analyses can be significantly reduced (Kaczmarek and Prather, 2021). Before applying production strains as screening hosts, however, they may have to be modified and tested with the respective biosensor for proper mutual function.

For this screening campaign, an industrial *C. glutamicum* ATCC 13032 L-histidine production strain was provided from the industry partner SenseUP GmbH. *C. glutamicum* CgHis1 harbored substitutions of the four operon promoters of the L-histidine biosynthesis pathway with strong constitutive promoters: P_{trc}-*hisEG*, P_{trc}-*hisDCB*, P_{tuf}-*hisHAFI* and P_{H36}-*hisN* (Kulis-Horn, Persicke and Kalinowski, 2014). Furthermore, the *hisEG* operon was substituted with a feedback-deregulated variant of *hisG*, comprising the S143F substitution and C-terminal truncation (Kulis-Horn, Persicke and Kalinowski, 2015). Since *hisEG* overexpression is key in L-histidine production (Cheng *et al.*, 2013b; Kulis-Horn, Persicke and Kalinowski, 2015; Schwentner *et al.*, 2019; Wu *et al.*, 2020a), the feedback-deregulated operon was additionally overexpressed from a medium copy number plasmid (pHisOP1). To increase pentose phosphate pathway flux towards L-histidine biosynthesis, this strain harbored a promoter exchange of fructose biphosphatase (P_{tuf}-*fbp*), a gluconeogenetic enzyme. Uptake of D-glucose was increased by in-frame deletion of *iolR*. This regulator of *myo*-inositol utilization genes in *C. glutamicum* represses the expression of D-xylose and D-glucose transporter IolT1 (Klaffl et al., 2013; Brüsseler et al., 2018). In micro-scale cultivations, *C. glutamicum* CgHis1 pHisOP1 produced 11 mM L-histidine from 2 % D-glucose as sole source of carbon and energy (referred to as standard cultivation conditions, Materials & Methods Chapter 3.4). The growth rate of *C. glutamicum* CgHis1 (0.31 h^{-1}) was significantly reduced in comparison to the *C. glutamicum* wild type (0.41 h^{-1}), however, the cell morphology was found wild-type-like (Appendix Chapter 8.1).

Since the pSenHis biosensor is based on the transcriptional regulator LysG, the endogenous *lysG* copy was removed in *C. glutamicum* CgHis1 by in-frame deletion of *lysEG* to avoid the isolation of false-positives (Binder *et al.*, 2012). Hence, interaction of native LysG with L-lysine was prevented in *C. glutamicum* CgHis1. Despite LysE exporting basic amino acids

such as L-lysine and L-arginine (but not L-histidine), L-histidine itself still acts as a co-inducer of LysG-mediated native *lysE* expression. With the deletion of *lysEG*, potential effects of overexpressed LysE from high intracellular L-histidine concentrations were prevented as well (Bellmann *et al.*, 2001). Then, the LysG-A291L-based pSenHis biosensor module was subcloned onto the *hisEG* overexpression plasmid (pHisOP1), with which the otherwise plasmid-free *C. glutamicum* CgHis1 strain was subsequently transformed. Both plasmids, pSenHis and pHisOP1, share the pJC1-backbone, such that no difference in biosensor functionality was expected. The generated screening strain CgHis1 *ΔlysEG* pSenHis[*hisEG*], designated as *C. glutamicum* CgHis2 (or simply reference strain), accumulated up to 11.5 mM L-histidine from 2 % D-glucose in the supernatant during micro-scale cultivations and grew to a final OD_{600} of 19. These results were comparable to the *C. glutamicum* CgHis1 pHisOP1 starting strain, showing that the biosensor-carrying plasmid does not represent a larger burden for the L-histidine production strain (Appendix Chapter 8.1).

Hitherto, the pSenHis biosensor enabled the isolation of L-histidine-producing variants from a mutagenized *C. glutamicum* wild type culture, which produced up to 0.7 mM L-histidine. In pSenHis-based screening of mutagenized *C. glutamicum* CgHis2 cultures already producing high amounts of L-histidine, the basal fluorescence of the biosensor (pSenHis[*hisEG*]) is already on a significantly higher level. Hence, it had to be demonstrated that the operational range of pSenHis is still broad enough to translate any increased productivity of mutated *C. glutamicum* CgHis2 strain variants into an increased fluorescence in comparison to the *C. glutamicum* CgHis2 starting variant and allow for a successful isolation of improved variants. Therefore, the applicability of pSenHis[*hisEG*] in *C. glutamicum* CgHis2 screenings was investigated by simulating an increased intracellular L-histidine productivity with supplementation of L-His-L-Ala-dipeptides to *C. glutamicum* CgHis2 cultures. As a result, specific fluorescence of *C. glutamicum* CgHis2 cultures increased in correlation to the supplemented L-His-containing dipeptides up to 15 mM L-His-L-Ala (Figure 5), indicating increased biosensor induction and hence increased fluorescence response to be possible. However, the intracellular accumulation of L-histidine of *C. glutamicum* CgHis2 exerted a significantly higher effect on the overall fluorescence compared to the dipeptide supplementation. This became obvious when comparing the fluorescence response of the L-histidine-producing *C. glutamicum* CgHis2 variant and the more wild-type-like *C. glutamicum* pSenHis[*hisEG*] variant (Figure 5), which was not followed further.
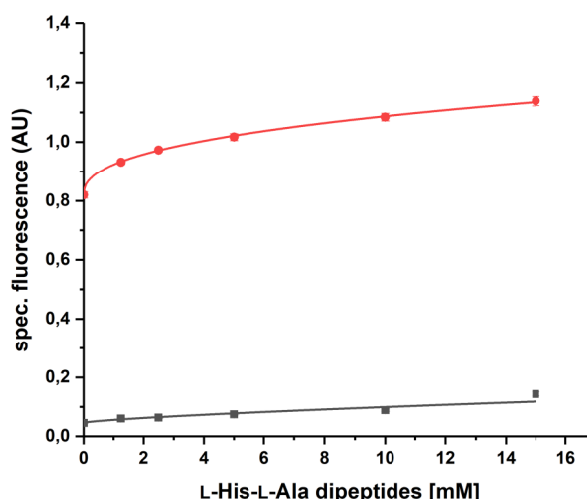
Figure 5: Specific fluorescence response upon dipeptide supplementation of biosensor-harboring *C. glutamicum* strains. An increased L-histidine productivity of *C. glutamicum* CgHis2 (red) and *C. glutamicum* pSenHis[*hisEG*] (grey) was simulated using supplementations of 0, 1.25, 2.5, 5, 10 and 15 mM L-His-L-Ala dipeptides. In all cases, dipeptides were supplemented to a final concentration of 15 mM using L-Ala-L-Ala dipeptides as co-supplement. Standard deviations represent three independent biological replicates under standard cultivation conditions (not visible).

More relevant for FACS-screenings, however, are fluorescence distributions of the cultures monitored at the single cell level. Therefore, a second experiment investigated dipeptide-supplemented *C. glutamicum* CgHis2 cultures with respect to increased biosensor induction states using FACS analysis and up to 30 mM dipeptide concentrations. In accordance to the first experiment, in which the biosensor response on the culture level was determined, the observed shift of the fluorescent populations indicated higher induction states of the biosensor in L-His-L-Ala-dipeptide-supplemented cultures also at the single cell level (Figure 6).
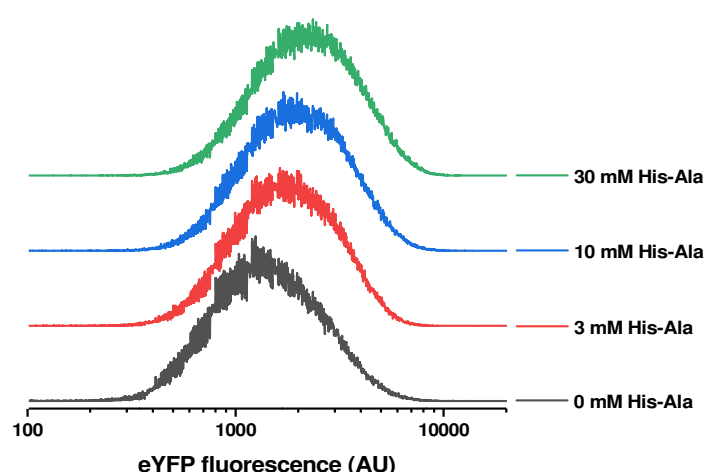
Figure 6: Fluorescence distribution of *C. glutamicum* CgHis2 at the single cell level upon dipeptide supplementation. 0, 3, 10 and 30 mM L-His-L-Ala dipeptides were supplemented to *C. glutamicum* CgHis2 cultivations under standard conditions. In all cases, dipeptides were supplemented to a final concentration of 30 mM using L-Ala-L-Ala dipeptides as co-supplement. After 6 h of cultivation, cultures were analyzed by FACS as if they were used for cell sorting. Thereby a shift of the fluorescence populations from low to high L-His-L-Ala supplementation was observed.

Hence, the pSenHis-based biosensor in *C. glutamicum* CgHis2 could be induced by ʟ-histidine beyond the endogenous *C. glutamicum* CgHis2 production capabilities and should therefore be applicable for the isolation of *C. glutamicum* CgHis2 strain variants with improved ʟ-histidine production.

4.4 Investigation of potential biosensor crosstalk

Biosensor crosstalk represents a major challenge in biosensor-based screenings (Flachbart, Sokolowsky and Marienhagen, 2019). Biosensor crosstalk occurs when less- or non-productive cells can import the respective biosensor-inducing ligand secreted by other cells with higher productivity. Hence, they adopt a fluorescent phenotype similar or identical to overproducing variants. This can lead to the isolation of many false-positives during FACS-screenings. Even though ʟ-histidine import systems are not identified for *C. glutamicum* yet, ʟ-histidine uptake was determined during the characterization of histidine auxotrophic mutants (Kulis-Horn, Persicke and Kalinowski, 2014). Hence, co-cultivations of *C. glutamicum* CgHis2 as producer strain with pSenHis-harboring *C. glutamicum* wild type cells as reporter strain were performed in different mixing ratios in the microtiter plate format and analyzed using FACS over a seven-hour time course (Figure 7 and Appendix Chapter 8.2). In none of the co-cultivations, the reporter strain exhibited significantly increased fluorescence. As a result, biosensor crosstalk was considered irrelevant for this study.
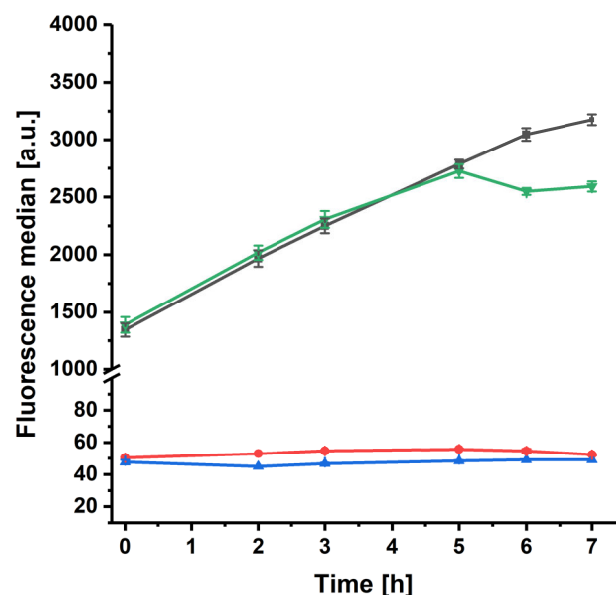
Figure 7: Monitored fluorescence of *C. glutamicum* CgHis2 as producer strain and *C. glutamicum* pSenHis as reporter strain from respective co-cultivations and FACS analysis. The co-cultivation with initial mixing ratio of 1:2 was monitored over a seven-hour time course. The respective populations' fluorescence medians were determined from 100,000 events of the *C. glutamicum* CgHis2 control (green), *C. glutamicum* WT pSenHis control (blue), *C. glutamicum* CgHis2 subpopulation of co-cultivation (black) and *C. glutamicum* WT pSenHis subpopulation of co-cultivation (red).

4.5   Multiplexed random genome mutagenesis of *C. glutamicum* CgHis2

To achieve the isolation of 100 independently evolved *C. glutamicum* CgHis2 strain variants with increased product formation, 100 independent random mutagenesis experiments, pSenHis-based FACS-screenings and strain characterizations were conducted. The isolation of variants from independent mutagenesis batches was considered necessary since growth periods between mutagenesis and FACS-screening can lead to isolation of genetically identical sister clones from the same batch, as it was observed in earlier studies (Binder *et al.*, 2012). One hundred strain variants, on the other hand, were estimated to yield good statistical significance of genetic hotspots, which were expected from subsequent comparative genome analyses. Genetic hotspots were defined as genes, in which randomly induced mutations from a significant number of improved strain variants are clustering.

Therefore, the random mutagenesis protocol using MNNG was modified in order to generate independently mutagenized *C. glutamicum* CgHis2 batches in a multiplexed manner. Since MNNG induces high mortality in treated cultures, which was found to exhibit high variance between separate multiplex runs, a MNNG dilution series was applied to generate mutagenized batches of differing culture mortality. The mortality of treated cultures was subsequently determined by isolating 100 single cells of the whole population randomly using FACS. Prior to this work, a culture mortality of 90 % was considered most beneficial for efficient FACS-based isolation of improved variants (Binder *et al.*, 2012). Along the screening process, improved *C. glutamicum* CgHis2 variants were isolated from independently mutagenized cultures with mortalities ranging 41-99 % with no significant optimum (Appendix Chapter 8.3). In subsequent mutagenesis experiments, the induced culture mortality shifted for no apparent reason, such that lower MNNG concentrations were sufficient to yield appropriately mutagenized cultures. In the end, more than 600 *C. glutamicum* CgHis2 cultures were independently mutagenized, but only 212 cultures were screened at the single cell level due to the high variance of mortality in treated cultures. Before this could be achieved, however, a FACS workflow suitable for the required throughput had to be identified.

4.6   Development of a high-throughput FACS workflow

In order to find most suitable conditions for high-throughput isolation of improved *C. glutamicum* CgHis2 strain variants, multiple variables of a potential FACS workflow were investigated. First, suitable parameters for prior-to-FACS cultivations were determined. Both, cultivation and screening in defined CGXII medium turned out most reliable since resulting FACS phenotypes showed less variance compared to cultivation and screening in undefined BHI medium (not shown). In addition, isolation from equal cultivation conditions as later

production conditions appeared most reasonable, too. Several seedtrain variations such as direct inoculation (from glycerol stocks) or using several precultures for inoculation of the main culture before screening were investigated as well (Appendix Chapter 8.4). As a result, inoculation of the screening culture directly from glycerol stock not only turned out to be the most straight-forward approach, but also ensured minimized dilution of each *C. glutamicum* CgHis2 mutant library due to reduced growth periods between mutagenesis and screening.

Initially, a cultivation of a *C. glutamicum* CgHis2 mutant library was monitored regarding EYFP maturation-dependent formation of fluorescence in the exponential (and stationary) growth phase using hourly FACS samples to determine a suitable screening time window. As a result, cell sorting was considered most beneficial in mid-exponential growth phase when glucose uptake, growth-coupled L-histidine production and (hence) *eyfp* expression are at high rates. At this point in time, differences in genotype-dependent productivity should be reflected in differences in the specific fluorescence at the single cell level. In fact, *C. glutamicum* CgHis2 mutant libraries were typically characterized by an increased fluorescence heterogeneity, such that a ratio could be determined by electronic gating of highly fluorescent cells in FACS analysis and comparison to non-mutated *C. glutamicum* CgHis2 cells as control. This ratio was termed *separation efficiency* (Figure 8). A high separation efficiency was initially thought to reduce the probability of isolating false-positives since only a minority of cells of the control population fell in that gate. This was found most beneficial in exponential growth phase after 5-7 hours of cultivation time, when the culture was inoculated directly from glycerol stock (Figure 8).
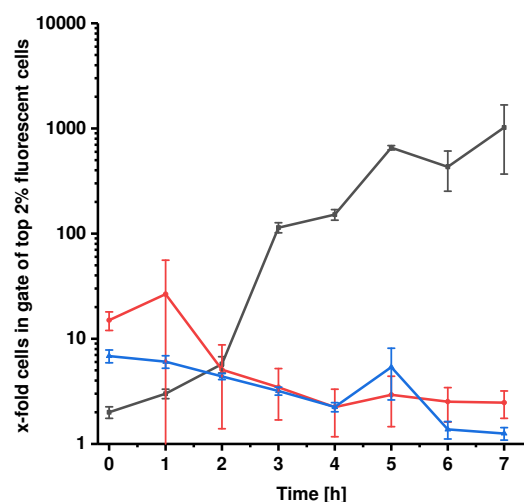
Figure 8: Investigation of the *separation efficiency* of *C. glutamicum* CgHis2 cells in FACS analysis. Depicted are three seedtrain variants of the cultivation prior to FACS analysis: inoculated from glycerol stock (grey), inoculated from agar plate (red) and inoculated following the standard seedtrain including undefined and defined medium precultures (blue). Based on an electronic gate comprising the top 2 % of highly fluorescent cells (of the MNNG-treated batch) in FACS analysis, the ratio of events of the MNNG-treated batch (MNNG$^+$) to its control (MMNG$^-$) in that gate was determined.

This time window was hence used for the screening process. Along this process however, cells with highest fluorescence usually turned not out to be the best L-histidine producers. Mutations acquired on the biosensor module were hypothesized to cause this false-positive phenotype, but not investigated in more detail as this has been already found earlier (Binder *et al.*, 2012).

In addition, the impact of the gating window dimensions on the isolation of significantly improved strain variants was investigated as well. In the end, isolation of the top 5 % of fluorescent cells in the determined time window of 5-7 hours after inoculation proved most reliable as tradeoff for minimizing false-positives, when the separation efficiency was high, and ideally, when the fluorescent outliers (cells with highest fluorescence) were excluded during FACS.

Considering the FACS-screening workflow itself, FACS-based enrichment strategies using repetitive cell sorting were useful in earlier studies for reducing the isolation of false-positives (Cheng *et al.*, 2012; Wagner *et al.*, 2018; Flachbart, Sokolowsky and Marienhagen, 2019). Therefore, a two-step FACS-based enrichment strategy was intended to further eliminate false-positives in screening of *C. glutamicum* CgHis2 cultures. Several workflow variations were investigated concerning combinations of gating strategy, number of cells to be sorted in each FACS step and the type of medium used for re-cultivation and re-screening. Despite many tested conditions, sorting into liquid medium was unsuccessful with regard to the growth characteristics of *C. glutamicum* CgHis2 and the intended throughput since most sorted cultures exhibited multiday-long lag phases or no growth at all. Solving the growth issue by increasing the number of sorted cells from 20,000 to 100,000 was also not successful. Sorting an even higher number, however, was found unfeasible considering the demand of time per FACS screening.

Finally, the second FACS-screening step (considered as enrichment step) was abandoned in favor of a cultivation-based rescreening step, which, due to the FACS-rescreening issues, was less time consuming and easier to perform. Therefore, cell sorting of the top 5 % of highly fluorescent cells was performed by FACS onto solid agar plates such that 240 clones were isolated from each of the 212 screened *C. glutamicum* CgHis2 libraries, accounting for more than 50,000 strain variants in total (Figure 9). Isolated strain variants, which were able to grow on solid undefined medium were randomly picked and subsequently applied to a two-step cultivation workflow for rescreening and characterization of improved *C. glutamicum* CgHis2 variants.
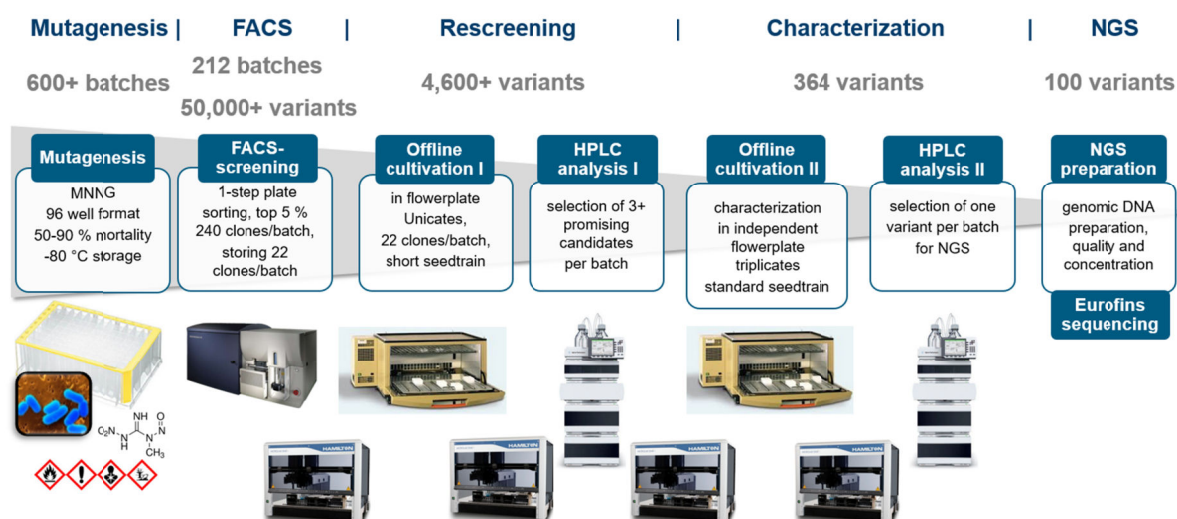
Figure 9: Schematic overview of the screening workflow for the identification of significantly improved ʟ-histidine producing *C. glutamicum* CgHis2 strain variants. More than 600 *C. glutamicum* CgHis2 cultures were treated with MNNG in multiplexed random genome mutagenesis. Subsequently, 212 mutagenized cultures were screened by FACS for variants with increased fluorescence, thereby isolating more than 50,000 *C. glutamicum* CgHis2 strain variants. Then, rescreenings using individual cultivations of more than 4,600 *C. glutamicum* CgHis2 strain variants were performed. Of these, 364 identified strain variants with improved production were subjected to more detailed characterizations to verify the increased production phenotype. For high accuracy, all cultivation and HPLC analysis steps were performed with robotics. Finally, significantly improved ʟ-histidine production was verified for 100 *C. glutamicum* CgHis2 strain variants. Their genomes were sequenced and comparative and combinatorial genome analyses were performed.

4.7 Rescreening and characterization of isolated *C. glutamicum* CgHis2 strain variants

For the realization of a medium- to high-throughput characterization of isolated *C. glutamicum* CgHis2 strain variants, a two-step cultivation workflow was developed for quick analysis of randomly selected variants in the first step (rescreening) and accurate analysis of improved variants in the second step (characterization) (Figure 9). Cultivations were performed with end-point determination of the ʟ-histidine titer and also the culture´s optical density as a measure for biomass formation at microscale. (Ideally) unaltered biomass formation in comparison to the starting variant was considered as additional parameter to predominantly identify novel mutations with a benefit for ʟ-histidine production but only low impact on cell growth.

To achieve high accuracy, cultivation and HPLC analysis steps included robotic sample preparation, since only minor improvements in ʟ-histidine production of the already highly engineered *C. glutamicum* CgHis2 variants were expected. Twenty-two strain variants of every FACS-screened *C. glutamicum* CgHis2 culture were cultivated in the first step (rescreening), accounting for over 4,600 strain variants. From these, 364 strain variants with improved ʟ-histidine production were subjected to a second cultivation for a more detailed characterization of the individual growth and ʟ-histidine production capabilities. Thus, the

proposed 100 independently evolved variants with improved production phenotype could be reliably isolated and characterized (Figure 10 and Appendix Chapter 8.5).
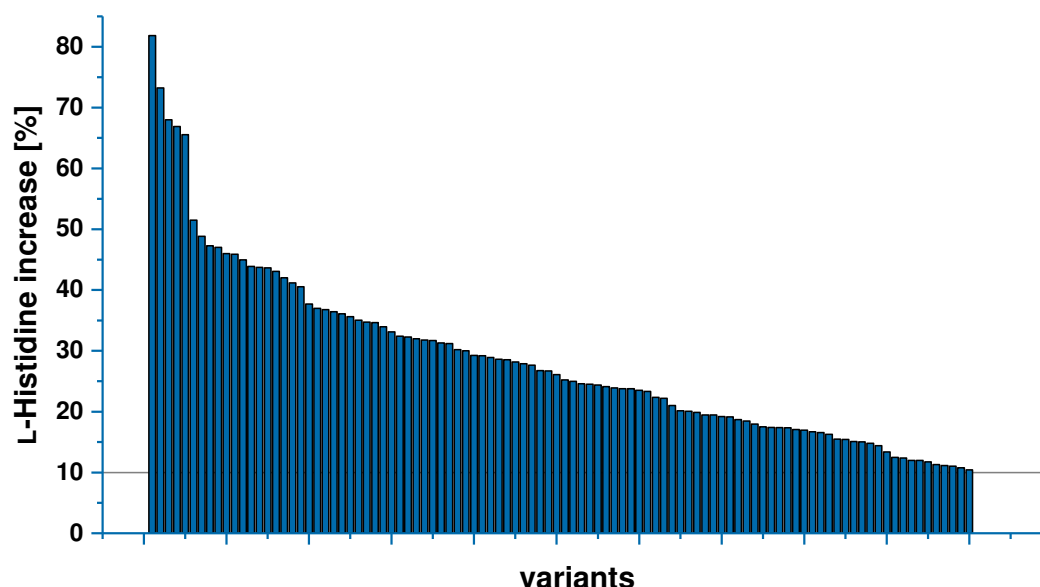

Figure 10: 100 independently evolved *C. glutamicum* CgHis2 strain variants with significantly improved L-histidine production. Every of the 100 *C. glutamicum* CgHis2 strain variants was isolated from an independently mutagenized *C. glutamicum* CgHis2 culture using pSenHis-based FACS-screening and characterized by improved L-histidine production of at least 10 % in comparison the *C. glutamicum* CgHis2 reference strain. Results are collected from multiple characterization runs with technical variance of humidity parameters, such that the relative improvement of L-histidine titers [%] to the respective control is used to compare all isolated variants. All absolute L-histidine titers and corresponding final optical density values, each from at least three independent technical replicates, are attached in the Appendix (Chapter 8.5).

For good statistical significance, only variants with at least 10 % improvement in L-histidine production were considered for whole genome sequencing. In addition, a threshold for biomass formation was set at a maximum decrease of 35 % in optical density (end point determination).

