Novel approaches for *in vivo* evolution, screening and characterization of enzymes for metabolic engineering of *Escherichia coli* as hyper L-tryptophan producer

Vom Promotionsausschuss der

Technischen Universität Hamburg

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften (Dr. rer. nat)

genehmigte Dissertation

von

Minliang Chen

aus

Fujian, China

2020

Gutachter:

Prof. Dr. An-Ping Zeng

Prof. Dr. Andreas Liese

Prüfungsausschussvorsitzender:

Prof. Dr. Frerich Keil

Tag der mündlichen Prüfung:

27. November 2020

Acknowledgement

First of all, I would like to express my sincere gratitude to my supervisor Prof. Dr. An-Ping Zeng for giving me the opportunity to do my PhD work at the Institute of Bioprocess and Biosystems Engineering (IBB), Hamburg University of Technology, and for his support, patience, and immense knowledge. His guidance and experience helped me in the past four years of research and writing of this thesis.

Also, I would like to express my genuine appreciation to Prof. Dr. Andreas Liese for being a member of my thesis committee and to Prof. Dr. Frerich Keil for being the Chair of the committee.

I would like to thank Dr. Lin Chen, for the stimulating discussions in pathway optimization, protein engineering, and microbial fermentation, also for his help with the review of this thesis. I also thank Dr. Chengwei Ma for his work in protein structure analysis and design of the automatic evolutionary system, and for his help in reviewing this thesis. In addition, I thank Shamini Ramkumar Thirumalasetty for working as master student together on pathway engineering. I would also like to thank Dr. Wei Wang who introduced me to the GC, also for her help in reviewing this thesis. Besides that, I would like to thank Anna Gorte who introduced me to the HPLC. I also would like to thank Dr. Wael Sabra, Yaeseong Hong, and Philipp Arbter who introduced me to the DASGIP bioreactor system and offer relative troubleshooting tips. Moreover, I would like to thank Dr. Johannes Möller who introduced me to online monitoring of fluorescence intensity. I thank Jan Sens for his help in preparing the gas supply for fermentations. Without their generous support, it would be harder for me to finish everything in this thesis. Meanwhile, I would like to take this opportunity to thank Philipp Arbter, Ludwig Selder, and Yaeseong Hong for their help with the German abstract of this thesis. Then, I would like to express my gratitude to other IBB members for their willingness to share their experiences and support me in many ways.

Last but not least, I would like to thank my friends and family, especially my wife Wan Fang, for her love, her selfless support, and her understanding. I owe you everything.

Contents

Abstract

Nowadays, microbes have been extensively optimized for production of L-tryptophan (Trp) from renewable feedstocks. Numerous strategies have been investigated, including tuning the gene expression levels and alleviation of negative regulations. However, it has become apparent that development of a more efficient Trp producer will be inconceivable without broader optimization of the corresponding enzymes. Thus, the present thesis aims to develop new *in vivo* evolution, screening and characterization methods for optimization of enzymes for Trp production.

A reliable *in vivo* screening approach is desired to link the mutations to cell growth or to couple the inconspicuous intracellular molecules with a biomarker, for example, the enhanced green fluorescent protein (eGFP). In the first part of this thesis, a novel enzyme screening approach, namely plasmid-assisted growth-coupled and sensor-guided *in vivo* screening (PGSS), is developed. This approach combines the advantages of complementary auxotrophy-coupled screening with biosensor-driven *in vivo* characterization. The efficiency of PGSS was first demonstrated for improving an anthranilate (ANTH)-inhibited enzyme TrpC from *E. coli* (*Ec*TrpC), which is composed of indole glycerol phosphate synthase and N-(5-phosphoribosyl) anthranilate isomerase. Based on a Trp-auxotrophic strain S028Δ*EctrpC*, a highly efficient ANTH-resistant candidate *Ec*TrpC$^{S58Q\text{-}P59V\text{-}S60F\text{-}K61Q}$ was identified by using the PGSS approach. Afterwards, the PGSS approach was employed to identify an ANTH-activated TrpC from *Aspergillus niger* (*An*TrpC). As a result, an enzyme variant (*An*TrpC^{R378F}) that showed increased ANTH activation was discovered. Fed-batch fermentation demonstrated that the strain S028Δ*EctrpC* containing *An*TrpC^{R378F} was able to produce more Trp (35.36 g/L) than the strain containing *An*TrpC^{WT} (31.15 g/L), indicating that the variant *An*TrpC^{R378F} is more efficient for the Trp pathway.

To overcome limitations of screening performed under non-representative conditions, PGSS is combined with the CRISPR/Cas9 technique, resulting in a novel strategy called CRISPR/Cas9-facilitated engineering with growth-coupled and sensor-guided *in vivo* screening (CGSS). The efficiency of this method was demonstrated for the optimization of a key enzyme in the chorismate pathway, namely 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase. *E. coli* possesses three isoenzymes of DAHP synthase:

AroG, AroF, and AroH. The aim was to obtain AroG variants with increased resistance against feedback inhibition by L-phenylalanine (Phe). Starting from a Trp-producing *E. coli* strain (harboring the reference variant AroG^{S180F}), all the endogenous DAHP synthases were removed and the growth of the subsequent strain exhibited dependence on the activity of introduced AroG variants. The different catalytic efficiencies of AroG variants will lead to different intracellular concentrations of Trp, which can be monitored by a Trp biosensor. Taking cell growth rate and the signal strength of a Trp biosensor as selection criteria, several novel Phe-resistant AroG variants with higher activities were identified. The replacement of AroG^{S180F} with the best variant AroG$^{D6G-D7A}$ in a Trp-producing strain significantly improved the Trp production by 38.50%.

A high glucose conversion yield is a key parameter for cost-effective Trp production. Theoretical analysis suggests that activation of galactose permease/glucokinase (GalP/Glk) in a PTS-defective strain could result in an *E. coli* strain with significantly increased Trp yield. To explore this possibility, a laboratory adaptive evolution (LAE) approach was applied. To this end, a potentially GalP/Glk-dependent *E. coli* strain G028 was developed, in which the *ptsI* gene was deleted and a tandem gene circuit with promoter mutation (*ptac*^{MT}-*galP*-*pJ23119*^{MT}-*glk*) was integrated. Batch LAE of this strain resulted in a promising candidate B3. However, B3 exhibited similar Trp yield and production as S028. One conceivable explanation is that the PTS-defective strain is forced to strengthen their growth rather than Trp synthesis in conventional LAE. Thus, a continuous LAE system (auto-CGSS) was developed which combines CGSS-facilitated *in vivo* mutagenesis with real-time measurement of cell growth and online monitoring of fluorescence intensity, leading to a new promising candidate strain D8. Fed-batch fermentation with D8 showed an increase of Trp yield by 23.07% compared with that by B3 (0.16 vs. 0.13 g/g).

Finally, two selected gene variants (*aroG*$^{D6G-D7A}$ and *AntrpC*^{R378F}) were integrated into the chromosome of Trp-producing strains S028G and D8 to establish highly producing strains S028AARF and D8AA, respectively. These strains were evaluated in fed-batch fermentations. Remarkably, S028AARF reached a very high Trp concentration (51.19 g/L) after 65h of fermentation, which is 19.20% higher than that of the previously reported strain S028GΔ*fruR*:*aroG*$^{D6G-D7A}$ (42.95 g/L). Fed-batch cultivations of D8AA clones showed strong variations in growth and Trp production. The reason for the variations is not clear. Nevertheless, one of the clones D8AA-1 exhibited a Trp yield as high as 0.20 g/g (vs. 0.19 g/g with S028AARF), representing the highest Trp yield reported in the literature so far and making it attractive for industrial-scale Trp bioproduction.

Zusammenfassung

In der Literatur wurden Mikroorganismen seit Jahren für die Produktion von L-Tryptophan (Trp) aus erneuerbaren Rohstoffen optimiert. Verschiedene Ansätze wurden dabei untersucht: die Anpassung der Genexpression sowie der Enzymkonzentration und die Abmilderung negativer regulatorischer Mechanismen. Hierbei wurde deutlich, dass die Entwicklung von effizienteren Trp-Produzenten ohne die umfassende Optimierung der zugehörigen Enzyme nicht möglich sein wird. Das Ziel dieser Arbeit ist folglich die Entwicklung von neuartigen *in vivo* Evolutions-, Screenings- und Charakterisierungsmethoden zur Optimierung von Enzymen der Trp-Synthese.

Ein zuverlässiger *in vivo* Screening-Ansatz muss Mutationen mit dem Zellwachstum in Verbindung bringen können oder unscheinbare intrazelluläre Moleküle mit Biomarkern, wie etwa dem „enhanced green fluorescent protein“ (eGFP), koppeln. Im ersten Teil dieser Arbeit wurde hierfür eine neuartige Genvarianten-Screening-Methode, die PGSS genannt wird (aus dem Englischen: *Plasmid-assisted Growth-coupled and Sensor-guided in vivo Screening*), entwickelt. Dieser Ansatz kombiniert die Vorteile von komplementären zell-autotroph-abhängigen Screenings mit biosensor-basierten *in vivo* Charakterisierungsmethoden. Die Effizienz von PGSS wurde hierbei zuerst bei der Verbesserung eines durch Anthranilat (ANTH) inhibierten Enzyms TrpC aus *E. coli* nachgewiesen, das aus der Glycerol-Phosphat-Synthase und der N-(5-phosphoribosyl) Anthranilat-Isomerase besteht. Ausgehend von einem Trp-auxotrophen Stamm S028Δ*EctrpC* konnte mit dem PGSS-Ansatz ein hocheffizienter ANTH-resistenter Kandidat *Ec*TrpC$^{S58Q\text{-}P59V\text{-}S60F\text{-}K61Q}$ identifiziert werden. Anschließend wurde PGSS genutzt, um eine durch ANTH aktivierte TrpC-Variante von *Aspergillus niger* zu identifizieren (*An*TrpC). Hieraus resultierte die Entdeckung einer weiteren Enzymvariante (*An*TrpC^{R378F}), die eine erhöhte Aktivierung durch ANTH besitzt. In anschließenden Fedbatch-Fermentationen konnte gezeigt werden, dass *An*TrpC^{R378F} in 51 h mehr Trp (35.36 g/L) als der Wildtyp *An*TrpC^{WT} (31.15 g/L) produzierte. Dies weist darauf hin, dass *An*TrpC^{R378F} in der Tat einen effizienteren Stoffwechsel zur Trp-Produktion besitzt.

Um Limitierungen des Screenings zu überwinden, die aus der Durchführung unter nicht-repräsentativen Versuchsbedingungen resultieren, wurde PGSS mit der CRISPR/Cas9-Methode kombiniert. Dies führte zu einer neuen Methode, die CGSS (aus dem Englischen: *CRISPR/Cas9-facilitated engineering with Growth-coupled and Sensor-guided in vivo Screening*) genannt wird. Die Effizient dieser Methode wurde zuerst bei der Optimierung der 3-deoxy-D-arabino-heptulosonat-7-phosphat (DAHP)-Synthase, einem Schlüsselenzym im Chorismatsftoffwechselweg, aufgezeigt. *E. coli* besitzt drei Isoenzyme der DAHP-Synthase: AroG, AroF, und AroH. Ziel war es, verschiedene AroG-Varianten mit erhöhter Resistenz gegenüber einer Feedbackinhibierung durch L-Phenylalanin (Phe) zu erhalten. Ausgehend von einem Trp-produzierenden *E. coli* Stamm (mit der AroG-Referenzvariante $AroG^{S180F}$), wurden alle endogenen DHAP-Synthasen entfernt, was das Wachstum dieses Stamms abhängig von den eingeschleusten AroG-Varianten machte. Die unterschiedlichen katalytischen Aktivitäten der verschiedenen AroG-Varianten führen zu unterschiedlich hohen intrazellulären Trp-Konzentrationen, die über den Trp-Biosensor nachverfolgt werden können. Unter Berücksichtigung der Wachstumsrate und der Signalstärke des Biosensorsignals als Selektionskriterien, konnten erfolgreich verschiedene neue Phe-resistente AroG-Kandidaten identifizieren werden. Die Ersetzung von $AroG^{S180F}$ mit der besten Variante $AroG^{D6G-D7A}$ in einem Trp-produzierenden Stamm konnte die Trp-Produktion signifikant um 38.50% steigern.

Ein hoher Glukoseumsatz und eine hohe Ausbeute sind Schlüsselparameter für eine kosteneffektive Trp-Produktion. Theoretische Überlegungen legen nahe, dass die Aktivierung der Galaktose-Permease/Glukose-Kinase (GalP/Glk) in einem PTS-negativen *E. coli* Stamm mit einer signifikant erhöhten Trp-Ausbeute resultieren kann. Um diese Überlegung experimentell zu überprüfen, wurde ein „Laboratory Adaptive Evolution“ (LAE) Ansatz verfolgt. Hieraus folgte ein GalP/Glk-abhängiger *E. coli* Stamm G028, in dem das *ptsI*-Gen deletiert wurde und ein Tandem-Gene-Circuit mit einer Promotermutation ($ptac^{MT}$-*galP*-$pJ23119^{MT}$-*glk*) integriert wurde. Die resultierenden Mutanten wurden unter satzweiser LAE einem Selektionsdruck ausgesetzt und es wurde eine verbesserte Variante (B3) erhalten. Jedoch zeigte sich in anschließenden Fedbatch-Fermentationen, dass die erreichten Produktionstiter des B3 Stamms nicht höher als die von S028 waren. Eine mögliche Erklärung hierfür könnte sein, dass die PTS-negative Variante durch das LAE eher dazu gezwungen wird, ihr Wachstum als die Trp-Produktion zu stärken. Deshalb wurde ein weiteres LAE-System entwickelt, das die CGSS-basierte *in vivo* Mutagenese mit einer Echtzeitmessung des Zellwachstums und der Fluoreszenzintensität kombiniert. Hierdurch wurde ein neuer

Kandidat (D8) erhalten. Weitere Fermentationen mit D8 zeigten, dass der Stamm eine um 23.07% erhöhte Trp-Ausbeute als D3 hatte (0.16 vs. 0.13 g Trp/g Glukose).

Zuletzt wurden zwei ausgewählte Genvarianten (*aroG*$^{D6G-D7A}$ and *AntrpC*^{R378F}) in das Chromosom der Trp-produzierenden Stämme S028G und D8 integriert. Hieraus resultierten die Stämme S028AARF und D8AA mit deutlich verbesserter Trp-Produktion. Die Stämme wurden im Folgenden im Zulaufverfahren unter definierten Bedingungen in Bioreaktoren näher charakterisiert. Bemerkenswerterweise erreichte der Stamm S028AARF eine sehr hohe Trp-Konzentration (51.19 g/L) nach 65 h, was etwa 19.2% höher ist als beim vorherigen Stamm S028GΔ*fruR*:*aroG*$^{D6G-D7A}$ (42.95 g/L). Die Fermentationen zeigten jedoch auch hohe Schwankungen beim Wachstum und der Trp-Produktion. Die Gründe hierfür sind unklar. Nichtsdestotrotz, eine Variante (D8AA-1) zeigte eine Trp-Ausbeute von 0.20 g/g (vs. 0.19 g/g bei S028AARF). Diese Ausbeute stellt den bisher berichteten höchsten Wert dar, was den Stamm für die Trp-Produktion im industriellen Maßstab attraktiv macht.

Abbreviations

AAAs	Aromatic amino acids
ALE	Adaptive laboratory evolution
ANTH	Anthranilate
CDW	Cell dry weight
CdRP	1-(2-Carboxyphenylamino)-l-deoxyribulose 5-phosphate
CRISPR	Clustered regularly interspaced short palindromic repeats
Cas9	CRISPR-associated protein 9
sgRNA	Single guide RNA
CGSS	CRISPR/Cas9-facilitated engineering with growth-coupled and sensor-guided *in vivo* screening
CHA	Chorismate
DAHP	3-deoxy-darabino-heptulosonate
DHAP	Dehydroxyacetone phosphate
DSA	3-dehydroshikimic acid
E4P	Erythrose-4-phosphate
F6P	Fructose 6-phosphate
G6P	Glucose 6-phosphate
Gln/L-Gln	L-Glutamine
MFU	Medium fluorescence unit
MP	Mutagenesis plasmid
IGP	Indole-3-glycerolphosphate
PEP	Phosphoenolpyruvate
PGSS	Plasmid-assisted growth-coupled and sensor-guided *in vivo* screening
Phe/L-Phe	L-Phenylalanine
PRA	N-(5-phosphoribosyl)-anthranilate
PRPP	5-phospho-α-D-ribose 1-diphosphate
PYR	Pyruvate
SA	Shikimic acid/Shikimate
Ser/L-Ser	L-Serine

Trp/L-Trp	L-Tryptophan
Tyr/L-Tyr	L-Tyrosine
Phe/L-Phe	L- Phenylalanine
E. coli	*Escherichia coli*
M. tuberculosis	*Mycobacterium tuberculosis*
S. solfataricus	*Sulfolobus solfataricus*
A. niger	*Aspergillus niger*
C. glutamicum	*Corynebacterium glutamicum*
S. cerevisiae	*Saccharomyces cerevisiae*
AroE	Shikimate dehydrogenase
AroF	DAHPs, Tyrosine repressible
AroG	DAHPs, Phenylalanine repressible
AroH	DAHPs, Tryptophan repressible
DAHPs	3-deoxy-D-arabino-heptulosonate-7-phosphate synthase
eIGPs	IGPs from *Escherichia coli*
IGPs	Indole-3-glycerolphosphate synthase
mIGPs	IGPs from *Mycobacterium tuberculosis*
TprC	N-(5-phosphoribosyl)anthranilate isomerase and indole-3-glycerolphosphate synthase
TrpA	Tryptophan synthase, alpha subunit
TrpB	Tryptophan synthase, beta subunit
TrpD	Anthranilate synthase component II
TrpE	Anthranilate synthase component I
TrpR	Tryptophan transcriptional repressor
trpL	Trp operon leader
CsrA	Carbon storage regulator
FruR	Fructose repressor
PP pathway	Pentose phosphate pathway
PTS	Phosphotransferase system
GalP	Galactose permease
Glk	Glucokinase

Chapter 1

1 Introduction and objectives

1.1 Uses of L-tryptophan

L-Tryptophan (Trp) is a non-polar aromatic amino acid and plays an essential role in the biosynthesis of proteins (Fig. 1.1). Trp is also a nutritionally essential amino acid in humans, which serves several specific purposes on keeping nitrogen balance in adults and contributing to growth in infants (Albanese et al., 1956; Singer, 2007).

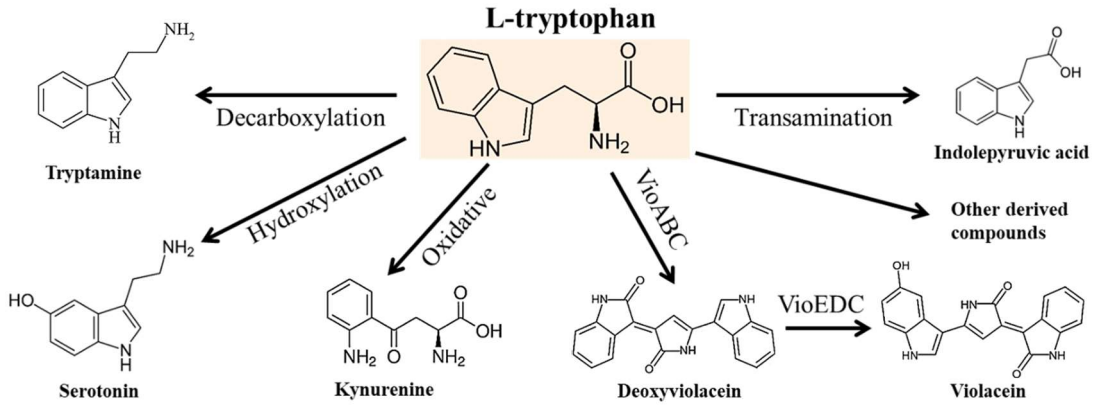

Figure 1.1: Structures of L-tryptophan and its derived bioactive compounds (adapted from Chen (2017)).

As shown in Fig. 1.1, Trp contains a α-carboxylic group, a α-amino group, and an indole group. Specifically, the functional side chain indole group makes Trp a fundamental precursor to a board range of biologically active compounds (Fig. 1.1). For instance, Trp can be metabolized into various neurotransmitters, such as tryptamine (Williams et al., 2014), serotonin (Mateos et al., 2009), and kynurenine (Badawy, 2017). Also, Trp is used as a starting compound for the biosynthesis of antitumor drugs violacein and deoxyviolacein (Fang et al., 2015; Rodrigues et al., 2013). It is not surprising that the demand for Trp is increasing remarkably, with annual global production of Trp being increased to 41,000 metric tons in 2019 (Wendisch, 2020). The remarkably increasing demand for Trp inevitably requires the development of more effective production methods. So far, Trp is mainly chemically synthesized from fossil feedstocks (Lee and

Wendisch, 2017). However, considering the non-renewable fossil resources, a promising alternative way is its bio-production, in which microbial cells produce Trp from renewable feedstocks, including carbohydrates, such as glucose and glycerol (Li et al., 2020). To date, bioproduction of Trp is mostly realized in *Escherichia coli* (*E. coli*) because of its rapid growth, well-known metabolism, and robust behavior during fermentation.

1.1.1 Bioproduction of L-tryptophan

In *E. coli*, the biosynthetic pathway of Trp can be divided into two parts: chorismate pathway and Trp branch (Fig. 1.2). As shown in Fig. 1.2, the chorismate pathway connects the glycolysis and the PP pathway and ends in the formation of chorismate. It begins with the condensation of two molecules of phosphoenolpyruvate (PEP) and one molecule of erythrose 4-phosphate (E4P) to form 3-deoxy-d-arabino-heptulosonate-7-phosphate (DAHP) by DAHP synthetase. The condensation product DAHP is subsequently converted into chorismate via seven steps catalyzed by the enzymes encoded by *aroB*, *aroD*, *aroE*, *aroK*, *aroL*, *aroA*, and *aroC*, respectively. In Trp branch, Trp is produced from chorismate in five steps catalyzed by the *trp* operon (*trpEDCBA*) encoded enzymes (Sprenger, 2006). Till now, considerable attention has been paid to the sustainable production of Trp in microbial cells using the strategies of metabolic engineering (Fujiwara et al., 2020; Huccetogullari et al., 2019a; Ikeda, 2006a).

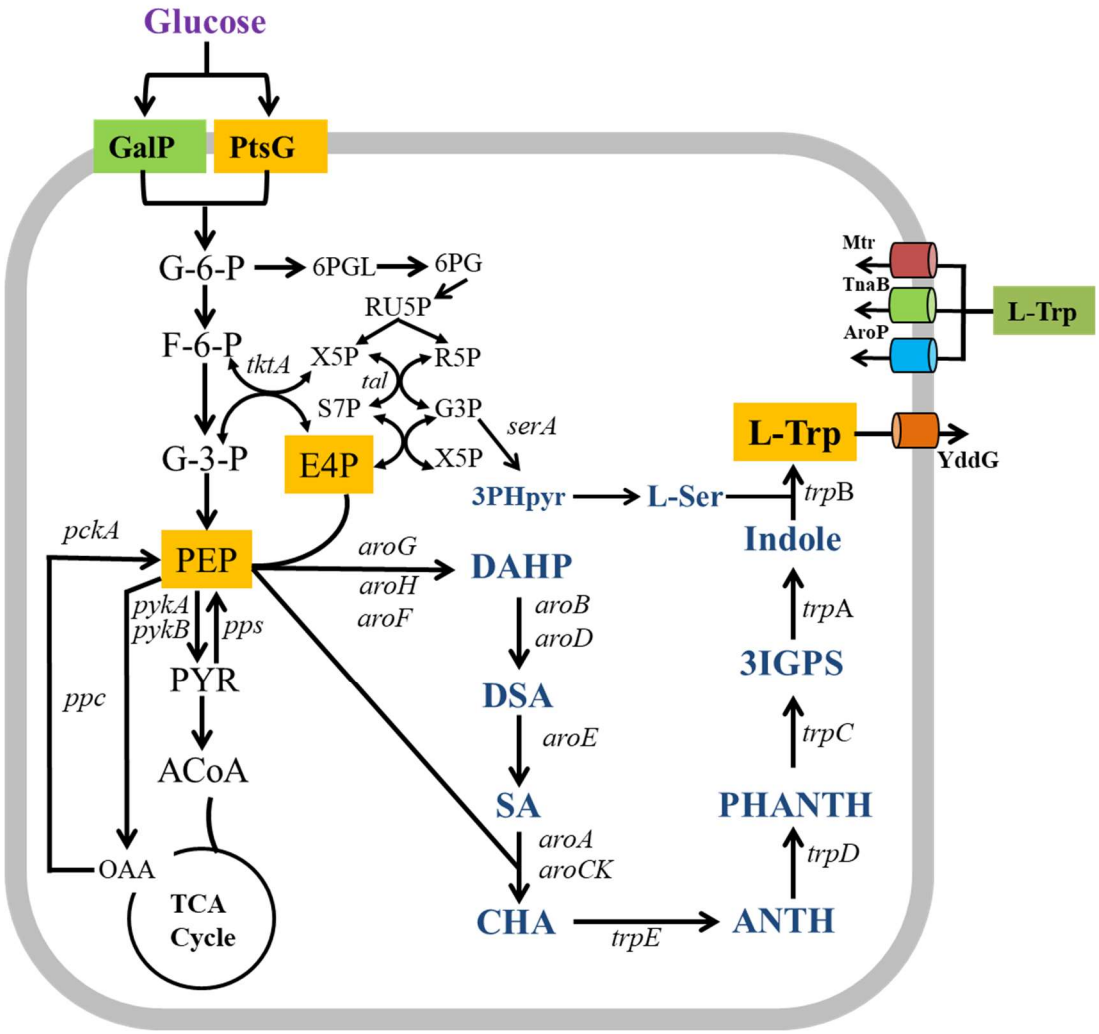

Figure 1.2: Simplified biosynthetic pathway of L-Trp in *E. coli* (adapted from Chen (2017) and Fig. 1 of Niu et al. (2018)). G6P glucose-6-phosphate, F6P fructose-6-phosphate, G3P glyceraldehyde-3-phosphate, 6PGL 6-phosphoglucono-lactone, 6PG 6-phosphogluconate, RU5P ribulose-5-phosphate, X5P xylulose-5-phosphate, R5P ribose-5-phosphate, S7P sedoheptulose-7-phosphate, E4P erythrose-4-phosphate, PEP phosphoenolpyruvate, PYR pyruvate, ACoA acetyl coenzyme A, OAA oxaloacetic acid, DAHP 3-deoxy-arabinoheptulosonate-7-phosphate, DHQ 3-dehydroquinate, DSA 3-dehydroshikimate, SA shikimate, CHA chorismate, ANTH anthranilate, PHANTH phosphoribosyl anthranilate, I3GP indol-3-glycerol phosphate, 3PHpyr 3-phospho hydroxypyruvate, L-Ser L-serine, L-Trp L-tryptophan.

Early attempts to develop efficient Trp-producing strains mainly relied on the traditional approaches of random mutagenesis to introduce genetic variants in Trp-producing strains so as to delete competing pathways or alleviate feedback inhibitions (Chan et al., 1993; Tribe and Pittard, 1979). For instance, one of the most effective production strains obtained by repeated random mutagenesis was able to form 54.6 g/L of Trp (Azuma et al., 1993), which represents one of the competitive candidates for industrial Trp bioproduction. A drawback of such approach is that unwanted mutations

are regularly generated in the bacterial genome. With the advancement of rational metabolic engineering, targeted modifications of a metabolic pathway, e.g., through deletion, overexpression, and deregulation of target genes, have become feasible, and these strategies have been widely applied in the development of many Trp-producing microbial cell factories (Zhao et al., 2012; Chen et al., 2018; Chen and Zeng, 2017; Chen et al., 2018b). Zhao et al. (2012) carried out rational metabolic engineering on an *E. coli* strain, and the resulting strain could produce 14.7 g/L of Trp with a glucose conversion yield of 0.12 g/g. In the study of Chen and Zeng (2017), they (i) disrupted the tryptophannase (TnaA)-catalyzed degradation pathway and the Trp-specific importers TnaB and Mtr (Fig. 1.2), (ii) deregulated the feedback regulation ($TrpC^{S60A}$), attenuation (P_{trc}-$trpE^{S40F}DCBA$), and repression (Δ*trpR*), (iii) enhanced enzymatic activities of rate-limiting steps ($P_{J23119\text{-}rpsL\text{-}tac}$-$aroG^{S180F}$), and (iv) increased the availability of the precursor L-serine in the *E. coli* strain DY330. The resulting strain S028 was capable of producing 40.3 g/L of Trp with a glucose conversion yield of 0.15 g/g. These studies demonstrate that rational metabolic engineering is preferentially applied to develop efficient and industrially attractive producers.

However, the titer and yield of Trp production from glucose in these rationally designed strains are often lower than those achieved in randomly mutated strains (Dodge and Gerstner, 2002). It is presumably due to the fact that unrecognized but favorable gene variants are generated during random mutagenesis. Floras et al. (1996) found that inactivation of the PEP:sugar phosphotransferase system (PTS) in a strain generated by conventional random mutagenesis conducive to higher glucose conversion yield (Floras et al., 1996). The reason behind this may be due to that PTS requires the consumption of 1 mol of PEP for each mol of internalized glucose. Inactivation of the PTS system leads to enhance supply of PEP as precursor for Trp biosynthesis (Carmona et al., 2015; Chen et al., 2018; Lu et al., 2012). So far, inactivation of the cytoplasmic components of PTS system (e.g., *ptsHIcrr* operon) has been the representative strategy to repeal the PTS-facilitated glucose transport system. Theoretically, over 80% of PEP can become available in a PTS-negative strain compared to a PTS-positive strain which can be used entirely for the synthesis of aromatic compounds (Chen, 2017) (Fig. 1.3B). However, the PTS-negative strain turned out to be seriously impaired in its growth capability due to impaired PTS-mediated glucose transport. As a promising strategy to restore the glucose transport, activation of the secondary carrier-type facilitators, the galactose

permease (GalP) and glucokinase (Glk) pathway, in PTS-negative strains has proven to contribute to supporting their growth and increase the glucose conversion yield towards Trp (Lu et al., 2012). As shown in Fig. 1.3B, the maximum theoretical Trp yield in a GalP/Glk-dependent strain is calculated to be 0.45 g/g, which is approximately twice that of a PTS-positive strain (0.23 g/g, Fig. 1.3A). It is, therefore, interesting to experimentally investigate the potentially achievable Trp yield in a GalP/Glk-dependent glucose utilization strain.

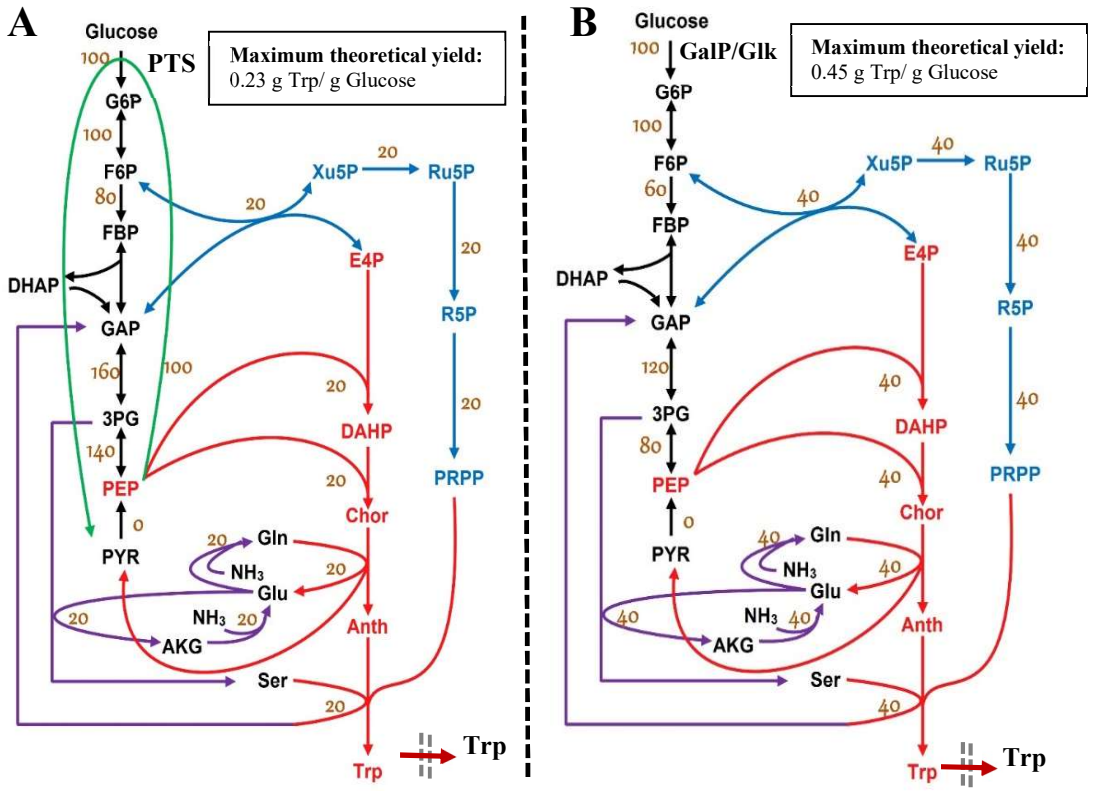

Figure 1.3: Optimal metabolic flux distributions calculated for achieving maximum yield of L-tryptophan (Trp) production in *E. coli*. (A) *E. coli* using the phosphoenolpyruvate:sugar phosphotransferase system (PTS) for glucose uptake. (B) *E. coli* using the GalP/Glk pathway for glucose uptake. (Chen, 2017).

1.1.2 Strategies for directed evolution and screening of enzyme variants

Protein engineering of rate-limiting enzymes for the overproduction of valuable aromatic compounds often depends on directed evolution techniques. Directed evolution of enzymes can be achieved either by introducing random mutagenesis *in vitro* (e.g., error-prone PCR and site saturation mutagenesis, Table 2.1) or by genome-wide mutagenesis *in vivo* (e.g., modified natural mutator, Table 2.2). In addition, existing approaches such as semi-rational design regularly utilize the interrelated

information among protein sequence, structure, and function, as well as computational predictive algorithms (Amrein et al., 2017; Chen et al., 2009). This allows us to preselect promising targets *in silico* and generate a small, high-quality library with higher functional properties by narrowing the diversity of amino acids. Therefore, two complementary strategies for engineering of enzymes TrpC and AroG are to be adopted: directed evolution and semi-rational design. In general, mutations obtained from this complementary approach are confined within the target sites that may not be enough for the development of industrial production strains, which often require a simultaneous modulation of multiple genes, such as the optimization of glucose uptake system in a PTS-negative strain. In principle, *in vivo* genome-wide mutagenesis enables simultaneous mutation and selection under living intracellular conditions, thereby minimizing possible discrepancies between the function of an enzyme variant developed *in vitro* and that generated under living cell state (d'Oelsnitz and Ellington, 2018). In addition, *in vitro* directed evolutionary approaches deal with the construction of a gene variant library and screening of candidate enzyme variants separately and generally require experimental manipulations, while continuous *in vivo* evolution seamlessly integrates these processes into an intact evolutionary cycle (d'Oelsnitz and Ellington, 2018). Therefore, *in vivo* continuous mutagenesis system is able to generate mutations on the genomic scale under intracellular conditions that are more promising in the development of high-performance producers.

Advanced metabolic engineering requires high-throughput screening methods to facilitate the characterization of a large number of microbial strains generated by directed evolution. Normally, cells that are auxotrophy for essential small molecules can be selected based on this property. This can be employed as an efficient selection and screening platform for directed evolution of target enzymes (Hall, 1981). Although this screening approach provides the most readily discernible phenotype, cell growth only represents an indirect indicator of the catalytic activity. Moreover, the growth-coupled screening approach generally cannot achieve high throughput if the target small-molecule, like the majority of intracellular metabolites, cannot be readily detected. It is therefore important to use biosensors that enable quantification of intracellular metabolites at high throughput by transducing the concentration of inconspicuous small molecules into a readily detectable output such as a fluorescence signal (Zeng et al., 2020). Therefore, this thesis is to propose a method to combine

plasmid-assisted engineering with growth-coupled *in vivo* screening and biosensor-driven *in vivo* characterization (PGSS) to ensure the selection of enzyme variants with desired functions. However, the plasmid-facilitated mutagenesis approach is introduced to confine the mutagenesis within the targeted plasmid so that the results, e.g., the Trp specific production rate, from this approach are not relevant to the intracellular environmental conditions of a host strain to be used for the bioproduction process. Recently, the CRISPR/Cas9 technique has become a powerful genome-editing tool to directly integrate target genes into the chromosome of a production strain (Fig. 2.2). Thus, coupling the CRISPR/Cas9 technique with the PGSS approach is to be a better alternative to overcome the above restrictions.

1.2 Objectives

The principal objective of this work is to design and construct a Trp producing strain with robust growth and high efficiency in Trp production and yield by engineering key enzymes involved in the Trp pathway. The targets include (1) anthranilate-resistant *Ec*TrpC enzyme and anthranilate-activated *An*TrpC enzyme; (2) Phe-resistant AroG enzyme; (3) the secondary carrier-type facilitators, GalP/Glk-facilitated glucose uptake system.

To effectively *in vivo* screen and characterize enzyme variants, an approach of linking plasmid-assisted protein engineering with growth-coupled and sensor-guided *in vivo* characterization (PGSS) requires to be established and first used to develop the enzymes *Ec*TrpC and *An*TrpC. In this approach, cell growth is directly linked to the catalytic efficiency of TrpC enzymes, and the intracellular concentration of Trp is monitored by a Trp sensor (P*tac-tnaC-eGFP*) developed by Fang et al., (2015).

To ensure that the screening and characterization of gene variants are performed under conditions that are relevant to the cultivation or intracellular conditions of the production strain, CRISPR/Cas9-facilitated engineering of gene variants integrated with growth-coupled and sensor-guided *in vivo* screening (CGSS) is to be developed for engineering of AroG enzyme. Then, to introduce the genome-scale mutagenesis in a continuous mode, a novel and scalable approach, CGSS-facilitate *in vivo* continuous mutagenesis system with real-time measurement of cell growth and online monitoring of the fluorescence intensity (auto-CGSS) is to be applied to continuously develop GalP/Glk-dependent glucose utilization strains to increase glucose conversion yield

further. Finally, integrative metabolic engineering should be performed for integrating all selected enzyme variants into previously developed Trp-producing strains S028 and D8.

Chapter 2

2 Theoretical and technological backgrounds

In this chapter, strategies of metabolic engineering for Trp production in *E. coli* are summarized. An overview regarding the development of *in vitro* directed evolution is then provided, followed by a discussion of *in vivo* directed evolution. In addition, development of colorimetric-, growth-, and biosensor-coupled screening techniques is also summarized, as well as the theoretical and technological backgrounds of CRISPR/Cas9-facilitated genome editing tools. Finally, the concept of CGSS-assisted *in vivo* continuous mutagenesis is introduced.

2.1 Metabolic engineering of *E. coli* for L-Trp bioproduction

Metabolic engineering for the production of Trp has been reviewed in several publications (Fujiwara et al., 2020; Huccetogullari et al., 2019a; Ikeda, 2006a). To realize Trp overproduction (Fig. 2.1), metabolic engineering of Trp pathway requires (i) alleviation of all restrictive regulations (repression, attenuation, feedback inhibition, and feed-forward regulation) (Chen et al., 2018b; Ikeda, 2003; Oldiges et al., 2004); (ii) deletion of competing pathways; (iii) enhancement and balancing of precursor supplements in the shikimate pathway as well as the specific branch, like chorismate (Ikeda, 2006a); and (iv) removal of Trp degradation pathway (Aiba et al., 1980).

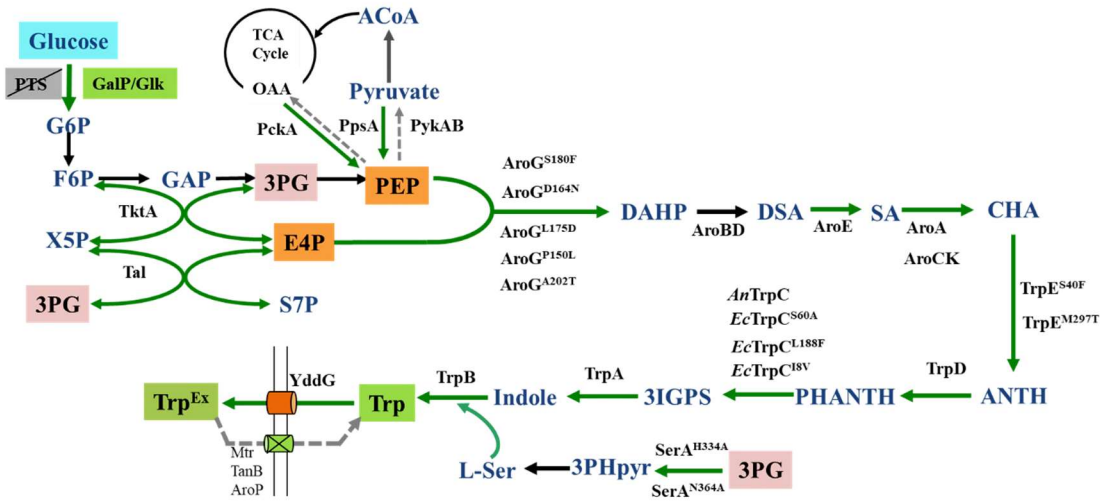

Figure 2.1: Main strategies applied in the metabolic engineering of *E. coli* for efficient biosynthesis of Trp. The corresponding feedback-resistant enzymes are shown beside the arrows. The thick green arrows indicate enhanced activity/flux of the corresponding reaction by gene overexpression and/or protein engineering. The gray dotted arrows show the corresponding enzymes and/or pathways need to be knocked out.

2.1.1 Metabolic engineering of restrictive regulations

Trp biosynthetic pathway in *E. coli* is highly regulated at multiple levels. So far, several types of regulations in the Trp biosynthetic pathway have been identified, including feedback and feed-forward regulations of enzyme activity, transcriptional repression and attenuation of gene expression, and global regulators (Fig. 2.2). In the shikimate pathway, 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase consists of three isoenzymes, e.g., AroG, AroH, and AroF, each of which is feedback-inhibited by L-tyrosine, L-phenylalanine, and L-tryptophan, respectively (Umbarger, 1978). Among them, feedback-resistant variants of AroG ($AroG^{fbr}$) have been frequently developed and applied for Trp synthesis in *E. coli* (Fig. 2.1), e.g., $AroG^{D146N}$ (Kikuchi et al., 1997), $AroG^{S180F}$ (Ger et al., 1994), $AroG^{L175D}$ (Hu et al., 2003), $AroG^{P150L}$ (Hu et al., 2003), and $AroG^{A202T/D146N}$ (Ding et al., 2014). Besides, the shikimate dehydrogenase (AroE) is inhibited by shikimate. To date, no feedback-resistant AroE enzyme has been reported, so a supplementary expression of AroE has been performed to enhance the catalytic activity of AroE (Díaz-Quiroz et al., 2018; Ghosh and Banerjee, 2015). In the Trp branch pathway, the anthranilate synthase TrpE suffers from strong feedback inhibition by the end-product Trp. Two feedback-resistant variants $TrpE^{S40F}$ and $TrpE^{M293T}$, have been developed for Trp synthesis (Fig. 2.1) (Caligiuri and Bauerle, 1991). Moreover, feed-forward regulation was discovered as a novel regulation in Trp

biosynthesis. Specifically, the indole glycerol phosphate synthase (IGPs) domain of the bifunctional *E. coli* TrpC enzyme is feed-forward inhibited by anthranilate (Fig. 2.2). Strikingly, the TrpC IGPs domain from *Saccharomyces cerevisiae* and *Aspergillus niger* is feed-forward activated by anthranilate (Chen et al., 2018). As expected, the expression of the anthranilate-activated *An*TrpC from *A. niger* in a previously engineered Trp-producing strain resulted in a more robust strain with efficient Trp production and improved Trp yield in the bioreactor.

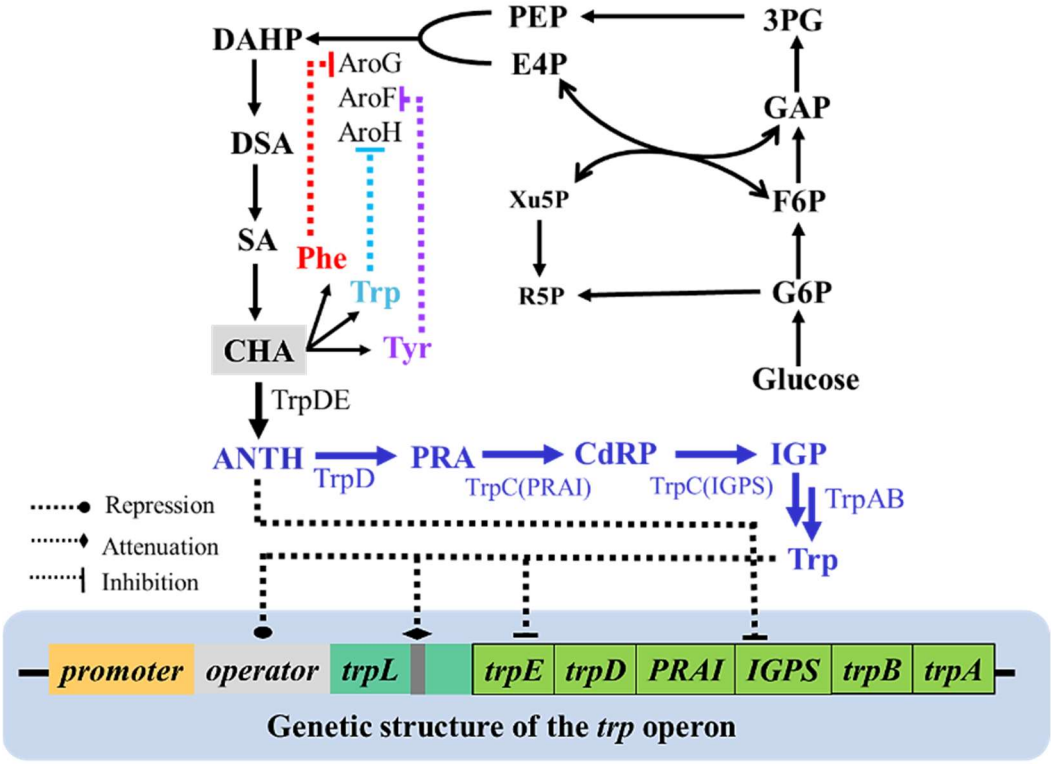

Figure 2.2: Selected parts of regulations in the Trp biosynthetic pathway.

Besides, the TrpR repressor facilitates the primary transcriptional repression and regulates the expression of genes involved in Trp pathway such as the *aroH* gene, the *aroL* gene, and the *trpEDCBA* operon (Fig. 2.2) (Klig et al., 1988; Lawley and Pittard, 1994). Deletion or mutation of TrpR is accordingly performed to eliminate the transcriptional repression. A Trp operon leader, *trpL*, also regulates transcription of the *trpEDCBA* operon by sensing the intracellular Trp concentration. In this regard, *trpL* leader was knocked out in a genetically modified *E. coli* strain for the production of Trp derivatives (Henkin and Yanofsky, 2002). Except for these two types of regulations, early publications found that several global transcriptional regulators determine the expression of genes involved in central metabolism. A global transcriptional regulatory network plays an extraordinary role by enabling an organism to modulate the

expression of numerous genes in response to environmental and genetic perturbations. For instance, the carbon storage regulator (CsrA) regulates about 25 genes in central metabolism (Sabnis et al., 1995). Specifically, this regulator negatively modulates the expression of PEP carboxykinase encoded by *pckA* and PEP synthetase encoded by *ppsA*, and positively controls the expression of pyruvate kinases encoded by *pykF* (Tatarko and Romeo, 2001). Therefore, deletion of the *csrA* gene could lead to accumulation of the precursor, PEP (Tatarko and Romeo, 2001; Yakandawala et al., 2008). Moreover, the fructose repressor FruR, a DNA-binding transcriptional dual regulator (catabolite repressor and activator) (Plumbridge, 1995), forces the direction of carbon flow by activating transcription of genes encoding enzymes such as pyruvate kinases. Inactivation of the global regulator FruR leads to PEP accumulation and improves Trp production (Liu et al., 2016). Engineering of global transcriptional regulators Fis (Nilsson et al., 1990; Ross et al., 1990), leucine-responsive protein (Lrp) (Calvo and Matthews, 1994), cyclic AMP (cAMP) receptor protein-cAMP complex (CRP-cAMP) (Man et al., 1997), and copper-responsive regulatory system (CpxA/CpxR) (Yamamoto and Ishihama, 2006) have been reported and are considered as a promising way to improve the metabolic flux in Trp pathway. Considering the mechanism of restrictive regulations in the Trp biosynthetic pathway, combinatorial modifications of these regulations are occasionally performed.

2.1.2 Metabolic engineering to enhance precursor supply

The biosynthetic pathway contributing to Trp biosynthesis is omnipresent in all microorganisms. In principle, enhancing the supply of precursors leads to the common synthesis pathway and the Trp branching pathway. The accumulation of PEP and E4P precursors has been proposed as a restriction for the production of Trp (Li et al., 2020). An early attempt was performed to enhance PEP precursor by disrupting the competing pathways (Fig. 2.1), like PEP carboxylase encoded by the *ppc* gene and pyruvate kinases encoded by the *pykAB* genes (Ikeda, 2006b), or phosphoglucose isomerase encoded by the *pgi* gene (Mascarenhas et al., 1991). Modulations of the most extensive PEP consumption system, PEP:carbohydrate phosphotransferase (PTS) system, were also exploited to improve the availability of PEP (Fig. 2.1) (Postma et al., 1993). In wild-type *E. coli*, almost 50% of PEP synthesized in glycolysis is simultaneously converted into pyruvate (Flores et al., 2002), and merely 3% of PEP entered into the biosynthetic pathway of aromatic compounds (Valle et al., 1996). In theory, switching

off the PTS system in *E. coli* should convert up to 50% of the available PEP into the aromatic compounds pathway (Báez et al., 2001; Floras et al., 1996). Due to the physiological effects suffered from inactivation of PTS system such as defection of cell growth, activation of a galactose permease (GalP) and glucokinase (Glk)-dependent glucose utilization system (Lu et al., 2012), or heterologous expression of a glucose facilitator (encoded by the *glf* gene) and glucokinase from *Zymomonas mobilis* (Chandran et al., 2003) is suspected to be an alternative strategy for enhancing the glucose utilization. In this regard, a PTS-negative strain was constructed, resulting in an improvement in formation of L-phenylalanine (Gosset et al., 1996). Pathway engineering also succeeds in Trp production by increasing supply of E4P precursor, which is realized by overexpressing *tktA* gene-encoded transketolase or *tal* gene-encoded transaldolase in the pentose phosphate pathway (Fig. 2.1) (Rüffer et al., 2004; Wendisch, 2007).

Besides, the supply of intracellular L-serine and phosphoribosyl 5-pyrophosphate (PRPP) likeward represents one bottleneck in Trp synthesis (Fig. 2.1) (Ikeda, 2006b). In this regard, additional genetic modification of the genes involved in the serine pathway (Ikeda et al., 1994) and the gene *prs* encoding phosphoribosyl 5-pyrophosphate synthetase (Yajima et al., 1990) can make contributions to the increase in Trp production (Fig. 2.1). Chorismate, the direct precursor to the Trp-branched pathway, could be another bottleneck. Undoubtedly, in *E. coli*, the shikimate-sensitive shikimate dehydrogenase (encoded by the *aroE* gene) can lead to a reduced supply of chorismate (Dell and Frost, 1993). Indeed, an enormous amount of intracellular shikimate was observed in our previous strains (Chen and Zeng, 2017; Chen et al., 2018). The conversion of shikimate to chorismate is carried out by overexpression of the genes *aroCK* and *aroL* (Fig. 2.1) (Zhang et al., 2015).

2.1.3 Metabolic engineering to improve microbial tolerance

During microbial production of solvent-related compounds, the accumulation of toxic metabolites can negatively affect the cell homeostasis and limit metabolite production. Mechanisms of solvent toxicity have been explored over the last decades (Mukhopadhyay, 2015; Qi et al., 2019). For instance, the accumulation of by-products, such as acetic acid and alcohol, could disturb the integrity of cell membrane and lead to cell leakage (Heipieper et al., 1994). Consequently, improving the solvent tolerance

of the microbial production strain through metabolic engineering has turned into an essential aspect of the strategy.

Fundamentally, one of the methods to avoid cell membrane damage and to enhance membrane integrity is alleviation of by-products formation, such as acetic acid, alcohol, and lactic acid, omnipresent in *E. coli* during prolonged fermentation or under stress conditions. In recent years, global transcription machinery engineering (gTME) has been used to engineer transcriptional factors in prokaryotic and eukaryotic organisms to enhance solvent tolerance (Guo et al., 2018c). Several transcriptional factors, such as TATA-box binding protein (TBP)-associated factors (Spt15), RNA polymerase sigma subunit (σ^{70}), cAMP receptor protein (CRP), and histone-like nucleoid structuring protein (H-NS) have been modified by gTME to improve strain tolerance (Alper et al., 2006; Hong et al., 2010; Xue et al., 2019; Zhang et al., 2012). For example, random mutagenesis was introduced into the *rpoD* gene, which encodes the primary sigma factor σ^{70}, to obtain *E. coli* σ^{70} mutants with enhanced exogenous ethanol tolerance (Guo et al., 2018a).

Although the development of specific targets such as membrane-modifying enzymes, redox enzymes, and transcriptional factors (Mukhopadhyay, 2015) has proven to be a competent strategy to improve solvent tolerance, efflux pumps are also useful in improving microbial tolerance. Efflux pumps provide a direct channel for draining the intracellular solvents to the extracellular space and potentially force a pull on the biosynthetic pathway. For instance, one of the solvent-tolerance transporters, hydrophobic/amphiphile efflux family of resistance-nodulation-division (RND) pumps, has been encountered in *E. coli* and used for the optimization of microbial tolerance (Nikaido and Takatsuka, 2009; Ramos et al., 2002). In particular, the AcrAB-TolC efflux pump, which belongs to the RND pump family, has been extensively investigated and identified as a compelling candidate for an increase of the tolerance toward several monoterpene hydrocarbons (Du et al., 2014). For instance, Doukyu and Iida found that an AcrAB-TolC efflux pump is required for the microbial tolerance and maximal production of the aromatic compounds such as styrene oxide (Doukyu and Iida, 2020) in *E. coli*. It was also identified that the native *E. coli* aromatic acid exporters AaeAB export aromatic compound *p*-hydroxybenzoic acid (pHBA) (Van Dyk et al., 2004). Moreover, the efflux pump YddG from *E. coli* is an inner membrane protein that exports aromatic compounds (Fig. 2.1). It is essential to eliminate the growth inhibition

caused by its excessive cytosolic accumulation and thereby contribute to cellular homeostasis (Doroshenko et al., 2007).

2.2 Optimization of biomolecules by laboratory-directed evolution

Optimization of the biosynthetic pathway for overproduction of valuable aromatic compounds relies upon techniques of directed evolution. Laboratory-directed evolution is a general term to describe the various techniques used to generate enzyme variants and to screen desired functions (Yuan et al., 2005). Usually, directed evolution can be carried out either *in vitro* or *in vivo*. *In vitro* mutagenesis can be achieved by error-prone PCR, DNA shuffling, or site-directed saturation mutagenesis (Sen et al., 2007), whereas *in vivo* directed evolution is relied on intracellular modification of target genes, i.e., modified natural mutators (Bridges and Woodgate, 1985), plasmid-facilitated mutagenesis (Badran and Liu, 2015b), and target-specific mutagenesis (Jakočiūnas et al., 2018).

2.2.1 Laboratory-directed *in vitro* protein evolution

One of the primary goals of protein engineering is to confer a protein with desired activities and functions. In the past few decades, chemical modifications and nucleobase analogs have been the most commonly used mutagenesis approaches (Lai et al., 2004; Pfeifer et al., 2005). These methods have generated many variants with desired properties, but these methods only generate a narrow sequence of preference with a low mutational potency. *In vitro* directed evolution has emerged as an impressive technology in the development of biomolecules (Yuan et al., 2005). Directed evolution, different from the chemical modification and ultraviolet irradiation, is able to obtain the fittest variant under the controlled evolutionary pressure. Here the conventional techniques and strategies of DNA mutagenesis and recombination (Sen et al., 2007) and structure-based enzyme redesign (Lutz, 2010) for *in vitro* directed evolution of enzymes are briefly summarized in Table 2.1.

Traditionally, *in vitro* directed evolution depends on a two-step protocol: (1) generation of gene variant libraries by random mutagenesis, and (2) high-throughput screening and selection of desired candidates (Fig. 2.3). However, screening such immense libraries is a time-consuming process. Besides undesirable candidates are hard to be excluded even with advanced screening approaches. Another strategy, rational protein design is used at the molecular level to create a new or activity-enhanced protein (Sen et al.,

2007). Rational protein design normally requires the availability of both the structure of the enzyme and the knowledge of the relationship between its structure and function (Korendovych, 2018) (Fig. 2.3). Recently, as more information about structure and function in protein becomes available, combinatorial protein engineering through directed evolution and rational design (semi-rational design) has been widely recognized (Korendovych, 2018; Lutz, 2010) (Fig. 2.3). A semi-rational design approach can generate a small, high-quality library through narrowing the diversity of amino acids, which leads to libraries with more excellent functional properties (Amrein et al., 2017; Chen et al., 2009). For instance, a semi-rational strategy has been adopted to alter the Phe binding site in the AroG enzyme to alleviate its inhibition, by taking advantage of the crystal structure of AroG complexed with its inhibitor Phe (PDB: 1KFL) (Ding et al., 2014).

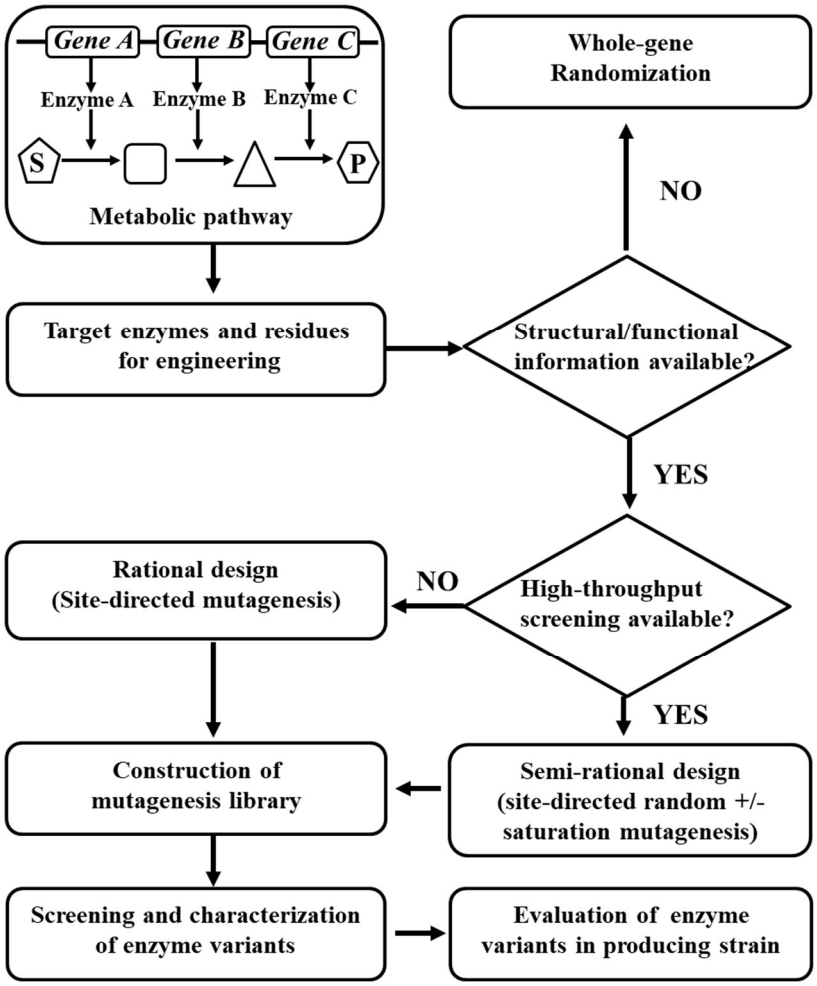

Figure 2.3: Schematic overview for choices of random mutagenesis, rational design, or semi-rational design for protein engineering. Application of the preferred approaches for the development of biomolecules based on the prior knowledge of its structure and function and the availability of screening techniques (Adapted from Fig. 1 in Chica et al., (2005)).