4.8 Genome sequencing and comparative genome analysis

Whole genome sequencing of the 100 improved *C. glutamicum* CgHis2 variants was outsourced to Eurofins SE. For comparative genome analysis and identification of hotspot genes, Dr. Michael Dal Molin as bioinformatician of the group, developed the Fast Automated Analysis of Multiple Sequences (FAAMS) software tool. This tool processed all sequences simultaneously and mapped them to the *C. glutamicum* ATCC 13032 wild type sequence. All *C. glutamicum* CgHis2 single nucleotide polymorphisms (SNPs) of the 100 strain variants were summarized and attributed to the respective coding sequence (gene) or artificial 200 bp promoter region upstream of respective genes, if applicable. Rationally engineered mutations of the *C. glutamicum* CgHis2 reference strain were removed from the data sets such that only newly acquired mutations were displayed. On average, strain variants acquired 180

mutations by random genome mutagenesis with a high variance of 47-470 mutations. The total number of mutations per gene was plotted to illustrate the emerging hotspot genes (Figure 11). Genes with high mutation rates could be linked to the improved production phenotypes of isolated *C. glutamicum* CgHis2 strain variants.
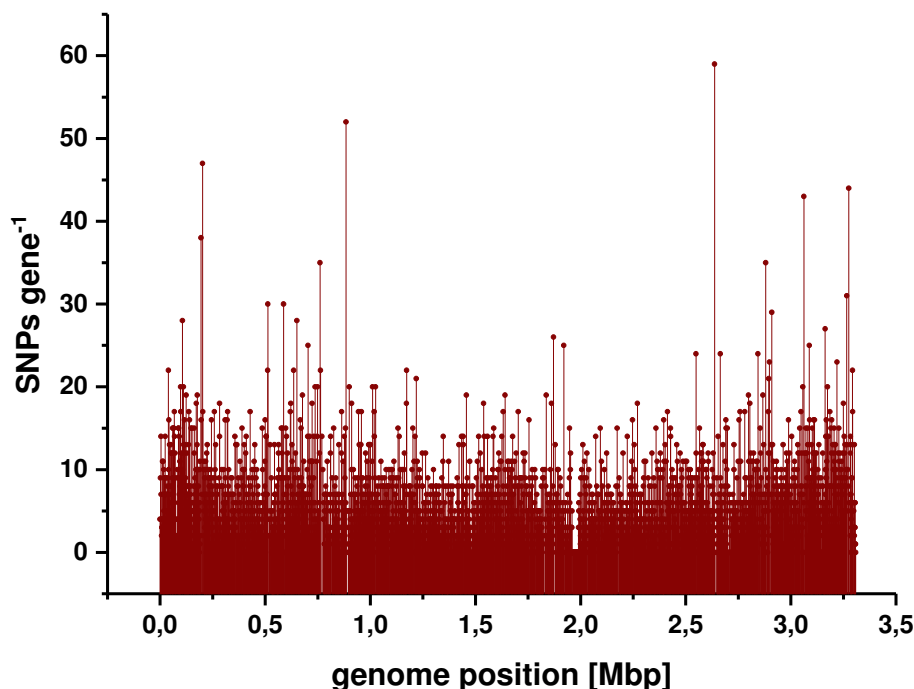
Figure 11: Total number of mutations per gene according to the position in the *C. glutamicum* genome. Synonymous and non-synonymous mutations (SNPs, single nucleotide polymorphisms) from all 100 isolated *C. glutamicum* CgHis2 strain variants with increased L-histidine production phenotype are depicted by needle height of each *C. glutamicum* gene. Genes with high mutation rates may be linked to the improved L-histidine production phenotypes.

4.9 Combinatorial genome analysis: Evaluation of hotspot genes

Having the mutational hotspots identified by comparative genome analysis, the aim of the combinatorial analysis was the identification of mutations in hotspot genes with a high probability of contribution to the observed improved production of L-histidine. Most significant mutations were supposed to be introduced in the *C. glutamicum* CgHis2 reference strain to investigate their individual effects on the overall L-histidine production performance (reverse engineering).

At first, due to the high number of mutations, synonymous mutations were filtered out from the comparative genome analysis data set such that only nonsynonymous mutations, leading to amino acid substitutions, were analyzed. In addition, nonsynonymous mutations with a frequency below 50 % (percentage of sequencing reads exhibiting the mutation) were also excluded. To evaluate the significance of specific hotspot genes, the number of *C. glutamicum* CgHis2 strain variants, which harbor nonsynonymous mutations in the respective gene, was determined and termed *hotspot significance* (Table 2).

Table 1: Most significant hotspot genes in *C. glutamicum* CgHis2 according to the number of variants harboring nonsynonymous mutations in the particular gene (*hotspot significance*).

#	Locus tag	Gene	Enzyme	*Hotspot significance*
1	NCgl2409	*fasB*	fatty acid synthase	24
2	NCgl2964		DEAD/DEAH box helicase	23
3	NCgl0802	*fasA*	fatty acid synthase	22
4	NCgl2773	*pks*	type 1 polyketide synthase	21
5	NCgl0184	*emb*	arabinosyl transferase C	20
6	NCgl0181	*gltB*	alpha subunit of glutamate synthase	19
7	NCgl0552		DNA segregation ATPase	19
8	NCgl2618	*cps*	non-ribosomal peptide synthase	18
9	NCgl0040	*pknB*	serine/threonine protein kinase	17
10	NCgl2959		phosphoesterase	17
11	NCgl1737		hypothetical membrane protein	17
12	NCgl2433	*dinG*	probable ATP-dependent DNA helicase-related protein	15
13	NCgl2981		hypothetical protein	15
14	NCgl0916	*ggtB*	gamma-glutamyltranspeptidase	14
15	NCgl0705		probable ATP-dependent helicase	13

The manual identification process of significant hotspot genes was verified by a subsequent computational processing and analysis of the FAAMS data set by Dr. Stephan Noack (Appendix Chapter 8.6). With the *hotspot significance* number, thresholds for combinatorial analysis were defined. Due to the consecutive nature of the screening and genome sequencing process, the comparative genome analysis data set grew over time and hence thresholds were modified along the process. This led to the final threshold of 10 % of strain variants with nonsynonymous mutations in a specific gene required, so that this particular gene can be considered a significant hotspot. In the end, more hotspot genes than could ultimately analyzed in detail within this study have reached this threshold (Appendix Chapter 8.6). Interestingly, promoter mutations were found very rarely and, in addition, deletions or insertions could not be identified at all.

All mutations of hotspot genes differed in their specific genetic position and amino acid substitution they induced, meaning that no mutation was found more than once. In addition to the *hotspot significance*, three additional criteria were defined to finally select specific mutations for reverse engineering of the *C. glutamicum* CgHis2 starting strain.

i)   *Hotspot abundance*: The number of mutations in other hotspot genes than the one selected in the respective strain variant must be low, such that the probability of contribution of the respective mutation to the increased production performance can be regarded as higher.

ii) *Variant performance*: L-histidine production of the *C. glutamicum* CgHis2 strain variant, from which the selected mutation is deduced, must be high, whereas biomass formation should be comparable to the *C. glutamicum* CgHis2 reference strain.

iii) *Network effects*: Potential effects of the resulting amino acid substitution on the protein-, cellular- and metabolic level were considered if information of the protein in question was available.

Based on these criteria, 30 point mutations from 20 hotspot genes were initially selected for individual reverse engineering of the *C. glutamicum* CgHis2 reference strain (Table 3). Two originally considered mutations in *NCgl1815* and *NCgl1774* were eventually neglected, since these prophage genes were found sequence-identical and shared several identical mutations over many isolated *C. glutamicum* CgHis2 variants. These mutations could not be verified by targeted sequencing and were hence regarded as technical error. As a control, in-frame deletion of the whole prophage area including *NCgl1815* and *NCgl1774* was performed, which resulted in L-histidine titers and biomass formation unchanged to the *C. glutamicum* CgHis2 reference.

Table 3: Single nucleotide polymorphisms from *C. glutamicum* CgHis2 hotspot genes selected for reverse engineering, ordered by *hotspot significance*. From hotspot genes with high *hotspot significance* (>17), a second mutation was selected for individual reverse engineering, if the first re-engineered mutation had no beneficial impact on L-histidine production.

#	Locus tag	*Gene*	Enzyme	*Hotspot significance*	His increase of resp. FACS-isolated variant [%]	Amino acid substitution
1	NCgl2409	*fasB*	fatty acid synthase B	**24**	36%	G2762D
2	NCgl2409	*fasB*	fatty acid synthase B	**24**	24%	G1921E
3	NCgl2964		DEAD/DEAH box helicase	**23**	21%	E512K
4	NCgl2964		DEAD/DEAH box helicase	**23**	26%	P863S
5	NCgl0802	*fasA*	fatty acid synthase A	**22**	38%	A2702T
6	NCgl0802	*fasA*	fatty acid synthase A	**22**	36%	P783S
7	NCgl2773	*pks*	type 1 polyketide synthase	**21**	18%	A1525V
8	NCgl2773	*pks*	type 1 polyketide synthase	**21**	31%	D1186N
9	NCgl0184	*emb*	arabinosyl transferase C	**20**	32%	T529I T539I
10	NCgl0184	*emb*	arabinosyl transferase C	**20**	47%	G477E
11	NCgl0181	*gltB*	alpha subunit of glutamate synthase	**19**	68%	P988S
12	NCgl0181	*gltB*	alpha subunit of glutamate synthase	**19**	19%	G1106D
13	NCgl0552		type VII secretion protein eccC / DNA segregation ATPase	**19**	11%	P823S
14	NCgl0552		type VII secretion protein eccC / DNA segregation ATPase	**19**	29%	G432D
15	NCgl2618	*cps*	non-ribosomal peptide synthase	**18**	34%	G987D

Table 3 continued: Single nucleotide polymorphisms from *C. glutamicum* CgHis2 hotspot genes selected for reverse engineering, ordered by *hotspot significance*. From hotspot genes with high *hotspot significance* (>17), a second mutation was selected for individual reverse engineering, if the first re-engineered mutation had no beneficial impact on L-histidine production.

#	Locus tag	*Gene*	Enzyme	*Hotspot significance*	His increase of resp. FACS-isolated variant [%]	Amino acid substitution
16	NCgl2959		phosphoesterase	**17**	43%	D1453N
17	NCgl2959		phosphoesterase	**17**	43%	G870D
18	NCgl2981		hypothetical protein	**15**	45%	D735G
19	NCgl0705		probable ATP-dependent helicase	**13**	34%	S1847N
20	NCgl0098	*putA*	proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase	**13**	15%	P217S
21	NCgl2633	*mrpA*	NADH ubiquinone oxidoreductase subunit 5 (chain L)/multisubunit Na+/H+ antiporter, A subunit	**12**	45%	L42F
22	NCgl0159	*iolD*	putative acetolactate synthase protein	**12**	31%	S481F
23	NCgl2859		probable cation-transporting ATPase transmembrane protein	**12**	16%	S372F
24	NCgl2809	*pyk2*	pyruvate kinase 2	**11**	21%	T357I
25	NCgl2789		hypothetical protein	**10**	18%	S265N
26	NCgl0111	*xylB*	xylulose kinase	**10**	16%	G55R
27	NCgl0659	*pyc*	pyruvate carboxylase	**10**	35%	A764V
28	NCgl2933	*ulaA*	ascorbate-specific PTS system enzyme IIC	**10**	28%	V219I

4.10 Reverse engineering of *C. glutamicum* CgHis2

Twenty-eight point mutations were introduced individually in the *C. glutamicum* CgHis2 reference strain (Table 3). Among these, three point mutations, one in the non-ribosomal peptide synthetase encoding gene *cps*, one in mycolic acid synthesis encoding gene *pks* and one in the unknown *NCgl2981* gene, were found to increase L-histidine production in *C. glutamicum* CgHis2 (Figure 12).

A key enzyme of mycolic acid synthesis in *C. glutamicum* is a polyketide synthase encoded by *pks* (Takeno *et al.*, 2018; Toyoda and Inui, 2018). The reverse engineered D1186N substitution in Pks increased L-histidine production in *C. glutamicum* CgHis2 by 8 % with p=0.06, according to a student's t-test. Hence, independent repetitions of the cultivation were performed, which led to increased L-histidine titers by 6 % and 9 % with p=0.01 and p=0.02, respectively.

Non-ribosomal peptide synthetases are known to synthesize secondary metabolites such as antibiotics (Strieker, Tanović and Marahiel, 2010). In mycobacteria, however, these enzyme complexes are involved in synthesis of cell wall components (Chalut, 2016). In

C. glutamicum, the *cps* gene (*NCgl2618*) encodes the only non-ribosomal peptide synthetase of this organism with so far unknown cellular function. The G987D substitution in Cps increased L-histidine production in *C. glutamicum* CgHis2 significantly by 12 % with p=0.009, according to a student's t-test (Figure 12).
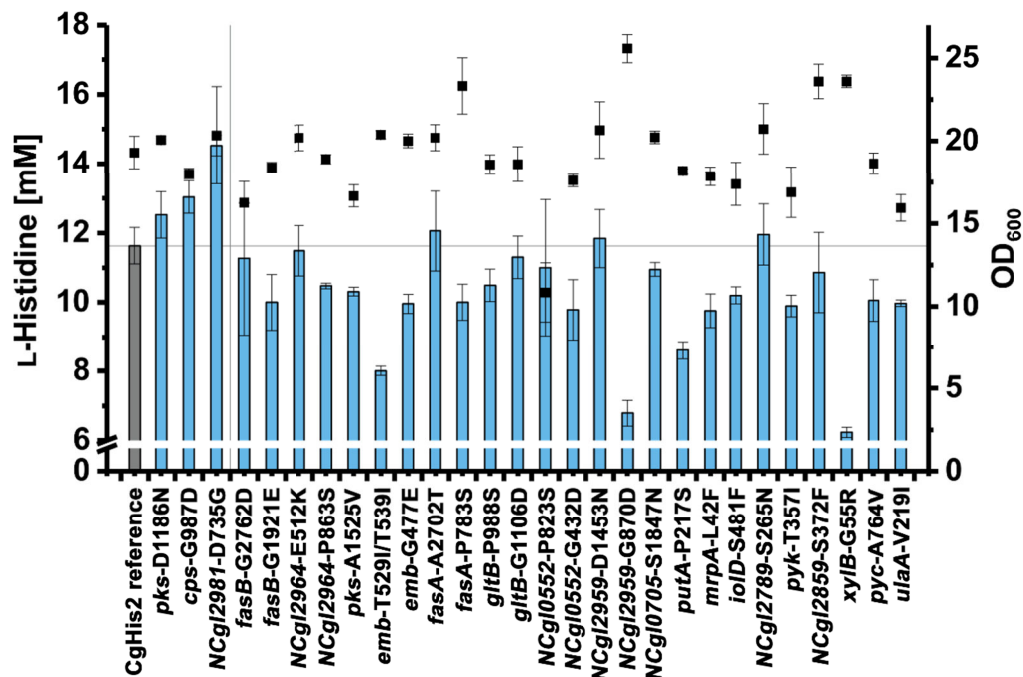

Figure 12: L-histidine accumulation in the supernatant of 28 reverse engineered *C. glutamicum* CgHis2 variants. The introduced point mutations resulted in the depicted amino acid substitutions in the respective *C. glutamicum* proteins. The substitutions *pks*-D1186N, *cps*-G987D and *NCgl2981*-D735G increased L-histidine production in *C. glutamicum* CgHis2.

NCgl2981 presumably encodes a hypothetical membrane glycoprotein. Its structure or cellular function in *C. glutamicum* is unresolved yet. Accessible databases could not reveal any other information on this gene. The reverse engineered D735G substitution increased L-histidine production in *C. glutamicum* CgHis2 by 26 % with p<0.0001, according to a student's t-test (Figure 12). Interestingly, this substitution position was predicted on an outside loop of the protein by AlphaFold (Figure 13) (Jumper *et al.*, 2021).
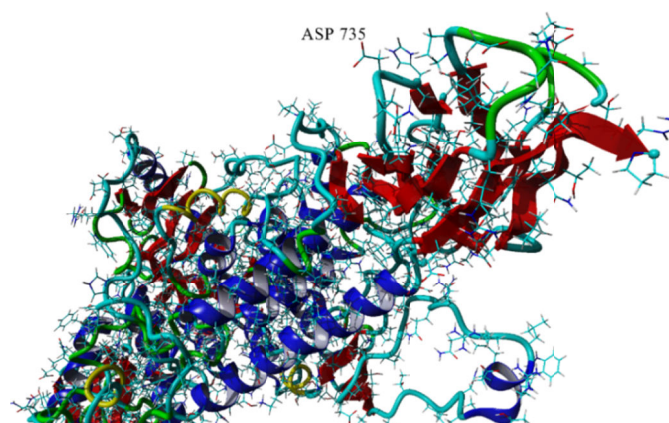

Figure 13: Structure prediction of the NCgl2981 protein and its D735 substitution. Structure calculated with AlfaFold (Jumper *et al.*, 2021).

4.11 In-frame deletion of selected hotspot genes

Random mutagenesis is expected to cause detrimental effects on gene function with higher probability than gain-of-function effects (Foster, 1991). In case of MNNG, a strong mutational bias towards GC→AT transitions has been described (Ohta, Watanabe-Akanuma and Yamagata, 2000; B Singer and Kusmierek, 2003; Ohnishi *et al.*, 2008). Transitions are often regarded as more conservative mutations as they either induce no amino acid substitution (silent mutations) or only substitutions to similar amino acids (Stoltzfus and Norris, 2016). Besides the many silent mutations (synonymous SNPs) removed from the dataset, many conservative amino acid substitutions such as L-alanine → L-valine were found (Figure 14).
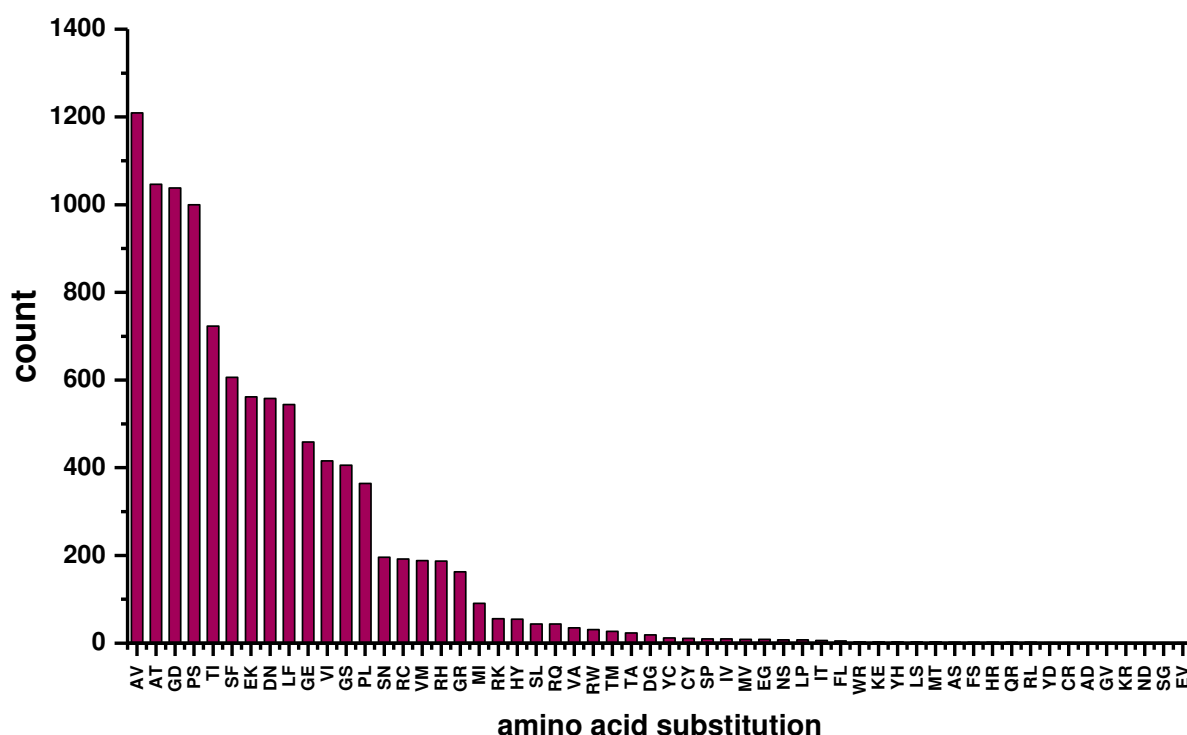
Figure 14: Total number and type of amino acid substitutions from all nonsynonymous mutations of 100 isolated *C. glutamicum* CgHis2 strain variants with improved L-histidine production. Random MNNG mutagenesis induced mutations leading to the depicted amino acid substitution distribution. Besides conservative amino acid substitutions such as L-alanine → L-valine (AV), many substitutions with significant steric, polar or conductive change such as L-glycine → L-aspartate (GD), L-proline → L-serine (PS), L-serine → L-phenylalanine (SF) or L-glutamate → L-lysine (EK) were identified.

Interestingly, also a high number of more drastic amino acid substitutions such as L-glycine → L-aspartate, L-proline → L-serine, L-serine → L-phenylalanine or L-glutamate → L-lysine could be deduced from the 100 *C. glutamicum* CgHis2 genome dataset (Figure 14). Consequently, in-frame deletions of significant hotspot genes were performed to investigate the individual maximum loss-of-function effect on the L-histidine production performance in *C. glutamicum* CgHis2. From 18 in-frame deletions desired, 13 *C. glutamicum* CgHis2 deletion strains could be generated (Table 4).

Table 4: *C. glutamicum* CgHis2 strains with respective in-frame deletions and accumulated L-histidine titer in microtiter plate cultivations.

#	*C. glutamicum* strain	Locus tag	Enzyme	His-Titer [mM]	His-Titer [%]
1	CgHis2 reference			11,5 ± 0.6	-
2	CgHis2 *ΔfasB*	NCgl2409	fatty acid synthase	14.3 ± 0.4	+24
3	CgHis2 *ΔNCgl2964*	NCgl2964	DEAD/DEAH box helicase	10.6 ± 0.3	-8
4	CgHis2 *ΔfasA*	NCgl0802	fatty acid synthase	no deletion possible	
5	CgHis2 *Δpks*	NCgl2773	type 1 polyketide synthase	no deletion achieved	
6	CgHis2 *Δemb*	NCgl0184	arabinosyl transferase C	no pSenHis[*hisEG*] transformation achieved	
7	CgHis2 *ΔgltB*	NCgl0181	alpha subunit of glutamate synthase	11.6 ± 0.2	0
8	CgHis2 *ΔNCgl0552*	NCgl0552	DNA segregation ATPase	11.1 ± 0.6	-4
9	CgHis2 *Δcps*	NCgl2618	non-ribosomal peptide synthase	13.6 ± 0.2	+18
10	CgHis2 *ΔpknB*	NCgl0040	eukaryotic-type serine/threonine kinase	no deletion achieved	
11	CgHis2 *ΔNCgl2959*	NCgl2959	phosphoesterase	10.9 ± 0.2	-6
12	CgHis2 *ΔNCgl1737*	NCgl1737	hypothetical membrane protein	10.8 ± 0.3	-8
13	CgHis2 *ΔNCgl2981*	NCgl2981	hypothetical protein	no deletion achieved	
14	CgHis2 *ΔggtB*	NCgl0916	probable gamma-glutamyltranspeptidase precursor PR	10.3 ± 0.4	-12
15	CgHis2 *ΔNCgl0705*	NCgl0705	probable ATP-dependent helicase	11.8 ± 0.5	+2
16	CgHis2 *ΔputA*	NCgl0098	proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase	9.7 ± 0.4	-17
17	CgHis2 *Δpyk2*	NCgl2809	pyruvate kinase	10.6 ± 0.3	-8
18	CgHis2 *Δpyk1*	NCgl2008	pyruvate kinase	12.1 ± 0.4	+5

Fatty acid synthase gene *fasA* is known to be essential in *C. glutamicum* and hence an in-frame deletion was not attempted because the resulting need for an oleic acid-supplementation for cell growth (and L-histidine production) was considered unfeasible for any future amino acid production process (Radmacher *et al.*, 2005). In-frame deletion of mycolic acid synthesis key player *pks* and kinase gene *pknB* could not be achieved in *C. glutamicum* CgHis2, even though a successful deletion was reported for other *C. glutamicum* strains (Gande *et al.*, 2004; Schultz *et al.*, 2009). Interestingly, *NCgl2981*-D735G could be reverse engineered whereas deletion of *NCgl2981* was not successful in *C. glutamicum* CgHis2. Furthermore, deletion of *emb,* key player of arabinogalactan synthesis in *C. glutamicum*, was successful in *C. glutamicum* CgHis2, but plasmid-based *hisEG* overexpression lead to significant growth defects such that no characterization of this strain could be performed (Alderwick *et al.*, 2005).