Table 2.1 Summary of methodologies for *in vitro*-directed evolution to facilitate protein engineering

Methodology	Method summary	Example	Pros	Cons	Ref.
Error-prone PCR	A method for introducing random mutations by reducing the fidelity of DNA polymerase	*β*-Lactamase; Lipase;	It is easily applied in nearly any laboratory by modifying PCR conditions	It makes the generation of a library accompanied by a codon bias	(Fujii et al., 2004; Pritchard et al., 2005)
Staggered extension process	A method for generating a library of chimeric sequences combined with error-prone PCR	*cry9Ca1* gene	It enables the generation of retroviral populations by template-switching recombination	It requires incredibly abbreviated annealing and DNA elongation conditions	(Vanhercke et al., 2005; Zhao et al., 1998)
Site saturation mutagenesis	A method for introducing all amino acids' codons at a target position	Xylanase; Glutamate decarboxylase	It is easily applied in introducing selected mutations to target sequences in a precise, site-specific manner	It is inflexible regarding the categories of generated mutants	(Fan et al., 2018; Wang et al., 2013)
DNA shuffling	A method for introducing *in vitro* recombination of homologous and randomly fragmented genes	Thymidine kinase; Biphenyl dioxygenase	It enables the removal of neutral mutations by backcrossing with parental DNA	It is limited in the efficiency of recombination that results in a reduction of diversity	(Christians et al., 1999; Kumamaru et al., 1998)
Whole-genome shuffling	A method for developing an organism at the whole-genome level by DNA shuffling	*Streptomyces fradiae*; *lactobacilli. Lactobacillus*	It enables the acceleration of evolution by recursive recombination of multiple parents	It is non-available in difficult-editing organisms	(Patnaik et al., 2002; Zhang et al., 2002)
Exon shuffling	A method for introducing *in vitro* recombination of non-homologous genes to generate new genes	Haptoglobin; Hemostatic proteases	It enables the generation of new combinations of exons by intronic recombination	It involves genetically compatible	(Kolkman and Stemmer, 2001; Patthy, 1999)

Table to be continued on next page

Table 2.1 Summary of methodologies for *in vitro*-directed evolution to facilitate protein engineering

Methodology	Method summary	Example	Pros	Cons	Ref.
Heteroduplex recombination	A method for introducing non-homologous recombination and chimeragnesis by *in vitro* heteroduplex formation and *in vivo* repair	Truncated green fluorescence protein (GFP)	It neither suffers the limitations of PCR-based approaches nor requires transformation with multiple gene fragments	It could be only useful for recombining abundant genes or the entire operon	(Maresca et al., 2010; Volkov et al., 1999)
Degenerate Oligonucleotide Gene Shuffling (DOGS)	A method for generating gene shuffling using degenerate primers that reduces the regeneration of unshuffled parental gene	Beta-xylanase	It avoids the use of endonucleases for gene fragmentation and allows the use of random mutagenesis of selected segments	It requires the design of perfectly complementary pairs of primers	(Bergquist et al., 2005; Gibbs et al., 2001)
Random drift mutagenesis	A method for enabling the screening of mutants from libraries where no adaptive selection has been imposed on the cells	*β*-Glucosidase	It combines with DOGS for a broader exploration of the sequence space of shuffled genes	It requires a specific colorimetric or fluorescence indicator to high-throughput screening of mutants	(Bergquist et al., 2005; Hardiman et al., 2010)
Structure-based enzyme redesign	A method based on the interrelated information among protein sequence, structure and function to pre-select promising target sites for enzyme redesign	Monooxygenase ω-Transaminase	It facilitates protein redesign to locate key residues near active sites effectively and at domain interfaces or hinge regions	Rare-known 3D structure pools provide less useful information for the rare-studied protein redesign	(Savile et al., 2010; Wu et al., 2010)

The fact is that semi-rational design is restricted with a small, high-quality library and advanced high-throughput screening methods. Only a fraction of target variants can be generated, and the information for unimproved variants is wiped out during directed evolution. However, protein engineering through machine-learning-guided directed evolution enables further optimization of protein functions by exploiting more information. This approach also expands the number of properties for selection with a higher fitness level (Yang et al., 2019). Theoretically, machine-learning approaches predict how sequences or structures can be mapped to their functions in a data-driven manner without the need for detailed information on metabolic pathways. By learning the properties of characterized enzyme variants, such an approach could speed up the process of directed evolution, and the selected target variants are expected to exhibit improved properties (Wu et al., 2019). The efficiency of this method has been demonstrated in development of human GB1 binding protein, and the results showed that machine-learning-guided directed evolution could generate variants with higher fitness. Together, these strategies offer promising tools and predictors for altering protein functions such as substrate specificity, stereo-selectivity, and stability through enzyme redesign.

2.2.2 Laboratory-directed *in vivo* protein evolution

In vitro directed evolution allows the control of mutagenesis rate and mutational spectrum, whereas *in vivo* directed evolution enables the simultaneous performance of mutation and selection under intracellular conditions (d'Oelsnitz and Ellington, 2018). Theoretically, an ideal *in vivo* mutagenesis that enables efficient evolution is supported by two significant features: (1) development of endogenous mutagenesis and (2) fitness-coupled selection platform and screening threshold (Badran and Liu, 2015a). This part highlights the latest developments in endogenous mutagenesis systems (Table 2.2), followed by comparisons of these technologies, including the modified natural mutators, plasmid-facilitated mutagenesis, and target-specific mutagenesis.

2.2.2.1 Method of modified natural mutators

The method of modified natural mutators is based on modifications of naturally existing mutation mechanisms (Table 2.2). Organism controls the fidelity of DNA amplification through a series of overlapping proofreading, mismatch repair, and base selection. These redundant DNA replications and precise error-prone repair systems have jointly

produced a base substitution rate of chromosomal DNA of ~ 10^{-9} to 10^{-10} substitutions per base pair per generation (s.p.bp.p.g) (Schaaper, 1993). Therefore, modification of the DNA replication system or inactivation of the DNA repair system can offer a broad mutation spectrum and a high mutagenesis rate. To this end, the DNA polymerase dominant-negative proofreading subunit *dnaQ926* and the error-prone repair DNA polymerase Pol V are expressed in target platforms to suppress proofreading and enhance error-prone lesion bypass, and the mutagenesis rate increases approximately by 100-fold (Carlson et al., 2014). This mutation rate, however, is still several orders of magnitude below the mutation rates achieved by the conventional *in vitro* mutagenesis technologies.

More aspects of the error-prone repair system are considered to be modified to strengthen the mutation rate. In theory, mismatch repair system reduces the mutation rate of bacterial DNA replication by around 10^3 times (Schaaper, 1993), and it is not wondering that the interruption of MutSL (specialized proteins monitor) and MutH (excise mismatched nucleotides following DNA replication) and also dynamic expression of Dam (DNA methylation) through a SeqA domain (hemimethylated GATC-binding domain) leads to an average of 4.4×10^{-7} s.p.bp.p.g. Heterologous expression of the cytidine deaminase CDA1 from *Petromyzon marinu* and impairment of the uracil-DNA glycosylase (Ung) synergize by introducing the inhibitor of Ung, Ugi, are beneficial for the mutation rate (2.0×10^{-6} s.p.bp.p.g). Besides that, impairment of the export of mutagenic nuclease is an alternative strategy to advance the mutator effect. Two significant determinants of base preference during DNA replication are the catalytic alpha subunit of DNA Pol III and the concentration of intracellular dNTP. It generally assumes perturbations affecting intracellular dNTP to affect the mutational spectrum (Badran and Liu, 2015a). For instance, Badran et al., found that the activation of transcriptional repressor EmrR leads to the down-regulation of *emrAB*, which is acknowledged as the putative export of mutagenic nucleobase intermediates (Badran and Liu, 2015a; Yang et al., 2004).

Table 2.2 *In vivo* genotype diversification strategies for *in vivo* directed evolution

Method	Targetability	Example	Description	Ref.
Modification of natural mutators	NO	*dnaQ926*/*mutSL*	Modification of naturally existing mutation mechanisms, e.g., proofreading subunit *dnaQ926* and the error-prone repair DNA polymerase pol V, to increase the mutation rate	(Badran and Liu, 2015a)
PACE	NO	Split T7 RNA polymerase	A plasmid-facilitated mutagenesis method links the activity of target enzymes to the conditional phage replication	(Packer et al., 2017)
OrthoRep	NO	Dihydrofolate reductases	Targeted gene in the TP plasmid is targeted and mutated by an orthogonal DNA polymerase pair extranuclear replication system	(Ravikumar et al., 2018)
MutaT7	YES	Kanamycin resistance gene (Kan^R)	It enables continuous, direct mutation within specific, well-defined DNA regions of any relevant length by utilization of a DNA-damaging cytidine deaminase that fused to a processive RNA polymerase	(Moore et al., 2018)
TaGTEAM	YES	Glycosylases	It enables the generation of point mutations occurs through error-prone homologous recombination (HR) and depend on re-section and the error-prone polymerase Pol ζ	(Finney-Manchester and Maheshri, 2013)
EvolvR	YES	Ribosomal protein subunit 5 gene (*rpsE*)	It is achieved by directly generating mutations using engineered DNA polymerases DNAP *Pol*3 M targeted to loci via CRISPR-guided nickases	(Halperin et al., 2018)
CRISPR/Cas9-mediated base editing	YES	Mammalian cells (HEK293)	It enables direct, irreversible conversion of one target DNA base into another using CRISPR/Cas9 and a cytidine deaminase enzyme, without requiring a donor template	(Kleinstiver et al., 2019)
CasPER	YES	Isoprenoid production	It enables CRISPR/Cas9-mediated directed evolution of large sequences in genomic contexts	(Jakočiūnas et al., 2018)
CREATE	YES	AcrB efflux pump	CRISPR-enabled trackable genome engineering method links each guide RNA to homologous repair cassettes to track genotype-phenotype relationships	(Garst et al., 2017)

2.2.2.2 Plasmid-facilitated mutagenesis

The plasmid-facilitated mutagenesis is introduced to confine the mutagenesis within the targeted plasmid. Until now, most conventional directed evolution methods have primarily linked the activity of target biomolecules to the growth of auxotrophic cells. More recent developments, however, are relied on conditional phage replication, i.e., phage-assisted continuous evolution (PACE, Fig. 2.4A) (Packer et al., 2017; Song and Zeng, 2017). In theory, target enzyme variants are fused to a phage coat protein and then displayed on phage particles to establish a connection between genotype and phenotype. This system employs a mutagenesis plasmid (MP) that enables the mutation rate at ~2.3 substitutions per kbp per generation of *E. coli* XL1-blue cells. During the continuous evolution, the activity of target enzyme variants on the selection phage (SP; contains the gene of interest and all the phage genes without *gene III*) is linked to the accessory plasmid (AP; contains an essential phage gene (*gIII*) expression circuit) (Fig. 2.4A). In this approach, the enzyme variants with the desired activity activate the expression of *gene III* and accordingly increase phage propagation. Only the enzyme variants with desired properties can be enriched after dozens of reinfection rounds (Brödel et al., 2018; Packer et al., 2017). So far, some derivative approaches from the PACE method, such as split T7 RNA polymerase variant-based PACE (activity-dependent activation of the expression of *gene III*) (Pu et al., 2017), transcription activator-based PACE (promoter-dependent activation of the expression of *gene VI*) (Brödel et al., 2016), DNA-binding PACE (evolution of transcription activator-like effector nucleases) (Hubbard et al., 2015), protease-PACE (evolution of protease against desirable cleavage sites) (Dickinson et al., 2014; Stano and Patel, 2004), and protein-binding PACE (evolution of protein-protein interactions) (Badran et al., 2016) have been developed and applied to evolve a broad category of biomolecules. The advanced approaches lead to new biomolecules with tailor-made properties and a defined function that shed light on basic evolutionary processes.

Existing approaches, such as the PACE-assisted continuous evolution approach, require specialized devices to maintain a continuous culture. These approaches are primarily limited by the mutation rate, which is realized by the mutagenesis plasmid. These shortcomings also hinder the extensive parallelization of directed evolution experiments to mine enzyme variants with multiple functions. Another method, known as OrthoRep, has been proposed using a highly error-prone orthogonal DNA polymerase (DNAP1)-DNA plasmid pair that mutates user-defined genes at rates of ~1×10^{-5} s.p.bp.p.g (Ravikumar et al., 2018) (Fig. 2.4B). The OrthoRep approach composes of an orthogonal p1 plasmid (contains target genes) and another

mutagenesis plasmid (contains all essential genes for the expression of orthogonal TP-DNAP1) (Fig. 2.4B). The mutagenesis plasmid with the strict orthogonal TP-DNAP1 autonomous replication process is realized by modifying an ep-DNAP to the target p1 plasmid. Using the OrthoRep, Arjun et al. developed drug-resistant malarial dihydrofolate reductases (DHFRs) in 90 independent replicates. The results showed that the OrthoRep system bypasses the frequent first-step adaptive mutations and paves the path for the selection of rare outcomes with new biomolecular and cellular functions, especially in the investigation of drug resistance and adaptation. To the best of our knowledge, the OrthoRep approach also supports a higher engineerable mutation rate (~100,000-fold higher than the host genome *in vivo*) and is a fully scalable platform without *in vitro* library construction. The OrthoRep could also easily handle more sophisticated *in vivo* mutagenesis and selections, including dominant-negative selections for new orthogonal tRNA/aaRS pairs or sequence-specific DNA binding proteins (Arzumanyan et al., 2018; Tan et al., 2019).

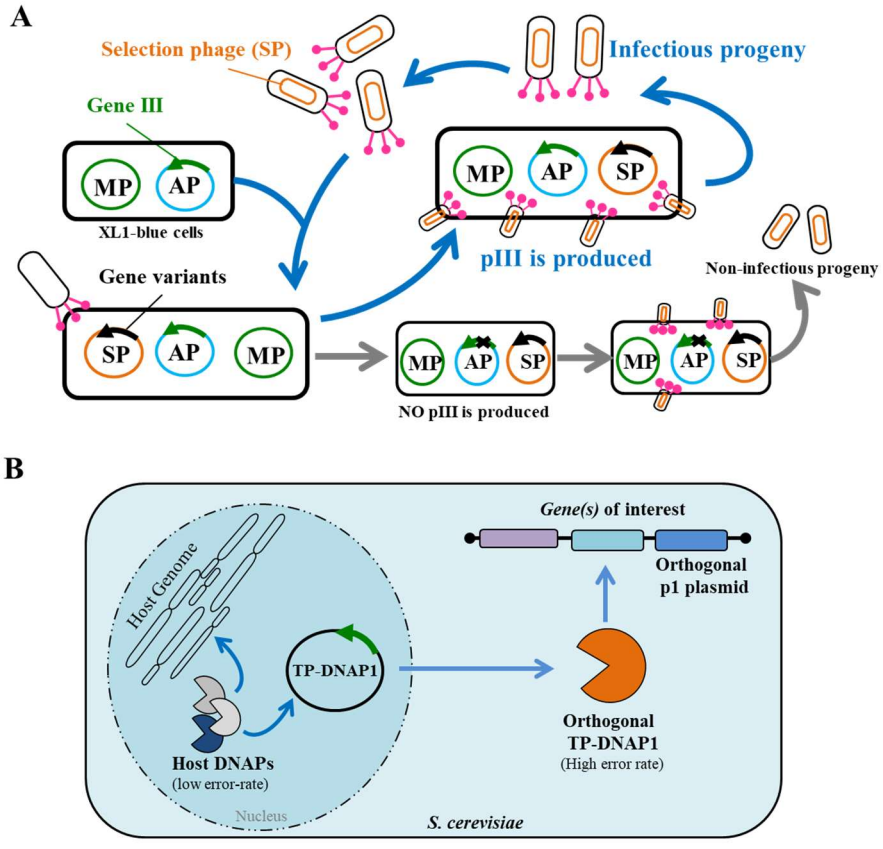

Figure 2.4: Schematic overview of the PACE and OrthoRep approaches. (A) Phage-assisted continuous evolution (PACE). MP represents a mutagenesis plasmid and AP represents an accessory plasmid. (Adapted from Figure 1 in Packer et al., (2017)). (B) OrthoRep: a system for the scalable, continuous evolution of user-defined genes *in vivo*. The basis of OrthoRep is a DNA polymerase (TP-DNAP1) that replicates a cytoplasmic DNA plasmid p1, exemplified for *S. cerevisiae*. (Taken from Figure 1 in (Ravikumar et al., 2018)).

2.2.2.3 Target-specific mutagenesis

Although the efficiency of the PACE and OrthoRep systems for the development of enzyme variants with desired performances has been demonstrated, both approaches are limited to specific host cells because of their unique mechanisms (Fig. 2.4). With the emergence of gene-targeting technologies, more precise technologies for targeting specific locus are being developed. For instance, TaGTEAM, namely targeting glycosylases to embedded arrays for mutagenesis (Table 2.2), is a method based on resection and error-prone polymerase Pol ζ (ep-*Pol* ζ) to create mutagenic ep-homologous recombination for targeted mutagenesis *in vivo* (Finney-Manchester and Maheshri, 2013). By fusing a DNA glycosylase (MAG1) with a DNA binding protein (*tetR*) and locating it at the targeted modified sites, the authors generated a 20-kb region with increased point mutagenesis, and the mutation rate was calculated to be ~10^{-7} per bp per generation. TaGTEAM represents a state-of-the-art method for target-specific mutagenesis in *S. cerevisiae*, where *in vivo* mutagenesis is continuously performed without the demand for rounds of genetic transformation.

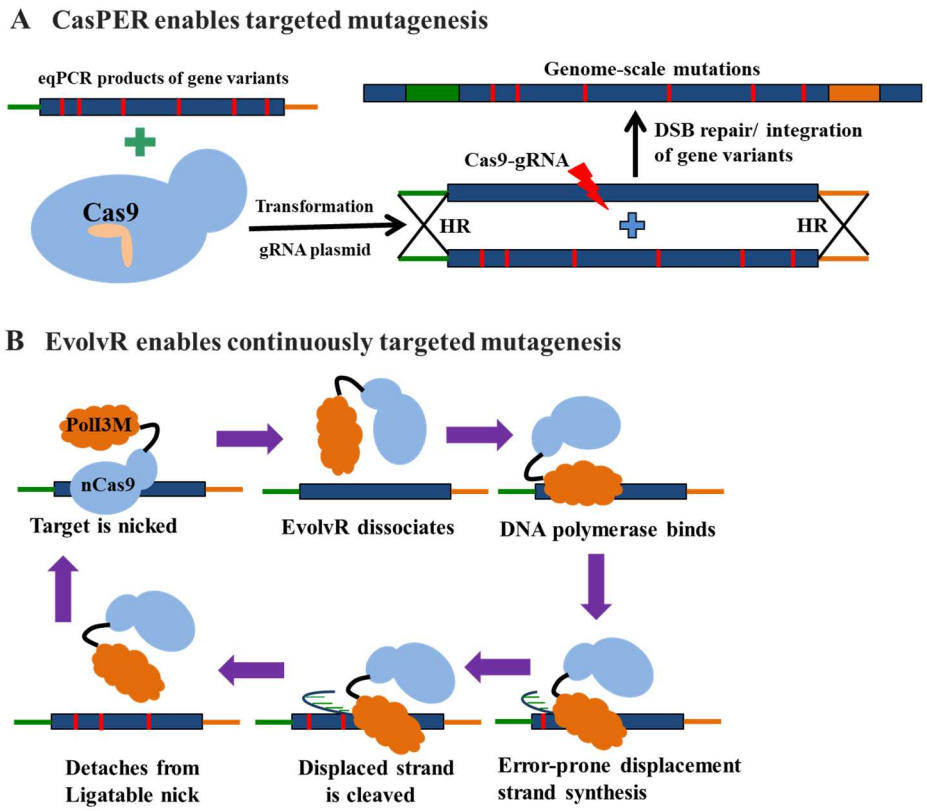

Figure 2.5: Conceptual illustration of the CasPER and EvolvR approaches for targeted mutagenesis. (A) CasPER is based on the generation of mutagenized linear DNA fragments and the expression of gRNA plasmid for genome integration. (Taken from Figure 1 in Jakočiūnas et al. (2018)). (B) EvolvR system consists of a Cas9-guided nickase (nCAs9) that nicks targeted locus and a fused DNA polymerase (PolI3M) that performs error-prone nick translation. (Taken from Figure 1 in Halperin et al. (2018)).

More recently, application of the CRISPR/Cas9 technology has received board approval for genome engineering in a broad category of organisms. With the ability to target and integrate the target DNA fragments through homologous recombination (HR), the technique of CRISPR/Cas9 has been applied to integrate the gene variant library into the chromosome of host strain, such as CasPER (Jakočiūnas et al., 2018). The efficiency of the CasPER approach was demonstrated for directed evolution of two essential enzymes in the mevalonate pathway of *S. cerevisiae*, and the results illustrated that the production of isoprenoids of selected mutants could be increased by up to 11-fold (Fig. 2.5A). The CRISPR/Cas9 technique is also a game-changer for *in vivo* genetic diversification technology such as EvolvR (Halperin et al., 2018) and CRISPR/Cas-mediated base editing (Kleinstiver et al., 2019; Molla and Yang, 2019). EvolvR is a system that can continuously *in vivo* diversify all nucleotides at user-defined loci, which can be achieved by directly generating mutations using manipulated DNA polymerases ep-DNAP *Pol*I targeted to loci via CRISPR-nickase Cas9 protein (Fig. 2.5B). This technology has also demonstrated that the CRISPR-guided DNA polymerases enable multiplexed and continuous diversification of user-defined genes. So far, various types of CRISPR/Cas-mediated base editing are being developed, which can be subdivided into cytosine base editors and adenosine base editors (Kleinstiver et al., 2019; Molla and Yang, 2019), and this approach relies on the recruitment of cytidine deaminases to introduce mutations without double-stranded breaks and donor templates (Hess et al., 2017).

2.3 Advances in biomolecule screening and selection

Most small biomolecules targeted for overproduction cannot be readily detected, and metabolic engineers require advanced high-throughput screening methods to facilitate the directed evolution of microbial strains. In this part, an overview of the development in growth, colorimetric, and biosensor-coupled high-throughput screening techniques is provided (Fig. 2.6).

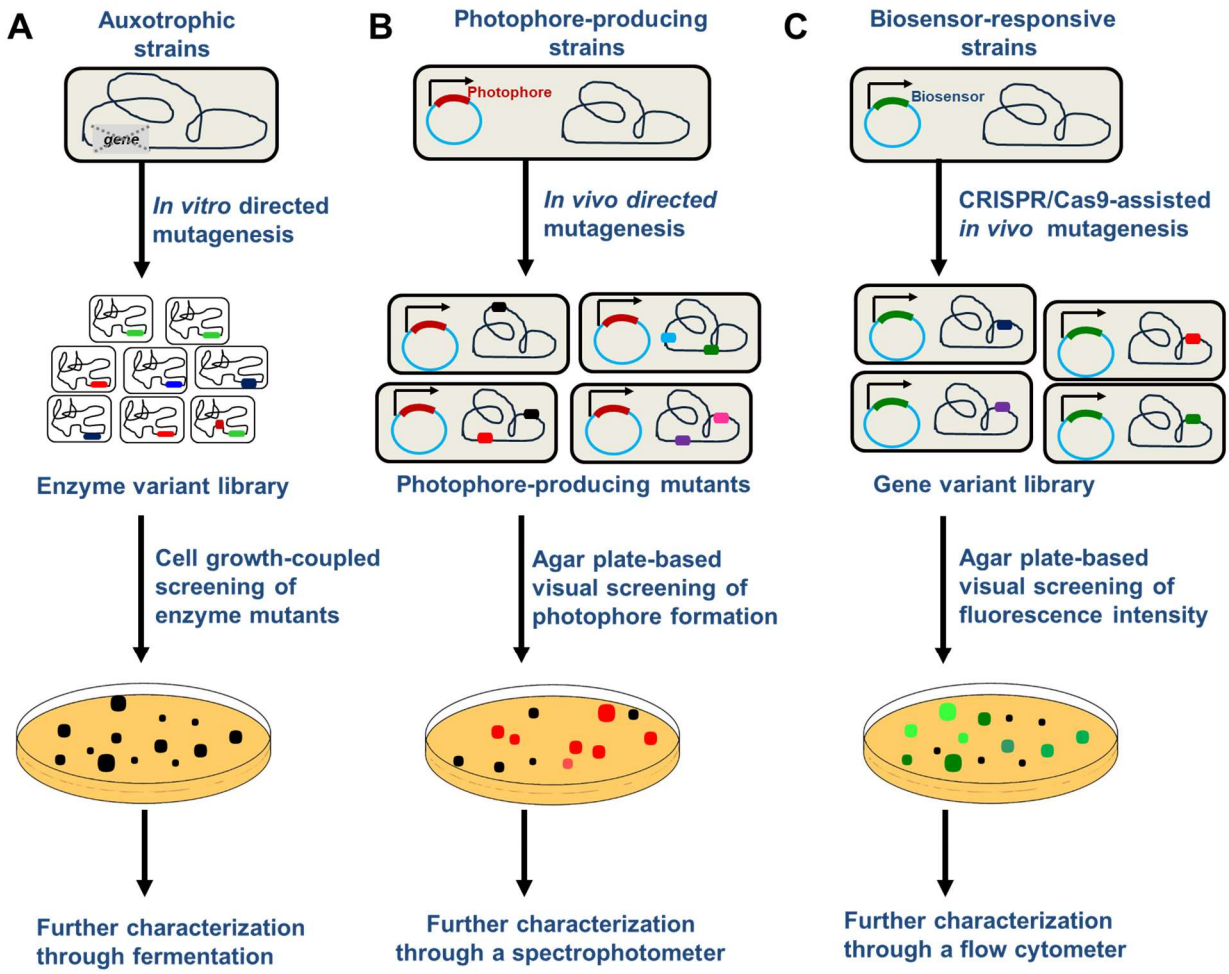

Figure 2.6: Examples of three types of library screening and characterization techniques. (A) Growth complementation-assisted library screening approach. (B) Colorimetric-based high-throughput screening technique. (C) Biosensor-driven library screening and characterization approach.

2.3.1 Growth-coupled high-throughput screening

In bacterial or other unicellular organisms, cells need favorable nutrient conditions to support their growth, such as carbon, nitrogen, phosphate, or other necessary compounds. Therefore, strains that are auxotrophy for certain essential molecules can be efficient selection and screening host cells for directed evolution of target enzymes (Hall, 1981) (Fig. 2.6A).

The auxotrophy for terminal components of the amino acid biosynthetic pathway has been successfully applied for protein evolution because these pathways are strongly coupled to cell growth (Fig. 2.6A) (Dietrich et al., 2010). An early study used aromatic amino acids-deficient cells with disrupted chorismate mutase as a screening strain. Chorismate mutase variants were introduced into the aromatic amino acids-deficient strains and then the cells were subjected to directed evolution. After several rounds of error-prone PCR and DNA shuffling with fine-tuned selection pressure, the catalytic efficiency of chorismate mutase was improved by tenfold over the best variant characterized previously (Neuenschwander et al., 2007). The same strategies

were also applied to develop the key enzymes of amino acid biosynthesis, such as aspartate aminotransferase (Rothman and Kirsch, 2003) or alanine racemase (Ju et al., 2005). The auxotrophy for intermediates in amino acid biosynthesis has been recently utilized for applications in protein evolution. In *E. coli*, 2-ketobutyrate derived from threonine degradation is essential for cell growth, but a more direct pathway to the 2-ketobutyrate is the citramalate pathway. In theory, an isoleucine auxotrophic *E. coli* strain can convert fluxes toward the citramalate biosynthesis pathway for the synthesis of 2-ketobutyrate. In this regard, an isoleucine-deficient *E. coli* strain was constructed to express a heterologous citramalate pathway, and then the cells were evolved to produce 2-ketobutyrate for the support of their growth. The evolved strain exhibited a 9- and 22-fold increase in 1-propanol and 1-butanol production, respectively (Atsumi and Liao, 2008).

2.3.2 Colorimetric-based high-throughput screening

Although cell auxotrophy-facilitated high-throughput screening is highly successful in protein evolution, cell growth represents an indirect indicator for the catalytic activity. Even mediocre performance variants can provide sufficient activity for the cells to restore their growth to wild-type levels, and it is laborious to distinguish the best variants from mediocre ones. The usefulness of this strategy seriously depends upon the end-products or intermediates that are essential for cell growth. Fortunately, the relative cases of conducting colorimetric-based screening approaches have been established (Fig. 2.6B) (Sylvestre et al., 2006).

Normally, photometric assays can be performed either through liquid cultures or on solid media. Photometric assays through liquid cultures are robust and offer a higher sensitivity with a more comprehensive linear detection range by dilution or concentration of samples, whereas screening of colonies on solid media provides an increased throughput (Dietrich et al., 2010). Interestingly, many end products or intermediate metabolites provide natural photophores as a direct indicator for the detection of production titer. These photophores include L-phenylalanine derivatives (phenyl azide, benzophenone, and trifluoromethylphenyldiazirine) (Wilson et al., 1997), *Monascus* yellow hydrophobic pigments (Jůzlová et al., 1996), and the well-known carotenoids (lycopene, *β*-carotene, and astaxanthin) (Di Mascio et al., 1989), etc. To date, lycopene is the primary carotenoid of focus for production in microbial organisms. Using lycopene-derived red chromophore as an indicator, the directed evolution in lycopene-produced *E. coli* strain has been conducted to improve lycopene synthesis (Kim and Keasling, 2001).

Unlike lycopene, the majority of primary and secondary metabolites are not natural chromophores. For these molecules, enzyme-mediated catalysis can offer an alternative

approach for indirect product detection. In this method, target molecules react with the exogenously added regents and the specific enzymes to yield detectable photophores (Dietrich et al., 2010) or traceable cofactors [e.g., ATP and ADP (Koresawa and Okabe, 2004), NAD and NADP (Klingenberg, 1974), and free coenzyme A (Molnos et al., 2003)]. For instance, *S*-adenosyl-L-methionine (SAM) as a universal methyl group donor can be catalyzed by multi-enzymatic biotransformation to homocysteine, and the final product can be quantified with the Ellman's reagent (Hendricks et al., 2004). In the case of carboxylic acid reductases (CARs), which catalyze the reduction of carboxylic acid substrates to the corresponding aldehyde with the consumption of NADPH cofactor, the consumption of NADH or NADPH in CARs-catalyzed reactions is an indirect indicator of substrate oxidation (Moura et al., 2016).