Pyruvate kinase is known as key enzyme in *C. glutamicum* for regulation of glycolytic flux and phosphoenolpyruvate (PEP) metabolism in response to the energy state of a cell (Sawada *et al.*, 2010, 2015). Besides *pyk1* (*NCgl2008*), another pyruvate kinase gene, *pyk2* (*NCgl2809*), was more recently discovered in *C. glutamicum* and found highly expressed

under oxygen-deprived conditions (Chai *et al.*, 2016). Since *pyk2* was identified in the comparative genome analysis of *C. glutamicum* CgHis2 strain variants, the respective deletion strain was generated but characterized by decreased L-histidine production (Figure 15). Hence, *pyk1*, which is known to be expressed under the cultivation conditions used for L-histidine production, was deleted in *C. glutamicum* CgHis2 (Sawada *et al.*, 2015). Interestingly, *C. glutamicum* CgHis2 Δ*pyk1* was characterized by improved L-histidine production, whereas the double mutant *C. glutamicum* CgHis2 Δ*pyk1* Δ*pyk2* was not. According to a student's t-test, the depicted 5 % increase in L-histidine titer was not significant ($p=0.096$). However, the improved L-histidine production phenotype could be observed over repeated cultivations.
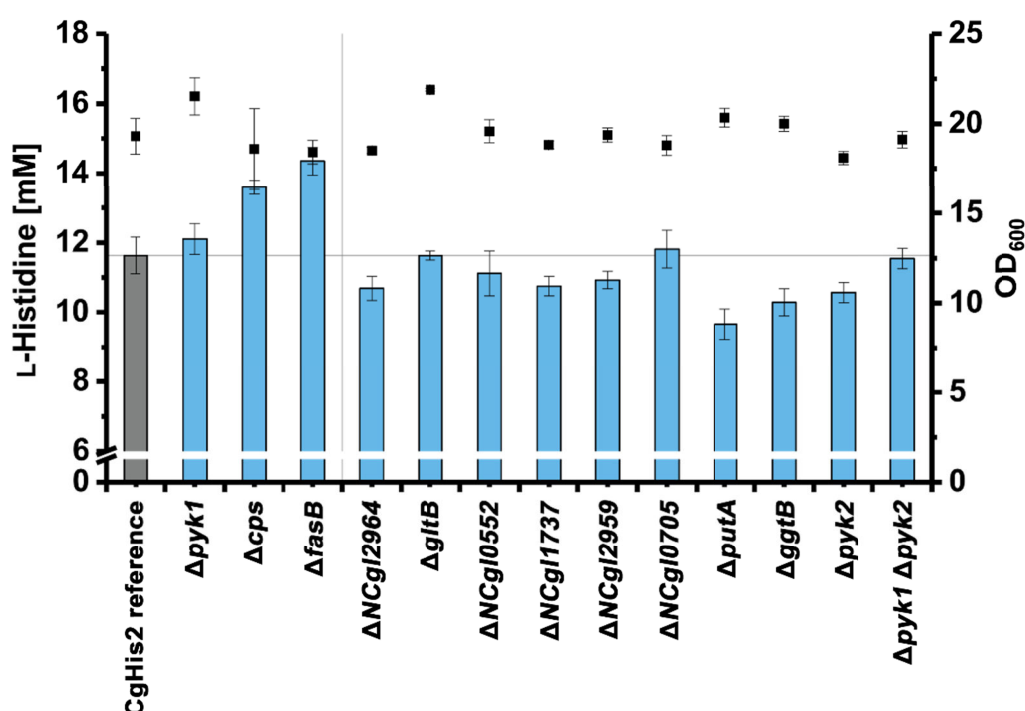
Figure 15: Production performance of *C. glutamicum* CgHis2 strains harboring in-frame deletions of significant hotspot genes. *C. glutamicum* CgHis2 strains with either Δ*pyk*, Δ*cps* or Δ*fasB* strains were characterized by improved L-histidine production.

In comparison to the identified beneficial mutation in *cps* (inducing *cps*-G987D and increasing L-histidine production by 12 %), the generated *C. glutamicum* CgHis2 Δ*cps* strain was characterized by 18 % increased L-histidine production. Hence, the *C. glutamicum* CgHis2 *cps*-G987D variant harbored a less productive loss-of-function mutation.

In contrast to *fasA*, *fasB* encodes the non-essential fatty acid synthase of *C. glutamicum*, primarily producing palmitate (Radmacher *et al.*, 2005). In the comparative genome analysis performed, *fasB* was identified as most significant hotspot gene with nonsynonymous mutations present in 24 % of *C. glutamicum* CgHis2 strain variants. Whereas the reverse engineered G2762D and G1921E substitutions led to decreased L-histidine titers,

C. glutamicum CgHis2 Δ*fasB* was characterized by 24 % improved L-histidine production (Figure 15).

In summary, three beneficial point mutations (*cps*-G987D, *pks*-D1186N, *NCgl2981*-D735G) and three beneficial in-frame deletions (Δ*pyk1,* Δ*cps,* Δ*fasB*) were identified to increase L-histidine production in *C. glutamicum* CgHis2. Omitting *cps*-G987D, further strain engineering aimed at combining beneficial mutations to search for potentially additive effects and hence obtain considerably improved L-histidine producers.

4.12 Combinatorial effects of individual mutations and gene deletions

To investigate any combinatorial effects of the beneficial mutations identified, several *C. glutamicum* CgHis2 double mutants were constructed. Interestingly, Δ*fasB* turned out as key engineering target since introduction of Δ*pyk1*, Δ*cps* or *pks*-D1186N in the background of the *C. glutamicum* CgHis2 Δ*fasB* variant led to improved L-histidine titers in comparison to the single mutant strains (Figure 16).
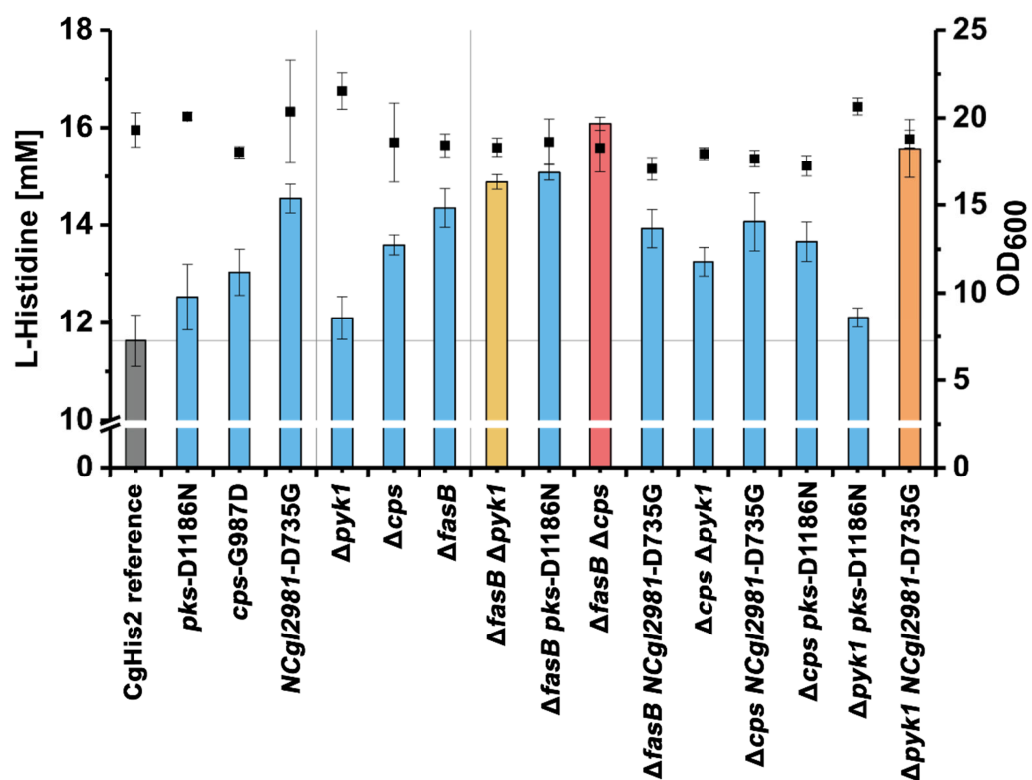
Figure 16: Production performance of *C. glutamicum* CgHis2 single mutant and double mutant strains harboring beneficial mutations and in-frame deletions of significant hotspot genes. The beneficial mutations *pks*-D1186N, *NCgl2981*-D735G and beneficial in-frame deletions Δ*pyk1*, Δ*cps* and Δ*fasB* were combined to several *C. glutamicum* CgHis2 double mutant strains. Only Δ*fasB* Δ*pyk1*, Δ*fasB pks*-D1186N, Δ*fasB* Δ*cps* and Δ*pyk1 NCgl2981*-D735G were characterized by improved L-histidine production in *C. glutamicum* CgHis2 in comparison to the respective single mutant strains. Δ*fasB* Δ*pyk1* (yellow), Δ*fasB* Δ*cps* (red) and Δ*pyk1 NCgl2981*-D735G (orange) were employed as *C. glutamicum* CgHis2 starting variants for further strain engineering.

C. glutamicum CgHis2 Δ*fasB* Δ*pyk1*, *C. glutamicum* CgHis2 Δ*fasB pks*-D1186N and *C. glutamicum* CgHis2 Δ*fasB* Δ*cps* were characterized by increased L-histidine titers of 29 %, 30 % and 40 % with regard to the reference strain, respectively, whereas biomass formation of these strains was reduced only by 5 %. Conversely, *C. glutamicum* CgHis2 Δ*pyk1* *NCgl2981*-D735G was characterized by a 34 % improvement in L-histidine titer, which was higher than expected (based on the positive assumption that the effects are simply additive).

In addition, 41 % of isolated *C. glutamicum* CgHis2 strain variants harboring nonsynonymous mutations in *fasA* were found to also harbor nonsynonymous mutations in *fasB*. Several *fasAB* double mutant construction attempts were performed, however, none of them was successful.

Next, the best-performing double mutant *C. glutamicum* CgHis2 Δ*fasB* Δ*cps* was used as starting variant for the introduction of additional mutations. However, this strain lineage showed no additional improvements (Figure 17).
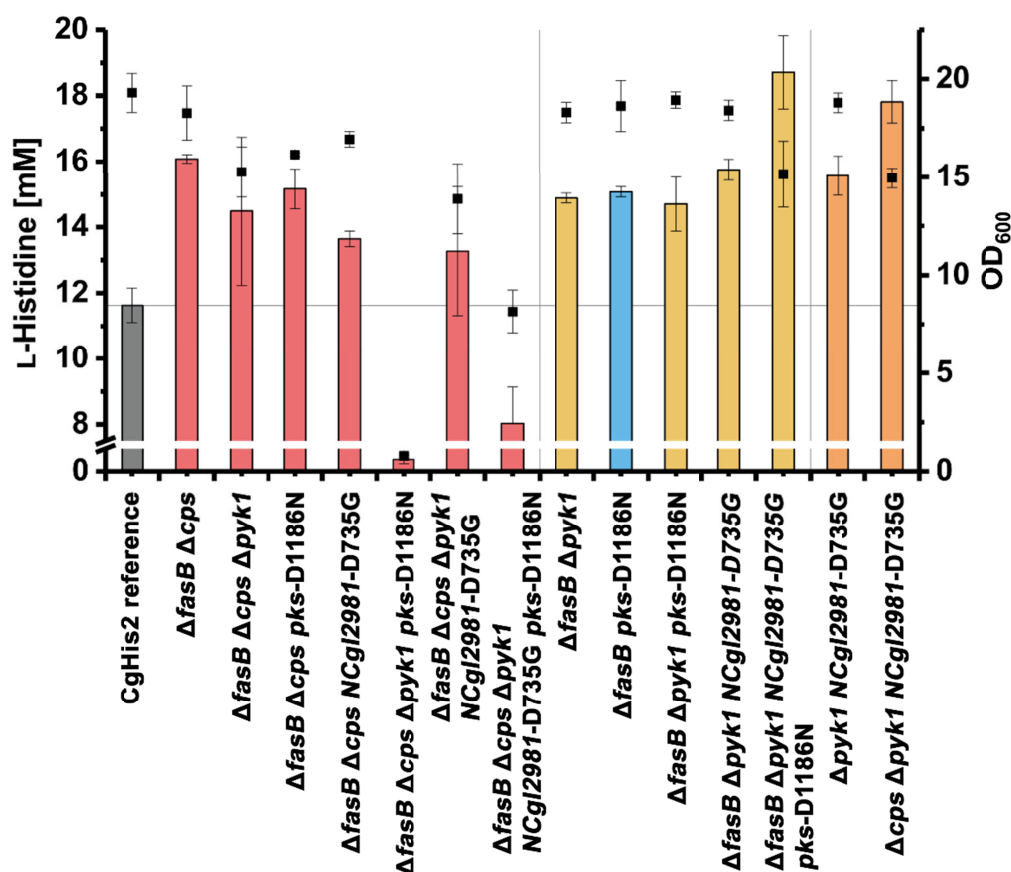
Figure 17: Production performance of *C. glutamicum* CgHis2 strains harboring multiple combinations of beneficial mutations and in-frame deletions. *C. glutamicum* CgHis2 Δ*fasB* Δ*cps* was employed as starting variant for derived *C. glutamicum* CgHis2 triple, quadruple and quintuple mutants (red strain lineage), but no further improvement in L-histidine production could be achieved. Introduction of *NCgl2981*-D735G and *pks*-D1186N in *C. glutamicum* CgHis2 Δ*fasB* Δ*pyk1* (yellow strain lineage) resulted in the most productive *C. glutamicum* CgHis2 strain, characterized by 69 % improved L-histidine titer. In-frame deletion of *cps* in *C. glutamicum* CgHis2 Δ*pyk1* *NCgl2981*-D735G (orange strain lineage) yielded a total improvement of 51 % in L-histidine titer.

Instead, *C. glutamicum* CgHis2 Δ*fasB* Δ*pyk1* turned out to be the best-performing *C. glutamicum* CgHis2 strain after introducing *NCgl2981*-D735G and *pks*-D1186N. Interestingly, only the combinatorial effect of these point mutations resulted in significantly improved production of 69 % in L-histidine titer. Biomass formation of *C. glutamicum* CgHis2 Δ*fasB* Δ*pyk1 NCgl2981*-D735G *pks*-D1186N, on the other hand, was reduced by 21 % (Figure 17).

Based on the other double mutant *C. glutamicum* CgHis2 Δ*pyk1 NCgl2981*-D735G, in-frame deletion of *cps* generated the best-performing triple mutant *C. glutamicum* CgHis2 Δ*cps* Δ*pyk1 NCgl2981*-D735G, which was characterized by a 51 % increase in L-histidine titer and 23 % decrease in biomass (Figure 17). Since this strain lineage took more time for construction, the final introduction of *pks*-D1186N could not be achieved in time. The respective quadruple strain *C. glutamicum* CgHis2 Δ*cps* Δ*pyk1 NCgl2981*-D735G *pks*-D1186N, however, would have been an interesting comparison to the best-performing variant, *C. glutamicum* CgHis2 Δ*fasB* Δ*pyk1 NCgl2981*-D735G *pks*-D1186N.

Apparently, simultaneous in-frame deletion of *fasB* and *cps* is detrimental for L-histidine production in *C. glutamicum* CgHis2. The combinatorial effect of Δ*fasB*, Δ*cps* and *pks*-D1186N induced significant growth defects, such that only strain lineages harboring either Δ*fasB* or Δ*cps* led to superior *C. glutamicum* CgHis2 production strains.

4.13 Lab-scale bioreactor fermentations

To investigate the suitability of isolated and engineered *C. glutamicum* CgHis2 variants for L-histidine production in a bioreactor setup, batch fermentations in a 0.5 L laboratory-scale format were performed. Hereby, the best reverse engineered quadruple variant *C. glutamicum* CgHis2 Δ*fasB* Δ*pyk1 NCgl2981*-D735G *pks*-D1186N was cultivated and characterized in comparison to the *C. glutamicum* CgHis2 reference strain (Figure 18). In addition, the *C. glutamicum* CgHis2 12-10-5-6 strain variant, as best isolate among the 100 randomly mutagenized and FACS-screened *C. glutamicum* CgHis2 variants (Figure 10) was cultivated and characterized as well. This variant acquired 120 SNPs during MNNG mutagenesis. From these experiments, growth rates, D-glucose uptake rates and L-histidine production rates were determined via bioprocess modelling (Table 5).

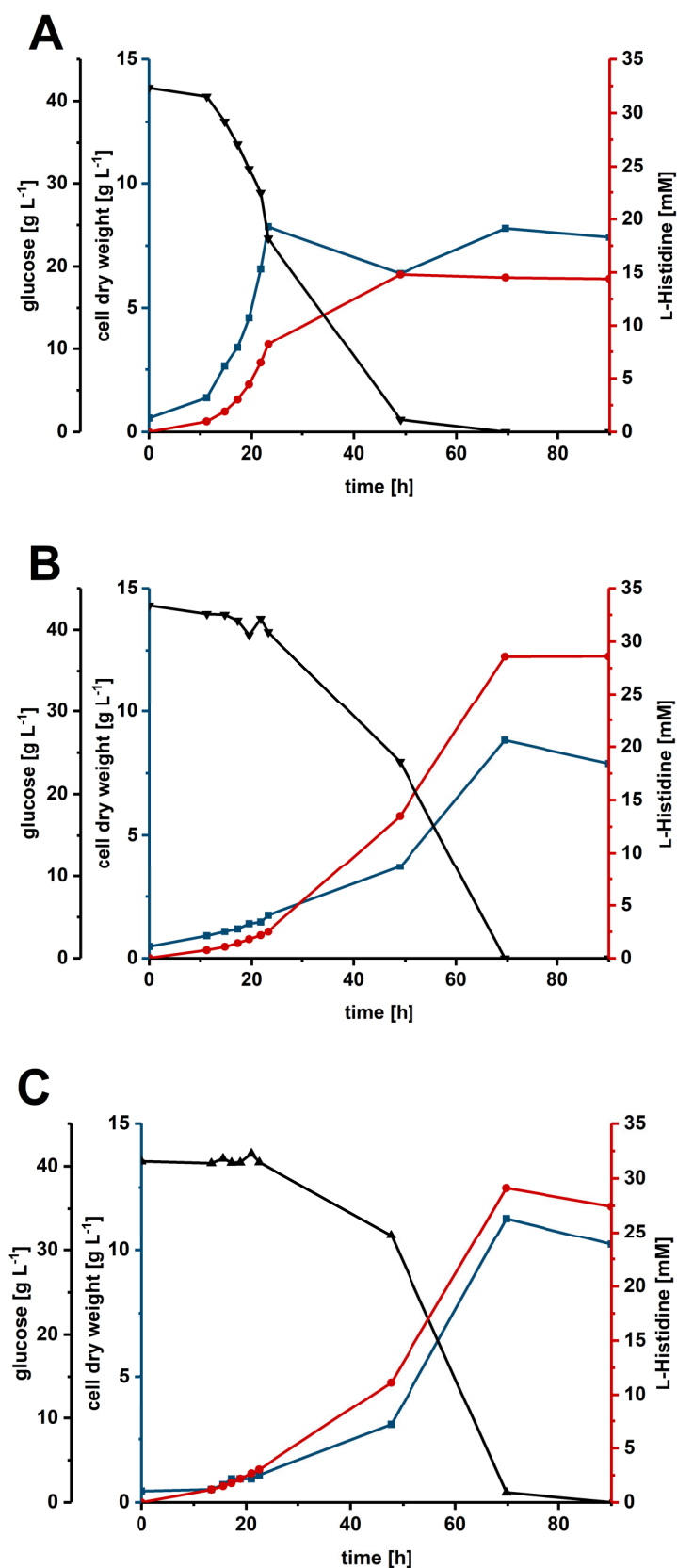**Figure 18: Production performance of (A) the *C. glutamicum* CgHis2 reference strain, (B) the FACS-isolated *C. glutamicum* CgHis2 12-10-5-6 variant and (C) the reverse engineered *C. glutamicum* CgHis2 Δ*fasB* Δ*pyk1 NCgl2981*-D735G *pks*-D1186N variant in lab-scale batch fermentations.** Depicted are the glucose concentration (black), the cell dry weight (blue) and the L-histidine titer (red) over 90 h fermentation time.

Table 5: Key Performance Indicators derived from lab-scale batch fermentations and bioprocess modelling for the *C. glutamicum* CgHis2 reference strain, the *C. glutamicum* CgHis2 12-10-5-6 variant and the *C. glutamicum* CgHis2 *ΔfasB Δpyk1* NCgl2981-D735G *pks*-D1186N variant. For bioprocess modelling, a parametric bootstrapping approach was used to estimate asymmetric confidence bounds. Results of the individual bioreactor cultivations are depicted in brackets.

C. glutamicum strain	CgHis2 reference	CgHis2 variant 12-10-5-6	CgHis2 *ΔfasB Δpyk1* NCgl2981-D735G *pks*-D1186N
OD_{600} [-]	30.4	30.6	41.5
CDW [g L^{-1}]	8.3	8.8	11.3
L-histidine titer [mM]	14.8	28.6	29.1
L-histidine yield [C-mol C-mol^{-1}]	0.07	0.13	0.13
Growth rate [h^{-1}]	0.15	0.06	0.05
	[0.155, 0.160]	[0.056, 0.058]	[0.049, 0.052]
D-glucose uptake rate	1.96	1.28	1.07
[mmol g_{CDW}^{-1} h^{-1}]	[1.960, 2.139]	[1.277, 1.383]	[1.078, 1.137]
L-histidine production rate	0.13	0.14	0.17
[mmol g_{CDW}^{-1} h^{-1}]	[0.131, 0.142]	[0.140, 0.154]	[0.175, 0.188]

Surprisingly, the isolated strain *C. glutamicum* CgHis2 12-10-5-6 and the reverse engineered strain *C. glutamicum* CgHis2 *ΔfasB Δpyk1* NCgl2981-D735G *pks*-D1186N produced similar amounts of L-histidine from 4 % D-glucose as sole source of carbon and energy (28.6 mM and 29.1 mM, respectively). This represented an almost doubled product titer compared to the already highly engineered *C. glutamicum* CgHis2 reference strain, which accumulated up to 14.8 mM L-histidine (Figure 18). Both variants grew with significantly reduced growth rates and glucose uptake rates (Table 5). Interestingly, *C. glutamicum* CgHis2 *ΔfasB Δpyk1* NCgl2981-D735G *pks*-D1186N formed significantly more biomass (11.3 g L^{-1}) than *C. glutamicum* CgHis2 12-10-5-6 (8.8 g L^{-1}), which grew to cell densities comparable to the *C. glutamicum* CgHis2 reference strain (8.3 g L^{-1}). Due to the slow growth of both variants, the L-histidine production rates were only slightly improved, but L-histidine was produced over a longer period such that significantly more glucose was converted to L-histidine. This resulted in an almost doubled L-histidine yield of 0.13 C-mol C-mol^{-1} for both variants, even though *C. glutamicum* CgHis2 *ΔfasB Δpyk1* NCgl2981-D735G *pks*-D1186N also produced 28 % more biomass than *C. glutamicum* CgHis2 12-10-5-6. Hence, the beneficial combinations of identified beneficial point mutations and gene deletions in *C. glutamicum* CgHis2 even lead to an increased performance in the lab-scale bioreactor setup compared to the microtiter plate format.

5 Discussion

The challenges of targeted or rational metabolic engineering campaigns are often based on the limited knowledge of the complex interactions of global metabolic, regulatory and signaling networks, or yet unknown gene functions in microbial metabolism (Lee and Kim, 2015). Especially in already highly engineered industrial strains, the identification of novel beneficial engineering targets represents a major bottleneck to push industrial strains towards maximum product yields. Here, directed evolution offers big potential for improving these industrial strains, since a large number of modifications with a potential benefit can be induced by random genome mutagenesis. Since the probability of beneficial mutations is very low, the identification of only a few improved variants in vast randomly mutated libraries represents the major challenge. However, the immense genetic space of microbial genomes can be explored by a combination of random genome mutagenesis, FACS-based high-throughput screening, comparative genome analysis and reverse engineering.

In this thesis, microbial L-histidine production was selected as exemplary strain engineering goal, since industrial L-histidine production strains are already highly engineered but inefficient in production since engineering of the highly intertwined L-histidine biosynthesis pathway network represents a major challenge (Cheng *et al.*, 2013b; Kulis-Horn, Persicke and Kalinowski, 2014, 2015; Schwentner *et al.*, 2019; Feith *et al.*, 2020; Wu *et al.*, 2020b). The L-histidine production strain used as starting variant was already engineered, taking advantage of known metabolic engineering targets. The strain was able to convert D-glucose to L-histidine with a product yield of 0.07 C-mol C-mol^{-1} (or 0.06 g g^{-1}) in lab-scale batch fermentations. Hence, high improvement capacity was considered with regard to the theoretical maximum product yield of 0.44 g L-histidine g D-glucose^{-1} (Schwentner *et al.*, 2019).

In preceding studies on L-histidine production, much attention has been paid to the feedback-inhibited and thus rate-limiting key enzyme of the L-histidine biosynthesis pathway, HisG. With the proof of the applicability of pSenHis for isolation of L-histidine-producing strain variants in the first part of this thesis, known beneficial mutations in *hisG* were identified. Targeted amino acid positions such as S143, D213, G230, S232, G233 and T235 were in accordance to a preceding biosensor-based screening study as well as other *hisG*-targeting approaches (Zhang *et al.*, 2012; Schendzielorz, Dippong, Grünberger, *et al.*, 2014; Kulis-Horn, Persicke and Kalinowski, 2015). Hence, the subsequently used industrial production strain *C. glutamicum* CgHis2 (as screening host), has been equipped with the gene for a feedback-resistant HisG variant (bearing substitution S143F and a deletion of the

C-terminus) that was additionally overexpressed. Interestingly, additional *hisG* mutations were found by comparative genome analyses (FAAMS) in 9 % of the isolated 100 *C. glutamicum* CgHis2 strain variants. Besides T228, other amino acid positions were identified such as A20, L61, A152, E171, T190, D221, A223 and M250, indicating that further improvements of HisG function for L-histidine production could have been attained. However, considering the number of other apparently novel hotspot genes and the many HisG-related studies already performed, those *hisG* mutations were not of interest for this study.

With the overexpression of all L-histidine biosynthesis genes in *C. glutamicum* CgHis2, only few nonsynonymous mutations were found in these genes of the 100 isolated *C. glutamicum* CgHis2 strain variants, such as *hisD* and *hisN* in 4 %, *hisC* and *hisA* in 3 % and *HisH*, *hisF* and *hisE* in 1 % of variants. This indicated, that the potential for additional beneficial modifications within the L-histidine pathway was majorly exhausted in *C. glutamicum* CgHis2 (Schwentner *et al.*, 2019; Wu *et al.*, 2020a). Mutation frequency in other known engineering targets for L-histidine production such as purine biosynthesis was also found low with *purA* in 2 % and *purB* and *purH* in 5 % of the isolated *C. glutamicum* CgHis2 strain variants (Malykh *et al.*, 2018; Schwentner *et al.*, 2019; Feith *et al.*, 2020; Wu *et al.*, 2020a). Since no heterologous glycine-cleavage system was expressed in *C. glutamicum* CgHis2, one might expect an increased mutation frequency in *glyA*, encoding serine hydroxymethyltransferase. This gene is essential for providing C_1-compunds (mTHF, fTHF) needed for conversion of L-histidine pathway intermediate AICAR towards purine and hence ATP biosynthesis as L-histidine precursor (Kulis-Horn, Persicke and Kalinowski, 2014; Malykh *et al.*, 2018; Schwentner *et al.*, 2019). However, nonsynonymous *glyA* mutations were found in only 3 % of *C. glutamicum* CgHis2 strain variants. Hence, the approach of leveraging the rationally modified *C. glutamicum* CgHis2 strain for the identification of additional beneficial off-site mutations within this study can be regarded as successful.

Overall, 18,000 randomly induced mutations in the 100 isolated *C. glutamicum* CgHis2 strain variants were identified and processed by automated-comparative and combinatorial genome analysis. The mutations were judged with regard to their individual hotspot significance, hotspot abundance, variant performance and hypothesized network effects. Due to the consecutive nature of the screening process and the high number of 71 identified hotspot genes, not every target could be analyzed in detail. Wherever possible, in-frame deletions of hotspot genes were performed and compared to reverse engineered point mutations in order to investigate the respective maximum loss-of-function effect on L-histidine production.

Only *cps*-G987D could be reliably identified as loss-of-function mutation, since Δ*cps* exhibited a higher contributive effect to improved L-histidine production in *C. glutamicum* CgHis2. In *C. glutamicum*, *cps* represents the only gene encoding a non-ribosomal peptide synthetase (NRPS) and was not mentioned in connection to any strain engineering efforts, yet. In general, NRPSs are known to synthesize peptides considered as secondary metabolites such as antibiotics, pigments or siderophores (Strieker, Tanović and Marahiel, 2010). In closely related mycobacteria, however, NRPS are involved in the synthesis of cell wall components and respective genes are located in direct proximity to other cell wall-related genes (Chalut, 2016). Using the compendium of expression profiles, no significant conditions for regulation of *cps* and its MarR-type regulator in *C. glutamicum* could be identified (Kranz *et al.*, 2022). Both genes, however, are highly conserved in many *Corynebacterium* species (Appendix Chapter 8.7). For *C. glutamicum* CgHis2, a reduced ATP consumption by mutated or deleted *cps* could also contribute to the improved production phenotype, since the first step of NRPS catalysis involves ATP for activation of amino acids via adenylation (Strieker, Tanović and Marahiel, 2010). A high ATP regeneration capacity was previously identified by Flux Balance Analysis to be necessary for efficient L-histidine production, which might account for prevention of ATP hydrolysis as well (Schwentner *et al.*, 2019). Interestingly, a series of other ATP-dependent enzymes such as helicases (*NCgl2964*, *NCgl0302*, *NCgl0705*, *NCgl2433*), ATPase components (*NCgl0552*, *NCgl2859*, *NCgl1085*) and a protease (*clpC*) were identified as significant hotspot genes as well.

NCgl2981 encodes a hypothetical membrane protein of yet unknown function or structure and yet unexplored potential for strain engineering. Its potential involvement in cell wall processes is supported by at least one transmembrane helix predicted by TMHMM 2.0 (Krogh *et al.*, 2001). In addition, a gene encoding *N*-acetylmuramoyl-L-alanine amidase involved in cell wall synthesis could be identified in close proximity (Appendix Chapter 8.7). The beneficial *NCgl2981*-D735G mutation thereby was located close to the C-terminal transmembrane helix and, according to the AlphaFold structure prediction, on an outside loop of the protein (Jumper *et al.*, 2021). *NCgl2981* was found highly conserved among *Corynebacterium* and *Mycobacterium* species (Appendix Chapter 8.7). However, no significant data for any regulation of gene expression could be obtained when referring to the compendium of expression profiles of *C. glutamicum* (Kranz *et al.*, 2022). Since the deletion of *NCgl2981* in *C. glutamicum* CgHis2 failed, effects of the identified D735G mutation on protein function and the link to L-histidine metabolism could not be evaluated, but might indicate an essential function of this gene.

As indicated, cell wall synthesis represented a metabolic module of *C. glutamicum*, in which a series of significant *C. glutamicum* CgHis2 hotspot genes were identified. The polyketide synthase encoded by *pks* is essential for the condensation reaction of mycolic acid synthesis, which represent major constituents of the *C. glutamicum* cell wall (Takeno *et al.*, 2018; Toyoda and Inui, 2018). Hence, *C. glutamicum* Δ*pks* strains were reported deficient of mycolic acids and exhibited an altered cell envelope (Portevin *et al.*, 2004). This may have reduced carbon sinks and improved L-histidine synthesis in *C. glutamicum* CgHis2. Moreover, an efflux of L-histidine may have been facilitated as well, as it was reported for L-lysine or L-glutamate production with *C. glutamicum* (Gebhardt *et al.*, 2007b; Lanéelle, Tropis and Daffé, 2013). Hence, *pks*-D1186N might represent a loss-of-function mutation, even though a *C. glutamicum* CgHis2 Δ*pks* strain could not be constructed.

Similar to *pks*, *emb* was identified with high hotspot significance in cell wall synthesis as well. This non-essential gene encodes arabinosyl transferase C, which is involved in arabinan synthesis (Jankute *et al.*, 2018). *C. glutamicum* Δ*emb* strains were reported viable before, but characterized by a highly truncated arabinogalactan and reduced growth (Alderwick *et al.*, 2005). In *C. glutamicum* CgHis2, deletion of *emb* was successful but plasmid-based overexpression of *hisEG* led to significant growth defects such that no experiments for strain characterization could be performed. Interestingly, growth of *C. glutamicum* CgHis2 strains harboring *emb*-T529I/T539I or *emb*-G477E substitutions was not reduced, indicating them not to be directly crucial for the overall functionality of the enzyme. However, there is a high probability that other *emb* mutations may lead to similar effects as discussed for *pks* (Seidel *et al.*, 2007).

Both fatty acid synthase genes, *fasA* and *fasB*, were identified as highly significant hotspot genes, which supply C_{18} oleic acid and C_{16} palmitic acid for cell wall synthesis in *C. glutamicum*, respectively (Radmacher *et al.*, 2005; Takeno *et al.*, 2018). Even though the reverse engineered *fasA* mutations did not result in increased L-histidine titers, other identified *fasA* mutations might significantly affect L-histidine production since *fasA* is much higher expressed than *fasB* in *C. glutamicum* (Radmacher *et al.*, 2005).

In contrast to essential fatty acid synthase FasA, FasB is dispensable for growth. Already constructed *C. glutamicum* Δ*fasB* mutants were characterized by an altered mycolic acid as well as phospholipid profile before (Radmacher *et al.*, 2005). In improved *C. glutamicum* CgHis2 strain variants isolated in this study, stop codon-inducing mutations were found in *fasB*, which indicated beneficial effects of a *fasB* deletion. In addition, comparative genome analysis characterized *fasB* with the highest hotspot significance of 24 %. In fact, *fasB* deletion significantly improved L-histidine production in *C. glutamicum* CgHis2 and turned out

as key engineering target for further *C. glutamicum* CgHis2 strain engineering as part of this study. Besides the reduced carbon sink hypothesis, an altered cell membrane and cell wall structure may have benefited L-histidine production by facilitating L-histidine efflux as well. Moreover, *fasB* was reported to be linked to C_1-metabolism via cofactor lipoic acid (Marquet, Tse Sum Bui and Florentin, 2001; Cronan, Zhao and Jiang, 2005; Ikeda *et al.*, 2017). Interestingly, *fasB* deletion contributed significantly to L-glutamate excretion in a *C. glutamicum* Δ*fasAB* double mutant (Radmacher *et al.*, 2005). In the comparative analysis performed, 41 % of *C. glutamicum* CgHis2 strain variants harboring nonsynonymous mutations in *fasA* were found to also harbor nonsynonymous *fasB* mutations. One of them, the best-performing isolated *C. glutamicum* CgHis2 strain variant 12-10-5-6, was characterized by a 93 % increase in L-histidine titer in later batch fermentations. Concerning hotspot abundance, 12-10-5-6 was characterized by only one more nonsynonymous mutation in investigated hotspot genes, which was the phosphatase-encoding gene *NCgl2959*. Thus, its production phenotype may be majorly attributed to its acquired *fasAB* mutations. Reverse engineering of several *fasAB* double mutants, however, was not successful in *C. glutamicum* CgHis2.

The recently discovered *pyk2*, encoding an alternative pyruvate kinase, was found as hotspot gene in the comparative genome analysis of *C. glutamicum* CgHis2 strain variants, which is mostly expressed under oxygen deprivation conditions (Chai *et al.*, 2016). However, reverse engineered *pyk2*-T357I or *pyk2* deletion exhibited no contributive effect on L-histidine production. Gene deletion of pyruvate kinase gene *pyk1*, expressed under *C. glutamicum* CgHis2 production conditions instead, was reported to result in an increased glucose uptake, increased metabolite pools of the pentose phosphate pathway and increased L-glutamate production in *C. glutamicum* (Sawada *et al.*, 2010, 2015). The contributive effect of Δ*pyk1* in *C. glutamicum* CgHis2 should be linked to a perturbation of glycolytic flux due to the key role of *pyk1* in glycolysis regulation (Sawada *et al.*, 2010, 2015).

Interestingly, reverse engineering of *C. glutamicum* CgHis2 revealed in-frame deletions of very large genes to contribute significantly to improved L-histidine production in *C. glutamicum* CgHis2, such as *cps* (encoding for a protein of 1295 amino acids) and *fasB* (encoding for protein of 2996 amino acids). Regarding recent genome reduction approaches, one could argue their beneficial effect to be simply attributed to their gene size. Genome reduction aims at removing non-essential enzymatic machinery not necessary for applied purposes in order to reduce an investment of carbon and energy in unnecessary proteins/enzymes and reactions (Choe *et al.*, 2016; Martínez-García and de Lorenzo, 2016; Baumgart *et al.*, 2018). Especially for L-histidine production, the reduction of ATP sinks (as L-histidine precursor) could be highly beneficial. In L-lysine biosynthesis, however, production

titers could not be increased by deletion of large non-essential gene clusters (Unthan *et al.*, 2015). In addition, in isolated *C. glutamicum* CgHis2 strain variants large genes such as *fasA*, *fasB*, *cps* and *pks* were found mutated frequently but no significant regulatory mutations (as in promoter areas) or recombination-induced deletions were found, which would decrease carbon and energy sinks as well. Hence, the significantly improved L-histidine production in strains such as *C. glutamicum* CgHis2 12-10-5-6, was a result of combinatorial effects of beneficial mutations without indications for genome reduction. Nevertheless, gene deletions may contribute to a higher degree to L-histidine production in comparison to beneficial loss-of-function mutations.

Furthermore, the number of large genes identified as hotspot genes (such as *fasA*, *fasB*, *cps* and *pks)* could have also been simply the result of the mutation frequency, which should be the same across the genome. Hence, the probability of finding mutations in longer genes is simply higher. Therefore, when mutation numbers are normalized by gene length, overrepresented genes could indicate promising strain engineering targets. By this alternative approach, respective genes could have been found more frequently mutated than randomly expected such as *cps* (1.7-fold), *pks* (1.6-fold) and *fasB* (1.1-fold).

In addition, the strong mutational bias of MNNG towards GC→AT transitions has to be considered (Ohta, Watanabe-Akanuma and Yamagata, 2000; B Singer and Kusmierek, 2003; Ohnishi *et al.*, 2008). GC-rich genes could exhibit an increased mutation frequency due to the GC content and not only sequence length. Indeed, respective genes exhibit an increased GC content (*cps*, 58 %; *pks*, 56 %; *fasB*, 58 %) in comparison to the *C. glutamicum* average of 53.8 % (Nishio *et al.*, 2003). Hence, bias from above-average GC-content or long gene sequences could have been factors for overrepresentation of respective genes in the dataset of this study (such as *fasB*, *cps* and *pks)*, but do not sufficiently explain their high hotspot significance, since the introduced modifications lead to significant improvements in L-histidine production.

When the identified beneficial modifications were combined in *C. glutamicum* CgHis2, additive effects on L-histidine production were observed for *C. glutamicum* CgHis2 Δ*fasB* Δ*pyk1*, *C. glutamicum* CgHis2 Δ*fasB* Δ*cps* and *C. glutamicum* CgHis2 Δ*fasB pks*-D1186N, but not *C. glutamicum* CgHis2 Δ*fasB NCgl2981*-D735G. Conversely, *C. glutamicum* CgHis2 Δ*pyk1 NCgl2981*-D735G was characterized by a more synergistic effect than additively expected. Interestingly, *C. glutamicum* CgHis2 Δ*pyk1 NCgl2981*-D735G could be further improved by in-frame deletion of *cps*, even though *C. glutamicum* CgHis2 Δ*cps* Δ*pyk1* or *C. glutamicum* CgHis2 Δ*cps NCgl2981*-D735G performed worse than the respective single mutations in *C. glutamicum* CgHis2. Hence, the synergistic effect of Δ*pyk1* and

NCgl2981-D735G is necessary for Δ*cps* to exhibit its contributive effect on L-histidine production. On the contrary, the synergistic effect of Δ*pyk1* and *NCgl2981*-D735G was significantly lower in combination with Δ*fasB*, which indicates significant differences in the cellular effects of *fasB* and *cps* deletions. Here, the synergistic effect of Δ*pyk1, NCgl2981*-D735G and *pks*-D1186N (in combination with Δ*fasB)* led to the superior *C. glutamicum* CgHis2 Δ*fasB* Δ*pyk1* *NCgl2981*-D735G *pks*-D1186N quadruple mutant, such that *C. glutamicum* CgHis2 Δ*cps* Δ*pyk1* *NCgl2981*-D735G *pks*-D1186N could have resulted in even higher L-histidine production.

In the end, only strain lineages harboring either Δ*fasB* or Δ*cps* led to superior L-histidine production strains. Especially the combinatorial effects of Δ*fasB, Δcps* and *pks*-D1186N resulted in severe growth defects, which might indicate *cps* to be involved in cell wall processes or cell wall synthesis of *C. glutamicum* CgHis2 as well. This would be consistent to mycobacterial species, in which cell wall lipids are synthesized by polyketide synthases as well as non-ribosomal peptide synthetases (Chalut, 2016).

In batch-mode bioreactor fermentations, the quadruple mutant *C. glutamicum* CgHis2 Δ*fasB* Δ*pyk1* *NCgl2981*-D735G *pks*-D1186N was characterized by a doubled L-histidine titer (29 mM or 4.5 g L^{-1}) and doubled L-histidine yield from glucose (0.13 mol mol^{-1} or 0.11 g g $^{-1}$). Hence, the integration of the novel beneficial modifications described also represents a further improvement in production performance compared to earlier *C. glutamicum* strains, but not *E. coli* strains, which also were highly engineered in known central engineering targets for L-histidine production (Schwentner *et al.*, 2019; Wu *et al.*, 2020b). Contrary to the latter strains, however, no fed-batch fermentation was performed in this thesis, with which even higher product titer and yields could have been achieved by a subsequent glucose-fed phase with growth-restricted cells. Nevertheless, there is still more room for improvement of these strains when considering the theoretical product yield of 0.44 g L-histidine g glucose^{-1} (Schwentner *et al.*, 2019). Here, more reconstruction attempts of identified *C. glutamicum* CgHis2 SNPs in hotspot genes such as *fasA* or *emb* could result in additional improvements on the road towards maximum L-histidine yields.

Interestingly, the isolated *C. glutamicum* CgHis2 12-10-5-6 variant performed similar in the batch-mode fermentations with regard to L-histidine titer and yield but without the laborious reconstruction performed for *C. glutamicum* CgHis2 Δ*fasB* Δ*pyk1* *NCgl2981*-D735G *pks*-D1186N. However, the *C. glutamicum* CgHis2 quadruple mutant is most likely more suitable for an industrial setting, since more robustness can be expected for reverse engineered strains when random mutations are omitted (Ikeda *et al.*, 2009). Moreover, the *C. glutamicum* CgHis2 quadruple mutant forms significantly more biomass than *C. glutamicum* CgHis2 12-

10-5-6, which makes it more efficient and suitable for a fed-batch bioprocess or scale-up of bioprocesses in general (Hewitt and Nienow, 2007).

As demonstrated in this thesis, the identification of novel beneficial modifications especially in already highly engineered industrial strains remains challenging, but is key to maximizing the performance of production strains. Sequencing of known pathway genes (as performed in the first part of this thesis) is apparently straightforward if strains were evolved from microbial wild types (Binder *et al.*, 2012; Schendzielorz, Dippong, Grünberger, *et al.*, 2014). In several other studies, which used biosensor-based FACS-screening of genetically diverse strain libraries, links of identified mutations to the amino acid production purpose were only hypothesized, omitting reverse engineering and hence verification (Mahr *et al.*, 2015; Zhang *et al.*, 2018; Hernandez-Valdes *et al.*, 2020; Liu *et al.*, 2021). Ikeda and co-workers instead, verified beneficial mutations for L-arginine production by reverse engineering, which, however, were identified by sequencing of the key *arg* operons of classically derived L-arginine producers (Ikeda *et al.*, 2009).

Identifying beneficial mutations without physiological or reasonable connection to the product metabolite is more challenging and mostly unexplored. Random gene knockout and gene overexpression libraries represented other approaches, which yielded improved strains. Here, the identification of beneficial genes was performed by PCR procedures or microarray-based methods (Alper, Miyaoku and Stephanopoulos, 2005; Santos and Stephanopoulos, 2008; Warner, Patnaik and Gill, 2009). The disadvantage of these libraries, however, is the limited number of mutations, which can be assessed (Liu and Jiang, 2015). A more combinatorial solution space for improved variants can be achieved by engineering global transcription or translation processes. Therefore, artificial transcription factors, global transcription machinery engineering (gTME), ribosome engineering or whole genome shuffling was tried and discussed before (Santos and Stephanopoulos, 2008; Warner, Patnaik and Gill, 2009; Liu and Jiang, 2015). Despite the success of these combinatorial methods, their aim was not the identification of beneficial genetic modifications, from which unnecessary or even detrimental modifications can be omitted by reverse engineering. As one solution, Payen and co-workers investigated a prediction-based approach for attractive genetic targets in experimentally evolved point mutations by analyzing related knockout and overexpression libraries with barcode sequencing in advance (Payen *et al.*, 2016).

On the bigger picture, random genome mutagenesis offers an immense genetic solution space to obtain improved microbial production strains from. Not only loss-of-function or gain-of-function effects, but also effects on gene expression levels could be acquired for improvement of industrial production strains. Comparative genome analysis represents a

valuable tool to realize a more systematic approach for the identification of beneficial mutations (Ohnishi *et al.*, 2002; Albert *et al.*, 2005; Herring *et al.*, 2006). Here, high-throughput can be achieved by automation using FAAMS, by which significantly higher numbers of FACS-isolated strain variants could be analyzed in the future. That way, also unknown genes or genes with unknown contribution (such as *cps*, *NCgl2981* or *fasB*, *pks* and *pyk1* for L-histidine production) can be identified. In addition, due to their assumed product unspecific nature, discovered beneficial modifications in this study might not only be useful for L-histidine synthesis, but also for engineering *C. glutamicum* towards the synthesis of other valuable small molecules.

6 Conclusion & Outlook

The identification of novel genomic targets for improvement of production strains still represents a major challenge in microbial strain engineering. The combinatorial application of high-throughput methods presented in this thesis may serve as a blueprint to reveal beneficial genetic modifications quicker and in a more systematic manner. Thus, especially highly engineered industrial production strains could benefit from this approach since novel targets without any rational physiological connection to the product of interest can be discovered. In addition, more robustness is achieved when randomly acquired mutations with detrimental effects are omitted by reverse engineering. For high-throughput screening, suitable biosensors need to be available or can be engineered towards the desired specificities. Then, the advantages of classical and rational strain engineering can be combined with high-throughput FACS and automated comparative genome analyses to push industrial production strains quickly towards maximum yields. In the near future, self-improving machine learning algorithms will speed up the identification process with more precise predictions of mutational effects and, more importantly, will help to find suitable combinations of individual beneficial mutations.

7 References

Albert, T. J., Dailidiene, D., Dailide, G., Norton, J. E., Kalia, A., Richmond, T. A., Molla, M., Singh, J. G., Roland D., Berg, D. E. (2005) 'Mutation discovery in bacterial genomes: metronidazole resistance in *Helicobacter pylori*', *Nature Methods* 2(12), pp. 951–953.

Alderwick, L. J., Radmacher, E., Seidel, M., Gande, R., Hitchen, P. G., Morris, H. R., Dell, A., Sahm, H., Eggeling, L., Besra, G. S. (2005) 'Deletion of Cg-*emb* in Corynebacterianeae leads to a novel truncated cell wall arabinogalactan, whereas inactivation of Cg-*ubiA* results in an arabinan-deficient mutant with a cell wall galactan core', *Journal of Biological Chemistry* 280(37), pp. 32362–32371.

Alifano, P., Fani, R., Li, P., Lazcano, A., Bazzicalupo, M., Carlomagno, M., Stella, C., Bruni, B. (1996) 'Histidine biosynthetic pathway and genes: structure, regulation, and evolution', *Microbiological reviews* 60(1), pp. 44–69.

Alper, H., Miyaoku, K. and Stephanopoulos, G. (2005) 'Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets', *Nature Biotechnology* 23(5), pp. 612–616.

An, G. H., Bielich, J., Auerbach, R., Johnson, E. A. (1991) 'Isolation and characterization of carotenoid hyperproducing mutants of yeast by flow cytometry and cell Sorting', *Bio/Technology* 9(1), pp. 70–73.

Araki, K. and Nakayama, K. (1974) 'Feedback-resistant phosphoribosyl-ATP pyrophosphorylase in L-Histidine-producing mutants of *Corynebacterium glutarnicum*', *Agricultural and Biological Chemistry* 38(11), pp. 2209–2218.

Araki, K., Shimojo, S. and Nakayama, K. (1974) 'Histidine production by *Corynebacterium glutamicum* mutants, multiresistant to analogs of histidine, tryptophan, purine and pyrimidine', *Agricultural and Biological Chemistry* 38(4), pp. 837–846.

Bailey, J. E. (1991) 'Toward a science of Metabolic Engineering', *Science* 252(5013), pp. 1668–1675.

Bailey, J. E., Sburlati, A., Hatzimanikatis, V., Lee, K., Renner, W. A., Tsai, P. S. (1996) 'Inverse Metabolic Engineering: A strategy for directed genetic engineering of useful phenotypes', *Biotechnology and Bioengineering* 52(1), pp. 109–121.

Baumgart, M., Unthan, S., Kloß, R., Radek, A., Polen, T., Tenhaef, N., Müller, M. F., Küberl, A., Siebert, D., Brühl, N., Marin, K., Hans, S., Krämer, R., Bott, M., Kalinowski, J., Wiechert, W., Seibold, G., Frunzke, J., Rückert, C., Wendisch, V. F., Noack, S. (2018) '**Corynebacterium glutamicum** Chassis C1∗: Building and testing a novel platform host for synthetic biology and industrial biotechnology', *ACS Synthetic Biology* 7(1), pp. 132–144.

Becker, J. and Wittmann, C. (2012) 'Bio-based production of chemicals, materials and fuels – *Corynebacterium glutamicum* as versatile cell factory', *Current Opinion in Biotechnology* 23(4), pp. 631–640.

Bellmann, M. Vrljic, M., Patek, M., Sahm, H., Krämer, R., Eggeling, L. (2001) 'Expression control and specificity of the basic amino acid exporter LysE of *Corynebacterium glutamicum*', *Microbiology* 147, pp. 1765–1774.

Berens, C. and Suess, B. (2015) 'Riboswitch engineering — making the all-important second and third steps', *Current Opinion in Biotechnology* 31, pp. 10–15.

Bertani, G. (1951) 'Studies on Lysogenesis I', *Journal of Bacteriology* 62(3), pp. 293–300.

Binder, S., Schendzielorz, G., Stäbler, N., Krumbach, K., Hoffmann, K., Bott, M., Eggeling, L. (2012) 'A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level', *Genome Biology* 13(R40), pp. 1–12.

Blombach, B., Schreiner, M. E., Bartek, T., Oldiges, M., Eikmanns, B. J. (2008) '*Corynebacterium glutamicum* tailored for high-yield L-valine production', *Applied Microbiology and Biotechnology*

79(3), pp. 471–479.

Bott, M. (2015) 'Need for speed - finding productive mutations using transcription factor-based biosensors, fluorescence-activated cell sorting and recombineering', *Microbial Biotechnology* 8(1), pp. 8–10.

Brockmann-Gretza, O. and Kalinowski, J. (2006) 'Global gene expression during stringent response in *Corynebacterium glutamicum* in presence and absence of the *rel* gene encoding (p)ppGpp synthase', *BMC Genomics* 7(1), pp. 1–15.

Brüsseler, C., Radek, A., Tenhaef, N., Krumbach, K., Noack, S., Marienhagen, J. (2018) 'The *myo*-inositol/proton symporter IolT1 contributes to D-xylose uptake in *Corynebacterium glutamicum*', *Bioresource Technology* 249, pp. 953–961.

Buchholz, J., Graf, M., Freund, A., Busche, T., Kalinowski, J., Blombach, B., Takors, R. (2014) 'CO_2/HCO_3– perturbations of simulated large scale gradients in a scale-down device cause fast transcriptional responses in *Corynebacterium glutamicum*', *Applied Microbiology and Biotechnology* 98(20), pp. 8563–8572.

Bunn, A. and Korpela, M. (2019) 'An Introduction to dplR'. *Dendrochronologia* 26(2), pp. 115-124.

Chai, X., Shang, X., Zhang, Y., Liu, S., Liang, Y., Zhang, Y., Wen, T. (2016) 'A novel pyruvate kinase and its application in lactic acid production under oxygen deprivation in *Corynebacterium glutamicum*', *BMC Biotechnology* 16(1), pp. 1–12.

Chalut, C. (2016) 'MmpL transporter-mediated export of cell-wall associated lipids and siderophores in *mycobacteria*', *Tuberculosis* 100, pp. 32–45.

Chen, S., Zhou, Y., Chen, Y., Gu, J. (2018) 'fastp: an ultra-fast all-in-one FASTQ preprocessor', *Bioinformatics* 34(17), pp. i884–i890.

Cheng, Q., Ruebling-Jass, K., Zhang, J., Chen, Q., Croker, K. M. (2012) 'Use FACS sorting in Metabolic Engineering of *Escherichia coli* for increased peptide production', *Methods in Molecular Biology* 834, pp. 177–196.

Cheng, Y., Zhou, Y., Yang, L., Zhang, C., Xu, Q., Xie, X., Chen, N. (2013) 'Modification of histidine biosynthesis pathway genes and the impact on production of L-histidine in *Corynebacterium glutamicum*', *Biotechnology Letters* 35(5), pp. 735–741.

Choe, D., Cho, S., Kim, Sun C., Cho, B. K. (2016) 'Minimal genome: Worthwhile or worthless efforts toward being smaller?', *Biotechnology Journal* 11(2), pp. 199–211.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., Zhang, F. (2013) 'Multiplex genome engineering using CRISPR/Cas systems', *Science* 339(6121), pp. 819–823.

Della Corte, D., van Beek, H. L., Syberg, F., Schallmey, M., Tobola, F., Cormann, K. U., Schlicker, C., Baumann, P. T., Krumbach, K., Sokolowsky, S., Morris, C. J., Grünberger, A., Hofmann, E., Schröder, G. F., Marienhagen, J. (2020) 'Engineering and application of a biosensor with focused ligand specificity', *Nature Communications* 11(1), pp. 1–11.