Although the enzymatic catalysis-facilitated indirect product detection is often highly robust and more accurate than direct product detection, there are still some shortcomings in colorimetric-based screening approaches:

1. Limitation of the linear range of detection. The linear detection range of the colorimetric detection approach is typically limited to the milligram level. This detection range may not be available for *de novo* synthesis products that have constantly been detected at the microgram level;
2. Colorimetric-based detection assays are performed directly in the medium with cell removal, production extraction, and sample dilution; these manipulations are at the expense of diminished sensitivity;
3. Enzymatic catalysis-facilitated indirect product detection requires complicated reaction reagents and specific enzymes. Selecting the most promising reaction conditions and enzymes is highly desirable to complete a reaction.

2.3.3 Biosensor-driven high-throughput screening

Conventional work in directed evolution for small-molecule production concentrates on conspicuous components; these molecules can be optically detected by the cell growth-dependent methods or the colorimetric-based methods (Dietrich et al., 2010). In most cases, growth-dependent detections are either non-specific or not optimal due to the required coupling of growth and production, and the linear detection range of the colorimetric-based method is limited to the conspicuous level (e.g., milligram level). These approaches cannot achieve a higher throughput if the target small-molecules are inconspicuous compounds. In contrast to the aforementioned selection approaches, biosensors enable the quantification with a higher

throughput by transmitting the concentration of inconspicuous molecules into an easily detectable output such as fluorescence signals (Fig. 2.6C) (Zeng et al., 2020).

Biosensors typically use a diversity of modes to detect many inconspicuous molecules of interest, such as transcription factors-based biosensors, RNA riboswitches-based biosensors, or coupled enzyme reactions-based biosensors. Besides, these natural biosensors and their constituent domains are supposed to be edited and modularized for the design of novel biosensors (Coulet and Blum, 2019). These novel biosensors, therefore, can monitor the concentration of intracellular metabolites and are widely applied in protein evolution (Williams et al., 2016). Individual variants expressed with metabolites-induced biosensors are initially observed as colonies on the solid agarose medium, and the detection of fluorescence intensity further confirms the variants with better performance (e.g., larger colony size or more chromophore synthesis) through flow cytometry. By coupling biosensor-based screening of enzyme variants, the screening of variants on solid medium ensures increased sensitivity. However, these approaches rarely achieve the throughput (e.g., 10^6 variants per experiment) required for effective screening in an extensive library (Dietrich et al., 2010). To date, this gap has been partially minimized by massive automation, parallelization, and high-throughput of library screening approaches, such as fluorescence-activated cell sorting (FACS) or microfluidic fluorescence-activated droplet sorting (FADS).

FACS is a single-cell analysis method and can analyze the distribution of cellular properties within a broad set of mutation libraries (e.g., 10^9~10^{10} variants per assay) (Naeem et al., 2017; Olsen et al., 2003). The cells are sorted and followed by a secondary screening if the fluorescence intensity of the single cell is over the preset threshold valve. For instance, a FACS platform was applied for high-throughput screening and sorting of the monoamine oxidase mutant library, and the result showed that the FACS approach could enrich the library with functional variants at a higher rate than other methods such as the growth-based method (Sadler et al., 2018). Although FACS is an ultra-high-throughput approach, overlapping profiles and aberrant fluorescence can lead to a high rate of false positives during sorting. This shortcoming of the conventional FACS approach can be remedied by the microfluidic flow sorting system known as microfluidic fluorescence-activated droplet sorting (mFADS) (Baret et al., 2009; Vallejo et al., 2019). To date, the microfluidic flow sorting system has become one of the most flexible and widely used screening platforms for the compartmentalization assays in droplets. An individual variant is compartmentalized in an emulsion droplet, and the variants can be sorted using di-electrophoresis in a fluorescence-activated manner (as in FACS) with the sorting

rates up to 2000 droplets s^{-1}. In this context, the mixtures of *E. coli* strain that express the reported enzyme β-galactosidase variants were compartmentalized with a fluorogenic substrate. The analysis of the sorting variants revealed that the false positive error rate of the sorted variants is much lower than that of the FACS approach (≤ 1 in 10^4 droplets). Such an efficient and effective microfluidic sorting approach will offer an opportunity in the development of biomolecules, where comprehensive libraries are functionally screened.

2.4 CRISPR/Cas9-mediated genomic DNA editing

The technique of CRISPR/Cas9, which stands for Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated protein 9, is one of the genomic editing technologies that represents an efficient tool to delete, insert, or alter the genomic DNA at specific locations (Fig. 2.7A) (Ran et al., 2013; Hsu et al., 2014).

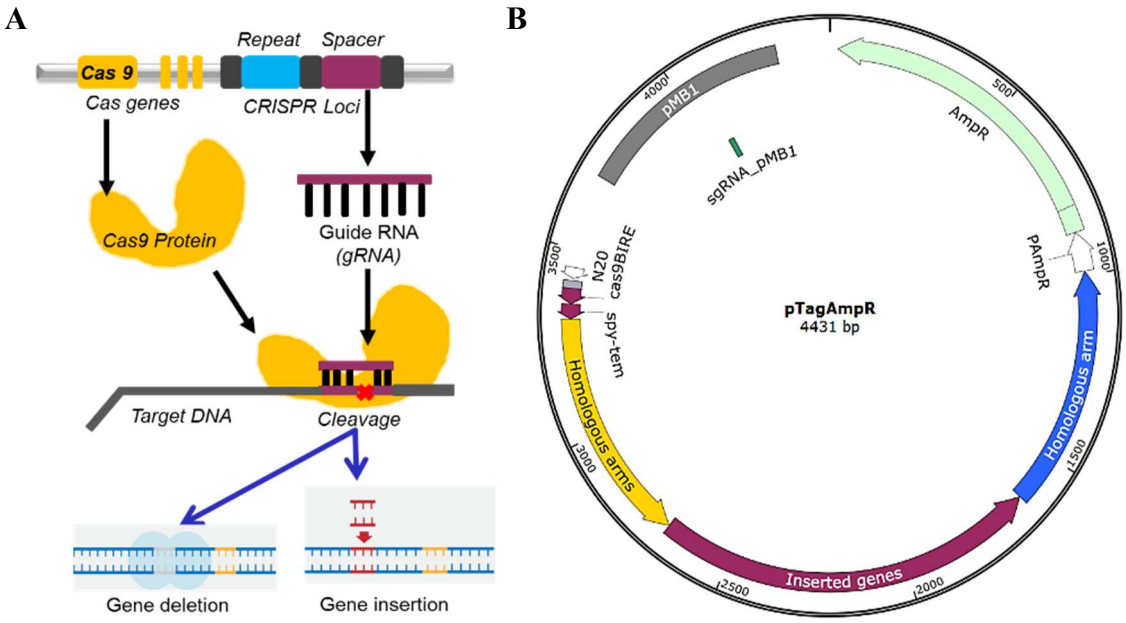

Figure 2.7: (A) CRISPR/Cas9 technique-facilitated genomic engineering (Adapted from Fig. 2 in Hsu et al., (2014)). (B) The gRNA plasmid pTagAmpR.

Usually, a two-plasmid based CRISPR/Cas9 technology is designed to apply a precise genome editing (Jiang et al., 2015). In this two-plasmid system, one is the pCas9 plasmid for expression of Cas9 protein, and the other is the plasmid pTagAmpR for expression of a single guide RNA and a DNA donor fragment (Fig. 2.7B). The Cas9, which belongs to the family of RNA-guided endonucleases, is guided and localized to specific DNA sequences via its gRNA sequence and direct base-pairs with the DNA target (Fig. 2.7A). The binding of Cas9 nuclease and the adjacent protospacer motif (PAM) within target locus helps to control Cas9-mediated double-strand breaks (DSBs). Using the cell's own DSBs repair system, genes are deleted or inserted

by homologous recombination. As shown in Fig. 2.7B, the pTagAmpR plasmid contains a short guide RNA sequence (N20 sequence) and a donor DNA fragment (e.g., homologous recombination arms and the genes to be inserted). Researchers can design short guide RNA sequence (N20 sequence, shown in Fig. 2.7B) with the aid of the web-based tool Cas-Designer (Bae et al., 2014; Park et al., 2015). Specifically, the pCas9 plasmid includes the sgRNA sequence of pMB1, and inoculation of IPTG can induce Cas9 nuclease to cleavage the pTagAmpR plasmid.

In practice, the pCas9 plasmid is first transferred into the host strains by electroporation. To prepare the electrocompetent cells, the cells expressed with pCas9 plasmid are cultivated with L-arabinose to induce Cas9 protein expression. Afterwards, the gRNA plasmid pTagAmpR is electroporated into the corresponding electrocompetent cells. On the following day, the resulting engineered cells are verified by using colony PCR. To cure the plasmid pTagAmpR in the positively engineered cells, the expression of Cas9 protein is induced with IPTG. The resulting engineered cells with pCas9 plasmid can be used for further genome editing, or be cured to remove the pCas9 plasmid by growing them overnight at 37 °C.

2.5 Auto-CGSS-assisted *in vivo* continuous mutagenesis

In vivo continuous mutagenesis approaches for the development of target molecules, such as the PACE and OrthoRep, have been extensively designed and applied. However, these approaches mostly focus on confining the mutations within the target molecules on the plasmid. In theory, the development of aromatic compounds-producing strains requires molecular engineering at the genome-scale, but these approaches are not available for genome engineering. To get rid of this shortcoming, an approach EvolvR was designed (Halperin et al., 2018). EvolvR is a system that can continuously diversify all nucleotides *in vivo* at user-defined loci, and it can be realized by the direct generation of mutations using manipulated DNA polymerases ep-DNAP *Pol*I, which targeted to loci via CRISPR-nickase Cas9 protein. However, especially for aromatic compounds-producing strains, it requires systematic engineering of the pathway for the overproduction of aromatic compounds. Thus, it is attractive to design an *in vivo* continuous mutagenesis approach by which the mutations can be introduced not only at user-defined loci but also the other genome-scale locus, such as the auto-CGSS approach in this thesis (Fig. 2.8).

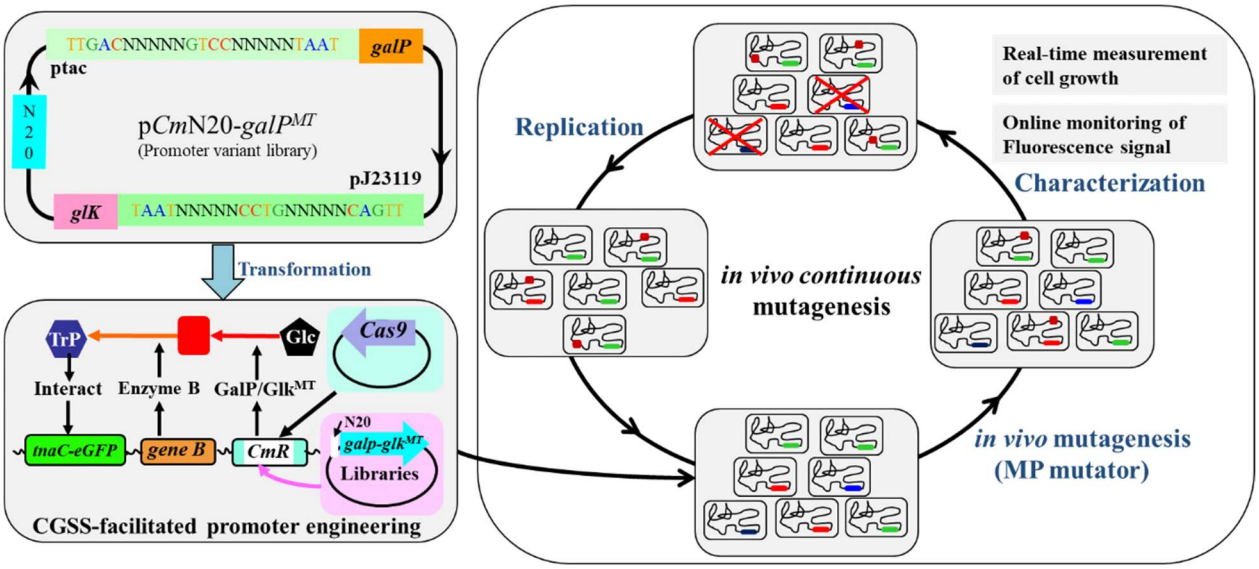

Figure 2.8: CGSS-facilitated *in vivo* continuous mutagenesis coupled with real-time measurement of cell growth and online monitoring of fluorescence signal (auto-CGSS) as a novel approach for strain development.

As shown in Fig. 2.8, as an example for elucidation of the concept, the auto-CGSS approach is applied to develop the GalP/Glk-dependent strain. In this approach, the promoter variant library of genes *galP* and *glk* is integrated into the chromosome of host strain by CRISPR/Cas9 technique. Afterwards, the promoter mutants are continuously evolved *in vivo* in an auxostat vessel by using a mutagenesis plasmid (MP, pJC184). To characterize promoter mutants, the cell growth is measured in real-time with a cell density meter (Ultrospec 10, Biochrom), and the fluorescence intensity is online monitored by a flow cytometer (CytoFLEX, Beckman Coulter). Theoretically, only the GalP/Glk-facilitated glucose utilization system regains its ability, and the corresponding mutants with desired performances will remain dominant during the continuous evolution.

Chapter 3

3. Materials and methods[1]

3.1 Strains, plasmids, and primers

3.1.1 Strains

The *E. coli* strains used in this study are listed in Table 3.1.

Table 3.1: *E. coli* strains used in this study

Strain	Characteristic	Source
DY330	W3110△*lacU169 gal490 λCI1857*△(*cro-bioA*)	(Yu et al., 2000)
S028	DY330::*rpsL*(StrR)△*aroF*△*aroH*△*mtr*△*tnaA*△*tnaB*△λ△*aroH*::$P_{J23119\text{-}rpsL\text{-}tac}$-$aroG^{S180F}$-$serA^{H344A/N364A}$$P_{trc}$-$trpE^{S40F}$*DCBA*	(Chen and Zeng, 2017)
S028*An*TrpC^{WT}	S028△*trpR*::$AntrpC^{WT}$	This study
S028*An*TrpC^{RF}	S028△*trpR*::$AntrpC^{R378F}$	This study
S028GM1 (S028G)	S028△$aroG^{S180F}$:: $aroG^{D6GD7A}$	This study
S028TS	S028△*trpR*:: P_{tac}-*tnaC-eGFP* (Trp sensor)	This study
G028	S028TS△*ptsI*::P_{tac}-*galP*-P_{J23119}-*glk*	This study
B3	S028TS△*ptsI*::P_{tacMT3}-*galP*-$P_{J23119MT3}$-*glk*	This study
G028JC	S028TSΔ*ptsI*Δ*glk*::Cm^R/pCas9^{MT}-pJC184	This study
D8	S028TS△*ptsI*::P_{tacMT1}-*galP*-$P_{J23119MT1}$-glk^{MT}	This study
S028GGMT	S028GM1△*fruR*:: $aroG^{D6GD7A}$	This study
S028AAWT	S028GGMT△*trpR*::$AntrpC^{WT}$	This study
S028AARF	S028GGMT△*trpR*::$AntrpC^{R378F}$	This study
D8AA	D8△*trpR*::$AntrpC^{R378F}$△*fruR*:: $aroG^{D6GD7A}$	This study

[1] Parts of this chapter were reported as the same or with minor changes in M. Chen et al., (2019).

3.1.2 Plasmids

Plasmids used in this study are listed in Table 3.2.

Table 3.2: Plasmids used in this study

Plasmid	Characteristic	Source
pJ23119	AmpR, P_{J23119}, multiple cloning sites	Lab stock
pJLK	AmpR, PMB1, $P_{J13119\text{-}rspL}$-*rspL-Km*	(Chen and Zeng, 2017)
pJLC	AmpR, PMB1, $P_{J23119\text{-}rpsL}$-*rpsL-CmR*	Lab stock
pET22b(+)	pET vector	Novagen
pTrc99A	Vector with *trc* promoter	(Amann et al., 1988)
pIBB24	Derived from pZS*24MCS with new multiple clone site	Lab stock
pJC184	*CmR*, L-arabinose-induced *in vivo* mutagenesis plasmid	(Badran et al., 2015)
pTrpSen	Trp biosensor, TnaC-eGFP	(Fang et al., 2015)
pRed-Cas9-*recA*	AmpR, all-in-one plasmid for the expression of Cas9 protein, RecA recombination protein, sgRNA sequence, and donor DNA fragment	(Zhao et al., 2016b)
pCas9	Expressing Cas9 protein and offering sgRNA-pMB1 for removing pTargetF donor plasmid	(Jiang et al., 2015)
pCas9^{MT}	Expressing Cas9 protein and offering sgRNA-Cm^R for removing pJC184 mutagenesis plasmid	This study
pTargetF	Plasmid for expressing sgRNA or with offering donor DNA, Spectinomycin resistance	(Jiang et al., 2015)
pTagAmpR	Plasmid for expressing sgRNA or with offering donor DNA, Ampicillin resistance	This study
p*Cas9*-*AntrpC*^{WT}	pRed-Cas9-*recA*-*trpR*-sgRNA [a] Δ*trpR*::P_{tac}-*AntrpC*^{WT}	This study
p*Cas9*-*AntrpC*^{R378F}	pRed-Cas9-*recA*-*trpR*-sgRNA Δ*trpR*::P_{tac}-*AntrpC*^{R378F}	This study
p*aroG*N20-*CmR*	pTagAmpR *aroG*-sgRNA Δ*aroG*::$P_{J23119\text{-}rpsL}$-*CmR*	This study
p*Cm*N20-*aroG*^{MT}	pTagAmpR *CmR*-sgRNA Δ*CmR*::$P_{J23119\text{-}rpsL\text{-}CmR}$-*aroG*^{D6XD7X}	This study
p*fruR*N20-*aroG*^{GA}	pTagAmpR *fruR*-sgRNA Δ*fruR*::$P_{J23119\text{-}rpsL\text{-}CmR}$-*aroG*^{D6GD7A}	This study
pN20-*ptsI*	pTagAmpR *ptsI*-sgRNA	This study
pN20-*glk*	pTagAmpR *glk*-sgRNA	This study
p*Cm*N20-*galP*^{MT}	pTagAmpR *CmR*-sgRNA Δ*CmR*::$P_{J23119MT}$ -*galP*- P_{tacMT} -*glk*	This study

Table to be continued on next page

Table 3.2: Plasmids used in this study

Plasmid	Characteristic	Source
pET-*aroG*^{S180F}	pET22b inserted with *aroG*^{S180F} gene	This study
pET-*aroG*^{D6XD7X}	pET22b inserted with *aroG*^{D6XD7X} gene variants	This study
pTrc-*EctrpC*	PTrc99A inserted with *E. coli trpC* gene	This study
pTrc-*AntrpC*^{WT}	PTrc99A inserted with *AntrpC*^{WT} gene	This study
pTrc-*AntrpC*^{RF}	PTrc99A inserted with *AntrpC*^{R378F} gene	This study

[a] *trpR*-sgRNA, sgRNA with an N20 sequence for targeting the *trpR* locus. The other sgRNAs were presented in the same way.

3.1.3 Primers

Primers used in this study are listed in Table 3.3.

Table 3.3: Primers used in this study

Primer	Sequence
lam-out-F	gaggtaccaggcgcggtttgatc
lam-out-R	gttgccgatgtgcgcgtactg
TrpSen-check-F	catccggctcgtataatgtgtgg
TrpSen-check-R	catggcaaacttgaagaagtcg
pTAC-C1	gccatccagtttactttgcagg
pTagC1	ttgagtgagctgataccgctcgc
pTargR	actagtattatacctaggactgagctagctgtcaag
EcTrpC-1F	cacgcattgccgccatttataaacattacgcttcggcaatttcggtgctgactgatgagaaatat
EcTrpC-1R	atggcggcaatgcgtgctggatcgaaatcatcacggatcacgccSNNSNNSNNSNNcgctttcttgcactcca
AnTrpC-FF	ttaaagaggagaaaggtaccatggcggactccggactcgtcg
AnTrpC-FR	tcgagggggggcccggtaccctagagtcccctcacggcctgcac
AnTrpC-T1R	cagactttggcgaacggcacgcaagtcg
AnTrpC-T1F	gagggcttgacgaacaggcccgccattctgNNKaaggaattcgttttcgacgaatatcag
AnTrpC-seq	accatcggagagccgctgttg
XbaI-6H-AnTrpC	ggatcctctagatgataattcagaaggagatatacatatgcaccaccaccaccaccacgcggactccggactcgtcgatc

Table to be continued on next page

Table 3.3: Primers used in this study

Primer	Sequence
AnTrpC-6H-SalI	atgatcgtcgacctactagtggtggtggtggtggagtcccctcacggcctgcac
N20-TrpR	gctcagtcctaggtataatactagtTCAGGTTTAACAACGGTAAAgttttagagctagaaatagcaag ttaaaataaggctagtccg
trpR-out-FF	caaacgctacaccagcggtaaggag
trpR-out-FR	tgcggatcagtaacgacgtccccattccgtggcgctc
trpR-out-RF	ctgcgccagtggctggaagaggtgttgctgaaaagcg
trpR-out-RR	gtggcgattgaagctggcatcgatg
dtrpR-AntrpC-F	ccgacgttgatgagcgccacggaatggggacgtcgttactgatccgcacgggcaattccgacgtctgtgtggaattg
dtrpR-AntrpC-R	cagcaacacctcttccagccactggcgcagctcgacgggcgcggctttcacaccgtaccctaggtctagggcggcg gatttg
N20-trpR-RR	cttaccgctggtgtagcgtttgaaaaaagcaccgactcggtgccacttttc
N20-trpR-FF	catcgatgccagcttcaatcgccacaagacgaaagggcctcgtgatacg
trpR-check-F	gattaccagactatcgcctcggcaatc
aroG-fus-R	ggcatcggtgatgctcttaccgtagg
D67X-F	ggaacagacatgaattatcagaacNNKNNKttacgcatcaaagaaatcaaagagttacttcctcctgtcgc
D67X-R	gttctgataattcatgtctgttccagtgttgcc
aroG-S180F-F	gcatcagggctttttgtccggtcggcttc
aroG-S180F-R	gaagccgaccggacaaaaaagccctgatgc
XbaI-serA	gagcggataacaattcccctc
aroG-His-HindIII	cgccggaagctttcattagtggtggtggtggtggcccgcgacgcgcttttactgcattc
D67M-F	ttacgcatcaaagaaatcaaagacttacttcc
D67M-LPR	ctctttgatttctttgatgcgtaaaggcaagttctgataattcatatgtatatctcc
D67M-PIR	ctctttgatttctttgatgcgtaaaatagggttctgataattcatatgtatatctcc
D67M-GAR	ctctttgatttctttgatgcgtaaagccccgttctgataattcatatgtatatctcc

Table to be continued on next page

Table 3.3: Primers used in this study

Primer	Sequence
Tsen-trpR-IF	ggataaaccgacgttgatgagcgccacggaatggggacgtcgttactgatccgcacggctgttgacaattaatcatccggc tcg
Tsen-trpR-IR	tcaatcgcttttcagcaacacctcttccagccactggcgcagctcgacgggcgcggctttcttttacttgtacagctcgtccat gcc
up-aroH-out	ggggcgttggtgtaaagattattgccctcaccctgtacgggtgagggcgtagagagattacgcggccgcttctagagttg
Cm-delG-R	gtatcttcccagcctatgcaggcatcggtgatgctcttaccgtaggccagcacctgaagtcagccccatacg
NotI-pTac-aroG	agtgcagcggccgctgttgacaattaatcatcggctcgtataatgtgtaggggaattgtgagcggataac
aroG-speI	tgcggcactagtttattacccgcgacgcgcttttac
u-rpsLp-tac	ttgtgtgaggacgttttattacgtgtttacgaagcaaaagctaaaaccaggagctatttactgttgacaattaatcatcggctcg
fruR-N20	gctcagtcctaggtataatactagtAGCCCAGCTGCCACGGCGTTgttttagagctagaaatagcaagtta aaataaggctagtccg
fruR-out-FF	agtcggtgctttttttccggccgtcattcacaatc
fruR-out-FR	gaacgggagtgcgcctgccgttaataacatagcttg
fruR-out-RF	cctacggtaagagcatcaccgatgcc
fruR-out-RR	cagcatcacttcttttgtcagcccggcatg
FruRN20-F	aaagaagtgatgctgaagacgaaagggcctcgtg
FruRN20-R	aaaaaagcaccgactcggtgccac
pstI-out-1	tcccgggttcttttaaaaatcagtcacaagtaaggtagggttatgatttcagttgacagctagctcagtcc
pstI-out-2	tcttctcctaagcagtaaattgggccgcatctcgtggattagcagattgtatttgatgcctgggcatgcg
pstI-N20	gctcagtcctaggtataatactagtCTGCCAGCTATTACGCTGGAgttttagagctagaaatagcaagtta aaataaggctagtccg
pstI-C1N	acgtacgaaacgtcagcggtc
pstI-C2N	acggttacgctaccggacag
glk-N20	gctcagtcctaggtataatactagtCTCCCTGTAAATATCGATCTgttttagagctagaaatagcaagttaa aataaggctagtccg
u-glk-out	atttacagtgtgagaaagaattattttgactttagcggagcagttgaagacgcggccgcttctagagttg
glk-out-d	agacgtgcgttggtgccgcccacatcaccgactaatgcatactttgtcatatttgatgcctgggcatgcg

Table to be continued on next page

Table 3.3: Primers used in this study

Primer	Sequence
glk-C1	ctattccttatgcggggtcag
glk-C2	atcgggatcgccatcgatac
glk-out-FF	tcatttgatgcatcatcgcgtgcaaccgtgg
glk-out-FR	cttttctagactcgagtcttcaactgctccgctaaagtc
glk-out-RF	ctcgagtctagaaaagaggagaaatactagatgacaaagtatgcattagtcggtg
glk-out-RR	tcgtcttcacctcgactggcgtaaatgtgcaccggaaccgag
Cm20glk-FOR	tgatgcatcaaatgaaaaaaagcaccgactcggtgccac
CmN20glk-REV	tcgaggtgaagacgaaagggcctc
XhoI-galP-MT	ctttagcggagcagttgaagactcgagNNNNNNattaatcatcggctcgNNNNNtgtgtaggcgggagaaggagatatac
XhoI-galP-WT	ctttagcggagcagttgaagactcgagttgacaattaatcatcggctcgtataatgtgtaggcgggagaaggagatatac
galP-XbaI-WT	gtatttctcctcttttctagagctagcaatttacctaggactgagctagctgtcaattaatcgtgagcgcctatttcgcg
galP-XbaI-MT	gtatttctcctcttttctagagctagcaNNNNNcctaggactgagctagcNNNNNNttaatcgtgagcgcctatttcgcg
pBla	gtctcatgagcggatacatatttg
glk-check-F	aacagccatgccggaagcatgacg
glk-check-R	aagttcagcagttgcgtcggctg
tac-galP-check	ccgaggcgattttcatcaccacaatcag
J23119-glk-check	cggctctgaacgtactgtttatcctg
galP-innercheck-R	attgccgggatgataatcacacccag

3.2 Media

3.2.1 LB, SOB, synthetic medium, and LB agar plates

LB medium: 10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl. The pH was 7.0 (adjusted with 5 M NaOH). It was sterilized by autoclaving at 121°C for 20 min. For preparing LB agar plates, 15 g/L agar was used. Unless otherwise stated, LB liquid medium was used for overnight cultivations.

SOB (Super Optimal Broth) medium: 20 g/L tryptone, 5 g/L yeast extract, 0.5 g/L NaCl, 1 g/L $MgCl_2$ and 1 g/L $MgSO_4$. The pH was adjusted to 7.2 with 5 M NaOH. It was sterilized by autoclaving at 121°C for 20 min. The SOB medium was used for the recovery of *E. coli* cells after heat shock transformation or electroporation.

Synthetic medium (SynM): 30 g/L glucose, 20 g/L $(NH_4)_2SO_4$, 0.5 g/L $MgSO_4$, 3 g/L KH_2PO_4, 12 g/L K_2HPO_4, 2 g/L NaCl, 2 mg/L biotin, 5 mg/L DL-calcium pantothenate, and 100× trace elements (Table 3.4). The pH was adjusted to 7.2 with 5 M NaOH. For preparation of SynM, salt solution (10×), 60% glucose stock solution, and concentrated trace element solution (100×) were autoclaved separately, and the components of biotin and DL-calcium pantothenate were sterilized by filtration. These components were then mixed, and autoclaved H_2O was added for adjusting the volume. The SynM was used for cultivation of cells during continuous evolution.

3.2.2 M9 minimal medium

M9 minimal medium was used for the selection and screening of *E. coli* mutants. The preparation was done as follows:

1. A 10× M9 salt solution was prepared by dissolving 75.2 g $Na_2HPO_4 \cdot 2H_2O$, 30 g KH_2PO_4, 5 g NaCl, 5 g NH_4Cl in 1 L ddH_2O. After adjusting the pH to 7.2, it was then autoclaved.
2. 100 ml of 10× M9 salt solution from step 1 were mixed with 7 mL of 60% glucose (autoclaved), 1 mL of 1 M $MgSO_4$ (autoclaved), 0.3 mL of 1 M $CaCl_2$ (autoclaved), 1 mL of 1 g/L biotin (filter sterilized), 1 mL of 1 g/L thiamin (filter sterilized), and 10 mL of 100× trace elements solution (Table 3.4.).
3. The volume was adjusted to 1L with sterilized ddH_2O.

For preparing M9 agar plates, 100 mL of 10× M9 stock solution with 15 g/L agar were autoclaved. Once the medium was cooled down to about 50 °C, other components as

mentioned above were added into the medium immediately and then mixed them completely and poured it into Petri-dishes.

Table 3.4: Components of 100× trace element solution.

No.	Components	Work Con. (mg/L)
1	$MnSO_4 \cdot 2H_2O$	12.5
2	$AlCl_3 \cdot 6H_2O$	12.5
3	$FeSO_4 \cdot 7H_2O$	50
4	$CoCl_3 \cdot 6H_2O$	8.75
5	$CaCl_2 \cdot 2H_2O$	50
6	$ZnSO_4 \cdot 7H_2O$	2.5
7	$CuCl_2 \cdot 2H_2O$	1.25
8	H_3BO_3	0.625
9	$Na_2MoO_4 \cdot 2H_2O$	2.5

3.2.3 Fermentation media

Seed medium (Table 3.5) and fermentation medium (FM-III, Table 3.6) were prepared according to those described by Gu et al., (2012) and Chen and Zeng (2017). To prepare 1 L of the seed medium (Table 3.5), a certain amount of each of the following components: $MgSO_4 \cdot 7H_2O$, KH_2PO_4, $(NH_4)_2SO_4$, yeast extract, and monosodium citrate was dissolved in 900 ml of ddH_2O. After autoclaving, 50 mL of 60 % glucose (autoclaved), 1.5 mL of 10 g/L $FeSO_4 \cdot 7H_2O$ (filter sterilized), and 10 mL of 10 g/L vitamin B_1 (filter sterilized), were added. The final volume was set to 1 L with autoclaved ddH_2O, and the pH was adjusted to 7.0 with 5 M NaOH.