Court, D. L., Sawitzke, J. A. and Thomason, L. C. (2003) 'Genetic engineering using homologous recombination', *Annual Review of Genetics* 36, pp. 361–388.

Cronan, J. E., Zhao, X. and Jiang, Y. (2005) 'Function, Attachment and Synthesis of Lipoic Acid in *Escherichia coli*', *Advances in Microbial Physiology* 50, pp. 103–146.

Dahl, R. H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T. S., Redding-Johanson, A. M., Petzold, C. J., Mukhopadhyay, A., Lee, T. S., Adams, P. D., Keasling, J. D. (2013) 'Engineering dynamic pathway regulation using stress-response promoters', *Nature Biotechnology* 31(11), pp. 1039–1046.

Demain, A. L., Vandamme, E. J., Collins, J., Buchholz, K. (2016) 'History of Industrial Biotechnology', *Industrial Biotechnology* 2, pp. 1–84.

Dietrich, J. A., Shis, D. L., Alikhani, A., Keasling, J. D. (2013) 'Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis', *ACS Synthetic Biology* 2(1), pp. 47–58.

Doong, S. J., Gupta, A. and Prather, K. L. J. (2018) 'Layered dynamic regulation for improving metabolic pathway productivity in *Escherichia coli*', *Proceedings of the National Academy of Sciences of the United States of America* 115(12), pp. 2964–2969.

Eggeling, L. and Bott, M. (2005) 'Handbook of *Corynebacterium glutamicum*', CRC Press, pp. 1–620.

Eggeling, L. and Bott, M. (2015) 'A giant market and a powerful metabolism: L-lysine provided by *Corynebacterium glutamicum*', *Applied Microbiology and Biotechnology* 99(8), pp. 3387–3394.

Eggeling, L., Bott, M. and Marienhagen, J. (2015) 'Novel screening methods — biosensors', *Current Opinion in Biotechnology* 35, pp. 30–36.

Eggeling, L. and Sahm, H. (2001) 'The cell wall barrier of *Corynebacterium glutamicum* and amino acid efflux', *Journal of Bioscience and Bioengineering* 92(3), pp. 201–213.

Eggeling, L. and Sahm, H. (2003) 'New ubiquitous translocators: Amino acid export by *Corynebacterium glutamicum* and *Escherichia coli*', *Archives of Microbiology* 180(3), pp. 155–160.

Eikmanns, B. J., Thum-Schmitz, N., Eggeling, L., Lüdtke, K., Sahm, H. (1994) 'Nucleotide sequence, expression and transcriptional analysis of the *Corynebacterium glutamicum gltA* gene encoding citrate synthase', *Microbiology* 140(8), pp. 1817–1828.

Erdmann, A., Weil, B. and Krämer, R. (1993) 'Lysine secretion by wild-type *Corynebacterium glutamicum* triggered by dipeptide uptake', *Microbiology* 139(12), pp. 3115–3122.

Fang, X., Lloyd, C. J. and Palsson, B. O. (2020) 'Reconstructing organisms in silico: genome-scale models and their emerging applications', *Nature Reviews Microbiology* 18(12), pp. 731–743.

Feith, A., Schwentner, A., Teleki, A., Favilli, L., Blombach, B., Takors, R. (2020) 'Streamlining the Analysis of Dynamic ^{13}C-Labeling Patterns for the Metabolic Engineering of *Corynebacterium glutamicum* as L-Histidine Production Host', *Metabolites* 10(11), p. 458.

Feng, R. N., Niu, Y. C., Sun, X. W., Li, Q., Zhao, C., Wang, C., Guo, F. C., Sun, C. H., Li, Y. (2013) 'Histidine supplementation improves insulin resistance through suppressed inflammation in obese women with the metabolic syndrome: A randomised controlled trial', *Diabetologia* 56(5), pp. 985–994.

Fernandez-López, R., Ruiz, R., de la Cruz, F., Moncalián, G. (2015) 'Transcription factor-based biosensors enlightened by the analyte', *Frontiers in Microbiology* 6, p. 648.

Fior Markets (2020) *Global Histidine Market Growth 2019-2024 - Fior Markets.* Available at: https://www.fiormarkets.com/report/global-histidine-market-growth-2019-2024-372503.html (Accessed: 4 January 2022).

Flachbart, L. K., Gertzen, C. G. W., Gohlke, H., Marienhagen, J. (2021) 'Development of a biosensor platform for phenolic compounds using a transition ligand strategy', *ACS Synthetic Biology* 10(8), pp. 2002–2014.

Flachbart, L. K., Sokolowsky, S. and Marienhagen, J. (2019) 'Displaced by deceivers: Prevention of biosensor cross-talk is pivotal for successful biosensor-based high-throughput screening campaigns', *ACS Synthetic Biology* 8(8), pp. 1847–1857.

Fondi, M., Emiliani, G., Liò, P., Gribaldo, S., Fani, R. (2009) 'The evolution of histidine biosynthesis in archaea: Insights into the his genes structure and organization in luca', *Journal of Molecular Evolution* 69(5), pp. 512–526.

Foster, P. L. (1991) 'In vivo mutagenesis', *Methods in Enzymology* 204(C), pp. 114–125.

Gande, R., Gibson, K. J.C., Brown, A. K., Krumbach, K., Dover, L. G., Sahm, H., Shioyama, S., Oikawa, T., Besra, G. S., Eggeling, L. (2004) 'Acyl-CoA carboxylases (accD2 and accD3), together with a unique polyketide synthase (Cg-*pks*), are key to mycolic acid biosynthesis in

Corynebacterianeae such as *Corynebacterium glutamicum* and *Mycobacterium tuberculosis'*, *Journal of Biological Chemistry* 279(43), pp. 44847–44857.

Gebhardt, H., Meniche, X., Tropis, M., Krämer, R., Daffé, M., Morbach, S. (2007) 'The key role of the mycolic acid content in the functionality of the cell wall permeability barrier in *Corynebacterineae'*, *Microbiology* 153(5), pp. 1424–1434.

Gibson, D. G., Young, L., Chuang, R., Venter, J. C., Hutchison, C. A., Smith, H. O. (2009) 'Enzymatic assembly of DNA molecules up to several hundred kilobases', *Nature Methods* 6(5), pp. 343–345.

Goodwin, S., McPherson, J. D. and McCombie, W. R. (2016) 'Coming of age: ten years of next-generation sequencing technologies', *Nature Reviews Genetics* 17(6), pp. 333–351.

Grünberger, A., van Ooyen, J., Paczia, N., Rohe, P., Schiendzielorz, G., Eggeling, L., Wiechert, W., Kohlheyer, D., Noack, S. (2013) 'Beyond growth rate 0.6: *Corynebacterium glutamicum* cultivated in highly diluted environments', *Biotechnology and Bioengineering* 110(1), pp. 220–228.

Haerlin, R., Süssmuth, R. and Lingens, F. (1970) 'Mechanism of mutagenesis by *N*-methyl-*N'*-nitro-*N*-nitroso-guanidine (MNNG)', *FEBS Letters* 9(3), pp. 175–176.

Harms, H., Wells, M. C. and Van Der Meer, J. R. (2006) 'Whole-cell living biosensors - Are they ready for environmental application?', *Applied Microbiology and Biotechnology* 70(3), pp. 273–280.

Hasegawa, S., Ichiyama, T., Sonaka, I., Ohsaki, A., Okada, S., Wakiguchi, H., Kudo, K., Kittaka, S., Hara, M., Furukawa, S. (2012) 'Cysteine, histidine and glycine exhibit anti-inflammatory effects in human coronary arterial endothelial cells', *Clinical and Experimental Immunology* 167(2), pp. 269–274.

Hernandez-Valdes, J. A., aan de Stegge, M., Hermans, J., Teunis, J., van Tatenhove-Pel, R. J., Teusink, B., Bachmann, H., Kuipers, O. P. (2020) 'Enhancement of amino acid production and secretion by *Lactococcus lactis* using a droplet-based biosensing and selection system', *Metabolic Engineering Communications* 11, p. e00133.

Herring, C. D., Raghunathan, A., Honisch, C., Patel, T., Applebee, M. K., Joyce, A. R., Albert, T. J., Blattner, F. R., Van Den Boom, D., Cantor, C. R., Palsson, B. (2006) 'Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale', *Nature Genetics* 38(12), pp. 1406–1412.

Hewitt, C. J. and Nienow, A. W. (2007) 'The Scale-Up of Microbial Batch and Fed-Batch Fermentation Processes', *Advances in Applied Microbiology* 62, pp. 105–135.

Hochreiter, B., Garcia, A. P. and Schmid, J. A. (2015) 'Fluorescent proteins as genetically encoded FRET biosensors in life sciences', *Sensors* 15(10), pp. 26281–26314.

Houssin, C., Sousa d'Auria, C., Constantinesco, F., Dietrich, C., Labarre, C., Bayan, N. (2020) 'Architecture and biogenesis of the cell envelope of *Corynebacterium glutamicum.*' *Microbiology Monographs* 23, pp. 25–60.

Hwang, H. G., Noh, M. H., Koffas, M. A.G., Jang, S., Jung, G. Y. (2021) 'Multi-level rebalancing of the naringenin pathway using riboswitch-guided high-throughput screening', *Metabolic Engineering* 67, pp. 417–427.

Hwang, H. J., Lee, S. Y. and Lee, P. C. (2018) 'Engineering and application of synthetic nar promoter for fine-tuning the expression of metabolic pathway genes in *Escherichia coli'*, *Biotechnology for Biofuels* 11(1), pp. 1–13.

Ikeda, M., Mitsuhashi, S., Tanaka, K., Hayashi, M. (2009) 'Reengineering of a *Corynebacterium glutamicum* L-arginine and L-citrulline producer', *Applied and Environmental Microbiology* 75(6), pp. 1635–1641.

Ikeda, M., Nagashima, T., Nakamura, E., Kato, R., Ohshita, M., Hayashi, M., Takeno, S. (2017) 'In vivo roles of fatty acid biosynthesis enzymes in biosynthesis of biotin and α-lipoic acid in *Corynebacterium glutamicum'*, *Applied and Environmental Microbiology* 83(19).

Ikeda, M. and Nakagawa, S. (2003) 'The *Corynebacterium glutamicum* genome: Features and impacts on biotechnological processes', *Applied Microbiology and Biotechnology* 62(2–3), pp. 99–109.

IMARC Group (2020) *White Biotechnology Market Size, Share, | Global Industry Report 2021-2026*. Available at: https://www.imarcgroup.com/white-biotechnology-market (Accessed: 30 November 2021).

IMARC Group (2021) *Amino Acids Market Size, Share, Trends and Forecast 2022-2027*. Available at: https://www.imarcgroup.com/amino-acid-technical-material-market-report (Accessed: 3 January 2022).

IPCC (2018) 'Global Warming of 1.5°C, Special Report 2018', *Cambridge University Press*.

Jankute, M., Alderwick, L. J., Moorey, A. R., Joe, M., Gurcha, S. S., Eggeling, L., Lowary, T. L., Dell, A., Pang, P. C., Yang, T., Haslam, S., Besra, G. S. (2018) 'The singular *Corynebacterium glutamicum* Emb arabinofuranosyltransferase polymerises the α(1 → 5) arabinan backbone in the early stages of cell wall arabinan biosynthesis', *The Cell Surface* 2, pp. 38–53.

Jorgenson, A. K., Fiske, S., Hubacek, K., Li, J., McGovern, T., Rick, T., Schor, J. B., Solecki, W., York, R., Zycherman, A. (2019) 'Social science perspectives on drivers of and responses to global climate change', *Wiley Interdisciplinary Reviews: Climate Change* 10(1), p. e554.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A.A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., Hassabis, D. (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature* 596(7873), pp. 583–589.

Jung, S., Chun, J. Y., Yim, S. H., Lee, S. S., Cheon, C., Song, E., Lee, M. S. (2010) 'Transcriptional regulation of histidine biosynthesis genes in *Corynebacterium glutamicum*', *Canadian Journal of Microbiology* 56(2), pp. 178–187.

Kaczmarek, J. A. and Prather, K. L. J. (2021) 'Effective use of biosensors for high-throughput library screening for metabolite production', *Journal of Industrial Microbiology and Biotechnology* 48(9-10), p. 49.

Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, B. J., Gaigalat, L., Goesmann, A., Hartmann, M., Huthmacher, K., Krämer, R., Linke, B., McHardy, A. C., Meyer, F., Möckel, B., Pfefferle, W., Pühler, A., Rey, D. A., Rückert, C., Rupp, O., Sahm, H., Wendisch, V. F., Wiegräbe, I., Tauch, A. (2003) 'The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins', *Journal of Biotechnology* 104(1–3), pp. 5–25.

Käß, F., Hariskos, I., Michel, A., Brandt, H. J., Spann, R., Junne, S., Wiechert, W., Neubauer, P., Oldiges, M. (2014) 'Assessment of robustness against dissolved oxygen/substrate oscillations for *C. glutamicum* DM1933 in two-compartment bioreactor', *Bioprocess and Biosystems Engineering*. 37(6), pp. 1151–1162.

Keasling, J. D. (2010) 'Manufacturing molecules through metabolic engineering', *Science* 330(6009), pp. 1355–1358.

Keilhauer, C., Eggeling, L. and Sahm, H. (1993) 'Isoleucine synthesis in *Corynebacterium glutamicum*: molecular analysis of the *ilvB-ilvN-ilvC* operon', *Journal of Bacteriology* 175(17), pp. 5595–5603.

Kim, G. B., Kim, W. J., Kim, H. U., Lee, S. Y. (2020) 'Machine learning applications in systems metabolic engineering', *Current Opinion in Biotechnology* 64, pp. 1–9.

Kinoshita, S., Udaka, S. and Shimono, M. (1957) 'Studies on the amino acid fermentation part I. Production of L-glutamic acid by various microorganisms', *The Journal of General and Applied Microbiology* 3(3), pp. 193–205.

Kirchner, O. and Tauch, A. (2003) 'Tools for genetic engineering in the amino acid-producing bacterium *Corynebacterium glutamicum*', *Journal of Biotechnology* 104(1–3), pp. 287–299.

Kiss, A. A., Grievink, J. and Rito-Palomares, M. (2015) 'A systems engineering perspective on process integration in industrial biotechnology', *Journal of Chemical Technology & Biotechnology* 90(3), pp. 349–355.

Klaffl, S., Brocker, M., Kalinowski, J., Eikmanns, B. J., Bott, M. (2013) 'Complex regulation of the phosphoenolpyruvate carboxykinase gene *pck* and characterization of its Gntr-type regulator IolR as a repressor of *myo*-inositol utilization genes in *Corynebacterium glutamicum'*, *Journal of Bacteriology* 195(18), pp. 4283–4296.

Koboldt, D. C., Chen, K., Wylie, T., Larson, D. E., McLellan, M. D., Mardis, E. R., Weinstock, G. M., Wilson, R. K., Ding, L. (2009) 'VarScan: variant detection in massively parallel sequencing of individual and pooled samples', *Bioinformatics* 25(17), pp. 2283–2285.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L., Wilson, R. K. (2012) 'VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing', *Genome Research* 22(3), pp. 568–576.

Koch, M., Pandi, A., Borkowski, O., Cardoso Batista, A., Faulon, J. L. (2019) 'Custom-made transcriptional biosensors for metabolic engineering', *Current Opinion in Biotechnology* 59, pp. 78–84.

Kranz, A., Polen, T., Kotulla, C., Arndt, A., Bosco, G., Bussmann, M., Chattopadhyay, A., Cramer, A., Davoudi, C. F., Degner, U., Diesveld, R., Freiherr von Boeselager, R., Gärtner, K., Gätgens, C., Georgi, T., Geraths, C., Haas, S., Heyer, A., Hünnefeld, M., Ishige, T., Kabus, A., Kallscheuer, N., Kever, L., Klaffl, S., Kleine, B., Kočan, M., Koch-Koerfges, A., Kraxner, K. J., Krug, A., Krüger, A., Küberl, A., Labib, M., Lange, C., Mack, C., Maeda, T., Mahr, R., Majda, S., Michel, A., Morosov, X., Müller, O., Nanda, A. M., Nickel, J., Pahlke, J., Pfeifer, E., Platzen, L., Ramp, P., Rittmann, D., Schaffer, S., Scheele, S., Spelberg, S., Schulte, J., Schweitzer, J. E., Sindelar, G., Sorger-Herrmann, U., Spelberg, M., Stansen, C., Tharmasothirajan, A., van Ooyen, J., van Summeren-Wesenhagen, P., Vogt, M., Witthoff, S., Zhu, L., Eikmanns, B. J., Oldiges, M., Schaumann, G., Baumgart, M., Brocker, M., Eggeling, L., Freudl, R., Frunzke, J., Marienhagen, J., Wendisch, V. F., Bott, M. (2022) 'A manually curated compendium of expression profiles for the microbial cell factory *Corynebacterium glutamicum*', *Scientific Data* 9(1), pp. 1–10.

Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E. L.L. (2001) 'Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes', *Journal of molecular biology* 305(3), pp. 567–580.

Kulis-Horn, R. K., Persicke, M. and Kalinowski, J. (2014) 'Histidine biosynthesis, its regulation and biotechnological application in *Corynebacterium glutamicum*', *Microbial Biotechnology* 7(1), pp. 5–25.

Kulis-Horn, R. K., Persicke, M. and Kalinowski, J. (2015) '*Corynebacterium glutamicum* ATP-phosphoribosyl transferases suitable for ʟ-histidine production – Strategies for the elimination of feedback inhibition', *Journal of Biotechnology* 206, pp. 26–37.

Lanéelle, M.-A., Tropis, M. and Daffé, M. (2013) 'Current knowledge on mycolic acids in *Corynebacterium glutamicum* and their relevance for biotechnological processes', *Applied Microbiology and Biotechnology 2013 97:23*. Springer, 97(23), pp. 9923–9930.

Lawson, C. E., Martí, J. M., Radivojevic, T., Jonnalagadda, S. V. R., Gentz, R., Hillson, N. J., Peisert, S., Kim, J., Simmons, B. A., Petzold, C. J., Singer, S. W., Mukhopadhyay, A., Tanjore, D., Dunn, J. G., Garcia M. H. (2021) 'Machine learning for metabolic engineering: A review', *Metabolic Engineering* 63, pp. 34–60.

Lee, J. H. and Wendisch, V. F. (2017) 'Production of amino acids – Genetic and metabolic engineering approaches', *Bioresource Technology* 245, pp. 1575–1587.

Lee, S. Y. and Kim, H. U. (2015) 'Systems strategies for developing industrial microbial strains', *Nature Biotechnology* 33(10), pp. 1061–1072.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics* 25(16), pp. 2078–2079.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows–Wheeler transform', *Bioinformatics* 25(14), pp. 1754–1760.

Li, L., Wei, K., Liu, X., Wu, Y., Zheng, G., Chen, S., Jiang, W., Lu, Y. (2019) 'aMSGE: advanced multiplex site-specific genome engineering with orthogonal modular recombinases in actinomycetes', *Metabolic Engineering* 52, pp. 153–167.

Liao, S. M., Du, Q. S., Meng, J. Z., Pang, Z. W., Huang, R. B. (2013) 'The multiple roles of histidine in protein interactions', *Chemistry Central Journal* 7(1), pp. 1–12.

Liu, D., Evans, T. and Zhang, F. (2015) 'Applications and advances of metabolite biosensors for metabolic engineering', *Metabolic Engineering* 31, pp. 35–43.

Liu, W. and Jiang, R. (2015) 'Combinatorial and high-throughput screening approaches for strain engineering', *Applied Microbiology and Biotechnology* 99(5), pp. 2093–2104.

Liu, Y., Yuan, H., Ding, D., Dong, H., Wang, Q., Zhang, D. (2021) 'Establishment of a biosensor-based high-throughput screening platform for tryptophan overproduction', *ACS Synthetic Biology* 10(6), pp. 1373–1383.

Long, Q., Liu, X., Yang, Y., Li, L., Harvey, L., McNeil, B., Bai, Z. (2014) 'The development and application of high throughput cultivation technology in bioprocess development', *Journal of Biotechnology* 192, pp. 323–338.

Mahr, R., Gätgens, C., Gätgens, J., Polen, T., Kalinowski, J., Frunzke, J. (2015) 'Biosensor-driven adaptive laboratory evolution of L-valine production in *Corynebacterium glutamicum*', *Metabolic Engineering* 32, pp. 184–194.

Malykh, E. A., Butov, I. A., Ravcheeva, A. B., Krylov, A. A., Mashko, S. V., Stoynova, N. V. (2018) 'Specific features of L-histidine production by *Escherichia coli* concerned with feedback control of AICAR formation and inorganic phosphate/metal transport', *Microbial Cell Factories* 17(1), p. 42.

MarketWatch (2021) *Histidine Market Share,Size 2021 – Global Trends, Market Demand, Industry Analysis, Growth, Opportunities and Forecast 2027 - MarketWatch*. Available at: https://www.marketwatch.com/press-release/histidine-market-sharesize-2021-global-trends-market-demand-industry-analysis-growth-opportunities-and-forecast-2027-2021-12-09 (Accessed: 4 January 2022).

Marquet, A., Tse Sum Bui, B. and Florentin, D. (2001) 'Biosynthesis of biotin and lipoic acid', *Vitamins and Hormones* 61, pp. 51–101.

Martínez-García, E. and de Lorenzo, V. (2016) 'The quest for the minimal bacterial genome', *Current Opinion in Biotechnology* 42, pp. 216–224.

Michener, J. K. and Smolke, C. D. (2012) 'Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways', *Metabolic Engineering* 14(3), pp. 212–222.

Michener, J. K. and Smolke, C. D. (2012) 'High-throughput enzyme evolution in *Saccharomyces cerevisiae* using a synthetic RNA switch', *Metabolic Engineering* 14(4), pp. 306–316.

Mizukami, T., Hamu, A., Ikeda, M., Oka, T., Katsumata, R. (1994) 'Cloning of the ATP phosphoribosyl transferase gene of *Corynebacterium glutamicum* and application of the gene to L-histidine production', *Bioscience, Biotechnology, and Biochemistry* 58(4), pp. 635–638.

Mohsin, M. and Ahmad, A. (2014) 'Genetically-encoded nanosensor for quantitative monitoring of methionine in bacterial and yeast cells', *Biosensors and Bioelectronics* 59, pp. 358–364.

Montaño López, J., Duran, L. and Avalos, J. L. (2021) 'Physiological limitations and opportunities in microbial metabolic engineering', *Nature Reviews Microbiology* 20(35-48)*,* pp. 1–14.

Mordor Intelligence (2020) *White Biotech Market | 2021 - 26 | Industry Share, Size, Growth - Mordor*

Intelligence. Available at: https://www.mordorintelligence.com/industry-reports/white-biotech-market (Accessed: 30 November 2021).

Mustafi, N., Grünberger, A., Mahr, R., Helfrich, S., Nöh, K., Blombach, B., Kohlheyer, D., Frunzke, J. (2014) 'Application of a genetically encoded biosensor for live cell imaging of L-valine production in pyruvate dehydrogenase complex-deficient *Corynebacterium glutamicum* strains', *PLOS ONE* 9(1), p. e85731.

Nielsen, J. (2001) 'Metabolic engineering', *Applied Microbiology and Biotechnology* 55(3), pp. 263–283.

Nielsen, J. and Keasling, J. D. (2016) 'Engineering Cellular Metabolism', *Cell* 164(6), pp. 1185–1197.

Nishio, Y., Nakamura, Y., Kawarabayasi, Y., Usuda, Y., Kimura, E., Sugimoto, S., Matsui, K., Yamagishi, A., Kikuchi, H., Ikeo, K., Gojobori, T. (2003) 'Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of *Corynebacterium efficiens*', *Genome Research* 13(7), p. 1572.

OECD (2011) 'Industrial Biotechnology and Climate Change. Organisation for Economic Co-operation and Development (OECD)'. Available at: www.oecd.org/sti/biotechnology (Accessed: 20 May 2022).

Ohnishi, J., Mitsuhashi, S., Hayashi, M., Ando, S., Yokoi, H., Ochiai, K., Ikeda, M. (2002) 'A novel methodology employing *Corynebacterium glutamicum* genome information to generate a new L-lysine-producing mutant', *Applied Microbiology and Biotechnology* 58(2), pp. 217–223.

Ohnishi, J., Mizoguchi, H., Takeno, S., Ikeda, M. (2008) 'Characterization of mutations induced by *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine in an industrial *Corynebacterium glutamicum* strain', *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 649(1–2), pp. 239–244.

Ohta, T., Watanabe-Akanuma, M. and Yamagata, H. (2000) 'A comparison of mutation spectra detected by the *Escherichia coli* Lac+ reversion assay and the *Salmonella typhimurium* His+ reversion assay', *Mutagenesis* 15(4), pp. 317–323.

Oku, M., Hoseki, J., Ichiki, Y., Sakai, Y. (2013) 'A fluorescence resonance energy transfer (FRET)-based redox sensor reveals physiological role of thioredoxin in the yeast *Saccharomyces cerevisiae*', *FEBS Letters* 587(6), pp. 793–798.

Parekh, S., Vinci, V. A. and Strobel, R. J. (2000) 'Improvement of microbial strains and fermentation processes', *Applied Microbiology and Biotechnology* 5*4*(3), pp. 287–301.

Payen, C., Sunshine, A. B., Ong, G. T., Pogachar, J. L., Zhao, W., Dunham, M. J. (2016) 'High-throughput identification of adaptive mutations in experimentally evolved yeast populations', *PLOS Genetics* 12(10), p. e1006339.

Peroza, E. A., Ewald, J. C., Parakkal, G., Skotheim, J. M., Zamboni, N. (2015) 'A genetically encoded Förster resonance energy transfer sensor for monitoring in vivo trehalose-6-phosphate dynamics', *Analytical Biochemistry* 474, pp. 1–7.

Polgár, L. (2005) 'The catalytic triad of serine peptidases', *Cellular and Molecular Life Sciences* 62(19), pp. 2161–2172.

Portevin, D., Sousa-D'Auria, C., Houssin, C., Grimaldi, C., Chami, M., Daffé, M., Guilhot, C. (2004) 'A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms', *Proceedings of the National Academy of Sciences* 101(1), pp. 314–319.

Prăvălie, R., Patriche, C., Borrelli, P., Panagos, P., Roșca, B., Dumitrașcu, M., Nita, I. A., Săvulescu, I., Birsan, M. V., Bandoc, G. (2021) 'Arable lands under the pressure of multiple land degradation processes. A global perspective', *Environmental Research* 194, p. 110697.