FM-III medium (Table 3.6) was mostly used in batch fermentation of *E. coli* strains in shake flasks. To buffer the pH during fermentation, 30 g/L $CaCO_3$ were used in the FM-III medium. It is noteworthy that the concentration of $(NH_4)_2SO_4$ (8 g/L) was doubled, compared to that (4 g/L) in the FM-II medium reported by Chen and Zeng (2017). The composition of the fermentation medium (FM-B) in bioreactors is given in Table 3.7. Preparation of the media FM-III and FM-B were similar to that of the seed medium as described above.

Table 3.5: Components of the seed medium.

No.	Components	Work Con. (g/L)
1	Glucose	30
2	$MgSO_4 \cdot 7H_2O$	0.5
3	KH_2PO_4	1.5
4	$(NH_4)_2SO_4$	10
5	Yeast extract	15
6	$FeSO_4 \cdot 7H_2O$	0.015
7	Monosodium citrate	0.5
8	Vitamin B_1	0.1

Table 3.6: Components of the fermentation medium (FM-III) in shake flasks.

No.	Components	Work Con. (g/L)
1	Glucose	30
2	$MgSO_4 \cdot 7H_2O$	0.5
3	KH_2PO_4	3
4	K_2HPO_4	12
4	$(NH_4)_2SO_4$	8
5	Yeast extract	1
6	Monosodium citrate	2
7	$CaCO_3$	30

Table 3.7: Components of the fermentation medium (FM-B) in bioreactors.

No.	Components	Work Con. (g/L)
1	Glucose	30
2	$MgSO_4 \cdot 7H_2O$	0.5
3	KH_2PO_4	2
3	$(NH_4)_2SO_4$	4
5	Yeast extract	1
6	Monosodium citrate	2
7	Trace elements	Table 3.4
8	Biotin	1×10^{-4}
9	DL-calcium pantothenate	5×10^{-4}

3.3 Molecular cloning methods

3.3.1 Genomic DNA and plasmid isolation

Isolation of genomic DNA from *E. coli* was performed using the genomic DNA isolation kit NucleoSpin® Tissue (Macherey-Nagel, Germany). Moreover, the extraction of plasmid DNA was carried out using the NucleoSpin® Plasmid kit (Macherey-Nagel, Germany). Unless otherwise stated, isolations were carried out by following the protocol outlined in the manual that comes with the kit.

3.3.2 Recombinant DNA techniques

DNA amplification, purification, digestion, and ligation were carried out according to the standard protocol (Russell and Sambrook, 2001) and the instructions of the kits. More specifically, Phusion High-Fidelity PCR Master Mix (Thermo Scientific) was used to amplify a DNA fragment with the requirement of higher fidelity, whereas, DreamTaq™ Hot Start Green PCR Master Mix (Thermo Scientific) was used for doing colony PCR. When the PCR product is very specific, a PCR clean-up was carried out according to the manual of the NucleoSpin gel and PCR clean-up kit (Macherey-Nagel, Germany). Otherwise, gel-purification was performed to recover desired DNA fragments from agarose gels after PCR amplification where the kit was also used. When necessary, the DNA product was treated with DpnI to get rid of PCR templates. In this study, plasmid construction was mostly carried out by using the In-Fusion HD Cloning Kit (Clontech Laboratories, Inc.).

3.3.3 Colony PCR

Colony PCR was used for pre-check if the transformants containing a desired plasmid construction or a desired DNA construction in the chromosome. The temperatures for the denaturation, annealing, and extension were set accordingly. In this study, unless otherwise stated, the product Hot Start Green PCR Master mix (2×) from Thermo Scientific was used for colony PCR. Each colony PCR reaction with a total volume of 10 µL was set up and performed as follows:

1. Cells from single colony on agar plate were picked and transferred them into 4.4 µL of ddH_2O in PCR tube.
2. 0.3 µL of each of the corresponding forward and reverse primers (10 µM) were added into the mixture from step 1.
3. 5 µL of Hot Start Green PCR Master mix (2×) was added into the mixture from step 2.

4. PCR reaction was run in a thermocycler with a program set according to the instruction of the product user guide.
5. PCR product was checked with agarose gels electrophoresis.

3.4 Transformation of *E. coli*

To do genome editing, the plasmid and/or linear dsDNA fragment has to be transferred into the competent cells of a host strain, which was done through electroporation in this study.

3.4.1 Preparation of electroporation-competent cells

The procedure for preparation of electroporation-competent cells was modified from the methods reported by Yu et al. (2000) and Chen and Zeng (2017). Cyro-stock cells or single colony were inoculated into LB medium and grown at 30 °C overnight. The overnight culture was then inoculated at a 50-fold dilution into 10 mL of fresh SOB medium with 20 mM of L-arabinose in a 50 mL conical tube. Then the cells were grown at 30 °C, 250 rpm. When the OD_{600} reached 0.4-0.6, induction was carried out by incubating the culture at 42 °C for 15 min with shaking. After that, the culture was chilled on ice for 10 min with occasional shaking. Cells were harvested from the culture by centrifugation at 4 °C, 5000 rpm for 10 min. After washing with 1 ml pre-cooled sterilized water three times, the pellets were re-suspended in 200 μL pre-cold 10% glycerol, and divided into two reactions for transformation or stored at -80 °C for further use.

3.4.2 Electroporation

For the electroporation, a certain amount of desired DNA (plasmid or linear dsDNA) eluted in H_2O was added to the electrocompetent cells and mixed sufficiently. The mixture was later transferred into a 0.2 cm gap sterile electroporation cuvette. The electroporation was performed at 2.5 kV, 1.8 μF with a pulse controller of 200 Ω. Afterwards, the reaction mixture was transferred from the cuvette into 1mL SOB medium. The cells were recovered by incubating at 30 °C for 2 hours. After that, 100 μL of the cells were spread on the agar plates with appropriate antibiotics and grown at 30 °C.

3.5 Cultivation conditions

During cultivation, if necessary, antibiotics were added at working concentrations as follows: 50 μg/L kanamycin (Kan), 100 μg/L ampicillin (Amp), 34 μg/L chloromycetin (Cm), and 50 μg/L streptomycin (Sm).

3.5.1 Cultivation conditions on 96-well deep well plate

The cultivation conditions for characterization of the enzyme variants on 96-well deep well plate were modified from those reported in (Zhang et al., 2019). Generally, 300 µL M9 medium or FM-III medium containing the appropriate antibiotics were prefilled into each well of a deep 96 well plate (Eppendorf® Deepwell Plate 96). After inoculating with transformants of interest, the plate was sealed with a gas-permeable seal and incubated at 37°C, 300 rpm for 24 hours. The cultures were then analyzed by measuring the OD600 and fluorescence intensity. Specifically, 100 µL of each of the cultures were used for the measurement of $OD_{600,}$ and 100 µL of each of the cultures were sampled for measuring the fluorescence signal. According to the results of both the OD and the specific fluorescence intensity, the candidates which had higher specific fluorescence intensity as well as high OD were selected. These candidates were then double-checked with another round of characterization in deep 96 well plate under the same conditions as used for the first round. Finally, the candidates that had the best growth and strongest fluorescence intensity were selected for further characterization, i.e., fed-batch fermentation.

3.5.2 Continuous evolution of strains with GalP/Glk-dependent glucose utilization

In vivo continuous mutagenesis experiment was performed as described in (Packer et al., 2017) with minor modifications. Briefly, the donor plasmid pCmN20-$galP^{MT}$ was first transformed into the R01 host cells (Table 3.1) by electroporation, and then the resultant cells (R01: pCmN20-$galP^{MT}$) were inoculated into the SOB medium with 10 mM L-arabinose and cultured for three hours at 30 °C. Auxostat vessel containing SynM with 15 µg/mL chloramphenicol and 25 µg/mL kanamycin was inoculated with the starter cultures (resuspended three times with SynM medium) and grown at 30 °C while mixing via a magnetic stir bar. Once the mutants grew to OD_{600} = 0.8, the inlet flow of fresh media (SynM with 10% w/v arabinose, 15 µg/mL Cm, and 25 µg/mL Kan) was set at a rate of 60-80 mL/h, and the outlet flow of the auxostat culture was set at a rate of 100 mL/h with the waste needle set at the height of 30 mL. To realize the real-time measurement of OD_{600} and monitoring of the fluorescence signal, the OD measurement device and the flow cytometer were connected tandem to the auxostat vessel. The total flow rate through each lagoon was automatically set based on cell growth to keep the cell growth at around 0.8 during the evolution process. The final evolutionary samples were selected from the auxostat and further screened on the SynM agar plates.

3.5.3 Fermentation with complex medium

The conditions for batch fermentation in shake flasks were reported previously (Chen and Zeng, 2017). All the batch fermentations were carried out at 37 °C and 250 rpm. In practice, an isolated clone was pre-cultured with 5 mL of LB medium in a conical flask (50 mL) overnight. The pre-culture was then inoculated into 10 mL of seed medium in a baffled shake flask (100 mL) with an initial OD_{600}=0.2. After growing it for 4-6 hours, the seed culture was transferred into 30 mL of fermentation medium (FM-III) at an initial OD_{600}=0.1.

For fed-batch fermentation in bioreactors, the pre-culture and seed culture were performed as done for batch fermentations but with 10 mL of LB medium (in 100 mL baffled shake flask) and 50 mL of seed medium (in 300 mL baffled shake flask), respectively. The fermentations were carried out in a highly instrumented and automated 4-parallel 1.5 L bioreactor system (DASGIP parallel bioreactor system, Eppendorf, Germany) with an initial working volume of 500 mL. The concentration of glucose was controlled during the fermentation via adding feeding solution at a variable feed rate. Especially, to avoid the case of glucose depletion (<1 g/L) or excess (>30 g/L), the feeding rate was automatically increased from 0 to 6 mL/L.h by DASGIP system during the exponential and early stationary phases. Afterward, the glucose concentration was measured manually with a YSI glucose analyzer, and then the feeding rate was adjusted accordingly to maintain the glucose concentration at around 10 g/L. Fermentations were performed at 37 °C and with pH 6.8 which was controlled with 25% NH_4OH. The dissolved oxygen was controlled at 30% of air saturation by orderly varying the agitation speed, the oxygen content, and the aeration rates. When required, antibiotic(s) and inducer(s) were added in medium(s) and feeding solution(s).

3.6 Analytic methods

3.6.1 Analysis of metabolites by HPLC

In this study, the accumulation of two intermediates of the tryptophan pathway, namely 3-dehydroshikimate (DSA) and shikimic acid (SA), were analyzed during fermentations. Their concentrations in the fermentation samples can be analyzed by using HPLC as reported by Lin (2017). In general, they are able to be separated by the Aminex HPX-87H column (300 × 7.8 mm column) (Bio-Rad, Hercules, USA) when operated at 60°C with 5 mM H_2SO_4 as mobile phase at a flow rate of 0.6 ml/min. All these two components have good absorbance at 210nm.

3.6.2 Analysis of extracellular Trp

The extracellular concentration of Trp was measured by a simple and sensitive spectrophotometric method reported by Nagaraja et al., (2003). Lin had scaled down the volume of each reactant in this method and developed a protocol making it easier to be applied (Chen, 2017). The preparation of each solution and a detailed protocol were also described here as follows:

Solutions:

1) 0.2 % PPDD (100 ml): dissolve 0.2 g PPDD in 100 ml of 5 % (V/V) HCl and store it in the dark at room temperature.
2) 0.5 % sodium nitrite ($NaNO_2$) in water: store it at -20°C for short term storage (if possible, prepare it freshly).
3) 3 % sulfamic acid (H_3NO_3S, 100 ml): dissolve 3g H_3NSO_3 in 100 ml of ddH_2O and store it at room temperature
4) 1:1 sulfuric acid (H_2SO_4) (in a glass container): mix 96-98% concentrated H_2SO_4 with an equal volume of deionized water.

Protocol: (total volume for each reaction is 1.0 ml, the reaction can be carried out in 1.5 EP tubes)

1) Mix 0.12 ml of 0.2 % PPDD and 0.24 ml of 1:1 H_2SO_4, vortex and chill it down on the ice.
2) Add 0.08 ml of pre-cooled 0.5 % $NaNO_2$ into the mixture from step 1 (on ice).
3) Add 0.04 ml of 3% H_3NSO_3 into the mixture from step 2 and mix them completely by vortex. Then keep it on ice for at least 10 min.
4) Add 0.1 ml of Trp standards or pre-diluted samples to the mixture from step 3 and mix them completely by vortex. Then keep it on ice for more than 20 min.
5) Add 0.42 ml of 1:1 H_2SO_4 to the mixture from step 4 and mix them completely by vortex.
6) Measure the absorption at 520 nm with a microplate reader in 96-well microplates (200 µl sample for each well).

3.6.3 Measurement of fluorescence intensity

Single colony (or 3 µl of cryo-stock) of each of the strains to be studied was inoculated into LB medium and grown at 37°C, 220 rpm overnight. Cells from each overnight culture were harvested by centrifugation (4 °C, 5000 rpm for 10 min) and followed by washing them with the M9 medium (Chapter 3.2.2) three times. After that, cells of each were re-suspended in fresh

M9 medium and inoculated with the same initial OD_{600} into 10 ml of fresh M9 medium in a 50 mL conical tube. After 10 hours of cultivation, cells were subjected to fluorescence analysis using CytoFLEX Flow Cytometer (Beckman Coulter, USA). To do so, after washed three times with PBS buffer, cells from each culture were diluted 100-fold, and then were used for measuring eGFP fluorescence (MFI of 10,000 events) using a flow cytometer at an excitation wavelength of 488 nm. All data were processed with the Beckman Flow software, and electronic gating was used to separate positive signals from the instrument and water sample background. For fluorescence intensities, a medium fluorescence unit (MFU) was calculated for each culture.

3.7 Enzyme assay

3.7.1 Protein expression and purification

Protein expression and purification protocols were performed in a manner similar previously published with minor modifications (Chen et al., 2018). Concentration of purified protein was analyzed according to Bradford's method using bovine serum albumin as standard and a prefabricated assay from Bio-Rad Laboratories (Hercules, CA).

3.7.2 Assay of indole-3-glycerol phosphate synthase activity

In *E. coli* and *A. niger*, TrpC shares the same subunit, indole-3-glycerol phosphate synthase (IGPs). Therefore, *in vitro* assay of IGPs activities of the enzymes *An*TrpC and *Ec*TrpC were performed as described in (Chen et al., 2018) but with minor modifications. The activity of IGPs was measured by monitoring the formation of IGP via absorbance at 278 nm (Zaccardi et al., 2012) with a molar extinction coefficient value of 5500 M^{-1} cm^{-1} (Kirschner et al., 1987). To investigate the activation or inhibition of anthranilate on the activity of IGPs, the activities were measured in the presence of different concentrations (from 0 to 2 mM) of anthranilate. The complete reaction mixture contained 4 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffer (pH 7.4), 180 μM 1-(2-Carboxyphenylamino)-l-deoxyribulose 5-phosphate (CdRP), and 80 μg purified enzyme with or without anthranilate. With a total volume of 0.2 mL, reactions were performed in a cuvette at 30°C. Enzyme and anthranilate were pre-mixed and equilibrated to 30°C for 2 min, after that, the substrate CdRP was added to start the reaction. The substrate CdRP was chemically synthesized by adopting the method reported by Kirschner et al., (1987).

3.7.3 Assay of 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase activity

3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase (L-phenylalanine repressible, AroG) is a key enzyme of the chorismate pathway for the synthesis of aromatic amino acids (AAA). Enzyme kinetics of DAHP synthase was performed as described in (Schoner and Herrmann, 1976) but with minor modifications. The enzyme activity was measured by monitoring the disappearance of PEP via absorbance at 232 nm with a molar extinction coefficient value of 2544.5 M^{-1} cm^{-1} (Fig. 3.1). To investigate the effect of Phe on the activity of AroG, the activities were measured in the presence of different concentrations (from 0 to 40 mM) of Phe. The complete reaction mixture contained 10mM Bis-tris propane (BTP) buffer (pH 7.0), 50 μM $MnSO_4$, 600 μM PEP, and 500 μM E4P with or without Phe. Reactions were carried out in a total volume of 0.2 mL in a cuvette at 25°C. The mixture (without PEP) and the substrate PEP were equilibrated to reaction temperature, separately, and the reaction was started by adding the substrate PEP.

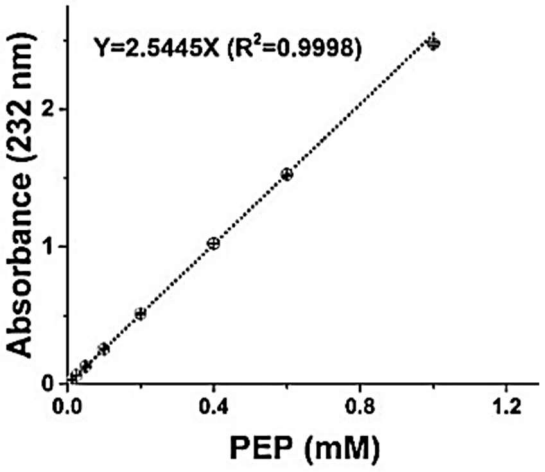

Figure 3.1: Standard curve of PEP with the absorbance at 232 nm measured in the cuvette.

Chapter 4

4. Growth-coupled and sensor-guided *in vivo* screening of multifunctional TrpC enzyme variants

4.1 Introduction

Development of an efficient Trp-producing microbial cell factory is a challenging work owing to complex and negative regulations of Trp pathway involved, such as attenuation, repression, and feedback inhibition (Fig. 2.2). Specifically, a feed-forward regulation (e.g., inhibition and activation) was recently discovered as a novel regulation of Trp operon in *E. coli* (Fig. 2.2) (Chen et al., 2018). To deregulate the Trp pathway, especially to develop a feed-forward-resistant or feed-forward- activation enzyme, the semi-rational design has been applied in this work for engineering of anthranilate-sensitive enzyme TrpC, making it more efficient for Trp pathway.

Anthranilate (ANTH) is the first intermediate of the Trp branch pathway in *E. coli*. As illustrated in Fig. 2.1, it is subsequently converted into the final product Trp by a series of enzymatic reactions (They are TrpD, TrpC, TrpB, and TrpA). A previous study discovered that the indole glycerol phosphate synthase (IGPs) of bifunctional enzyme TrpC of *E. coli* (*Ec*TrpC) is feed-forward inhibited by anthranilate, and another TrpC of *Aspergillus niger* (*An*TrpC) is feed-forward activated (Chen et al., 2018). In *E. coli*, *Ec*TrpC consists of two isoenzymes: N-(5-phosphoribosyl) anthranilate isomerase (PRAI) and IGPs, whereas, in *A. niger*, the trifunctional enzyme *An*TrpC is composed of three isoenzymes: anthranilate synthase component II (AS II), IGPs, and PARI, for which the AS II domain is essential for the activation (Chen et al., 2018). It was also found that, although the catalytic efficiency of *An*TrpC was not satisfactory in contrast to that of *Ec*TrpC, expressing *An*TrpC in a Trp-producing *E. coli* S028 made the strain more efficient in Trp production (Chen et al., 2018). One conceivable explanation is that the enzyme *Ec*TrpC suffers stronger feed-forward inhibition from anthranilate and cannot be conducive to Trp synthesis. In this regard, it was of great attraction to develop *Ec*TrpC and *An*TrpC using semi-rational design, making the anthranilate-resistant

enzyme variant *Ec*TrpC and enhanced anthranilate-activation enzyme variant are more efficiency for Trp pathway.

Semi-rational design is of elemental significance for development of efficient bioprocesses. While impressive progress has been made in the design of the enzyme variant library, the availability of reliable and highly effective screening methods is becoming increasingly essential (Dietrich et al., 2010). To date, enzyme variants obtained from conventional approaches can hardly be phenotypically distinguished (Ren et al., 2018). For example, insufficient phenotypic characterization, especially that based merely on the cell growth rate, makes it challenging to identify the best performer among the improved enzyme variants (Ren et al., 2015). Also, using a biosensor alone requires high-throughput screening equipment such as FACS, and the generated overlapping profiles and aberrant fluorescence from FACS can cause a high rate of false positives during sorting. In theory, a reliable *in vivo* screening approach is desired to link the change of interest caused by an induced mutation on a target gene to a change in cell growth (Lu et al., 2012; Zhu et al., 2017) and the expression strength of a reporter gene via a biosensor (Binder et al., 2013; Fang et al., 2016). In this chapter, a method of linking plasmid-assisted engineering with growth-coupled and sensor-guided *in vivo* screening (PGSS) was therefore proposed to ensure that enzyme variants are selected with desired functions. To implement the PGSS approach, the Trp synthesis pathway in the host strain was blocked by knocking-out the *EctrpC* gene to design a Trp-auxotrophic strain. Then this auxotrophic strain was employed for engineering of *Ec*TrpC and *An*TrpC enzymes (Fig. 4.1A).

4.2 Design of an auxotrophic host cell for library screening and *in vivo* characterization

In the part of this work, a screening platform was designed, in which the growth of the auxotrophic strain is directly coupled to the catalytic efficiency of TrpC variants, and the intracellular Trp concentration is monitored by the fluorescence intensity via a Trp sensor (Fig. 4.1A).

To this end, Trp synthesis was first blocked by knocking out the gene *EctrpC* to design a Trp auxotrophic strain S092. Subsequently, the Trp sensor (P_{tac}-*tnaC-eGFP*) was introduced into the auxotrophic strain to construct strain S093 (Fig. 4.1A). Afterward, the Trp necessity was tested for the cell growth by cultivating S093 in the M9 medium without amino acids. Without Trp synthesis, S093 was not able to grow in the synthetic medium (Fig. 4.1B). More results subsequently revealed that the auxotrophy is relieved by complementation of lost enzymatic

function after reintroducing the enzyme *Ec*TrpC in S093 (Fig. 4.1B). The fluorescence signal from the S094 cells (S093:p*tac-EctrpC*) could also be detected (Fig. 4.1C). These results suggested that the strain is auxotrophy for the essential Trp molecule. It provided what is the readily discernible phenotype and could be used for the library screening. To investigate the threshold of evolutionary pressure (anthranilate), the effects of varying concentrations of extracellular anthranilate on the growth and fluorescence intensity of S094 were also examined. It was found that the growth and fluorescence intensity were suppressed entirely when the concentration of anthranilate was increased above 45 mM (Fig. 4.1A and B). It suggested that the wild type strain will be wiped out during library selection when the threshold valve is set to 50 mM. Nevertheless, since *An*TrpC is feed-forward activated by anthranilate, no adaptive selection is imposed on the *An*TrpC variants. Thus, it was necessary to demonstrate whether the growth and the fluorescence intensity of S093 are strongly correlated with the relative enzyme activity of *An*TrpC.

A biosynthetic pathway expressed at different levels of a key enzyme can cause different concentrations of the final target. The growth of S093 was therefore determined at different concentrations of extracellular Trp addition (0-14 mM) in M9 medium. As indicated in Fig. 4.1D, the growth and fluorescence intensity of auxotrophic strain varied as a function of the concentrations of extracellular Trp. When the concentration of extracellular Trp ranged from 0 to 4 mM, the growth (R^2 = 0.9901) and the fluorescence intensity (R^2 = 0.9925) of S093 were correlated strongly with the concentration of extracellular Trp. With an increased concentration of extracellular Trp, however, S093 showed a comparable growth and fluorescence intensity (Fig. 4.1D). These results demonstrated that the strong correlation between the cell growth and fluorescence intensity with the catalytic efficiency of the enzyme TrpC could highlight the sensitivity and selectivity of the screening platform. It can be employed to facilitate the engineering of *Ec*TrpC with the desired performance.

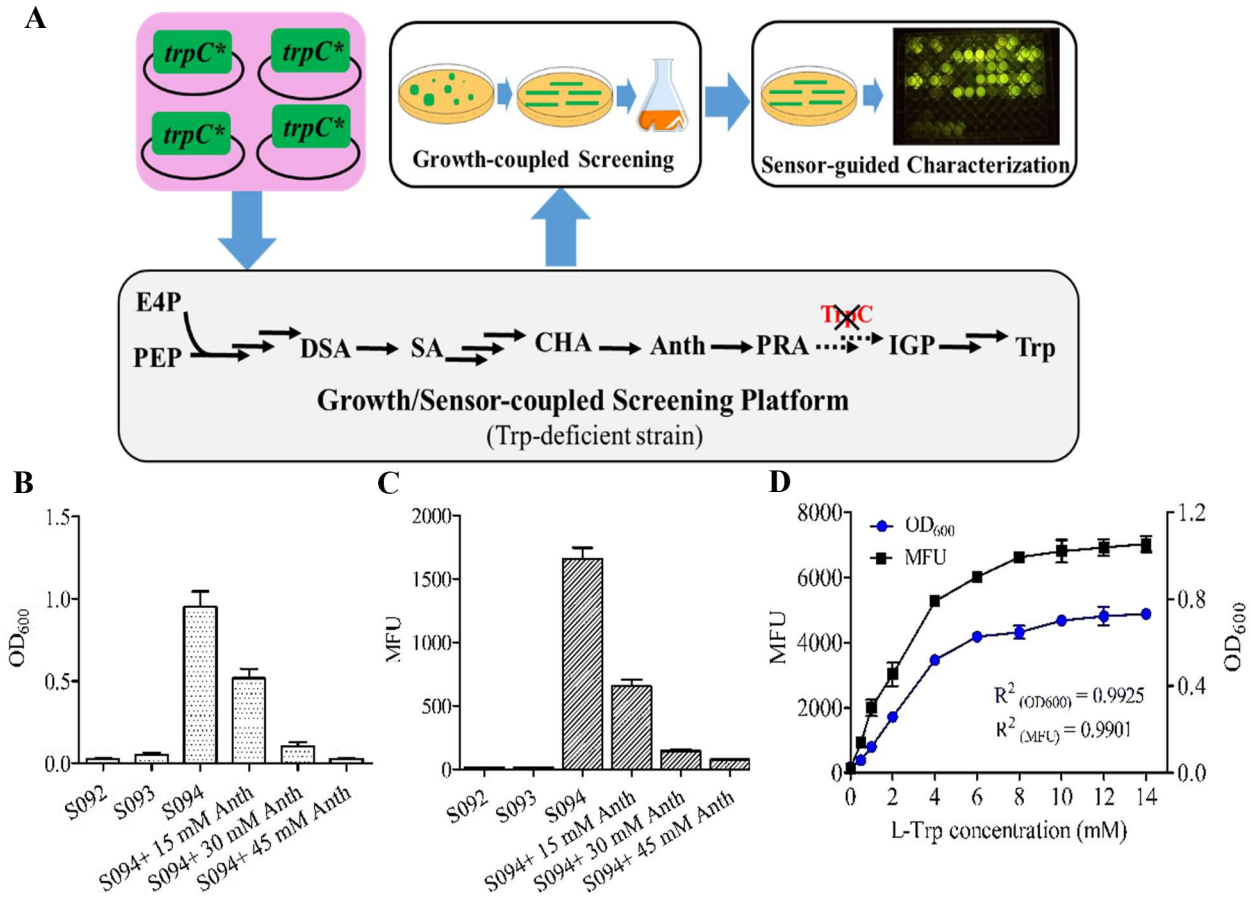

Figure 4.1: (A) Plasmid-assisted engineering of enzyme variants integrated with growth-coupled and sensor-guided *in vivo* screening of candidate enzyme(s) (PGSS) as an efficient approach for protein and pathway engineering; Effects of anthranilate on (B) the growth and (C) the fluorescence intensity of the Trp-auxotrophic strain S092, S092::P_{tac}-*tnaC-eGFP* (S093), and S093::*EctrpC* (S094); (D) The correlation of growth and fluorescence intensity of the strain S093 with the addition of extracellular Trp. The coefficient of determination (R^2) was calculated in the range of 0-4 mM Trp. Results were derived from three independent experiments.

4.3 PGSS-assisted engineering of anthranilate-inhibited *Ec*TrpC

4.3.1 Library construction and high-throughput screening

After testing the sensitivity and selectivity of the screening platform, one set of *EctrpC* gene variants was constructed using codon saturation mutagenesis on a few rational target residues, e.g., S58, P59, S60, and K61. As observed in Fig. 4.2A, these residues are actively involved in the anthranilate binding sites. Using the entire plasmid p*trc*-*Ec*TrpC as a template (Table 3.2), the mutagenesis libraries were first generated by using a pair of synthesized mutagenesis primers (*eIGPs-1-F* and *eIGPs-1-R*, Table 3.3), and then they were transferred into the strain S093. Afterwards, the resulting mutants were screened and characterized using the PGSS approach.

During library screening and selection, 96 mutants that produced a bigger colony and stronger

fluorescence intensity on the M9 agar plate supplemented with 50 mM anthranilate were selected for further characterization. After 30 hours of cultivation, it was discovered that 20 mutants exhibited a high Trp production than those of other mutants (data not shown). Thereby, the sequences of *EctrpC* gene variants in the selected mutants were examined. The alignment of sequences revealed that there were mainly three types of mutations at selected positions: 45% for the variant *Ec*TrpC$^{S58Q\text{-}P59V\text{-}S60F\text{-}K61Q}$ (QVFQ), 45% for the variant *Ec*TrpC$^{S58Q\text{-}P59L\text{-}S60R\text{-}K61C}$ (QLRC), only 10% for the variant *Ec*TrpC$^{S58A\text{-}P59F\text{-}S60V\text{-}K61R}$ (AFVR) (data not shown). These mutants were then subjected to a re-characterization using fermentation in shaking flasks. An increased Trp production was confirmed for the strain *Ec*TrpC^{QVFQ} after 20 hours of fermentation, while the strains *Ec*TrpC^{QLRC} and *Ec*TrpC^{AFVR} turned out to be false positives (data not shown). As given in Fig. 4.2B, the IGPs domain of *Ec*TrpC is connected with the PRAi domain by a helix. Also, the anthranilate-binding-sites in the IGPs domain are adjacent to this helix. Therefore, it was assumed that the mutations *Ec*TrpC^{QLRC} and *Ec*TrpC^{AFVR} generated in the IGPs domain might affect the stability of the protein structure. Indeed, as observed in Fig. 4.3A, only the isoenzyme IGPs of the enzyme variants QLRC and AFVR were expressed, while another isoenzyme, phosphoribosyl-anthranilate isomerase (PRAi), was lost. The truncated expression could neither lead to an intact catalytic activity of *Ec*TrpC, nor an increase in the Trp production. Thus, the enzyme variant QVFQ was selected as a better candidate for further studies.