Radmacher, E., Alderwick, L. J., Besra, G. S., Brown, A. K., Gibson, K. J. C., Sahm, H., Eggeling, L. (2005) 'Two functional FAS-I type fatty acid synthases in *Corynebacterium glutamicum*', *Microbiology* 151(7), pp. 2421–2427.

Raper, K. B. (1946) 'The development of improved penicillin-producing molds', *Annuals of the New York Academy of Sciences*, 48(2), pp. 41–56.

Rebek, J. (1990) 'On the structure of histidine and its role in enzyme active sites', *Structural Chemistry* 1(1), pp. 129–131.

Rogers, J. K. and Church, G. M. (2016) 'Genetically encoded sensors enable real-time observation of metabolite production', *Proceedings of the National Academy of Sciences of the United States of America* 113(9), pp. 2388–2393.

Rowlands, R. T. (1984) 'Industrial strain improvement: mutagenesis and random screening procedures'. *Enzyme and microbial technology* 6(1), pp. 3–10.

Sahm, H., Eggeling, L., Eikmanns, B., Krämer, R. (1995) 'Metabolic design in amino acid producing bacterium *Corynebacterium glutamicum*', *FEMS Microbiology Reviews* 16(2–3), pp. 243–252.

Sahm, H., Eggeling, L. and de Graaf, A. A. (2000) 'Pathway analysis and metabolic engineering in *Corynebacterium glutamicum*', *Biological Chemistry* 381(9–10), pp. 899–910.

Sambrook, J. and Russel, D. W. (2012) 'Molecular Cloning: A Laboratory Manual, 4th edition. *Cold Spring Harbor Laboratory Press*, pp. 1-34.

San Martín, A., Ceballo, S., Baeza-Lehnert, F., Lerchundi, R., Valdebenito, R., Contreras-Baeza, Y., Alegría, K., Barros, L. F. (2014) 'Imaging mitochondrial flux in single cells with a FRET sensor for pyruvate', *PLOS ONE* 9(1), p. e85780.

Sang Yup Lee and E. Terry Papoutsakis (1999) *Metabolic Engineering*. CRC Press, pp. 1-450.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) 'DNA sequencing with chain-terminating inhibitors.', *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), pp. 5463–5467.

Santos, C. N. S. and Stephanopoulos, G. (2008) 'Combinatorial engineering of microbes for optimizing cellular phenotype', *Current Opinion in Chemical Biology* 12(2), pp. 168–176.

Sawada, K., Zen-in, S., Wada, M., Yokota, A. (2010) 'Metabolic changes in a pyruvate kinase gene deletion mutant of *Corynebacterium glutamicum* ATCC 13032', *Metabolic Engineering* 12(4), pp. 401–407.

Sawada, K., Wada, M., Hagiwara, T., Zen-in, S., Imai, K., Yokota, A. (2015) 'Effect of pyruvate kinase gene deletion on the physiology of *Corynebacterium glutamicum* ATCC13032 under biotin-sufficient non-glutamate-producing conditions: Enhanced biomass production', *Metabolic Engineering Communications* 2, pp. 67–75.

Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G., Pühler, A. (1994) 'Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*', *Gene* 145(1), pp. 69–73.

Schallmey, M., Frunzke, J. and Eggeling, L. (2014) 'Looking for the pick of the bunch: high-throughput screening of producing microorganisms with biosensors', *Current Opinion in Biotechnology* 26, pp. 148–154.

Schendzielorz, G., Dippong, M., Grunberger, A., Kohlheyer, D., Yoshida, A., Binder, S., Nishiyama, C., Nishiyama, M., Bott, M., Eggeling, L. (2014) 'Taking control over control: Use of product sensing in single cells to remove flux control at key enzymes in biosynthesis pathways', *ACS Synthetic Biology*, 3(1), pp. 21–29.

Schultz, C., Niebisch, A., Schwaiger, A., Viets, U., Metzger, S., Bramkamp, M., Bott, M. (2009) 'Genetic and biochemical analysis of the serine/threonine protein kinases PknA, PknB, PknG and PknL of *Corynebacterium glutamicum*: evidence for non-essentiality and for phosphorylation of OdhI and FtsZ by multiple kinases', *Molecular Microbiology* 74(3), pp. 724–741.

Schwentner, A., Feith, A., Münch, E., Stiefelmaier, J., Lauer, I., Favilli, L., Massner, C., Öhrlein, J., Grund, B., Hüser, A., Takors, R., Blombach, B. (2019) 'Modular systems metabolic

engineering enables balancing of relevant pathways for L-histidine production with *Corynebacterium glutamicum*', *Biotechnology for Biofuels* 12(1), p. 65.

Seidel, M., Alderwick, L. J., Sahm, H., Besra, G. S., Eggeling, L. (2007) 'Topology and mutational analysis of the single Emb arabinofuranosyltransferase of *Corynebacterium glutamicum* as a model of Emb proteins of *Mycobacterium tuberculosis*', *Glycobiology* 17(2), pp. 210–219.

Serganov, A. and Nudler, E. (2013) 'A Decade of Riboswitches', *Cell* 152(1–2), pp. 17–24.

Singer, B. and Kusmierek, J. T. (1982) 'Chemical mutagenesis', *Annual review of Biochemistry* 51, pp. 135–158.

Sonntag, C. K., Flachbart, L. K., Maass, C., Vogt, M., Marienhagen, J. (2020) 'A unified design allows fine-tuning of biosensor parameters and application across bacterial species', *Metabolic Engineering Communications* 11, p. e00150.

Stephanopoulos, G., Aristidou, A. A. and Nielsen, J. (1998) 'Metabolic Engineering: Principles and Methodologies.' *Elsevier Acadamic Press*, pp. 1-752.

Stoltzfus, A. and Norris, R. W. (2016) 'On the causes of evolutionary transition:transversion bias', *Molecular Biology and Evolution* 33(3), p. 595.

Strieker, M., Tanović, A. and Marahiel, M. A. (2010) 'Nonribosomal peptide synthetases: structures and dynamics', *Current Opinion in Structural Biology* 20(2), pp. 234–240.

Taguchi, S., Ooi, T., Mizuno, K., Matsusaki, H. (2015) 'Advances and needs for endotoxin-free production strains', *Applied Microbiology and Biotechnology* 99(22), pp. 9349–9360.

Takeno, S., Murata, N., Kura, M., Takasaki, M., Hayashi, M., Ikeda, M. (2018) 'The *accD3* gene for mycolic acid biosynthesis as a target for improving fatty acid production by fatty acid-producing *Corynebacterium glutamicum* strains', *Applied Microbiology and Biotechnology* 102(24), pp. 10603–10612.

Takors, R., Bathe, B., Rieping, M., Hans, S., Kelle, R., Huthmacher, K. (2007) 'Systems biology for industrial strains and fermentation processes—Example: Amino acids', *Journal of Biotechnology* 129(2), pp. 181–190.

Tang, J. L., Li, C. Y., Li, Y. F., Zou, C. X. (2014) 'A ratiometric fluorescent probe with unexpected high selectivity for ATP and its application in cell imaging', *Chemical Communications* 50(97), pp. 15411–15414.

Tatsumi, N. and Massyuki, I. (2012) *Corynebacterium glutamicum:* Biology and Biotechnology. *Springer Business and Science Media*, pp. 1-416.

Teleki, A. and Takors, R. (2019) 'Quantitative profiling of endogenous metabolites using hydrophilic interaction liquid chromatography–tandem mass spectrometry (HILIC-MS/MS)', *Methods in Molecular Biology* 1859, pp. 185–207.

Tenhaef, N., Stella, R., Frunzke, J., Noack, S. (2021) 'Automated rational strain construction based on high-throughput conjugation', *ACS Synthetic Biology* 10(3), pp. 589–599.

Toya, Y. and Shimizu, H. (2013) 'Flux analysis and metabolomics for systematic metabolic engineering of microorganisms', *Biotechnology Advances* 31(6), pp. 818–826.

Toyoda, K. and Inui, M. (2018) 'Extracytoplasmic function sigma factor σD confers resistance to environmental stress by enhancing mycolate synthesis and modifying peptidoglycan structures in *Corynebacterium glutamicum*', *Molecular Microbiology* 107(3), pp. 312–329.

Unthan, S., Radek, A., Wiechert, W., Oldiges, M., Noack, S. (2015) 'Bioprocess automation on a Mini Pilot Plant enables fast quantitative microbial phenotyping', *Microbial Cell Factories* 14(1), pp. 1–11.

Vinkenborg, J. L., Nicolson, T. J., Bellomo, E. A., Koay, M. S., Rutter, G. A., Merkx, M. (2009) 'Imaging of intracellular free Zn^{2+} in real time using genetically-encoded FRET sensors', *Nature methods* 6(10), p. 737.

Waagbⱷ, R., Trⱷße, C., Koppe, W.,. Fontanillas, R., Breck, O. (2010) 'Dietary histidine supplementation prevents cataract development in adult Atlantic salmon, *Salmo salar* L., in seawater', *British Journal of Nutrition* 104(10), pp. 1460–1470.

Wade, A. M. and Tucker, H. N. (1998) 'Antioxidant characteristics of L-histidine', *The Journal of Nutritional Biochemistry* 9(6), pp. 308–315.

Wagner, J. M., Liu, L., Yuan, S. F., Venkataraman, M. V., Abate, A. R., Alper, H. S. (2018) 'A comparative analysis of single cell and droplet-based FACS for improving production phenotypes: Riboflavin overproduction in *Yarrowia lipolytica*', *Metabolic Engineering* 47, pp. 346–356.

Wang, H. H.,Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., Church, G. M. (2009) 'Programming cells by multiplex genome engineering and accelerated evolution', *Nature* 460(7257), pp. 894–898.

Warner, J. R., Patnaik, R. and Gill, R. T. (2009) 'Genomics enabled approaches in strain engineering', *Current Opinion in Microbiology* 12(3), pp. 223–230.

Wiechert, W. (2001) '^{13}C Metabolic Flux Analysis', *Metabolic Engineering* 3(3), pp. 195–206.

Wiechert, W. (2002) 'Modeling and simulation: tools for metabolic engineering', *Journal of Biotechnology* 94(1), pp. 37–63.

Wu, H., Tian, D., Fan, X., Fan, W., Zhang, Y., Jiang, S., Wen, C., Ma, Q., Chen, N., Xie, X. (2020) 'Highly efficient production of L-histidine from glucose by metabolically engineered *Escherichia coli*', *ACS Synthetic Biology* 9(7), pp. 1813–1822.

Xu, L., Liu, P., Dai, Z., Fan, F., Zhang, X. (2021) 'Fine-tuning the expression of pathway gene in yeast using a regulatory library formed by fusing a synthetic minimal promoter with different Kozak variants', *Microbial Cell Factories* 20(1), pp. 1–11.

Yoshida, T., Nakajima, H., Takahashi, S., Kakizuka, A., Imamura, H. (2019) 'OLIVe: A genetically encoded fluorescent biosensor for quantitative imaging of branched-chain amino acid levels inside single living cells', *ACS Sensors* 4(12), pp. 3333–3342.

Zhang, F., Carothers, J. M. and Keasling, J. D. (2012) 'Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids', *Nature Biotechnology* 30(4), pp. 354–359.

Zhang, J., Jensen, M. K. and Keasling, J. D. (2015) 'Development of biosensors and their application in metabolic engineering', *Current Opinion in Chemical Biology* 28, pp. 1–8.

Zhang, X., Zhang, X., Xu, G., Zhang, X., Shi, J., Xu, Z. (2018) 'Integration of ARTP mutagenesis with biosensor-mediated high-throughput screening to improve L-serine yield in *Corynebacterium glutamicum*', *Applied Microbiology and Biotechnology* 102(14), pp. 5939–5951.

Zhang, Y., Shang, X., Deng, A., Chai, X., Lai, S., Zhang, G., Wen, T. (2012) 'Genetic and biochemical characterization of *Corynebacterium glutamicum* ATP phosphoribosyltransferase and its three mutants resistant to feedback inhibition by histidine', *Biochimie* 94(3), pp. 829–838.

8   Appendix

8.1   Appendix to Chapter 4.3 – Engineering an industrial L-histidine production strain for pSenHis-based FACS-screening
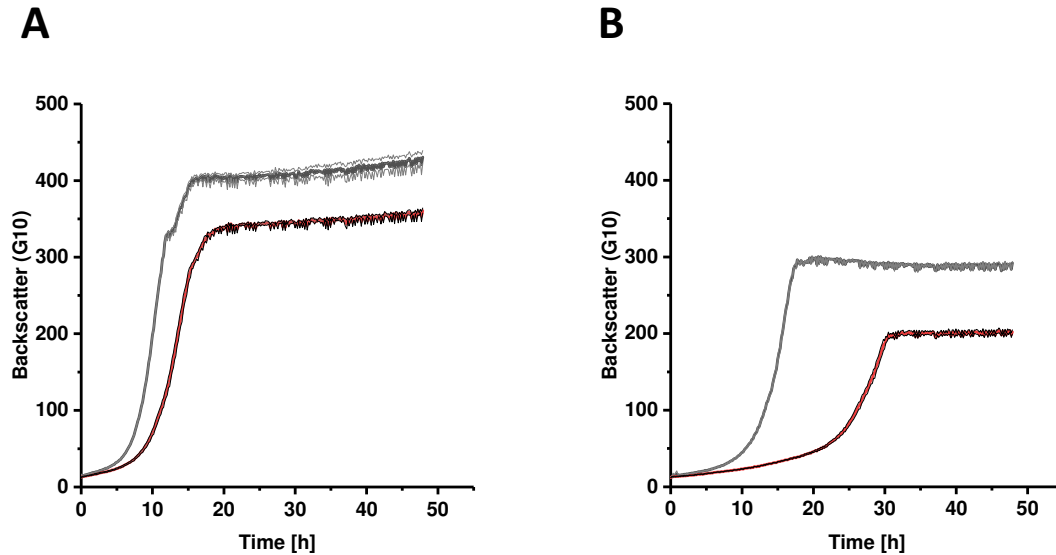
A

B
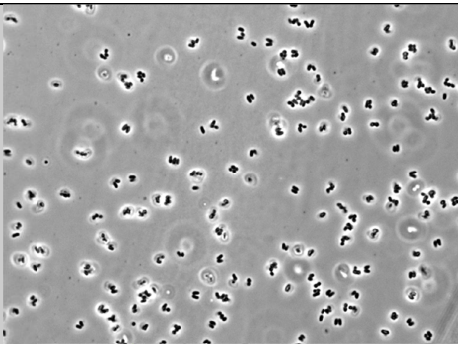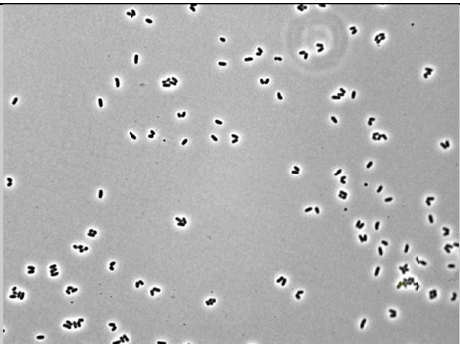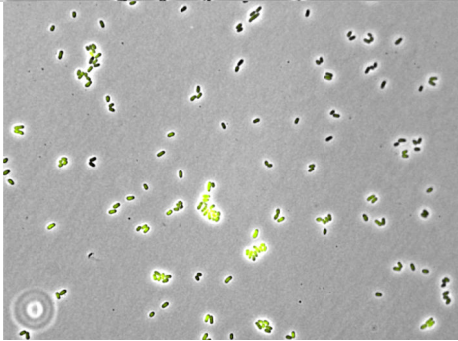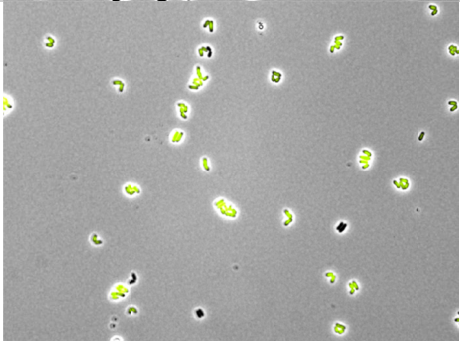
Figure 19: Growth of *C. glutamicum* wild type and *C. glutamicum* CgHis1 harboring no plasmid (A) or transformed with pSenHis[*hisEG*] (B). Depicted are *C. glutamicum* wild type (grey) and *C. glutamicum* CgHis1 (red). Error bars represent biological triplicates. Growth rates (µmax) were determined as follows: *C. glutamicum* wild type, 0.41 h^{-1}; *C. glutamicum* CgHis1, 0.31 h^{-1}; *C. glutamicum* WT pSenHis[*hisEG*], 0.30 h^{-1}; *C. glutamicum* CgHis1 pSenHis[*hisEG*], 0.19 h^{-1}.

Table 6: Microscopic pictures of *C. glutamicum* wild type and *C. glutamicum* CgHis1 harboring either pSenHis or pSenHis[*hisEG*].

strain	pSenHis	pSenHis[*hisEG*]
C. glutamicum wild type		
C. glutamicum CgHis1		

Figure 20: Comparison of *C. glutamicum* CgHis1 pHisOP1 as starting strain and *C. glutamicum* CgHis2 as screening host with results from microtiter cultivations regarding L-histidine titer (bars) and biomass formation (dots, OD$_{600}$). Error bars represent biological triplicates.

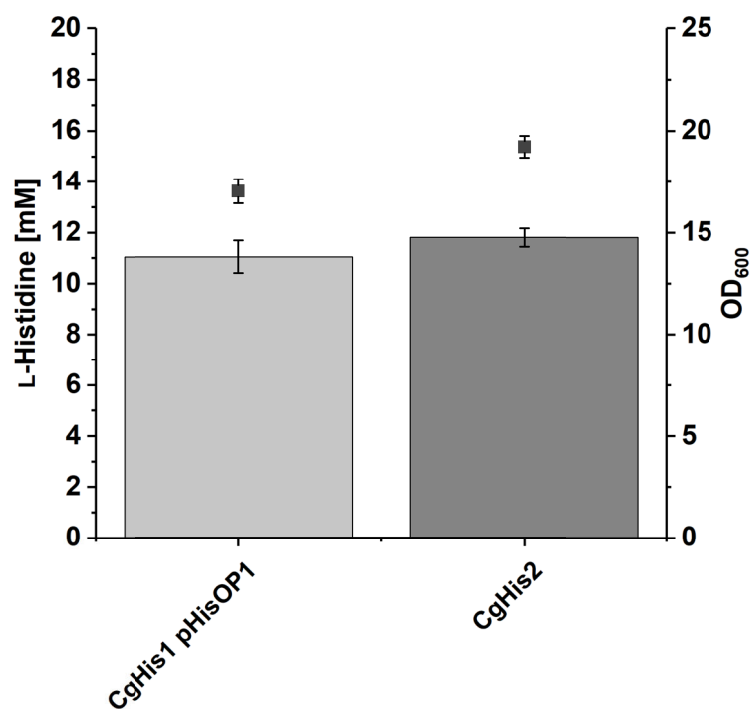## 8.2   Appendix to Chapter 4.4 – Investigation of potential biosensor crosstalk
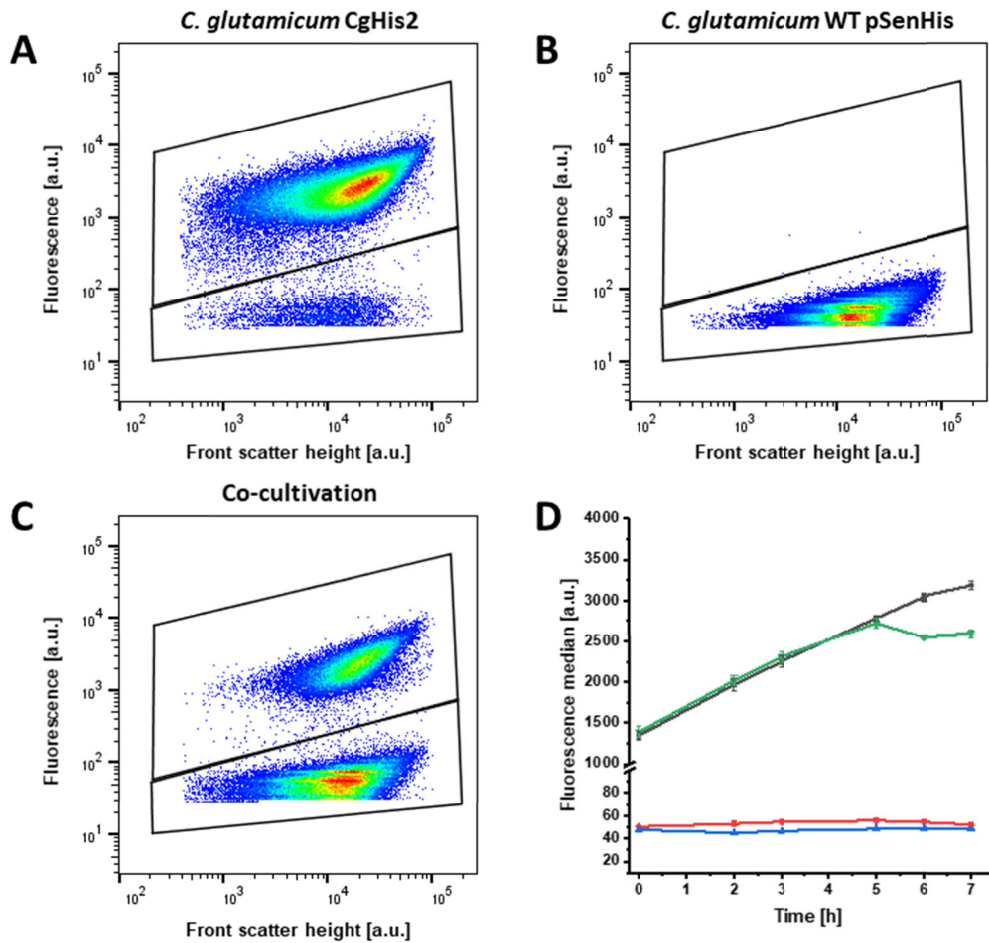
Figure 21: Biosensor crosstalk – Co-cultivation of *C. glutamicum* CgHis2 and *C. glutamicum* wild type pSenHis using an initial mixing ratio of 1:2. Scatter plots for (A) *C. glutamicum* CgHis2 as producer strain, (B) *C. glutamicum* wildtype pSenHis as reporter strain and (C) the co-cultivation of both strains were derived from FACS analysis of 100,000 events. The upper and lower gates were used to monitor the populations' fluorescence median, respectively, over a cultivation time of seven hours. (D): *C. glutamicum* CgHis2 control (green, from upper gate in A); *C. glutamicum* WT pSenHis control (blue, from lower gate in B); *C. glutamicum* CgHis2 subpopulation of co-cultivation (black, from upper gate in C); *C. glutamicum* WT pSenHis subpopulation of co-cultivation (red, from lower gate in C).Other mixing ratios (1:10; 10:1) were analyzed equally (not shown).

8.3 Appendix to Chapter 4.5 - Multiplexed random genome mutagenesis of *C. glutamicum* CgHis2
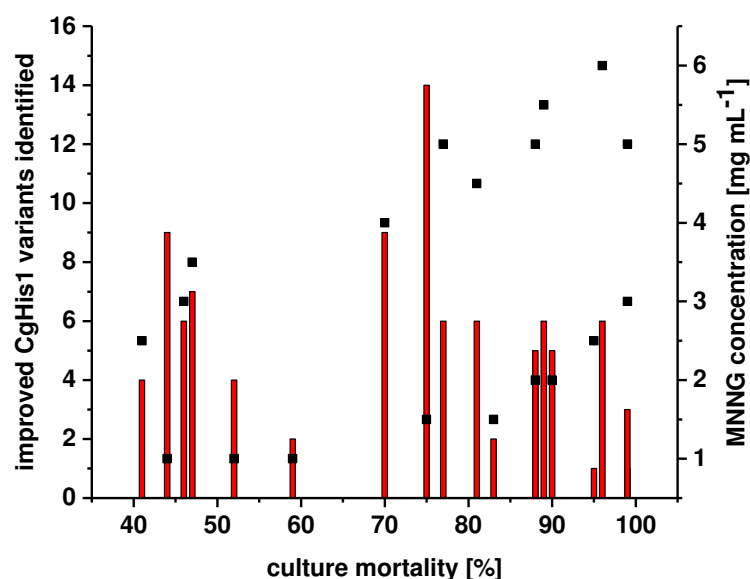
Figure 22: Impact of MNNG-dependent mortality of treated cultures and number of independently isolated *C. glutamicum* CgHis2 strain variants with improved L-histidine production. All isolated *C. glutamicum* CgHis2 variants (red bars) were isolated from independent *C. glutamicum* CgHis2 culture, respectively, which were mutated by multiplexed MNNG mutagenesis. Depending on the MNNG concentration applied (concentration shown as stock concentration of MNNG dilution series, black squares), the culture mortality varied. In later mutagenesis rounds, the MNNG-induced mortality shifted, such that lower MNNG concentrations were sufficient to yield appropriately mutagenized cultures for FACS-screening.

8.4 Appendix to Chapter 4.6 – Development of a high-throughput FACS workflow
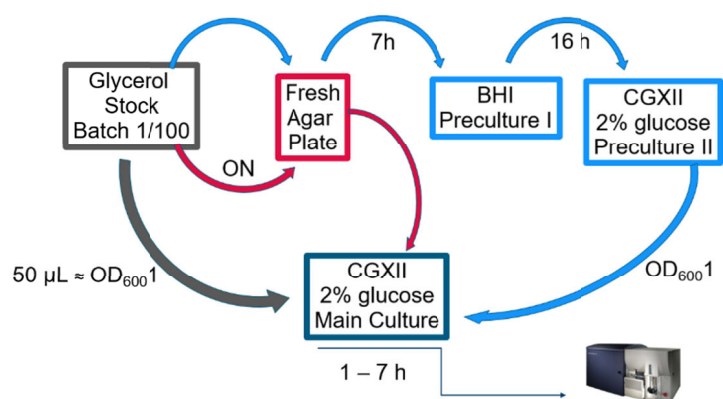
Figure 23: Seedtrain variations for inoculation of the main culture prior to FACS-screening. The inoculation of the main culture (CGXII, 2 % glucose, 15 µg/mL Kanamycin) was performed directly from glycerol culture (grey), from (BHI, 15 µg/mL Kanamycin) agar plate (red) or following the standard seedtrain (blue) which includes agar plate as well as undefined medium (BHI, 15 µg/mL Kanamycin) and defined medium (CGXII, 2 % glucose, 15 µg/mL Kanamycin) precultures.

8.5 Appendix to Chapter 4.7 - Rescreening and characterization of isolated *C. glutamicum* CgHis2 strain variants
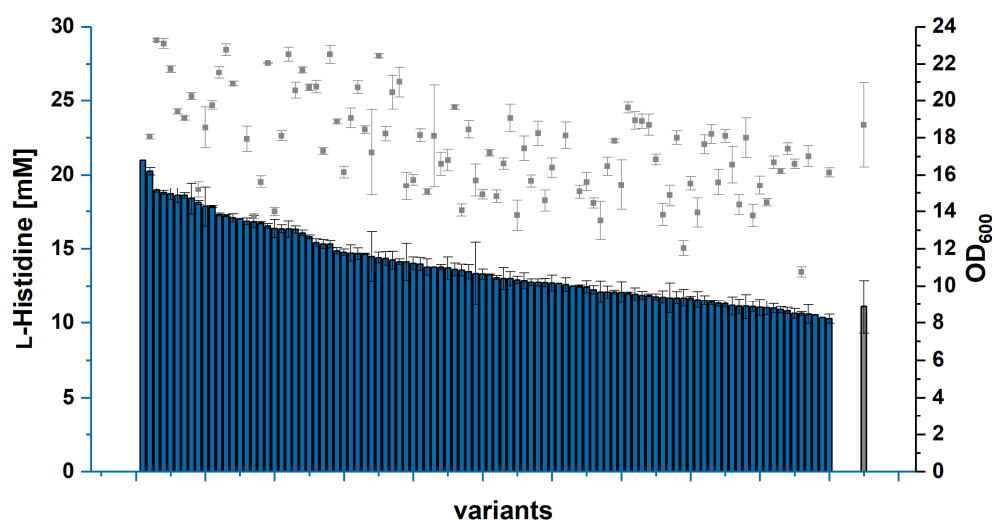
Figure 24: L-histidine titer and biomass (as final optical density, OD_{600}) of the 100 improved *C. glutamicum* CgHis2 strain variants after 48 h cultivation in the second characterization step. Error bars represent three independent technical replicates. The *C. glutamicum* CgHis2 reference strain (grey) is displayed as mean over ten independent cultivation rounds. Notably, due to technical variation of humidity parameters during the cultivation, strain variants were evaluated only in comparison to the respective triplicate of controls on the same microtiter plate.

8.6 Appendix to Chapter 4.9 - Combinatorial genome analysis: Evaluation of hotspot genes

Table 7: List of identified hotspot genes using computational analysis and the 10 % threshold level of hotspot significance. These results of the computational processing of the FAAMS dataset by Dr. Stephan Noack show the number of nonsynonymous SNPs per gene of the 100 *C. glutamicum* CgHis2 strain variants (numSNP) and the number of strain variants harboring nonsynonymous SNPs in the particular gene (numCLO, hotspot significance).

#	annotation	locus tag	gene	numSNP	numCLO
1	CGL_RS12390	NCgl2409	*fasB*	37	24
2	CGL_RS15265	NCgl2964		28	23
3	CGL_RS04205	NCgl0802	*fasA*	26	22
4	CGL_RS14305	NCgl2773	*pks*	27	21
5	CGL_RS00975	NCgl0184	*emb*	27	20
6	*gltB*	NCgl0181	*gltB*	22	19
7	CGL_RS02900	NCgl0552		21	19
8	CGL_RS13520	NCgl2618	*cps*	23	18
9	*pknB*	NCgl0040	*pknB*	20	17
10	CGL_RS15240	NCgl2959		19	17
11	CGL_RS09035	NCgl1737		18	17
12	CGL_RS12520	NCgl2433	*dinG*	16	15
13	CGL_RS15355	NCgl2981		16	15
14	*ggt*	NCgl0916	*ggtB*	15	14
15	CGL_RS03670	NCgl0705		18	13
16	CGL_RS11935	NCgl2324	*benR*	15	13
17	CGL_RS00545	NCgl0098	*putA*	14	13
18	CGL_RS00870	NCgl0163		14	13
19	CGL_RS12940	NCgl2503	*nuc*	14	13
20	CGL_RS13350	NCgl2585	*clpC*	14	13
21	CGL_RS14755	NCgl2859		17	12
22	CGL_RS02505	NCgl0472	*rpoC*	16	12
23	CGL_RS13580	NCgl2628		15	12
24	CGL_RS13605	NCgl2633	*mrpA*	14	12
25	CGL_RS05640	NCgl1085		13	12
26	*iolD*	NCgl0159	*iolD*	13	12
27	CGL_RS08860	NCgl1702		14	11
28	CGL_RS02075	NCgl0394		13	11
29	CGL_RS04840	NCgl0927		13	11
30	CGL_RS05800	NCgl1117		13	11
31	CGL_RS13165	NCgl2548a		13	11
32	CGL_RS03135	NCgl0599		12	11
33	CGL_RS03195	NCgl0611	*dnaE2*	12	11
34	CGL_RS04255	NCgl0812		12	11
35	CGL_RS07785	NCgl1494a		12	11
36	CGL_RS13380	NCgl2591		12	11

37	CGL_RS14975	NCgl2903		12	11
38	*glpK*	NCgl2790	*glpK*	12	11
39	*prpB*	NCgl0629	*prpB*	12	11
40	CGL_RS00515	NCgl0092		11	11
41	CGL_RS10665	NCgl2068	*ileS*	11	11
42	CGL_RS14485	NCgl2809	*pyk2*	11	11
43	*pepN*	NCgl2340	*pepN*	11	11
44	CGL_RS14385	NCgl2789	*psp5*	18	10
45	CGL_RS02500	NCgl0471	*rpoB*	15	10
46	CGL_RS07320	NCgl1407	*thiD1*	15	10
47	CGL_RS04190	NCgl0799	*mctC*	13	10
48	CGL_RS03440	NCgl0659	*pyc*	13	10
49	CGL_RS03060	NCgl0584		12	10
50	*mfd*	NCgl0924	*mfd*	12	10
51	CGL_RS02370	NCgl0450	*menD*	11	10
52	CGL_RS05435	NCgl1044		11	10
53	CGL_RS08810	NCgl1692		11	10
54	CGL_RS11220	NCgl2185	*phoD*	11	10
55	CGL_RS12955	NCgl2507	*ptrB*	11	10
56	CGL_RS14895	NCgl2887		11	10
57	CGL_RS14915	NCgl2891		11	10
58	*gabT*	NCgl0462	*gabT*	11	10
59	*hrpB*	NCgl0139	*hrpB*	11	10
60	*topA*	NCgl0304	*topA*	11	10
61	CGL_RS00340	NCgl0060		10	10
62	CGL_RS01805	NCgl0340	*capD*	10	10
63	CGL_RS03515	NCgl0674	*wbpC*	10	10
64	CGL_RS09170	NCgl1767		10	10
65	CGL_RS12690	NCgl2464		10	10
66	CGL_RS14285	NCgl2769	*mmpL1*	10	10
67	CGL_RS14880	NCgl2884	*mrcB*	10	10
68	CGL_RS14925	NCgl2893		10	10
69	*fusA*	NCgl0478	*fusA*	10	10
70	CGL_RS15125	NCgl2933	*ulaA*	10	10
71	*xylB*	NCgl0111	*xylB*	10	10

Figure 25: Computational processing of the FAAMS dataset for identification of hotspot genes in *C. glutamicum* CgHis2 by Dr. Stephan Noack.

```
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
```

```
In [2]: # Load FAAMS dataset into Pandas DataFrame
        df_faams = pd.read_csv('CodSNS_upd.txt', delimiter="\t")
        df_faams
```

Out[2]:

	CLO	POS	DP	ABQ	FREQ	TXID	CONSEQ	REFCO	VARCO	REFAA	VARAA	AAPOS	AACODE	LocusTag
1	NG-22595-M1-8	10839	412	36	99.51	1515	synonymous	GAG	GAA	E	E	111	E111E	CGL_RS00045
2	NG-22595-M1-8	15796	387	36	99.74	1518	nonsynonymous	GCA	ACA	A	T	150	A150T	CGL_RS00080
3	NG-22595-M1-8	19525	362	36	99.17	13	Promoter	-	-	-	-	60	-	CGL_RS00105
4	NG-22595-M1-8	23165	406	37	98.52	16	synonymous	ACC	ACT	T	T	334	T334T	CGL_RS00125
5	NG-22595-M1-8	34514	425	37	99.76	25	nonsynonymous	GCA	GTA	A	V	50	A50V	CGL_RS00200
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
18805	NG-27238-7-40-7-8	3281750	682	37	100.00	3048	nonsynonymous	CGC	TGC	R	C	142	R142C	CGL_RS15270
18806	NG-27238-7-40-7-8	3286087	648	36	100.00	1502	nonsynonymous	TCC	TTC	S	F	162	S162F	CGL_RS15295
18807	NG-27238-7-40-7-8	3286507	640	37	100.00	1502	nonsynonymous	GCC	GTC	A	V	302	A302V	CGL_RS15295
18808	NG-27238-7-40-7-8	3296492	718	36	100.00	1508	synonymous	GTC	GTT	V	V	143	V143V	CGL_RS15360
18809	NG-27238-7-40-7-8	3305820	593	36	99.83	-	-	-	-	-	-	-	-	-

18809 rows × 14 columns

```
In [3]: # Load histidine dataset into Pandas DataFrame
        df_his = pd.read_csv('HIS_data.txt', delimiter="\t")
        df_his
```

Out[3]:

	CLO	HIS
1	NG-22595-M1-8	0.420
2	NG-22595-M15-6	0.120
3	NG-22595-M38-2	0.370
4	NG-22595-M39-12	0.120
5	NG-22595-M44-2	0.680
...	...	...
100	NG-27238-12-10-12-5	0.320
101	NG-27238-7-40-12-7	0.192
102	NG-27238-7-35-12-20	0.267
103	NG-27238-11-15-4-6	0.252
104	NG-27238-11-5-9-18	0.436

104 rows × 2 columns

```
In [4]: # Merge both datasets
        df_all = pd.merge(df_faams, df_his, on='CLO')
        df_all
```

Out[4]:

	CLO	POS	DP	ABQ	FREQ	TXID	CONSEQ	REFCO	VARCO	REFAA	VARAA	AAPOS	AACODE	LocusTag	
0	NG-22595-M1-8	10839	412	36	99.51	1515	synonymous	GAG	GAA	E	E	111	E111E	CGL_RS00045	0
1	NG-22595-M1-8	15796	387	36	99.74	1518	nonsynonymous	GCA	ACA	A	T	150	A150T	CGL_RS00080	0
2	NG-22595-M1-8	19525	362	36	99.17	13	Promoter	-	-	-	-	60	-	CGL_RS00105	0
3	NG-22595-M1-8	23165	406	37	98.52	16	synonymous	ACC	ACT	T	T	334	T334T	CGL_RS00125	0
4	NG-22595-M1-8	34514	425	37	99.76	25	nonsynonymous	GCA	GTA	A	V	50	A50V	CGL_RS00200	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
18804	NG-27238-7-40-7-8	3281750	682	37	100.00	3048	nonsynonymous	CGC	TGC	R	C	142	R142C	CGL_RS15270	0
18805	NG-27238-7-40-7-8	3286087	648	36	100.00	1502	nonsynonymous	TCC	TTC	S	F	162	S162F	CGL_RS15295	0
18806	NG-27238-7-40-7-8	3286507	640	37	100.00	1502	nonsynonymous	GCC	GTC	A	V	302	A302V	CGL_RS15295	0
18807	NG-27238-7-40-7-8	3296492	718	36	100.00	1508	synonymous	GTC	GTT	V	V	143	V143V	CGL_RS15360	0
18808	NG-27238-7-40-7-8	3305820	593	36	99.83	-	-	-	-	-	-	-	-	-	0

18809 rows × 15 columns

```
In [5]: # Remove non-annotated genes and some outliers
        indl = df_all[(df_all['LocusTag'] == '-')
                      | (df_all['LocusTag'] == 'CGL_RS09205')
                      | (df_all['LocusTag'] == 'CGL_RS09405')
                      | (df_all['CLO'] == 'NG-25372-12-13')
                      | (df_all['CLO'] == 'NG-25372-55r7c18')
                      | (df_all['CLO'] == 'NG-25517-10-20-8-8')
                      | (df_all['CLO'] == 'NG-26744-7-25-10-21')].index

        df_red1 = df_all
        df_red1.drop(indl, inplace=True)
        df_red1
```

Out[5]:

	CLO	POS	DP	ABQ	FREQ	TXID	CONSEQ	REFCO	VARCO	REFAA	VARAA	AAPOS	AACODE	LocusTag	
0	NG-22595-M1-8	10839	412	36	99.51	1515	synonymous	GAG	GAA	E	E	111	E111E	CGL_RS00045	0
1	NG-22595-M1-8	15796	387	36	99.74	1518	nonsynonymous	GCA	ACA	A	T	150	A150T	CGL_RS00080	0
2	NG-22595-M1-8	19525	362	36	99.17	13	Promoter	-	-	-	-	60	-	CGL_RS00105	0
3	NG-22595-M1-8	23165	406	37	98.52	16	synonymous	ACC	ACT	T	T	334	T334T	CGL_RS00125	0
4	NG-22595-M1-8	34514	425	37	99.76	25	nonsynonymous	GCA	GTA	A	V	50	A50V	CGL_RS00200	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
18803	NG-27238-7-40-7-8	3279529	503	37	100.00	3047	nonsynonymous	ACA	ATA	T	I	691	T691I	CGL_RS15265	0
18804	NG-27238-7-40-7-8	3281750	682	37	100.00	3048	nonsynonymous	CGC	TGC	R	C	142	R142C	CGL_RS15270	0

83

18805	NG-27238-7-40-7-8	3286087	648	36	100.00	1502	nonsynonymous	TCC	TTC	S	F	162	S162F	CGL_RS15295 0
18806	NG-27238-7-40-7-8	3286507	640	37	100.00	1502	nonsynonymous	GCC	GTC	A	V	302	A302V	CGL_RS15295 0
18807	NG-27238-7-40-7-8	3296492	718	36	100.00	1508	synonymous	GTC	GTT	V	V	143	V143V	CGL_RS15360 0

16059 rows × 15 columns

```
In [6]:  # Count number of SNPs per gene
         df_SNP = df_red1.groupby('LocusTag')['LocusTag'].count().reset_index(name='numSNP')

         # Average possition of genes
         df_POS = df_red1.groupby('LocusTag')['POS'].mean().reset_index(name='meanPOS')
```

```
In [7]:  # Plot some results
         fig, ax = plt.subplots(1, 2, figsize=(10,3), dpi=300)

         HIS_data = -np.sort(-df_red1.groupby('CLO').mean().HIS.values*100)

         ax[0].bar(np.arange(len(HIS_data)), HIS_data)
         ax[0].set_xlabel('variants [-]')
         ax[0].set_ylabel('L-histidine increase [%]')

         POS_data = df_POS['meanPOS'].values/1e6
         SNP_data = df_SNP['numSNP'].values

         markerline, stemline, baseline = ax[1].stem(POS_data, SNP_data, use_line_collection='true')
         plt.setp(stemline, linewidth = 0.5)
         plt.setp(markerline, markersize = 2)
         ax[1].set_xlabel('genome position [Mbp]')
         ax[1].set_ylabel('SNPs per gene [-]')

         fig.align_ylabels()
         fig.tight_layout()
         fig.savefig('Figure_1.png', dpi=300)
```
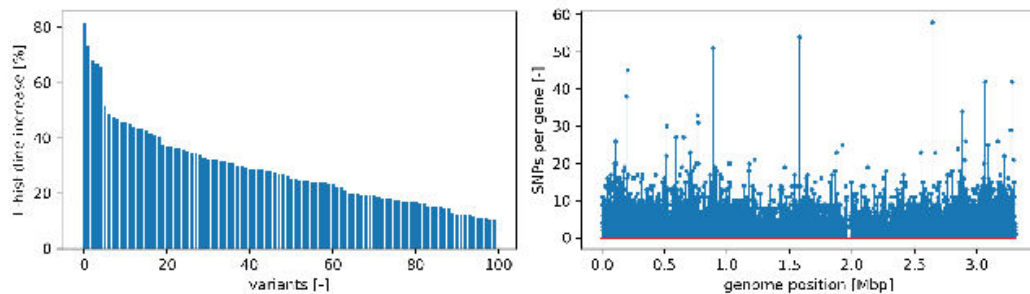


```
In [8]:  # Remove non-informative data
         ind2 = df_all[(df_all['CONSEQ'] == 'synonymous')
                       | (df_all['CONSEQ'] == '-')
                       | (df_all['CONSEQ'] == 'nonsense')
                       | (df_all['CONSEQ'] == 'Promoter')
                       | (df_all['LocusTag'] == 'CGL_RS09205')
                       | (df_all['LocusTag'] == 'CGL_RS09405')
                       | (df_all['CLO'] == 'NG-25372-12-13')
                       | (df_all['CLO'] == 'NG-25372-55r7c18')
                       | (df_all['CLO'] == 'NG-25517-10-20-8-8')
                       | (df_all['CLO'] == 'NG-26744-7-25-10-21')
                       | (df_all['FREQ'] <= 50)].index

         df_red2 = df_all
         df_red2.drop(ind2, inplace=True)
         df_red2
```

Out[8]:

	CLO	POS	DP	ABQ	FREQ	TXID	CONSEQ	REFCO	VARCO	REFAA	VARAA	AAPOS	AACODE	LocusTag
1	NG-22595-M1-8	15796	387	36	99.74	1518	nonsynonymous	GCA	ACA	A	T	150	A150T	CGL_RS00080 0
4	NG-22595-M1-8	34514	425	37	99.76	25	nonsynonymous	GCA	GTA	A	V	50	A50V	CGL_RS00200 0
5	NG-22595-M1-8	35570	382	37	99.21	26	nonsynonymous	CCA	TCA	P	S	187	P187S	CGL_RS00205 0

```
In [11]:  # Count number of unique clones
          numCLO = df_red2['CLO'].nunique()

          # Identify hot spot genes
          hot = 0.1*numCLO

          df_hs_red = df_hs.loc[(df_hs['numCLO'] >= hot)]
          pd.set_option('display.max_rows', None)
          df_hs_red
```

Out[11]:

	LocusTag	numSNP	numCLO	meanHIS
1695	CGL_RS12390	37	24	0.327973
2176	cadA	32	25	0.315469
2120	CGL_RS15265	28	23	0.280464
1961	CGL_RS14305	27	21	0.310889
153	CGL_RS00975	27	20	0.291000
611	CGL_RS04205	26	22	0.327885
1838	CGL_RS13520	23	18	0.242783
2257	gltB	22	19	0.239955
415	CGL_RS02900	21	19	0.276000
2380	pknB	20	17	0.233500
2115	CGL_RS15240	19	17	0.372105
1268	CGL_RS09035	18	17	0.260278
537	CGL_RS03670	18	13	0.284611
1975	CGL_RS14385	18	10	0.268389
2036	CGL_RS14755	17	12	0.282412
1713	CGL_RS12520	16	15	0.266813
2132	CGL_RS15355	16	15	0.280687
378	CGL_RS02505	16	12	0.412125
2244	ggt	15	14	0.322333
1630	CGL_RS11935	15	13	0.279533
1846	CGL_RS13580	15	12	0.355933
377	CGL_RS02500	15	10	0.288000
1026	CGL_RS07320	15	10	0.252267
81	CGL_RS00545	14	13	0.237000
137	CGL_RS00870	14	13	0.296143
1752	CGL_RS12940	14	13	0.309857
1817	CGL_RS13350	14	13	0.314714
1851	CGL_RS13605	14	12	0.287214
1239	CGL_RS08860	14	11	0.290071
814	CGL_RS05640	13	12	0.343462
2289	iolD	13	12	0.240154
324	CGL_RS02075	13	11	0.241077
699	CGL_RS04840	13	11	0.328462
839	CGL_RS05800	13	11	0.341615
1789	CGL_RS13165	13	11	0.263923
608	CGL_RS04190	13	10	0.399385
449	CGL_RS03135	12	11	0.309417
460	CGL_RS03195	12	11	0.384833
618	CGL_RS04255	12	11	0.335833
1093	CGL_RS07785	12	11	0.257250
1823	CGL_RS13380	12	11	0.269833
2071	CGL_RS14975	12	11	0.261583
2255	glpK	12	11	0.237833
2389	prpB	12	11	0.374667
435	CGL_RS03060	12	10	0.239417
2318	mfd	12	10	0.291333
76	CGL_RS00515	11	11	0.254455

1455	CGL_RS10665	11	11	0.328727
1990	CGL_RS14485	11	11	0.279455
2370	pepN	11	11	0.247364
366	CGL_RS02370	11	10	0.396636
786	CGL_RS05435	11	10	0.329364
1232	CGL_RS08810	11	10	0.304636
1534	CGL_RS11220	11	10	0.266545
1754	CGL_RS12955	11	10	0.303182
2057	CGL_RS14895	11	10	0.379091
2061	CGL_RS14915	11	10	0.246818
2237	gabT	11	10	0.267818
2281	hrpB	11	10	0.281818
2510	topA	11	10	0.276000
50	CGL_RS00340	10	10	0.272200
282	CGL_RS01805	10	10	0.249500
514	CGL_RS03515	10	10	0.334100
1284	CGL_RS09170	10	10	0.296000
1734	CGL_RS12690	10	10	0.255500
1958	CGL_RS14285	10	10	0.334400
2054	CGL_RS14880	10	10	0.281000
2063	CGL_RS14925	10	10	0.276500
2236	fusA	10	10	0.274600
2543	xylB	10	10	0.262000

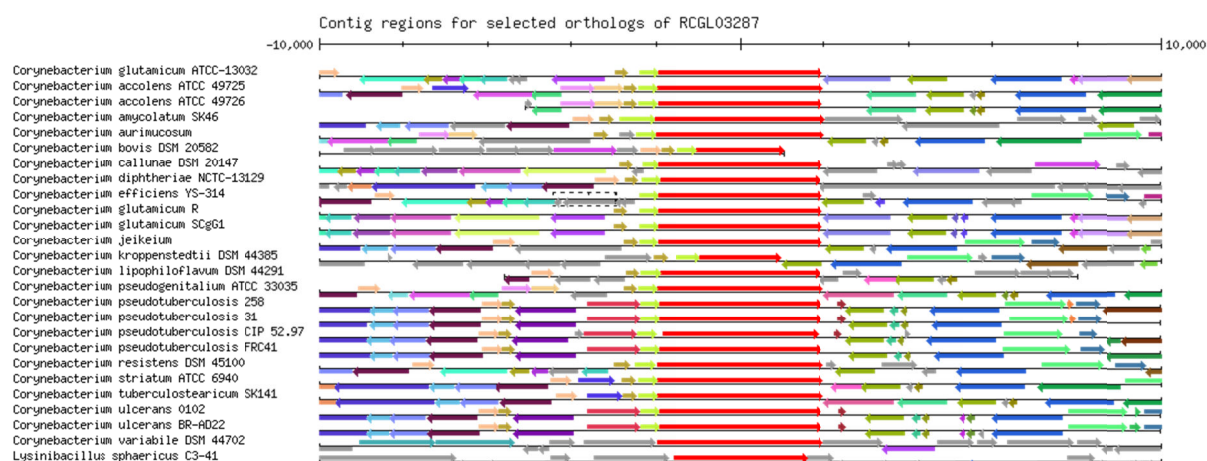## 8.7   Appendix to Chapter 5 – Discussion

A
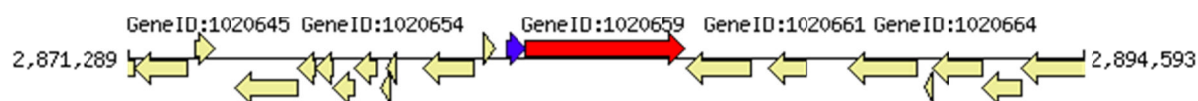
B

Figure 26: A) Position of NRPS-encoding gene *cps* (red) and its MarR-type regulator gene (blue) in the *C. glutamicum* genome and B) Conservation of these genes over many *Corynebacterium* species. Illustrations from ERGO systems biology informatics toolkit (Igenbio Inc., Chicago, IL, USA).
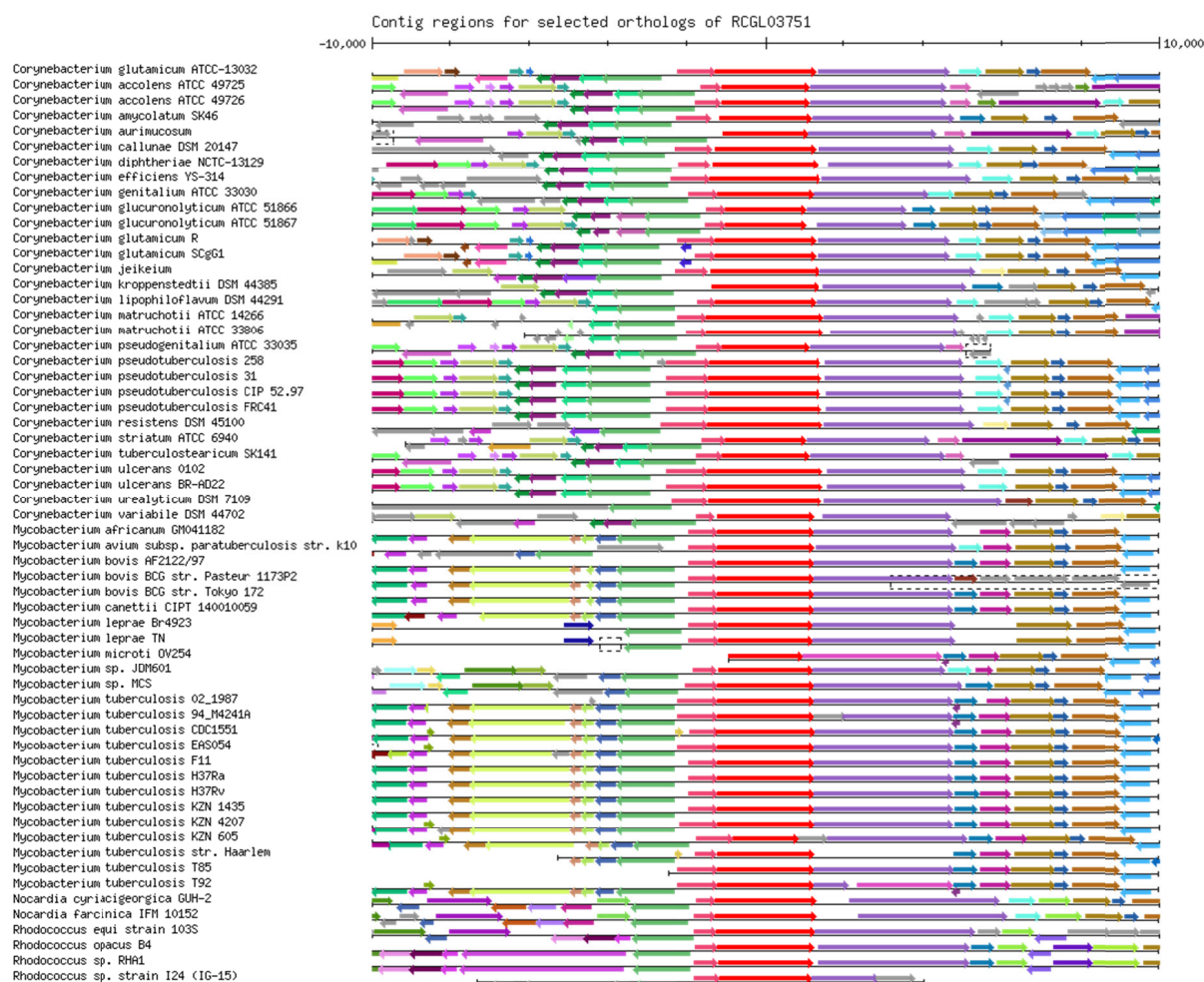
A

B

Figure 27: A) Position of *NCgl2981* in the genome of *C. glutamicum* and B) conservation of this gene over many *Cornyebacterium* and *Mycobacterium* species. Illustrations from ERGO systems biology informatics toolkit (Igenbio Inc., Chicago, IL, USA).