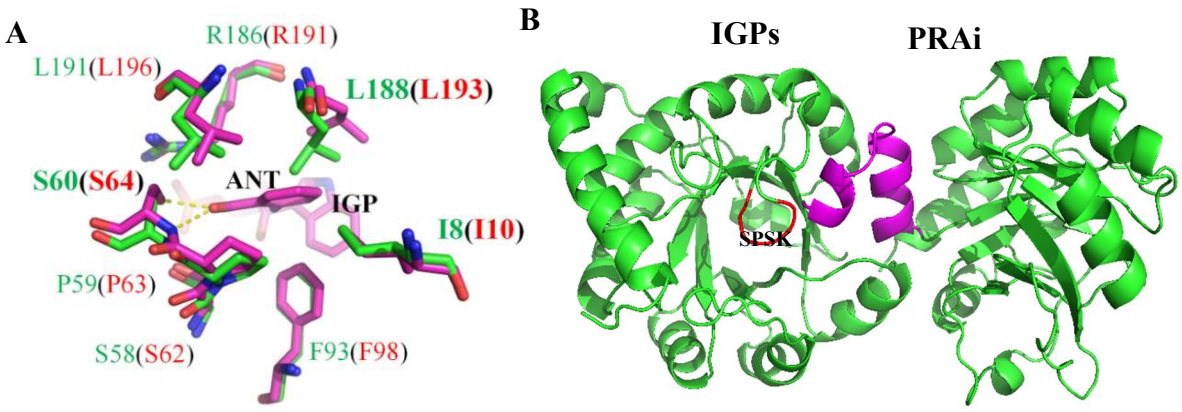

Figure 4.2: (A) Comparison of the anthranilate binding sites of eIGPs from *E. coli* (residues are shown in green) and *Mt*IGPs from *M. tuberculosis* (residues are shown in red) (Chen, 2017). ANT: anthranilate. (B) Crystal structure of TrpC from *E. coli*. The 58th to 61st residues (SPSK) for the wild-type enzyme *Ec*TrpC is shown in red. The helix for the connection of IGPs and PRAi domains is shown in pink.

4.3.2 Characterization of selected enzyme variant QVFQ

To elucidate the effect of anthranilate on the enzyme activity of the variant QVFQ, the catalytic activity of the purified enzyme variant QVFQ-6His (Fig. 4.3A) was investigated in the presence

of different concentrations of anthranilate. The plot of the activity of QVFQ against anthranilate showed that the activity was significantly decreased with the increase in anthranilate concentration (Fig. 4.3B). The specific activity of the variant QVFQ was 3.03±0.05 μmol/min/mg without the addition of anthranilate, which was 23.17% higher that of the wild-type (2.47±0.07 μmol/min/mg). Also, it was revealed that 52% of the activity of QVFQ was inhibited by 3 mM anthranilate, but the activity of the wild type was nearly completely lost under the same condition. The inhibition constant (IC_{50}, 50% inhibitory concentration) of QVFQ for anthranilate was estimated to be about 2.22 mM, while it was approximately 1.12 mM for the wild-type *Ec*TrpC. This result demonstrated that the enzyme variant QVFQ alleviated the inhibition against a higher concentration of anthranilate.

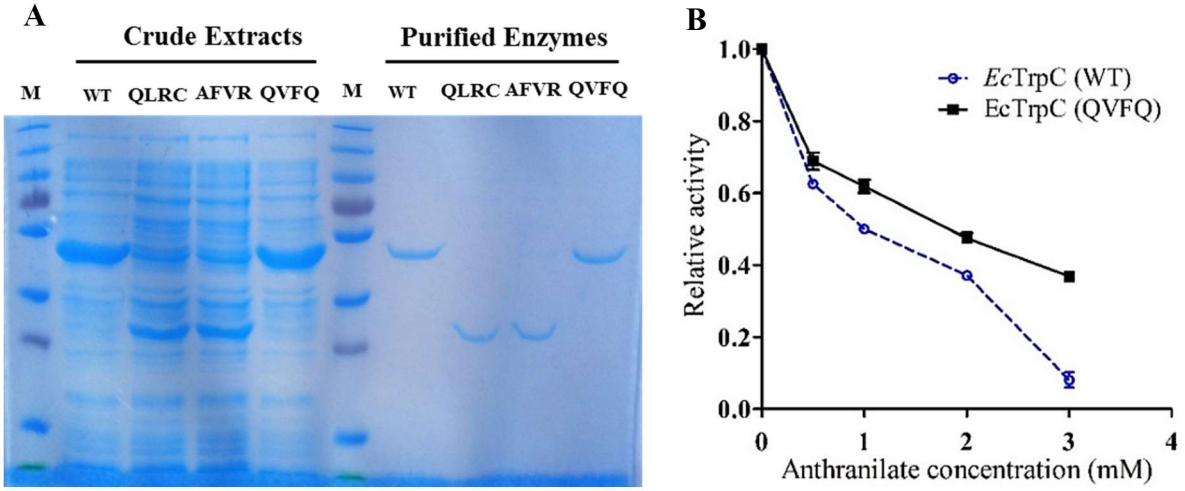

Figure 4.3: (A) SDS-PAGE of the wild-type and enzyme variants of the *Ec*TrpC. M: PageRuler^{TM} plus prestained protein ladder. (B) Effect of anthranilate on the wild-type enzyme *Ec*TrpC and enzyme variant QVFQ. The data were derived from three independent experiments.

The initial rates of substrate turnover were recorded over a substrate range of 0.004 to 0.3 mM CdRP for the enzyme variant QVFQ, and the kinetic data are summarized in Table 4.1. The catalytic constant (k_{cat}) for QVFQ decreased by about 25% compared to the wild-type enzyme, and the Michaelis constants (k_m) for QVFQ also decreased by about 27%. The decrease in k_{cat} and k_m for QVFQ is most probably because the anthranilate-binding-sites are adjacent to the catalytic sites that are surrounded by *β* sheets (Fig. 4.2B), and the alternation of sequences can influence the catalytic activity. Nevertheless, it was found the catalytic efficiency (k_{cat}/k_m) for QVFQ is similar to that of the wild type, and this result is consistent with the similar production of Trp between the strain *Ec*TrpC^{QVFQ} and the wild-type strain (data not shown). Although the catalytic efficiency of QVFQ showed no advantages, anthranilate-resistant enzyme variant QVFQ might be a better candidate for enhancing strain tolerance against a high concentration of anthranilate when the enzyme variant is applied to an industrial scale.

Table 4.1: Apparent kinetic constants of the wildtype and enzyme variant QVFQ of *Ec*TrpC enzymes

Enzyme	CdRP		
	k_m (μM)	k_{cat} (s^{-1})	k_{cat}/k_m ($M^{-1}s^{-1}$)
WT	8.37±0.15	0.04±0.01	430.11
QVFQ	6.14±0.04	0.03±0.01	439.74

Note: CdRP, 1-(2-Carboxyphenylamino)-1-deoxyribulose 5-phosphate

4.4 PGSS-assisted engineering of anthranilate-activated enzyme *An*TrpC

4.4.1 Library construction and high-throughput screening

After testing the feasibility of the PGSS-assisted screening platform, it was applied to develop another enzyme TrpC, anthranilate-activated enzyme TrpC of *A. niger* (*An*TrpC), to enhance its catalytic activity. One set of *AntrpC* gene variant libraries was first generated by codon saturation mutagenesis on a rational target residue (Figs. 4.4A and B). Although the crystal structure of *An*TrpC from *A. niger* with the substrate CdRP is not available, the protein sequence alignment of *An*TrpC protein with other TrpC proteins such as *Ec*TrpC from *E. coli*, *Mt*TrpC from *Mycobacterium tuberculosis*, *Ss*TrpC from *Sulfolobus solfataricus*, and *Sc*Trp31 from *Saccharomyces cerevisiae* revealed several parallels among their catalytic centers (Fig. 4.4A). Specifically, according to the available crystal structure of *Ss*TrpC with its substrate CdRP (PDB: 1LBL), the Arg378 residue in *An*TrpC was supposed to be a homologous residue with the Arg182 residue in the catalytic site of *Ss*TrpC (Figs. 4.4A and B). In this respect, the residue (Arg378) of enzyme *An*TrpC was selected for codon saturation mutagenesis with a pair of primers (*AntrpC-T1F* and *AntrpC-T1R,* Table 3.3) by amplification of the whole plasmid p*trc-AntrpC* (Table 3.2). Then the resulting plasmid was transferred to the strain S093.

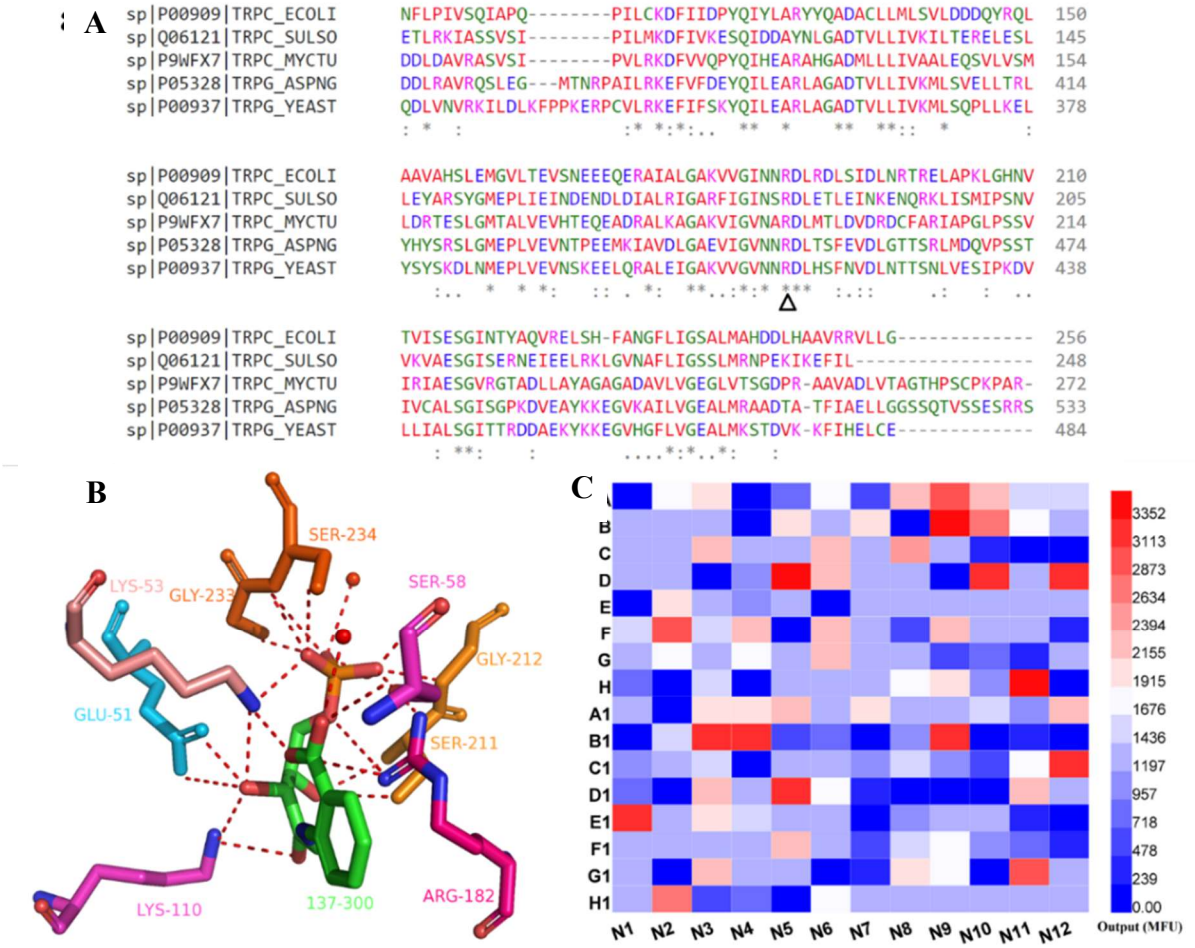

Figure 4.4: (A) Alignment of the TrpC sequences from *E. coli* (P00909), *Sulfolous solfatarious* (Q6121), *M. tuberculosis* (P9WFX7), *A. niger* (P05328), and *S. cerevisiae* (P00937). The residues of the conserved binding site of substrate CdRP is highlighted in a black triangle. (B) Key residues involved in the substrate CdRP binding sites of *Ss*TrpC from *S. solfatarious*. (C) Heat map of the fluorescence intensity (also known as medium fluorescence unit, MFU) of the selected *An*TrpC mutants. A total of six samples in H1:N7-N12 wells are presented as reference control (wild-type *An*TrpC). The sample in each well presents one mutant, and the wells in red color indicate that the mutants have higher activity.

After 15 hours of cultivation, it was found that the M9 agar plate yielded hundreds of mutants with different colony sizes and different strengths of fluorescence intensity (also known as medium fluorescence unit, MFU) (data not shown). After phenotypes confirmation on the M9 agar plate, a total of 186 candidates *An*TrpC^{R378X} were selected and then tested by cultivation with M9 medium in two 96 deep-well plates. After 10 hours of cultivation, all mutants *An*TrpC^{R378X} were subjected to the measurement of fluorescence intensity. All MFU intensities were represented by a heat map (Fig. 4.4C). As shown in Fig. 4.4C, a total of 16 mutants *An*TrpC^{R378X} produced higher MFU intensities (over 2400) than those obtained by other mutants and the wild-type strain *An*TrpC^{WT} (Fig. 4.4C, H1:N7~12). All 16 mutants giving higher fluorescence intensities were further investigated using batch-fermentation with FM-III medium in a 96 deep-well plate. It was noticed that the specific Trp production rate of 15 out

of the 16 candidates was significantly higher than that of the wild-type strain, while one of the candidates showed a decreased specific Trp production rate (Fig. 4.5). Overall, the increased MFU intensity and specific Trp production rate of the mutants demonstrated the selectivity and sensitivity of the screening platform. After evaluating the candidates, the mutated genes *AntrpC* from these candidates were isolated for sequencing. The sequencing results showed that the mutants with a higher Trp production rate had the same mutation Arg378Phe, whereas the mutation Arg378Ile in the A8 candidate resulted in a reduction of specific production rate (Fig. 4.5). To provide more direct evidence, an enzyme assay was performed on the purified protein of the enzyme variant *An*TrpC^{R378F}.

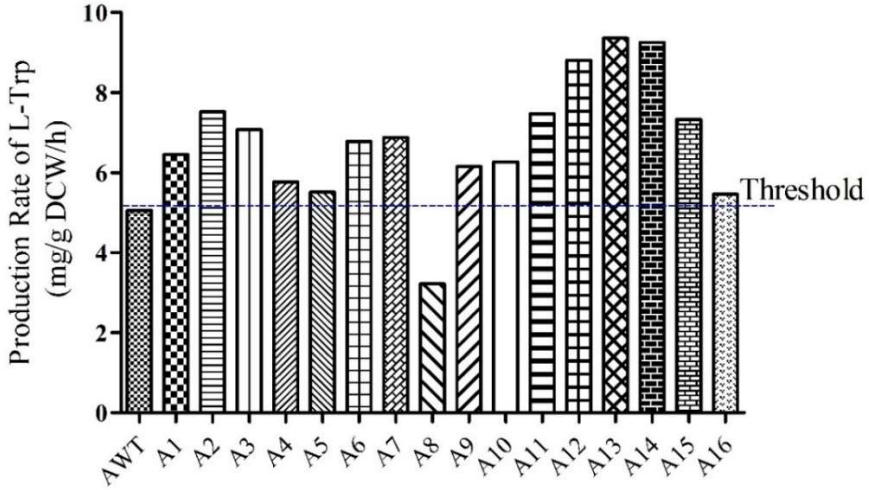

Figure 4.5: Results of the Trp production rate from all 16 candidates *An*TrpC^{R378X} as well as the wild-type strain. The specific production rate of the wild-type strain (5.05 mg/g DCW/h) was set as a screening threshold.

4.4.2 Characterization of the selected variant *An*TrpC^{R378F}

It was assumed that the higher specific production rate of Trp (q_{trp}) observed in the strain *An*TrpC^{R378F} indicated an enhancement in the anthranilate-activation of the corresponding enzyme variant *An*TrpC^{R378F}. To support this hypothesis, the strain *An*TrpC^{R378F} was cultivated at different extracellular concentrations of anthranilate, and the q_{trp} of the strain *An*TrpC^{R378F} under different conditions was determined. As observed in Fig. 4.6A, a scatter plot relating extracellular anthranilate to the q_{trp} shows a stronger linear correlation with $R^2 = 0.954$ for *An*TrpC^{R378F} when the concentrations of extracellular anthranilate are in the range of 0 to 0.5 mM. However, the q_{trp} of the strain remained constant with a further increase in the concentration of anthranilate (Fig. 4.6A). It was revealed that the q_{trp} depends mainly on the catalytic activity of variant *An*TrpC.

To characterize *An*TrpC^{R378F} *in vitro*, *An*TrpC^{WT} and *An*TrpC^{R378F} were selected for anthranilate

activation test. To this end, the enzyme activity was measured in a wide range of anthranilate concentrations up to 2.0 mM with 180 µM of substrate CdRP (Fig. 4.6B). In accordance with the previous study, the activity of *An*TrpC^{R378F} was also activated with the increasing concentration of anthranilate (range from 0 to 0.75 mM), but the activation effect was off when anthranilate concentration was over 1.0 mM (Fig. 4.6B). Interestingly, *An*TrpC^{R378F} exhibited a 1.67-fold higher specific activity than that of the wild-type in the absence of anthranilate but under 180 µM of substrate CdRP (Fig. 4.6B). Also, the enzyme activity of *An*TrpC^{R378F} was still around 1.68~2.88-times higher than that of the wild type at higher concentrations of anthranilate (range from 1.0 to 2.0 mM) (Fig. 4.6B). This result, therefore, indicated why the strain *An*TrpC^{R378F} has higher q_{trp} and MFU intensity than that achieved by the strain *An*TrpC^{WT}.

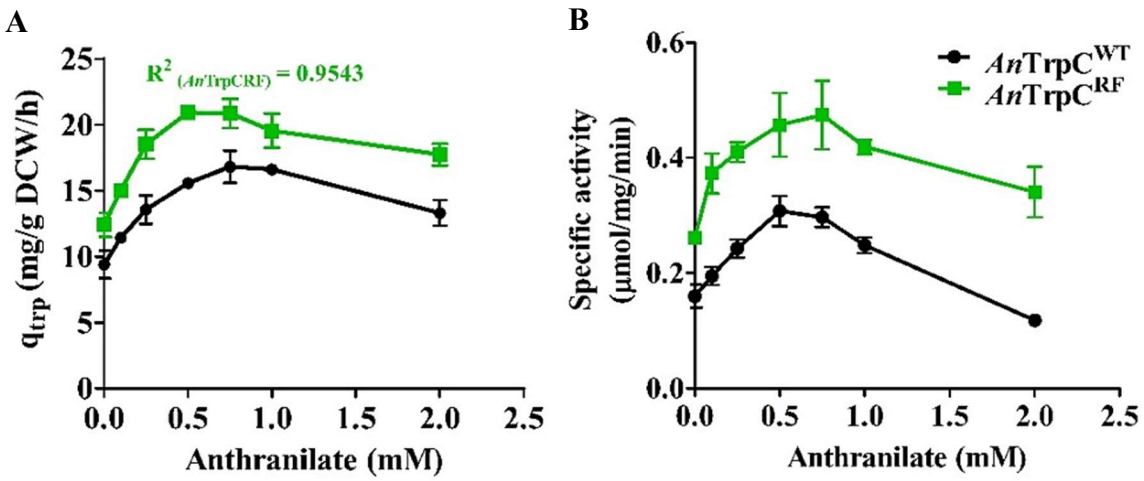

Figure 4.6: (A) A scatter plot relating extracellular anthranilate concentration to the Trp production rate of the strains *An*TrpC^{WT} and *An*TrpC^{RF}. (B) Effect of anthranilate on the specific activity of the variant *An*TrpC^{R378F} and the wild-type *An*TrpC enzyme. The concentration of substrate CdRP was 180 µM. The data were derived from three independent experiments.

4.4.3 Expression of the enzyme variant *An*TrpC^{R378F} in Trp producing strain

To investigate the impact of enzyme variant *An*TrpC^{R378F} on development of Trp-producing strain, the gene variant *AntrpC^{R378F}* and wild-type gene *AntrpC^{WT}* were individually integrated into the chromosome of the previously constructed Trp-producing strain S028 by substitution of the TrpR repressor (Fig. 2.2) with the aid of CRISPR/Cas9 technology, whereby the strains *An*TrpC^{R378F} and *An*TrpC^{WT} were generated. In practice, a pair of primers (*Cas9-N20-FOR* and *trpR-N20-REV*, Table 3.3) was initially used to amplify an all-in-one plasmid pRed-Cas9-*recA* backbone (Table 3.2) for introducing the target sgRNA sequence, from which the 20 bp spacer sequence specific for *trpR* gene was synthesized within the primers. To construct the donor dsDNA fragment, two homologous arms and the target gene sequence (*AntrpC*) were amplified separately and then fused by fusion PCR with the primers (*trpR-out-FF* and *trpR-out-RR*, Table

3.3). The gel purification of the fusion PCR fragment was integrated into the target sgRNA plasmid, resulting in two plasmids, pCas9-*AntrpC*^{WT} and pCas9-*AntrpC*^{R378F} (Fig. 4.7). Subsequently, both sgRNA plasmids were separately transferred into strain S028 to facilitate genome editing. Next, a comparison of fermentation parameters obtained by fed-batch fermentation of the strains *An*TrpC^{R378F}, *An*TrpC^{WT}, and S028 was performed in the part of this work.

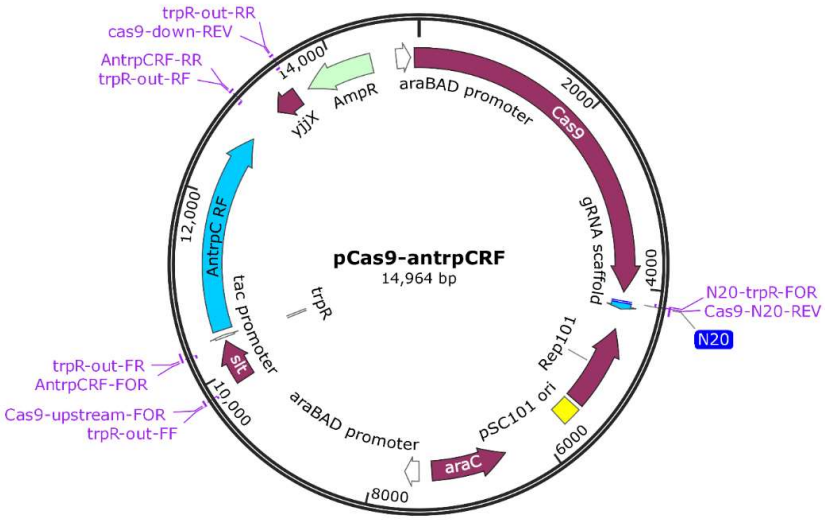

Figure 4.7: The plasmid pCas9-*AntrpC*^{R378F}. The donor DNA contained in this plasmid is composed of the whole gene variant *AntrpC*^{R378F}, the upper and the lower homologous recombination fragments.

Fig. 4.8C shows that fed-batch cultivation of both strains *An*TrpC^{R378F} and *An*TrpC^{WT} exhibited the same growth pattern before the post exponential phase (10-36h). Also, they had a significant growth rate than that of the strain S028 (data not shown). As predicted, a high formation of biomass could, therefore, directly lead to a high Trp production of the strains *An*TrpC^{R378F} and *An*TrpC^{WT} (Fig. 4.8A). It should be noted that *An*TrpC^{R378F} contains two copies of *trpC* genes, homologous *EctrpC* gene from *E. coli* and heterogeneous gene variant *AntrpC*^{R378F} form *A. niger*. The results from Figs. 4.8A and C indicate that, although both the strains *An*TrpC^{R378F} and *An*TrpC^{WT} had the same biomass formation (data not shown), *An*TrpC^{R378F} produced a higher amount of Trp than that of the reference strain *An*TrpC^{WT} during the log phase. Especially in the post log phase (in 30h), *An*TrpC^{R378F} produced 20.29±1.02 g/L of Trp, which is 45.34% higher than that of the reference strain (13.96±1.02 g/L). It was suggested that the increased Trp production of *An*TrpC^{R378F} is mainly due to the higher catalytic activity of the enzyme variant *An*TrpC^{R378F} rather than the increased gene copy number. Apart from this, during the mid-log phase (23-36h), the specific formation rate of Trp and the productivity of *An*TrpC^{R378F} were also remarkably higher than those of *An*TrpC^{WT} (Figs. 4.8E and F), e.g., 0.61±0.02 vs. 0.47±0.01 g/L/h for productivity and 22.25±1.32 vs. 19.61±0.32 mg/gDCW/h for

q_{trp} in 30h. An explanation is that anthranilate could well activate the catalytic efficiency of the variant *An*TrpC^{R378F}, and consequently enhanced the q_{trp} and productivity, as confirmed by the *in vitro* assay (Fig. 4.6).

It was also noticed that although the same amount of biomass (approx. 36.40±0.37 g/L, Table 4.2) and the same level of Trp yield (approx. 0.19 g/g, Table 4.2) were obtained in both strains *An*TrpC^{R378F} and *An*TrpC^{WT}, *An*TrpC^{R378F} formed a higher production of Trp (36.19±0.21 g/L, Table 4.2) at 51 h than that of *An*TrpC^{WT} (33.40±0.37 g/L, Table 4.2) because *An*TrpC^{R378F} exhibited high productivity of Trp during the stationary phase (Fig. 4.8F). It was further validated that the variant *An*TrpC^{R378F} is more efficient for Trp biosynthesis. Furthermore, the Trp yield of *An*TrpC^{R378F} was almost 1.27-fold higher than that of the parental strain S028 (Fig. 4.8D and Table 4.2), e.g., 0.19 g/g vs. 0.15 g/g in 54h, and it was consistent with the previous results that anthranilate-activated *An*TrpC enzyme contributes to the Trp yield (Chen et al., 2018). Notably, it was observed that, in comparison of *An*TrpC^{WT}, fed-batch fermentation of *An*TrpC^{R378F} resulted in the reduction of the production of SA and DSA (Figs. 4.8G and H). This result indicated that in *An*TrpC^{R378F}, more intermediates SA and DSA were redirected into the Trp pathway due to the higher catalytic efficiency of the variant *An*TrpC^{R378F}. Collectively protein engineering based on the PGSS approach was, therefore, more favorable to generate an anthranilate-activated variant *An*TrpC^{R378F}.

Table 4.2: Comparison of fermentation parameters obtained by fed-batch fermentation of S028 derivative strains

strains	OD_{600}	DCW (g/L)	GlcC* (g/L)	Trp (g/L)	q_{Trp} (mg/gDCW/h)	Yield (g Trp/g Glc.)	Vp (g/L/h)
S028	102.10±0.14	35.74±0.05	164.66±1.93	24.04±0.20	12.46±0.38	0.15	0.45±0.01
*An*TrpC^{WT}	106.30±0.42	37.21±0.05	176.72±2.76	33.40±0.37	16.62±0.13	0.19	0.62±0.01
*An*TrpC^{R378F}	101.40±1.97	35.49±0.69	195.62±1.61	36.19±0.21	18.88±0.16	0.19	0.67±0.01

The engineered Trp-producing strains were cultivated in a highly instrumented and automated 4-paralleled 1.5 L bioreactors system DASGIP for 54 hours; The initial glucose concentration was 30 g/L; The initial inoculation OD_{600} was 0.1. The data represents the mean ± SD from three measurements. GlcC* is the calculated cumulative consumption per reactor volume; q_{Trp}, specific production rate of Trp; Yield, Trp production vs. glucose consumed; Vp, volumetric productivity.

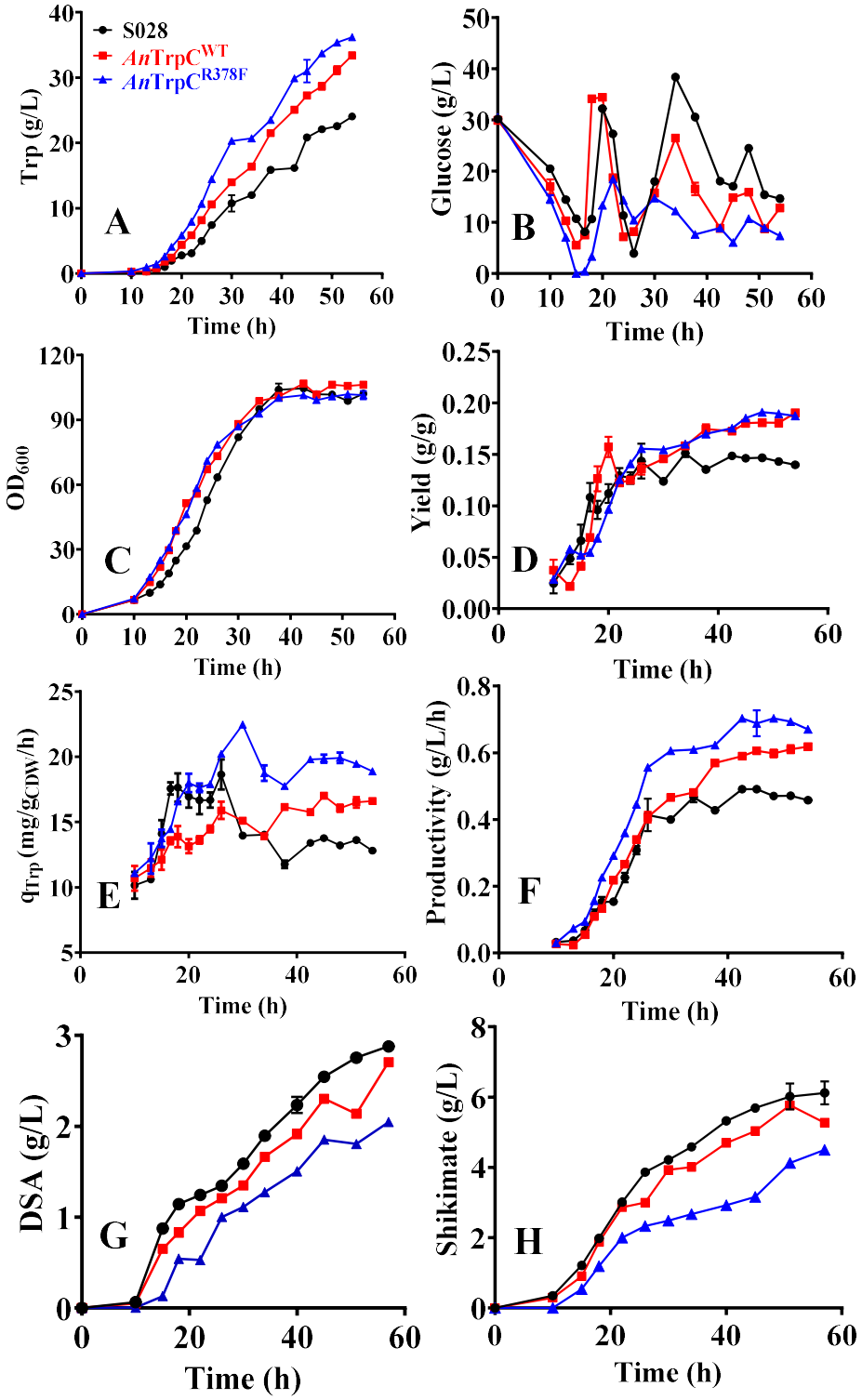

Figure 4.8: Fed-batch fermentation results of the strains S028 (Black and circle), *An*TrpC^{WT} (Red and square), and *An*TrpC^{RF} (Blue and triangle). (A) Trp concentration; (B) Residual glucose concentration; (C) Cell growth; (D) Overall Yield (g Trp/g glucose); (E) Specific formation rate of Trp (q_{Trp}); (F) Productivity; and Accumulation of the intermediates (G) dehydroshikimate (DSA) and (H) shikimate (SA).