8.8   Appendix – Oligonucleotides

Name	Type	Sequence
Rsp_pK19 Check Primer1	Sequencing Primer	CACAGGAAACAGCTATGACCATG
Univ_pK19 Check Primer 2	Sequencing Primer	CGCCAGGGTTTTCCCAGTCACGAC
1_FasA 893775 G-A (A-T) up fragment_fwd	Gibson Primer	ATCCCCGGGTACCGAGCTCGTCGACGCCTTCCTGTCCTC
1_Fas A 893775 G-A (A-T) up fragment_rev	Gibson Primer	GATGCACGTGTCAGGAGGACGGTACCGCC
1_FasA 893775 G-A (A-T) down_fwd	Gibson Primer	GTCCTCCTGACACGTGCATCCGTTGCTG
1_FasA 893775 G-A (A-T) down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATGGAGCCTGCGCCGAG
1_FasA 893775 Check fwd	Sequencing Primer	GACTTCGACCCTGCCAAGTG
1_FasA 893775 Check rev	Sequencing Primer	TGCGCATTGCCTGCTCGAAG
2_gltB 198201 up fragment_fwd	Gibson Primer	ATCCCCGGGTACCGAGCTCGGTTCTGGCAGCTACGAGATTTTC
2_gltB 198201 up fragment_rev	Gibson Primer	GTGGTGGAGAGGAAATCAGACCAACGCC
2_gltB 198201 down fragment_fwd	Gibson Primer	TCTGATTTCCTCTCCACCACACCACGATATTTAC
2_gltB 198201 down fragment_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGGTGTTCAGCCTTGCCGGTG
2_gltB 198201 Check fwd	Sequencing Primer	ATACAAGTGGCGCCGCGAAG
2_gltB 198201 Check rev	Sequencing Primer	GCACCTGGGCTTGTCCGACG
3_pyk 3111252 up fragment_fwd	Gibson Primer	ATCCCCGGGTACCGAGCTCGAGCCGAGTTCGACGGCGA
3_pyk 3111252 up fragment_rev	Gibson Primer	CCGGCCGCATACCACGCATCAGCTGCAC
3_pyk 3111252 down fragment_fwd	Gibson Primer	GATGCGTGGTATGCGGCCGGATCCGAGGG
3_pyk 3111252 down fragment _rev	Gibson Primer	TTGTAAAACGACGGCCAGTGCCGGGAAATCCGCGTCAGC
3_pyk 3111252 Check fwd	Sequencing Primer	GGACTTGGGTGGCCAAGATG
3_pyk 3111252 Check rev	Sequencing Primer	AAGCGGCATGAATCTAGCTC
4_iolD_174806_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCCACCACCGCATCTCGCAC
4_iolD_174806_up_rev	Gibson Primer	CATGAGGTAGAACCCATCACCAACCATGATCAC
4_iolD_174806_down_fwd	Gibson Primer	GTGATGGGTTCTACCTCATGCTCAACAC
4_iolD_174806_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCTAAATACGTTTTGGTTTAGCC
4_iolD_Check_fwd	Sequencing Primer	TTGCGGGTGATGCTGATGTG
4_iolD_Check_rev	Sequencing Primer	CAGCCACAGATGAAGCTTTG
5_ulaA_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCACCCGTGGCATTACCGTAAAC
5_ulaA_up_rev	Gibson Primer	GATATGAAGATTCCGGAAGGCTTGCGCT
5_ulaA_down_fwd	Gibson Primer	CCTTCCGGAATCTTCATATCCTCAGTGGAAGGG
5_ulaA_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCTGCTCATTGGTGCGGGTG
5_ulaA_Check_fwd	Sequencing Primer	CAGAAGACCGTTGGCAAAGG

5_ulaA_Check_rev	Sequencing Primer	ATCGGTGGAGCAATCAAAGC
6_fasB_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCACAAGATCTGCCTTTGGTGCCTTCG
6_fasB_up_rev	Gibson Primer	CAGGCCTCGATGCCCTTGGTGCTGCTCG
6_fasB_down_fwd	Gibson Primer	ACCAAGGGCATCGAGGCCTGGGGCTGGGA
6_fasB_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCCGGTGGCTACGGACAGATGATCC
6_fasB_Check_fwd	Sequencing Primer	ATCGCAACCAGAGCGGAGAC
6_fasB_Check_rev	Sequencing Primer	CAACGACGTTCTGCAGGAAG
7_pks_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCACACCGTCAGAGTAAGGC
7_pks_up_rev	Gibson Primer	GCATCCTGGTTAAACTTAATTTTGCAGACTGGG
7_pks_down_fwd	Gibson Primer	ATTAAGTTTAACCAGGATGCGGTTGTCCAC
7_pks_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCTTCTTCCGTGGTCTACCAACC
7_pks_Check_fwd	Sequencing Primer	ATGGGCGTAGATGGCCACTG
7_pks_Check_rev	Sequencing Primer	AGCCTGCAGTGTTCATGTTC
8_emb_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCTTTTCACATCGTGGACAAAG
8_emb_up_rev	Gibson Primer	TCTGGTACCACGAATATGTGCGCTACCAAATCGTCATGGAACAAACCGTTG
8_emb_down_fwd	Gibson Primer	ATTTGGTAGCGCACATATTCGTGGTACCAGATCAGTGCCGGGCCCTTCG
8_emb_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCCGTGATCACCAGAGAGGTCATGC
8_emb_Check_fwd	Sequencing Primer	GTTCGCCCATGGAGTGATTG
8_emb_Check_rev	Sequencing Primer	TTCGTCTACCCGCATTGCTC
9_NRPS_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCTGCAAAGCACCGCTTCCC
9_NRPS_up_rev	Gibson Primer	CTCATAGCCATCGATGAGGGTGTCATCGG
9_NRPS_down_fwd	Gibson Primer	CCCTCATCGATGGCTATGAGCTGGGTAATGG
9_NRPS_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCAGCGGCTGAAGAGAGGGTG
9_NRPS_Check_fwd	Sequencing Primer	GACCATCAAGCCTGGTTCTC
9_NRPS_Check_rev	Sequencing Primer	GATTCCACGAACGCATCTTG
10_DNSSeg_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCGTCGCCATCACCTCGCAC
10_DNSSeg_up_rev	Gibson Primer	ACTTTCTCTGAATCTTTGCGTCCTGCCAC
10_DNSSeg_down_fwd	Gibson Primer	CGCAAAGATTCAGAGAAAGTCCGCCGAG
10_DNSSeg_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCCCGTGCAATCGGGATGCC
10_DNSSeg_Check_fwd	Sequencing Primer	CCGCTTCCAAGCAAGTTATG
10_DNSSeg_Check_rev	Sequencing Primer	TCCCACGTCAGCGTGGATAG
13_helicase Check rev	Gibson Primer	CTGAGCCCTTTGATCTTGTC
13_helicase Check fwd	Gibson Primer	ATCGCTTCTTGCCACTGTTC
13_helicase up fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCTTCAGCACTGTCATAGGG
13_helicase up rev	Gibson Primer	ACTGAGCTTAAGGTTAAAAACGATGACG
13_helicase down fwd	Sequencing Primer	TTTTTAACCTTAAGCTCAGTCGTATAAGCATTG
13_helicase down rev	Sequencing Primer	TTGTAAAACGACGGCCAGTGAATTCCCACAGGCATCACCTTGG

14_hypo up fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCAGACGGCACCTGGTCACAG
14_hypo up rev	Gibson Primer	GCAGCGCAGATTTTCGTCCGCACCCCAAC
14 hypo_down fwd	Gibson Primer	CGGACGAAAATCTGCGCTGCTCAAACCC
14_hypo down rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCCAGCATCGCCAGGACATG
14_hypo Check fwd	Sequencing Primer	TCCAGCGTGACCAGCAATTC
14_hypo Check rev	Sequencing Primer	TAGTCGTGCCACCTGTGTTG
15_phosphatase up fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCACTTTGATAACCAAGGCCAAAAG
15_phosphatase up rev	Gibson Primer	GTCACATAGTTCCAAATTCCGGGGACAAATG
15_phosphatase down fwd	Gibson Primer	GGAATTTGGAACTATGTGACCAACGCATTC
15_phosphatase down rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCGTGCTGACTGGAGTGGTG
15_phosphatase Check fwd	Sequencing Primer	TTCGATTACCAAGGCCTACC
15_phosphatase Check rev	Sequencing Primer	GCGTTTGCGCTTGGATCTTC
16_emb up fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCACGCACCGATGGAAATGG
16_emb up rev	Gibson Primer	GCGTCGAGGGAGGCGTCGAAAAGCAAAGTC
16_emb down fwd	Gibson Primer	TTCGACGCCTCCCTCGACGCCCCAATAAG
16_emb down rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCCGCGGATTATATGGCCAAC
16_emb Check fwd	Sequencing Primer	GCTTCGATGCCCTTGATCTG
16_emb Check rev	Sequencing Primer	CAAGCTGAAGCCACTTGATG
17_DNA seg up fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCGTGGCTGAAATGGGTACCG
17_DNA seg up rev	Gibson Primer	TCCCATTCCGTCGTGCGCCGATTCCTTCAG
17_DNA seg down fwd	Gibson Primer	CGGCGCACGACGGAATGGGACCACATGG
17_DNA seg down rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCGGGAATCCAATCCGCGCAAC
17_DNA seg Check fwd	Sequencing Primer	CGCTGGTTGTGCAGTTGTTG
17_DNS seg Check rev	Sequencing Primer	CCGGGTTGGCTGGGTAATTG
18_fasB up fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCGATCCTTGTAGAACTCCAGG
18_fasB up rev	Gibson Primer	AGGTCACCGAACACGATGGTGTGCTTGC
18_fasB down fwd	Gibson Primer	ACCATCGTGTTCGGTGACCTGCTCTGCAAAC
18_fasB down rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCCTCTCGCCGTAACCAGCTC
18_fasB Check fwd	Sequencing Primer	ATCCGACCCAGTTGATGATG
18_fasB down rev	Sequencing Primer	CAGCTTCACTCCTTCTGATG
19_pyc up fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCTTCGCGGCCGCAACACCG
19_pyc up rev	Gibson Primer	CCGTCAACAACATCTGCACCAGCTTGAGCTGC
19_pyc down fwd	Gibson Primer	GGTGCAGATGTTGTTGACGGTGCTTCCGC
19_pyc down rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCATGACAGAGTCTGGGATGTCG
19_pyc Check fwd	Sequencing Primer	ACGATGTGGCGATGCGTTTC
19_pyc Check rev	Sequencing Primer	GGCTATTGCGACGTTCCTTG
20_pks_up fwd_new	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCAGTGGTTCGAGGGTTTCTG

20_pks_up rev_new	Gibson Primer	CATCGAAAATAACTTCCAGATCCCACCAC
20_pks_down fwd_new	Gibson Primer	TCTGGAAGTTATTTTCGATGCGGTTTTTGATGC
20_pks_down rev_new	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCCGACATGCTGCCACCAAG
20_pks_Check_fwd	Sequencing Primer	AGACGACGCATTAGTGGTTG
20_pks_Check_rev	Sequencing Primer	TCCGCGTGAATCTGCCAAAC
21_fasA up fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCCTGGCGGCTGGGTTGGCC
21_fasA up rev	Gibson Primer	GAACCCAAGACATAGGCTTGTGGTGCTTGCGG
21_fasA down fwd	Gibson Primer	CAAGCCTATGTCTTGGGTTCCAGCAATCG
21_fasA down rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCCTCACGGGAAACAACTGG
21_fasA Check fwd	Sequencing Primer	AAGCAGGCACTGGTCGACAC
21_fasA Check rev	Sequencing Primer	GCGGAAACAGATCCAACACC
22_mrpA_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCGCAATAACATTCATTCCC
22_mrpA_up_rev	Gibson Primer	ACTCCGAAAACACCCAGAAAAAACCAATG
22_mrpA_down_fwd	Gibson Primer	TTTCTGGGTGTTTTCGGAGTTCATCAAAGGCAC
22_mrpA_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCCCCAGGAGCTTGAGTAGG
22_mrpA_Check_fwd	Sequencing Primer	ACAACGTTTCCCGCACCTGG
22_mrpA_Check_rev	Sequencing Primer	CACATCTGCCCAGGAGCTTG
23_helicase_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCAACCTTAGCCCACTTCATC
23_helicase_up_rev	Gibson Primer	AAATCTACTTTCAGATCTTGCTATGTGGG
23_helicase_down_fwd	Gibson Primer	CAAGATCTGAAAGTAGATTTGTTGCAAGG
23_helicase_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCATAAGAAAAACAAATCTGCATTAAAG
23_heicase_Check_fwd	Sequencing Primer	GCGATGGCAGAACGTGAACC
23_helicase_Check_rev	Sequencing Primer	TGCCGTTGCCACAATTGAGG
24_phosphatase_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCTTTCTTCGATGCCACCTTC
24_phosphatase_up_rev	Gibson Primer	ACAGTATCGTCACTTGGATCACTTGCTC
24_phosphatase_down_fwd	Gibson Primer	GATCCAAGTGACGATACTGTCAAGATGACGTTCCTGGAAG
24_phosphatase_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCGGTGCGCTGCGGTCACGT
24_phosphatase_Check fwd	Sequencing Primer	CGACTAATTCTGCCTGATGG
24_phosphatase_Check rev	Sequencing Primer	GCACGTAGTCATGGATGTTG
25_gltB_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCCCACTACTTGAACAACTGC
25_gltB_up_rev	Gibson Primer	TTCTTCGGCATCGAGAAGAGCTGCGATAAC
25_gltB_down_fwd	Gibson Primer	CTCTTCTCGATGCCGAAGAATTCGGTTTTG
25_gltB_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCCAACTGAACGGTTGACGTTG
25_gltB_check_fwd	Sequencing Primer	CTGGGCGGTATGTCCAACTC
25_gltB_check_rev	Sequencing Primer	TGGGATGAACGCGCCAAAGG
26_hypo_up fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCCTCCGCCCGACCTTTTTC
26_hypo_up_rev	Gibson Primer	ATGTGCGCACCCTGATTTCCGGAATACAAAATCTG

26_hpyo_down_fwd	Gibson Primer	GGAAATCAGGGTGCGCACATCAATACTCCG
26_hypo_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCATCACCATCCGTGCGTGC
26_hypo_Check_fwd	Sequencing Primer	GGTGGACAACGGCGATGAAG
26_hypo_Check_rev	Sequencing Primer	AGTCGAGCGCACCACATCAG
27_putA_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCGTTTGGAATTCACGGTTGGTTTCGTGGATCG
27_putA_up_rev	Gibson Primer	TGCCTTTCGACGCCCGCGCGGCCTCTAT
27_putA_down_fwd	Gibson Primer	CGCGCGGGCGTCGAAAGGCACGAAGTTCATCAACCTGGAC
27_putA_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCGGCGCGCGCCTGATCGGA
27_putA_Check_fwd	Sequencing Primer	CTGACCAGGCTGTGGACAAG
27_putA_Check_rev	Sequencing Primer	GGCGCACGAGGTAAGAAATG
28_Cu-ATPase_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCATTCGGGGCGGATTTTGTCCTCCAC
28_Cu-ATPase_up_rev	Gibson Primer	TCGCGATCTTCTCCGAGCGCGCCGCGAA
28_Cu-ATPase_down_fwd	Gibson Primer	GCGCTCGGAGAAGATCGCGATGACCAGCG
28_Cu-ATPase_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCGGCCGAGAAGGTCGTCGAC
28_Cu-ATPase Check fwd	Sequencing Primer	GGCGAAGACCTCATCGATCC
28_Cu-ATPase_Check_rev	Sequencing Primer	ATCATGCTGTTGGGCCACTG
29_helicase_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCCCAAGATGGGAGCCCACAATTC
29_helicase_up_rev	Gibson Primer	ACCGCTTTAATAATCCCGCTGCACAAGG
29_helicase_down_fwd	Gibson Primer	AGCGGGATTATTAAAGCGGTCGAGGTCTAAG
29_helicase_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCCACGCTCCTCCAGACTTC
29_helicase_Check_fwd	Sequencing Primer	AACCAGGGCATGTTCCTTAG
29_helicase_Check_rev	Sequencing Primer	TGCCGATCCTACAAAGGTTC
30_xylB_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCGAATGAGCTGTCGGCGGATATTTC
30_xylB_up_rev	Gibson Primer	CTAACAACCTCTCGGTAGCTTGATCCAGC
30_xylB_down_fwd	Gibson Primer	AGCTACCGAGAGGTTGTTAGAACGCGCGG
30_xylB_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCATCTCCCGTGCCTGCAGC
30_xylB_Check_fwd	Sequencing Primer	AAGACAATGGCGGATCATCG
30_xylB_Check_rev	Sequencing Primer	GATGGATCGTGGACGCTATG
del fasB_Check fwd	Sequencing Primer	GGAGGATACATCCACGGTCATTG
del fasB_Check rev	Sequencing Primer	CGCTATGAGTTCAGGATGTTGATCG
del2964_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCAGGACAATGGGGAAGAGAC
del2964_up_rev	Gibson Primer	GATTACTGATGCCGTTGAGACCATGAGAATTG
del2964_down_fwd	Gibson Primer	TCTCAACGGCATCAGTAATCGTGGACATGCATTCC
del2964_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCAACCCAGCGATGGCCGTT
del2964_Check_fwd	Sequencing Primer	CCGCATCCCTGATATTTCGAACTG
del2964_Check_rev	Sequencing Primer	GTACACGTACTTGTACCACTGC

del pks_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCACGCGCGATCTCTGAGTAC
del pks_up_rev	Gibson Primer	GAGCCAATCGAGTCGCCGCATTGATGAG
del pks_down_fwd	Gibson Primer	TGCGGCGACTCGATTGGCTCTGTTCCATATTG
del pks_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCTGACATTGAAAAGCTCATCATC
del pks_Check_fwd	Sequencing Primer	CTGAGTACATCGCCAAGGAG
del pks_Check_rev	Sequencing Primer	TGGCACTGCACACGGTGTTG
del emb_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCCATATGCTTATCCACGAGGTTG
del emb_up_rev	Gibson Primer	AGTCGGTGGTCTCTGGAATCCAGGGCATATG
del emb_down_fed	Gibson Primer	GATTCCAGAGACCACCGACTTGGCGCATAG
del emb_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCACGCGCAAGATCTGCCGC
del emb_Check_fwd	Sequencing Primer	ATCTCCGACTGGTACTCATC
del emb_Check_rev	Sequencing Primer	CCGGAGCTGCACGTTATTAC
del gltB_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCCTCATCCCAATTGGCGGTG
del gltB_up_rev	Gibson Primer	TTGCTGGGTCGAGTCCTTGTGGTTTCATGC
del gltB_down_fed	Gibson Primer	ACAAGGACTCGACCCAGCAATCAAGATCATGGAGGCAGTG
del gltB_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCGCCAGCGGGGCCGGAACC
del gltB_Check_fwd	Sequencing Primer	AAGCTGCAACGCCTTCGATTTTTCC
del gltB_Check_rev	Sequencing Primer	GCCGTAGCGCATGAGGCCGCCGAGG
del NCgl0552_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCAAAGTCATCAATTCCACG
del NCgl0552_up_rev	Gibson Primer	TCTGTGCAACCTCCACTGATGTTGTCATTTTC
del NCgl0552_down_fwd	Gibson Primer	ATCAGTGGAGGTTGCACAGATGGCACGC
del NCgl0552_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCCTGCGTTGAAATCTCCGC
del NCgl0552_Check_fwd	Sequencing Primer	GTGTTTCTGCAAGCGGAATC
del NCgl0552_Check_rev	Sequencing Primer	GAAATCTCCGCCACCTTATC
del NCgl1737_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCTGAACATTTGCTGCACGAATTTC
del NCgl1737_up_rev	Gibson Primer	GACAGTTTTAGGCGTAACCGAGAAGGGTG
del NCgl1737_down_fwd	Gibson Primer	CGGTTACGCCTAAAACTGTCGCAGTCAC
del NCgl1737_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCAGGTAAGGGCTTTGATAAAG
del NCgl1737_Check_fw	Sequencing Primer	TCGACGAACCTCGACTTCAG

del NCgl1737_Check_rev	Sequencing Primer	TGACGAGGTGCGTAGTGATG
del NCgl2959_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCGCTGTGGCTGCGGCGATT
del NCgl2959_up_rev	Gibson Primer	AGTCCCAAATGAGCTTTGAGATGTTCATAATTTTTCCTAGATCCAATG
del NCgl2959_down_fwd	Gibson Primer	CTCAAAGCTCATTTGGGACTATGTGACCAAC
del NCgl2959_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCCTGGAGTGGTGGGTGAAATC
del NCgl2959_Check_fwd	Sequencing Primer	TTTACGCTCCCACCAAATCC
del NCgl2959_Check_rev	Sequencing Primer	TGCTTCGGTGGTGCTGACTG
del NRPS cps_Check fwd	Sequencing Primer	GCAAACATAGTTCTAGATCAG
del NRPS cps_Check rev	Sequencing Primer	TGGATCGGGTGAGCACCTTTG
del pknB_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCCTGGATCCGCCCGCACAA
del pknB_up_rev	Gibson Primer	CGTGATCGCTCTCTTCGAATTCGATCTCGCTGC
del pknB_down_fwd	Gibson Primer	ATTCGAAGAGAGCGATCACGAAGGTCAC
del pknB_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCGCTAGCTGTTTCTGCTGTG
del pknB_Check_fwd	Sequencing Primer	AACAACTGCCGCAGCACAAC
del pknB_Check_rev	Sequencing Primer	GAAACGCGGTTCCGGCATTG
del putA_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCTGGCCACTGGTGGTGACATC
del putA_up_rev	Gibson Primer	CGTGGAAGGCCAGATTCATCGACGTCATGGTG
del putA_down_fwd	Gibson Primer	GATGAATCTGGCCTTCCACGAGTTGGCG
del putA_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCTTACTGCCCCGAGCATAGG
del putA_Check_fwd	Sequencing Primer	ATGTCGAATACCGGATCGTC
del putA_Check_rev	Sequencing Primer	GTCACGCCGTGCTCCATTTC
del ggtB_up_fwd	Gibson Primer	CCTGCAGGTCGACTCTAGAGGATCCACTGCTTCCGAGGATCTG
del ggtB_up_rev	Gibson Primer	GACCCAAAGCGATCCACGTAGAGAAGGC
del ggtB_down_fwd	Gibson Primer	TACGTGGATCGCTTTGGGTCACTTTCATC
del ggtB_down_rev	Gibson Primer	TTGTAAAACGACGGCCAGTGAATTCCTGCGAATAAATAGGAATAGTTAAAAAC
del ggtB_Check_fwd	Sequencing Primer	GATGGTCAGCCAAGAAATCC
del ggtB_Check_rev	Sequencing Primer	TTGCGGAATCGGTTGATTCG
del NCGL2981_Check2_fwd	Sequencing Primer	CAGCAGTCCCGTTTGGTTAC
del	Sequencing Primer	ACCGCACCCAACACGATTTC

NCGL2981_Check2 _rev		
delCGP3_D20_Chec k fwd	Sequencing Primer	CTACCGTTACAGCAGCTCAG
delCGP3_D20_Chec k rev	Sequencing Primer	CCGCACTGGAATTAGCTTTG
del pyk1_up_fwd	Gibson Primer	CGACGCGAAGGCTAACTGCATCCATG
del pyk1_up_rev	Gibson Primer	AGACGGCATGGATTCACGTCCACCTTCTTG
del pyk1_down_fwd	Gibson Primer	ACGTGAATCCATGCCGTCTTCTACCAAAC
del pyk1_down_rev	Gibson Primer	GGCATCCGAATCAACACCATC
del pyk1_Check fwd	Sequencing Primer	CCATTGGTTCAACGCTAAGG
del pyk1_Check rev	Sequencing Primer	AGGGCATTGATGGAGAAACG
del pyk2 up fwd	Gibson Primer	CAGGTCGACTCTAGAGGATCCTTAGCCGAATGCTGGACAAAG
del pyk2 up rev	Gibson Primer	GGGTAGGTGATTTGAATTTGTCTGGTCAAACTCATTCATTTGA GCTCC
del pyk2 down fwd	Gibson Primer	ACAAATTCAAATCACCTACCCCTGCTGCTGCGCAAGGTGAAG
del pyk2 down rev	Gibson Primer	AAAACGACGGCCAGTGAATTCAATCCTTGAAGATCCAAATGTT G
del pyk2 Check fwd	Sequencing Primer	CAGACAGGGAGGACAAGAATC
del pyk2 Check rev	Sequencing Primer	ACACGGCGACACTGAACTTCC

9 Danksagung

10 Erklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf" erstellt worden ist. Die Dissertation wurde in der vorgelegten oder ähnlichen Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Jülich, den 30.03.2023

Philipp Tobias Baumann