4.5 Discussion

4.5.1 PGSS approach for generation of TrpC variants with desired performance

To date, a variety of approaches have been developed for screening and characterization of mutagenesis libraries, e.g., growth-coupled, colorimetric-based, and biosensor-driven high-throughput libraries screening (Arnold and Georgiou, 2003; Leemhuis et al., 2009; Dietrich et al., 2010). In the part of this study, an enzyme variant screening approach was proposed, namely Plasmid-assisted Growth-coupled and Sensor-guided *in vivo* Screening (PGSS). This approach combines the advantages of cellular auxotrophy and *in vivo* sensor-driven screening of enzyme variants and provides a higher threshold valve for identification of the feed-forward-resistant *Ec*TrpC variants and the feed-forward-activated *An*TrpC variants with higher catalytic efficiency.

In a previous study, *in vitro* assay revealed that approx. 60% of the activity of the enzyme *Ec*TrpC was inhibited under 1 mM anthranilate (Chen et al., 2018). After site-directed mutagenesis of *Ec*TrpC, the experimental results revealed that the anthranilate sensitivity of *Ec*TrpC variants (e.g., I8V, S60A, S60G, and L188F) was significantly reduced. However, concomitantly, the catalytic efficiency of these enzyme variants was remarkably decreased (Chen, 2017). An approach based on PGSS-assisted protein engineering for generation of anthranilate-resistant *Ec*TrpC variants with higher catalytic efficiency was therefore proposed in this study. Similarly, an enzyme variant *Ec*TrpC^{QVFQ} identified by the PGSS approach exhibited a stronger resistance to anthranilate compared to that of the enzyme variant *Ec*TrpC^{S60A} but also showed a negligible decrease in the activity of *Ec*TrpC^{QVFQ} (Fig. 4.3B). Nevertheless, since *Ec*TrpC^{QVFQ} had a similar catalytic efficiency (k_{cat}/k_m) to that of the wild-type *Ec*TrpC (Table 4.1), this variant could, therefore, offer a higher tolerance resistance to the higher enrichment of anthranilate when the fermentation is scaled up to the industrial level.

On the contrary, the specific activity of the enzyme *An*TrpC was determined to be 0.19± 0.04 µmol/min/mg (Fig. 4.6B), which is about 13 times slower than that of the wild-type enzyme *Ec*TrpC (2.54 ± 0.09 µmol/min/mg). Also, as observed in Fig. 4.6B, the specific production rate of Trp in this study showed a stronger linear correlation with the activity of the enzyme *An*TrpC. To increase the catalytic efficiency of *An*TrpC, PGSS-based protein engineering was thus employed to develop enzyme *An*TrpC. In theory, the anthranilate-activated enzyme *An*TrpC could, to some extent, lead to suffering no evolutionary inducers such as inhibitors during the engineering of the enzyme variants. To address this limitation, two criteria, the growth rate and the signal strength of biosensors, were introduced to create a stringent threshold

valve. The specific production rate of Trp was also adopted as a further criterion when characterizing the strains *An*TrpC^{R378X}. Using the PGSS approach, a novel anthranilate-activated enzyme variant *An*TrpC^{R378F} was successfully identified, and it exhibited a higher specific activity than that of the reference enzyme *An*TrpC^{WT} in the presence of 0.5 mM anthranilate (Fig. 4.6B). Significantly, replacing the enzyme *An*TrpC^{WT} with the newly identified enzyme variant *An*TrpC^{R378F} in the Trp producing strain improved the Trp production by 13.51% (36.19±0.21g/L) in a simple fed-batch fermentation (Fig. 4.8A and Table 4.2). As shown in Fig. 4.4B and Fig. 4.9, in addition to the residues R378, other residues such as T454-S455-F456, C477-A478-L479, and E503-A504, are likely to be exposed to the binding site of the substrate CdRP with the enzyme *An*TrpC. It is, therefore, interesting to carry out the random mutagenesis at these target binding sites and explore more desired candidates with higher catalytic efficiency by the PGSS approach.

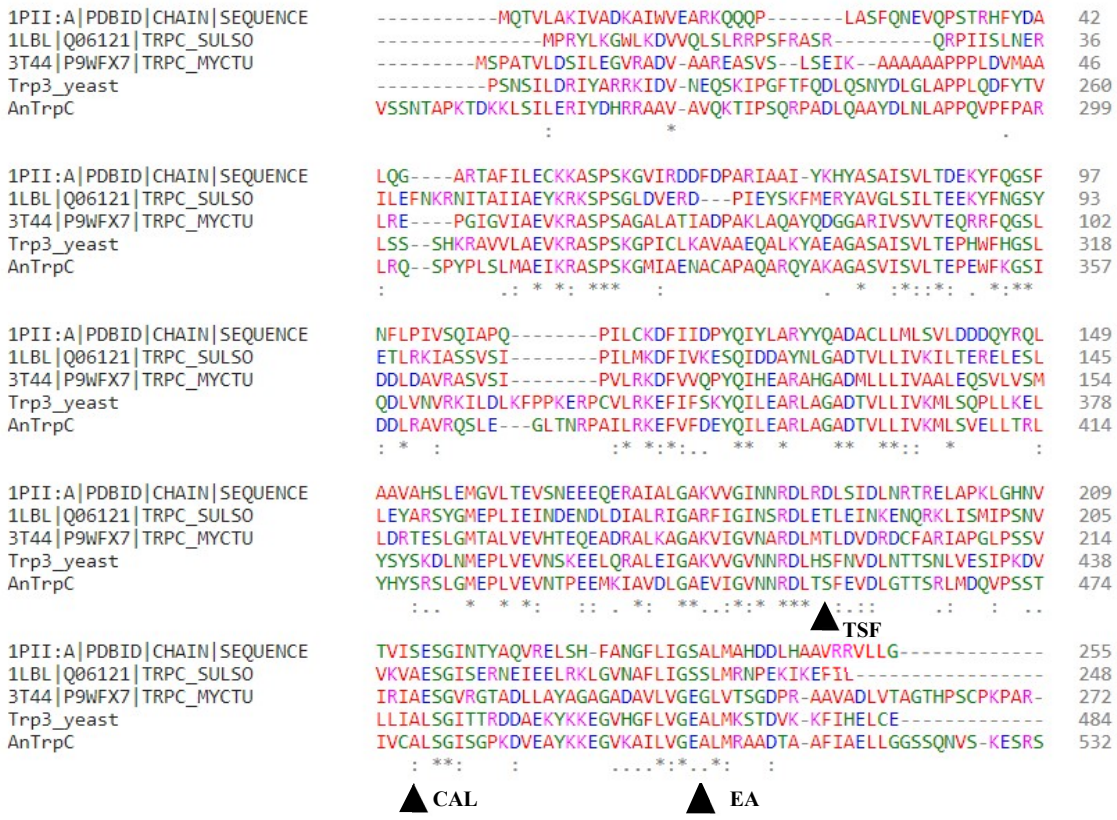

Figure 4.9 Alignment of TrpC sequences from *E. coli* (P00909), *Sulfolous solfatarious* (Q06121), *M. tuberculosis* (P9WFX7), *A. niger* (P05328), and *S. cerevisiae* (P00937). The residues of the conserved binding sites of substrate CdRP with the enzyme *An*TrpC are highlighted in a black triangle and capital letter.

4.5.2 Further prospects and optimizations of the PGSS approach

In this chapter, a plasmid-based approach to growth-coupled screening and sensor-guided characterization (PGSS) was developed and successfully applied to develop *Ec*TrpC and *An*TrpC enzymes with desired performances. However, PGSS approach still undergoes some

shortcomings: (1) the proposed threshold valve of this approach is insufficient. The experiments from this chapter indicated that some false-positive variants were selected during library screening, e.g., *Ec*TrpC^{QLRC}, *Ec*TrpC^{ACFR}, and *An*TrpC^{R378I}. This limitation, especially for target enzymes such as the enzyme *An*TrpC that are not subjected to evolutionary pressure, is more likely to be resolved with the help of *in vivo* continuous evolution. With the continuous evolution system, undesired candidates are diluted out, and the desired candidates are remained in the culture vessel because of the "Survival of the fittest". In theory, the efficiency of *in vivo* continuous evolution was supported by three major aspects (Tan et al., 2019):

1. Endogenous mutators, i.e., *dnaQ926* gene, a dominant-negative variant of DNA Pol III proofreading domain;
2. Fitness-coupled continuous evolutionary pressure;
3. Environment maintaining continuous culture, e.g., pH, especially for the pH-dependent Trp-producing strains.

(2) This approach largely depends on auxotrophic strains and also on the availability of biosensors. The majority of biomolecules, particularly the intermediates from the amino acid synthesis pathway, are not essential for cell growth. Accumulation of the intermediates even appears to be toxic to the cells. In this regard, a heterologous pathway can be introduced to transform the non-essential biomolecules into the molecules that are required for cell growth. For instance, chorismate, one of the intermediates from the aromatic pathway, has been widely used as a feedstock for the synthesis of the value-added compounds such as phenol, *p*-aminobenzoic acid, 2,3-hydroxybenzoic acid (Huccetogullari et al., 2019b). It is therefore assumed to convert the chorismate to the cell essential aromatic amino acid such as Trp by introducing Trp biosynthetic pathway. The concentration of Trp thus can be used as an indirect indicator for the evaluation of the engineered chorismate pathway. The PGSS approach also relies on the availability of biosensors. Although many natural or synthesized amino acids-responsive biosensors, e.g., tryptophan (Fang et al., 2016), arginine (Verma et al., 2017; Whitfield et al., 2015), lysine (Eggeling et al., 2015; Zhou and Zeng, 2015), glutamate (Wendisch et al., 2016), and branch amino acids (Mustafi et al., 2012)-responsive biosensors, have been widely applied in the protein engineering, other amino acids and also their derivatives-responsive biosensors are still rarely explored. Thus, it is imperative to develop other biosensors to the extent of the application of the PGSS approach. The dynamic range and specificity of biosensors must also be taken into account.

(3) This approach requires further integration of gene variants at the genomic scale to assess target candidates. Since the so-called desired enzyme variants are generated from the plasmid-

based variant library, the characterization results, e.g., the specific production rate of Trp, from this approach are not relevant to the intracellular environmental conditions of a host strain that is to be used for bio-production process. The instability of plasmid can affect the cell growth and provide false-positive candidates during library screening. Therefore, further time-consuming integration of gene variants on a genome level is required to evaluate target candidates. Recently, the CRISPR/Cas9 technique has been developed to become an ideal genome-editing tool to quickly and effectively integrate the target gene into the chromosome of a production strain. Thus, coupling the CRISPR/Cas9 technique with the PGSS approach will be a better alternative to overcome the restrictions mentioned above. With the assistance of the CRISPR/Cas9 technique, the gene variants can be directly integrated into the chromosome of a host strain for further *in vivo* characterization.

Chapter 5

5. CRISPR/Cas9-facilitated engineering with growth-coupled and sensor-guided *in vivo* screening of AroG enzyme variants[2]

5.1 Introduction

Developing an efficient *in vivo* library screening approach for target identification is one of the primary tasks of protein engineering and pathway optimization. In previous work, an enzyme variant library screening and selection approach, namely the PGSS approach, was developed for the identification of TrpC variants, *Ec*TrpC$^{S58Q\text{-}P59V\text{-}S60F\text{-}K61Q}$ and *An*TrpC^{R378F}. However, insufficient phenotypic characterization of this approach makes it challenging to identify the best performer among the improved enzyme variants. The subsequent *in vitro* and *in vivo* characterizations of the best performer identified from the screening may not apply to a real bio-production process with the host microorganism. For these reasons, it is interesting to develop an approach that combines the integration of gene variants of a targeted enzyme directly into the chromosome of the host microorganism. In this way, the *in vivo* screening and characterization of the enzyme variants are carried out under conditions that are more relevant to the cultivation and intracellular environments of a host strain.

Recently, the CRISPR/Cas9 technology for genome editing has made significant advances and received extensive attention (Cho et al., 2018; Zhang et al., 2018). Among others, it is applied for the engineering of microbial production strains (Jiang et al., 2015; Schuster et al., 2019). Because of its simplicity and efficiency, CRISPR/Cas9 is an ideal genome-editing tool for effectively integrating gene variants of a target enzyme into the chromosome of a production strain (Guo et al., 2018b). To cope with this technology, a reliable *in vivo* screening method based on the PGSS approach is desirable to link the change of interest caused by an induced mutation on a target gene to the change in cell growth (Lu et al., 2012; Zhu et al., 2017) and the expression strength of a reporter gene via a biosensor (Binder et al., 2013; Fang et al., 2016).

[2] The major results of this chapter have been published in Chen, M., Chen, L., & Zeng, A. P. (2019). CRISPR/Cas9-facilitated engineering with growth-coupled and sensor-guided *in vivo* screening of enzyme variants for a more efficient chorismate pathway in *E. coli*. Metab Eng Commun, 9, e00094.

In this study, a method of linking CRISPR/Cas9-facilitated engineering with growth-coupled and sensor-guided *in vivo* screening and characterization (CGSS) is proposed for protein engineering (Fig. 5.1). Generally, using CGSS, it is able to effectively evaluate the variants of the target gene with the relatively low and genetically stable expression in the chromosome of a particular host under relatively real production conditions. *In vitro* constructed mutation library of the target gene is integrated into the chromosome by lambda-Red recombination system with the facilitation of the CRISPR/Cas9 technology for higher efficiency. To make the mutation of the target gene position-regardless, an expression cassette of a marker gene, usually an antibiotic-resistant gene, is used to replace the whole endogenous target gene and to offer sgRNA targeting sites (Fig. 5.1). To maximize the biological effectiveness and reliability of this method, the special host is (a) preferably derived from a target chemical producer, and (b) its growth is somehow linked to the change of the target gene, which is also directly/indirectly controlled the expression of a reporter through a biosensor. In short, positive mutants will be selected according to the growth rate and the expression of the reporter gene at first and further evaluated in a real production process according to the performance of the corresponding strains.

To demonstrate this method, the optimization of a vital enzyme of the chorismate pathway for the biosynthesis of L-tryptophan was performed as an example. 3-Deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase is a key enzyme for engineering microbes for efficient biosynthesis of aromatic amino acids (AAAs): tryptophan (Trp), phenylalanine (Phe), and tyrosine (Tyr) (Chen and Zeng, 2017; Wu et al., 2018). It is under feedback inhibition by the end products (Ogino et al., 1982; Sprenger, 2006). For instance, in *E. coli*, all the wild-types of DAHP synthase, encoded by the genes *aroG*, *aroF*, and *aroH*, are subject to feedback inhibition by Phe, Tyr, and Trp, respectively (McCandliss et al., 1978; Schoner and Herrmann, 1976). Therefore, the engineering of feedback-resistant DAHP synthase is required for constructing efficient pathways for producing AAAs and their derivatives (Sprenger, 2006). In this study, CGSS is proposed and demonstrated for semi-rationally engineering and screening feedback-resistant DAHP synthase to obtain a more efficient chorismate pathway for Trp production in *E. coli* (Fig. 5.1).

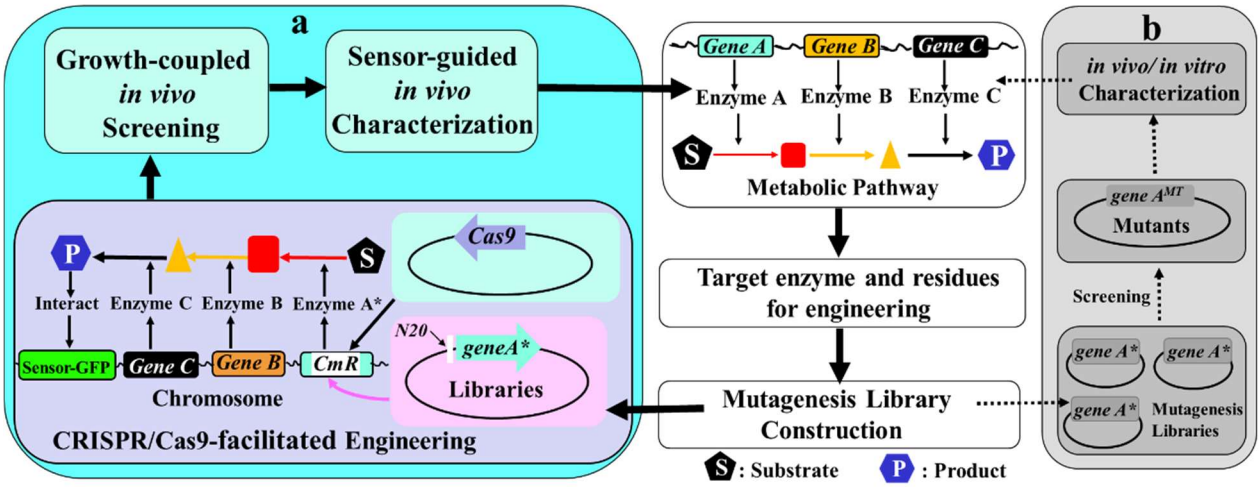

Figure 5.1: (A) CRISPR/Cas9-facilitated engineering of gene variants integrated with growth-coupled and sensor-guided *in vivo* screening and characterization of candidate enzyme(s) as a novel approach (CGSS) for protein and pathway engineering. (B) Conventional screening and characterization approach of protein engineering. The middle part illustrates the enzyme(s) in pathway engineering and the identification of key enzyme(s) and residues for the construction of the mutagenesis library. The two approaches share it. The significant advantages of CGSS are the integration of the gene variants (genotype) directly into the chromosome of the production strain and *in vivo* evaluation (screening and characterization) of the variants through the phenotype such as cell growth rate and concentration of the relative metabolite using a biomolecular sensor. It can be done under cultivation conditions related to the real conditions of the engineered enzyme and the corresponding production strain. The conventional approach involves expression typically (e.g., using plasmid) of the mutagenesis library in a host differently from the production strain, *in vitro* and *in vivo* screening and characterization under conditions which have little to do with the real culture and intracellular environment of the production strain.

5.2 Proof-of-concept of the CGSS method

For proof-of-concept of the proposed CGSS method, a screening strain that contains a Trp biosensor and lacks DAHP synthase activity was first constructed for screening variants of DAHP synthase. To this end, the genes *aroH* and *aroF* in our previously developed Trp-producing *E.coli* strain S028λ (Chen and Zeng, 2017) were knocked out. Deleting the $aroG^{S180F}$ from this mutant made the strain auxotrophic for aromatic amino acids (AAAs). Consequently, the growth of the mutant is linked to the DAHP synthase activity upon its re-introduction (Fig. 5.2). In principle, an engineered DAHP synthase with a higher activity should lead to a faster accumulation of Trp, which in turn stimulates the expression of a report gene regulated by the Trp biosensor (Fang et al., 2016).

5.2.1 Generation of aromatic amino acids-deficient strain

The Trp biosensor is composed of *tnaC*, which encodes the leader sequence of *tnaCAB* operon (Bischoff et al., 2014); the eGFP protein was fused to the upstream of *tnaC*. In order to avoid causing instability by using too many plasmids, the Trp-biosensor was integrated into the chromosome at the locus of the gene *trpR*, which led to the strain of WS001. To do so, the plasmid pN20-*trpR* was constructed from pTagAmpR plasmid (Fig. 2.7B) with the primers

pTargR and *TrpR-N20* (Table 3.3) for expressing gRNA targeting the *trpR* gene. The donor DNA fragment *Trp-Sensor* was amplified from the plasmid pSentrp (Fang M et al., 2017) with the primers *Tsen-trpR-IF* and *Tsen-trpR-IR* (Table 3.3).

In order to construct aromatic amino acids (AAA)-auxotrophic strain (WS002), the sole DAHP synthase of strain WS001 was removed by replacing the gene $aroG^{S180F}$ with the antibiotic resistance gene Cm^R which offers sgRNA targets for the CRISPR/Cas9 technology in further genome-editing, generating the strain WS002 (Fig. 5.2). In this regard, the plasmid pN20-*aroG* (Fig. 5.3) was constructed at first from the plasmid pTagAmpR plasmid with the primers *pTargR* and *AroG-N20* (Table 3.3) for expressing gRNA targeting the gene *aroG*. The donor DNA fragment $P_{J23119\text{-}rpsL}$-*cmR* was amplified from the plasmid pJLC (Table 3.2) with the primers *up-aroH-out* and *Cm-delG-R* (Table 3.3).

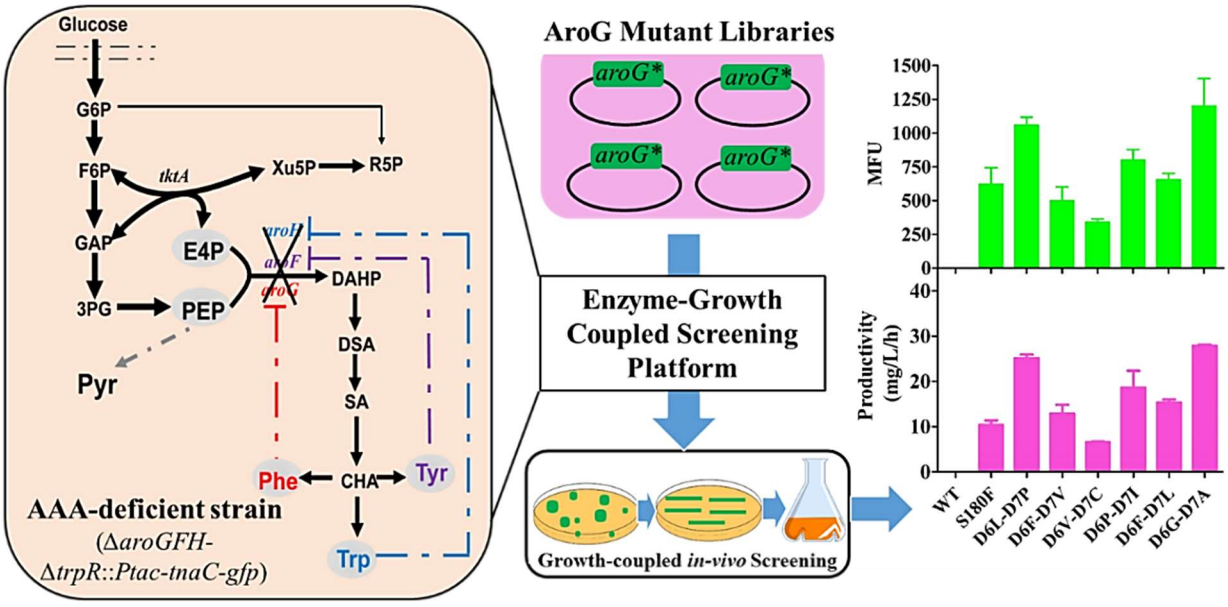

Figure 5.2: Design and implementation of CGSS in screening and characterization of feedback-resistant AroG ($AroG^{fbr}$) enzyme variants. In *E. coli*, DAHP synthase (AroG, AroF, and AroH) is a key rate-limiting enzyme of the pathways for aromatic amino acids (AAAs) biosynthesis. An AAAs-auxotrophic strain (strain WS002) was constructed by disrupting the DAHP synthase and used as a platform for screening *aroG* gene variants, which were individually integrated into the chromosome of *E. coli* using the CRISPR/Cas9 technology. In the presence of a high Phe concentration, only strains that express $AroG^{fbr}$ with good resistance to Phe can produce enough AAAs sustaining cell growth. These strains were further characterized using the strength of the fluorescence intensity (medium fluorescence unit, MFU) of Trp biosensor (P_{tac}-TnaC-eGFP) representing the productivity of Trp.

5.2.2 Characterization of the aromatic amino acids-auxotrophic strain

As expected, the auxotrophic strain WS002 was not able to grow in the M9 medium without the addition of any of the aromatic amino acids Phe, Tyr, and Trp (data not shown). Then, we tested if the strain with feedback resistant DAHP synthase will behave differently from the one

having the wild-type DAHP synthase in terms of cell growth and expression of the reporter gene in a defined medium containing a high concentration of enzyme inhibitor.

To this end, the wild-type gene *aroG* and the feedback-resistant gene $aroG^{S180F}$ (Ger et al., 1994) were introduced into the chromosome of the strain WS002 at the locus of the gene *CmR* using the CRISPR/Cas9 technique with the plasmids pCm-$aroG^{WT}$ and pCm-$aroG^{S180F}$ (Fig. 5.3), respectively. To construct pCm-$aroG^{WT}$ (Fig. 5.3), the plasmid pN20-*CmR* was constructed at first with the primers *pTargR* and *Cm-N20* (Table 3.3) by using the template pTagAmpR (Table 2.2). Afterward, the DNA fragments V-*N20CmR*, *F-CmR*, *F-aroG*, and *F-serA* were generated from the plasmid pN20-*CmR*, the plasmid pJLC, the genome DNA of *E.coli* W3110, and the plasmid strp015A (Chen and Zeng 2017), respectively. Then four fragments (*V-N20CmR*, *F-CmR*, *F-aroG*, and *F-serA*) were fused to construct the final plasmid pCm-$aroG^{WT}$ by using In-Fusion HD Cloning kits (Clontech® Laboratories, Inc.). The plasmid of pCm-$aroG^{S180F}$ was constructed using mutagenic primers (Table 3.3) to amplify the whole plasmid pCm-$aroG^{WT}$. The recombinants were spread on the M9-agar medium containing 25 mM Phe without Tyr and Trp. 0.1 mM IPTG was also supplemented into the medium for the following reasons. The first one is to repeal LacI-repression and thus induce the expression of genes for the *trp* biosynthetic pathway since it is regulated by the regulator LacI. The other is to induce the expression of sgRNA which guides Cas9 to cut the donor plasmids from which the gene of interest can also be expressed.

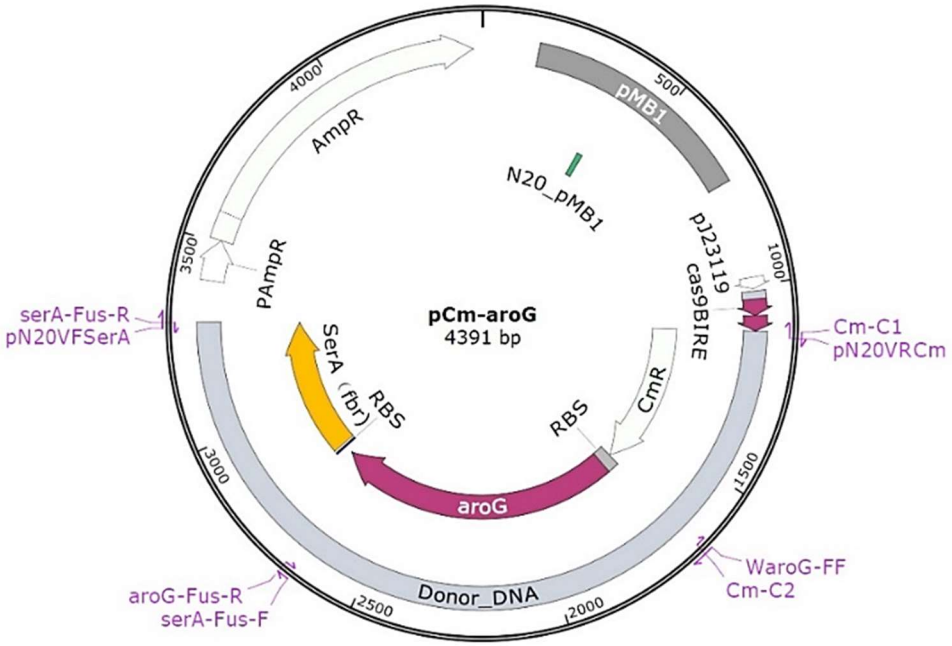

Figure 5.3: The plasmid pCm-*aroG*. The donor DNA contained in this plasmid is composed of a part of the gene Cm^R, the whole wildtype gene *aroG*, and a part of the gene *serA*.

The results showed that after introducing the gene *aroG*^{S180F} into the host, many recombinants grow up with a strong fluorescence intensity under the conditions mentioned above (Fig. 5.4). Several colonies were selected for further characterization. It turned out that these colonies all have the same mutation S180F. These recombinants were designated as WS004. No colony was observed when the wildtype gene *aroG* was integrated with the plasmid pCm-*aroG*^{WT} under the same conditions (Fig. 5.4). Presumably, the activity of the wild-type AroG is severely inhibited by Phe, resulting in the block of the biosynthesis of AAAs and leading to no growth. It could be, however, also possible that the recombineering efficiency is too low. To eliminate the latter possibility, the recombinants with the gene *aroG*^{WT} were also grown on LB-agar medium with IPTG. From the complex medium, many colonies were obtained (Fig. 5.4). They were further confirmed by colony PCR and designated as WS003. The strain WS003 and WS004 were re-checked on M9-agar medium with and without 25 mM Phe (Fig. 5.4). The strain WS003 was found to grow on the medium without Phe, but no growth was observed on the medium with Phe (Fig. 5.4). As expected, the growth of the strain WS004 did not show a notable difference in the media with or without Phe. These results suggested that CGSS is an approach useful to facilitate the engineering of enzymes with desired performance, such as a higher activity and a higher inhibitor tolerance. It is thus used in the following to obtain AroG variants with further improved tolerance against Phe.

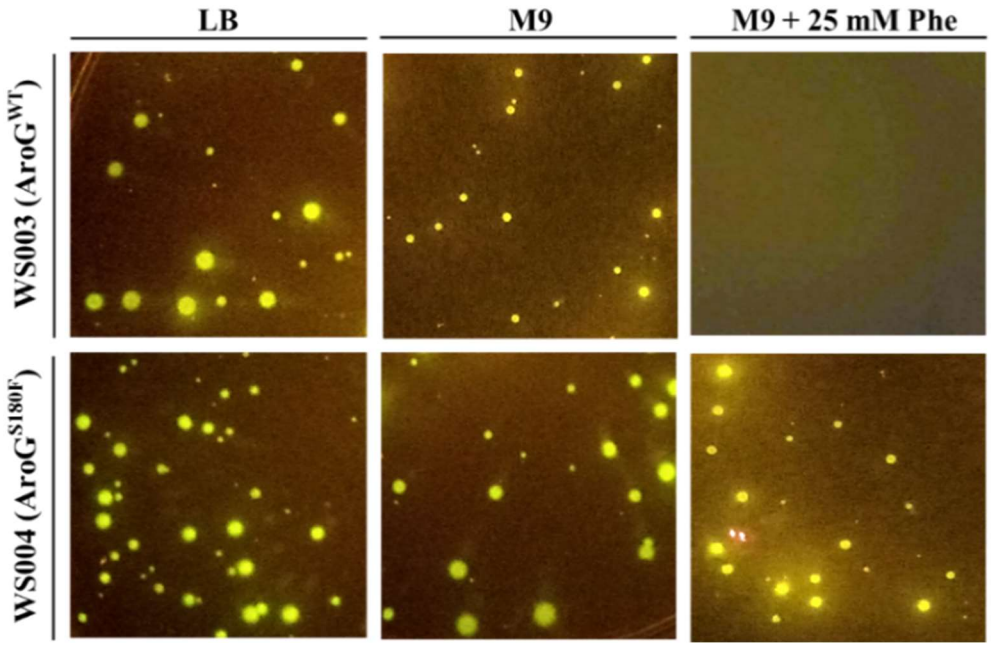

Figure 5.4: Comparison of the growth and fluorescence of the two strains WS003 and WS004 generated by introducing *aroG*^{WT} and *aroG*^{S180F}, respectively, into the chromosome of strain WS002 under different growth conditions. The fluorescence intensity of colonies grown on the agar plate was detected under UV light. Left, complex medium (LB-agar); middle, M9-agar (without any amino acids); right, M9-agar with 25 mM of Phe.

5.3 CGSS for screening Phe-resistant AroG variants

To demonstrate the usefulness of the CGSS method established above for obtaining more resistant AroG variants, a mutation library of AroG, was first generated. For this purpose, a semi-rational strategy was adopted, which makes use of information from the crystal structure of AroG complexed with its inhibitor Phe (PDB: 1KFL) (Fig. 5.5A). The residues D6 and D7 involved in the binding of Phe were selected as targets to perform saturation mutagenesis. To construct pCm-*aroG*^{fbr} (Fig. 5.3), the desired plasmid containing gene variants was constructed using mutagenic primers (Table 3.3) to amplify the whole plasmid pCm-*aroG*^{WT}. Finally, the gene variants were integrated into the chromosome of auxotrophic strain WS002, and then the mutants were screened with the CGSS.

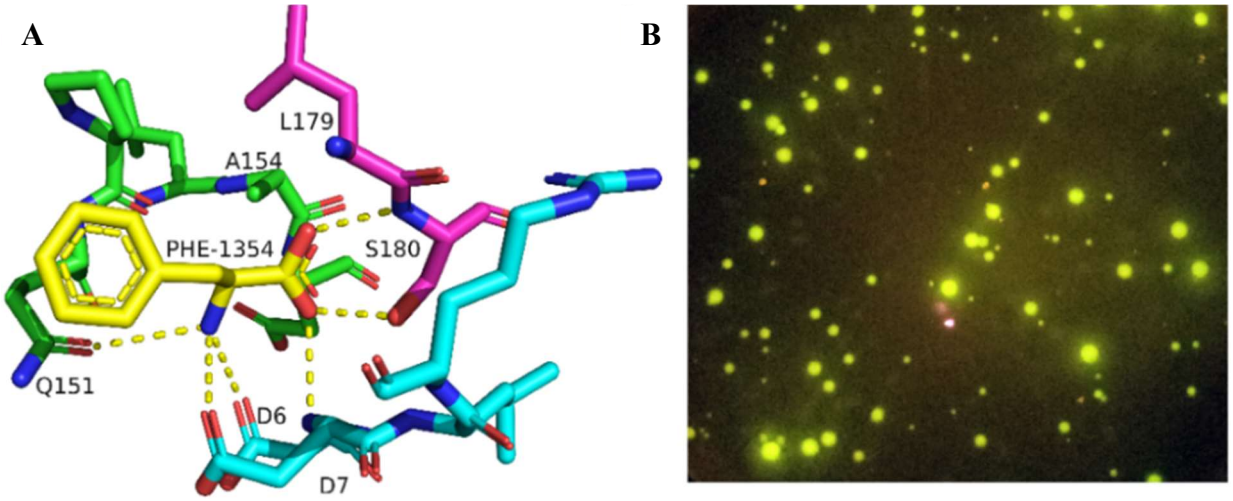

Figure 5.5: (A) Key residues involved in the Phe binding sites of AroG from *E. coli*. (B) The fluorescence induction for the variant AroG $^{D6X-D7X}$ on reduced M9-agar (without Tyr and Trp) with the addition of 25 mM Phe and 0.1 mM IPTG.

It was noteworthy that the procedures of screening and selection of AroG^{fbr} variants were similar to the PGSS approach, in which the variant AroG $^{D6X-D7X}$ that had a relatively a bigger size and a higher fluorescence intensity (Fig. 5.5B) was forced to further *in vivo* characterization. After confirmation of the phenotypes, the mutated genes *aroG* from 20 candidates were isolated for sequencing. The sequencing results showed that there are only 6 different variants of AroG among the 20 candidates (Table 5.1). They are AroG$^{D6G-D7A}$, AroG$^{D6L-D7P}$, AroG$^{D6P-D7I}$, AroG$^{D6F-D7V}$, AroG$^{D6V-D7C}$, and AroG$^{D6F-D7L}$, with the number of occurrences being 7, 6, 4, 1, 1, and 1, respectively (Table 5.1). Batch-fermentations were then carried out in FM-III medium in 50 mL conical tubes with the strains carrying these 6 recombinants and compared them with the wild-type strain WS003 and the strain WS004 with the variant AroG^{S180F}. As shown in

Table 4.1, 19 of the mutants had a higher Trp production compared to the strain WS004. It suggested that the efficiency of generated positive mutants is remarkable. Moreover, it was found that the strains carrying the variant AroG$^{D6G-D7A}$, AroG$^{D6L-D7P}$, or AroG$^{D6P-D7I}$, which had the higher frequency of occurrence in the 20 candidates (Table 5.1), had also a higher Trp productivity (Fig. 5.2). Notably, the first two mutants exhibited much higher productivity than the reference strain (AroG^{S180F}). Also, the relationship between Trp productivity and the strength of fluorescence intensity was investigated. The strain with a stronger fluorescence intensity also has a higher Trp productivity (Fig. 5.2). These results suggested that the variants AroG$^{D6G-D7A}$, AroG$^{D6L-D7V}$, and AroG$^{D6P-D7I}$ have a higher catalytic efficiency than the variant AroG^{S180F} under the test conditions. To provide more direct evidence, enzyme assay was done with the purified protein of these variants.

Table 5.1 Comparison of fermentation results with *E. coli* strains containing the variants AroG^{WT}, AroG^{S180F}, and AroG^{fbr} grown on FM-III with 25 mM of Phe.

ID	Mutated residues	Number of occurrence[a]	DCW (g/L)	Trp (g/L)
N1	WT	--	0.786±0.145	N. D.
N2	S180F	--	0.714±0.045	0.248±0.025
N3	D6L-D7P	6	0.983±0.118	0.604±0.018
N4	D6F-D7V	1	0.987±0.007	0.309±0.047
N5	D6V-D7C	1	1.071±0.094	0.157±0.004
N6	D6P-D7I	4	1.001±0.009	0.448±0.089
N7	D6F-D7L	1	1.015±0.156	0.368±0.017
N8	D6G-D7A	7	1.102±0.105	0.670±0.005

[a] "Number of variants" refers to all the 20 candidates examined; The fermentations were performed with a single clone for each AroG mutant; N. D., not detectable. The average value ± standard deviation is based on three independent experiments.

5.4 Characterization of selected AroG^{fbr} variants *in vitro*

To examine if the higher Trp productivity and stronger fluorescence intensity observed in the strains is due to improved Phe-tolerance of the corresponding AroG variants, the inhibition behavior of the variants AroG$^{D6G-D7A}$, AroG$^{D6L-D7P}$, and AroG$^{D6P-D7I}$ was investigated.

As shown in Fig. 5.6, all variants are significantly less sensitive to the inhibitor Phe, while the wild-type AroG is extremely sensitive to it. In the presence of 0.5 mM Phe, the wild-type AroG almost completely lost its activity, whereas all the variants retained more than 80% of their

activities under the same conditions. Compared to $AroG^{S180F}$, all three variants $AroG^{D6G-D7A}$, $AroG^{D6L-D7P}$, and $AroG^{D6P-D7I}$ have a weaker sensitivity to Phe when the concentration of Phe was higher than 10 mM (Fig. 5.6B). They also exhibited higher specific activities in the presence of more than 20 mM Phe (Fig. 5.6A). These results explain why the strains with these three variants performed better than the positive control in terms of the Trp production when they were cultivated in medium containing a very high Phe concentration (Table 5.1). As shown in Fig. 5.6A, the specific activities are significantly different between these three variants generated by the substitution of the same residues D6 and D7. Among them, the variant $AroG^{D6G-D7A}$ had the highest specific activity, which is nearly twice as high as that of the variant $AroG^{D6P-D7I}$ regardless of the Phe concentration. Besides, several feedback-resistant variants have been published so far, e.g., D146N, L175D, P150L, S180F, and A202T/D146N, but these enzyme variants were reported to be less sensitive to the lower concentration of Phe (1 mM), and the further information on how these variants act against the higher concentration of Phe (over 10 mM) is still rarely investigated. Among them, one of the reference enzyme variants S180F (Ger et al., 1994) has been applied by many engineers on an industrial scale for AAA biosynthesis because only 10% of enzyme activity is inhibited in the presence of 20 mM Phe (Chen and Zeng, 2017; Chen et al., 2018). Beyond our expectations, the performance of the variant S180F in this study was not consistent with the reported one, where 65% of the activity of S180F was lost at the same concentration of Phe, and in contrast, only 45% of the activity of $AroG^{D6G-D7A}$ was inhibited (Fig. 5.6B). These results suggested that the variant $AroG^{D6G-D7A}$ is more efficient for the chorismate pathway for the bio-production of aromatic amino acids and their derivatives.

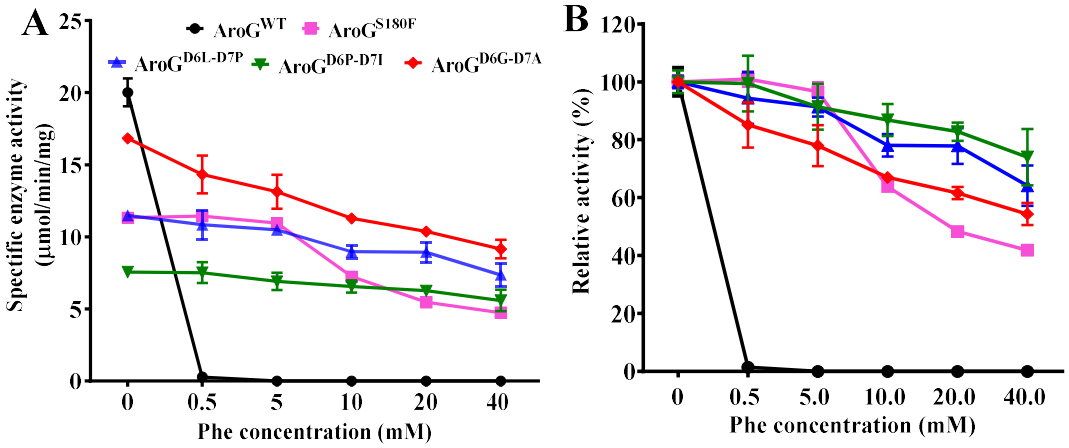

Figure 5.6: Effects of Phe on the activity of enzyme $AroG^{WT}$ and its variants $AroG^{S180F}$, $AroG^{D6L-D7P}$, $AroG^{D6P-D7I}$, and $AroG^{D6G-D7A}$. (A) Specific activities; (B) Relative activities. Results were derived from three independent experiments.

5.5 Improvement of the chorismate pathway and Trp biosynthesis

To explore the impact of our best variant AroG$^{D6G-D7A}$ on strain development for the biosynthesis of aromatic amino acids, the reference gene variant *aroG*^{S180F} in strain S028 was substituted with gene variant *aroG*$^{D6G-D7A}$, whereby the strain S028GM1 was generated. In practice, the sgRNA plasmid p*N202-CmR* was first constructed using the same strategy as other plasmids did (Fig. 2.7) with primers *pTargR* and *CmR-N202* (Table 3.3). The donor DNA fragment $P_{J23119\text{-}rpsL\text{-}tac}$-*aroG*$^{D6G-D7A}$ was obtained from two rounds of PCR. The first round was to add the *tac* promoter to the gene *aroG* $^{D6G-D7A}$ by amplifying the plasmid pET-AroG $^{D6G-D7A}$ with the primers *NotI-pTac-aroG* and *aroG-speI* (Table 3.3). The second round was to flank the upstream homologous arm using the primers *u-rpsLp-tac* and *aroG-fus-R* and the PCR product from the first round as the template to replace the promoter of $P_{J23119\text{-}rpsL\text{-}cmR}$-*aroG*$^{D6G-D7A}$ by the same one with S028 ($P_{J23119\text{-}rpsL\text{-}tac}$-*aroG*^{S180F}). The plasmid pN202-CmR and the fragment $P_{J23119\text{-}rpsL\text{-}tac}$-*aroG*$^{D6G-D7A}$ were co-transformed into the strain WS005/pCas9 to generate the strain S028GM1. The capacity of Trp production of strain S028GM1 was compared to that of the reference strain S028 by performing a simple fed-batch fermentation in bioreactors.

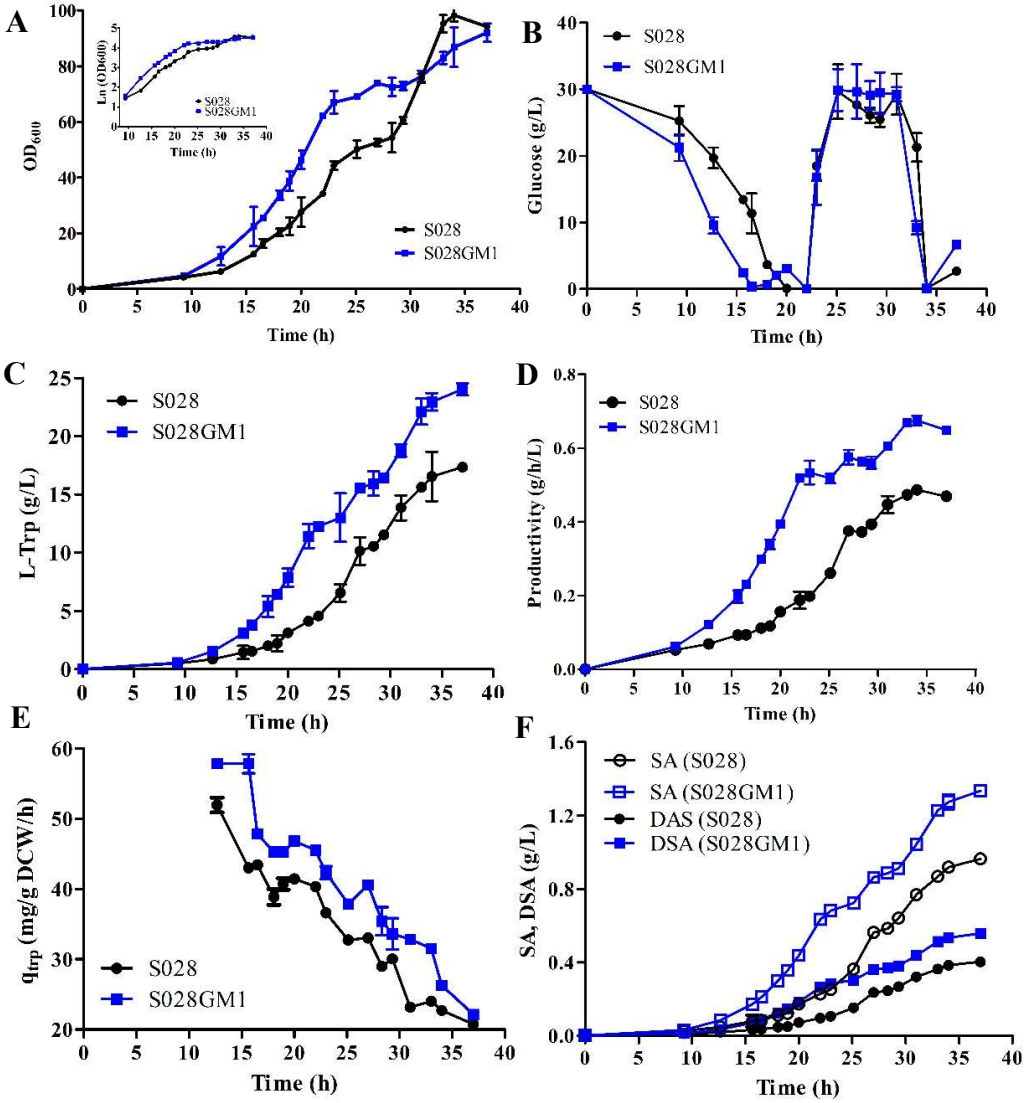

Figure 5.7: Fed-batch fermentation results of the strains S028 (black and circle) and S028GM1 (blue and square). (A) Cell growth; (B) Glucose concentration; (C) Trp production; (D) Overall productivity, (E) Formation rate of Trp (q_{Trp}), and (F) Accumulation of the intermediates shikimate (SA, open circle or square) and dehydroshikimate (DSA, solid circle or square). All results are based on two independent fed-batch fermentations.

Table 5.2: Comparison of the performances of the strains S028 and S028GM1 for L-Trp production in fed-batch fermentations

strains	OD_{600}	DCW (g/L)	GlcC* (g/L)	Trp (g/L)	q_{Trp} (mg/gDCW/h)	Yield (g Trp/g Glc.)	Vp (g/L/h)
S028	94.22±0.48	32.98±0.17	124.82±11.80	17.35±1.16	14.22±0.02	0.14	0.47±0.01
S028GM1	92.00±1.42	32.20±0.49	176.69±10.60	24.03±1.02	20.17±0.13	0.14	0.65±0.01

Fed-batch fermentations were performed in 1.5 L bioreactors at 37°C and pH 6.8; the initial glucose concentration was 30 g/L; the initial inoculation OD_{600} was 0.1. Results are given as means ± standard deviations. * GlcC is the calculated cumulative consumption per reactor volume.

As shown in Fig. 5.7C and Table 5.2, the strain S028GM1 produced a significantly higher amount of Trp than the reference strain after the lag phase (about 10 h) until the end of the fermentation. At the end of the fermentation (37h), S028GM1 produced 24.03±1.02 g/L, which is 38.50 % higher than that (17.35±1.16 g/L) of the strain S028 (Table 5.2). The concentrations of glucose were controlled at nearly the same level for both strains during the whole fermentation. It was found that S028GM1 had a slightly faster growth rate (0.211 h^{-1}) than the strain S028 (0.184 h^{-1}, Fig. 5.7A) at the beginning of the exponential growth phase. The enhanced DAHP synthase activity has obviously contributed to the higher growth rate to a certain extent. A higher biomass formation rate could reasonably result in higher productivity for the strain S028GM1 (Fig. 5.7D and Table 5.2). However, its higher Trp production can be considered to be mainly due to the enhanced DAHP synthase activity directly, because the specific Trp formation rate of the strain S028GM1 was remarkably higher than that of the strain S028 (20.17±0.13 vs. 14.22±0.02, Fig. 5.7E and Table 5.2) during the whole fermentation. Besides, during the fermentation, S028GM1 accumulated higher amounts of the intermediates SA and DSA (Fig. 5.7F) of the chorismate pathway than the reference strain S028. Both of the intermediates accumulated by the strain S028GM1 are about 36% higher than that of strain S028 at the end of fermentation. It suggests that more metabolic flux was redirected into the chorismate pathway in the strain S028GM1 than in the reference strain due to the difference between the variants $AroG^{D6G\text{-}D7A}$ and $AroG^{S180F}$. These results clearly showed that $AroG^{D6G\text{-}D7A}$ is more efficient for the chorismate pathway to the bio-production of aromatic amino acids and their derivatives.

5.6 Conclusions

Protein engineering in the context of metabolic engineering is of fundamental importance for the development of efficient bio-processes. While impressive progress has been made in the design of the enzyme variant library, the availability of reliable and highly effective screening methods is becoming increasingly important. In this work, a novel approach to integrate CRISPR/Cas9-facilitated engineering with growth-coupled and sensor-guided *in vivo* screening and characterization (CGSS) is developed to facilitate protein engineering. This approach has been successfully demonstrated in engineering and screening of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (AroG), with the aim to discover chromosome-context variants AroG that are more resistant to Phe and thus more suitable for the biosynthesis of aromatic amino acids like Trp. With two selected mutation points based on structural information, new variants ($AroG^{D6G\text{-}D7A}$, $AroG^{D6L\text{-}D7P}$, and $AroG^{D6P\text{-}D7I}$) were revealed to be

more resistant to Phe than the Phe-resistant variant AroG^{S180F} reported so far in the literature. The replacement of AroG^{S180F} with AroG$^{D6G-D7A}$ in a previously engineered Trp producing *E. coli* strain (S028) remarkably increased the Trp production by 38.5 % in simple fed-batch fermentations. Since CGSS is based on the integration of the AroG-encoding enzyme variants into the chromosome, it can also be simultaneously used to optimize the expression level of the engineered enzyme in the strain, i.e., by constructing the corresponding gene using different promoters and/or ribosome binding sites. Nevertheless, the CGSS approach relies on *in vitro* laborious construction of gene variant libraries followed by transformation into host cells for directed evolution, and these manipulations are at the expense of greatly diminishing the quantity and quality of gene variants. The *in vivo* continuous mutagenesis enables simultaneous mutation and selection under intracellular conditions without library construction and transformation, which has yielded extraordinary success in the identification of industrial and therapeutic enzymes as well as the optimization of valued compounds-producing industrial strains. Thus, a new approach was proposed by taking advantage of the *in vivo* continuous mutagenesis and the CGSS method in the next chapter.

Chapter 6

6. CGSS-facilitated optimization of GalP/Glk-dependent glucose utilization strain for high yield biosynthesis of Trp

6.1 Introduction

In principle, directed evolution is a general term to illustrate various approaches used to generate enzyme variants, screen and select the desired candidates, and characterize their functions and performances (Yuan et al., 2005). This approach can be carried out either *in vitro* or *in vivo*. Frequently, existing approaches (e.g., CGSS) rely on *in vitro* laborious construction of gene variant libraries followed by transformation into host cells for directed evolution, while *in vivo* continuous genome-wide mutagenesis enables simultaneous mutagenesis and selection under intracellular conditions (d'Oelsnitz and Ellington, 2018).

It is noteworthy that the plasmid-facilitated *in vivo* mutagenesis approach, e.g., phage-assisted continuous evolution (PACE) (Dickinson et al., 2014; Bryson et al., 2017), was introduced to confine the mutagenesis within the target plasmid. Nevertheless, applying the modified natural mutators (e.g., MP pJC184) allowed mutations to occur within or outside of user-defined genes (Badran et al., 2015a), and this could help to address a wide range of targets that require a simultaneous modulation of multiple genes, e.g., for the synthesis of metabolites like Trp that requires multiple precursors. Also, screening of such a plasmid library of enzyme variants is usually performed in a host strain under conditions not relevant to the actual conditions of the subsequent production strain. Thus, further genomic-context integration of gene variants in the production strain is required. With the advent of gene-targeting technologies (e.g., CRISPR/Cas9 technique), more precise technologies for engineering of genes or enzymes at a specific locus have been developed, such as CGSS, CasPER (Jakočiūnas et al., 2018), and EvolvR (Halperin et al., 2018) (Fig. 2.5). Therefore, with the advantage of genome-context integration of gene variants, the cell auxotrophy-coupled screening, and biosensor-driven *in vivo* characterization, a state-of-the-art approach, namely auto-CGSS, is proposed in this chapter. By implementing this approach, the CGSS approach is integrated into an *in vivo*

continuous mutagenesis system, of which the evolution process is coupled with real-time measurement of cell growth and online monitoring of the fluorescence intensity (Fig. 2.8 and Fig. 6.6).

Continuous evolution of the GalP/Glk-dependent glucose utilization system in PTS-negative strain was carried out as an example to demonstrate this method. The carbon source glucose is commonly used in the fermentation industry, and a high glucose conversion yield is a fundamental parameter for cost-effective production. So far, various strategies have been applied (Chen and Zeng, 2017; Panichkin et al., 2016) to achieve the maximum theoretical yield of *E. coli* (0.227 g Trp/g glucose, Fig. 1.3A), involving disruption of competing pathways, removal of feedback and feed-forward regulations, flux optimization of rate-limiting Trp biosynthetic pathway, and optimization of exporters' abilities (Fig. 2.1). Besides, inactivation of the phosphotransferase system (PTS) is a compelling strategy used in optimization of the pathway for overproduction of value-adding compounds derived from PEP, such as Trp (Fig. 6.1) (Carmona et al., 2015; Chen et al., 2018; Lu et al., 2012). Nevertheless, PTS-negative strains show severe impairments in their growing capabilities. Early studies have revealed that activation of the secondary carrier-type facilitators, the galactose permease (GalP) and glucokinase (Glk) pathway in the PTS-defective strain is able to regain glucose transport, and consequently restore its growth (Lu et al., 2012). Activation of the GalP/Glk-dependent glucose utilization pathway also contributes to the glucose conversion yield of Trp (Lu et al., 2012). As presented in Fig. 1.3B, the maximum theoretical yield for the GalP/Glk-dependent strain is calculated to be 0.454 g Trp/g glucose, which is approximately twice of that in the PTS-positive strain. It is, therefore, of great benefit to explore the potential achievable Trp yield in the GalP/Glk-dependent glucose utilization strain by employing direct evolution approach.

6.2 Design and characterization of GalP/Glk-dependent glucose utilization strain

To examine the potential Trp yield of the GalP/Glk-dependent strain, co-expression of enzymes GalP and Glk was performed in a PTS-defective strain. In theory, this engineered strain should lead to an accumulation of PEP, which stimulates the conversion of PEP to the Trp synthesis pathway (Fig. 6.1, shown in blue arrows).

In *E. coli*, the bacterial PTS system, which transports and phosphorylates its sugar substrates, depends on several cytoplasmic phosphoryl transfer proteins: PtsI protein (EI), Histidine protein (HPr), Enzyme IIA (EIIA), Enzyme IIB (EIIB), and the integral membrane sugar permease (EIIC) (Fig. 6.1). The phosphoryl group of PEP is transferred to the PtsI protein, HPr

protein, and Enzyme EII at a conserved histidine residue and subsequently forms glucose-6-phosphate (Postma, 1996) (Fig. 6.1). To block the translocation of the phosphoryl group in the PTS system, the PTS system was deactivated in a rationally constructed strain S028TS/pCas9 (TS: Trp sensor) by disrupting the PtsI enzyme (also known as Enzyme I). The resulting strain was named as S028TSΔ*pstI*/pCas9 (Table 3.1). Afterward, the donor DNA fragment of p*tac*-*galP*-p*J23119*-*glk* in the plasmid p*glk*N20-GG^{WT} (Table 3.2) was integrated into the chromosome of the strain S028TSΔ*pstI*/pCas9 using the CRISPR/Cas9 technique, resulting in the strain S028TSΔ*pstI*::*GG*^{WT} (G028).

To evaluate the glucose uptake rate and growth rate of the PTS-defective strain G028, it was cultivated in the M9 medium with glucose as the sole carbon source. After almost 60h of cultivation, the growth of G028 only reached an OD_{600} of 0.21±0.02 (the initial OD_{600} was 0.10). This result demonstrated that in the PTS-defective strain, the glucose uptake was seriously blocked and thus the cell growth was severely hindered. However, this result was inconsistent with the early report that overexpression of the glucose facilitator Glf from *Zymomonas mobilis* and the housekeeping enzyme Glk restored the glucose transport efficiently (Knop et al. 2001; Chandran et al. 2003). It was assumed that non-functional GalP/Glk-dependent glucose transporter was expressed in the G028 strain since the transcription of both genes *galP* and *glk* were controlled under two stronger promoters, p*tac* and p*J23119*, respectively. This resulted in the dysfunctional or unbalanced co-expression of the *galP* and *glk* genes. To validate this hypothesis, the plasmid pIBB24-*ptac-galp-pJ23119-glk* was transferred into the S028TSΔ*ptsI*/pCas9 to develop a strain S028TSΔ*ptsI*::pIBB24GG (Table 2.2). It should be mentioned that expression of the target genes on the pIBB24 plasmid ensures that the target genes are expressed at the corresponding level. After 60 hours of cultivation, the growth of the strain S028TSΔ*ptsI*::pIBB24GG was still severely hampered in the M9 medium. Nevertheless, the growth of the PTS-positive strain S028TS/pIBB24GG was determined at 1.24±0.21, which is 8% lower than that produced by the strain S028TS/pIBB24 (data not shown). The similar growth for both strains indirectly implied that maintaining the modulation expression of *galP* and *glk* genes is a better strategy to accomplish the biofunction of the GalP/Glk-dependent glucose transporter.

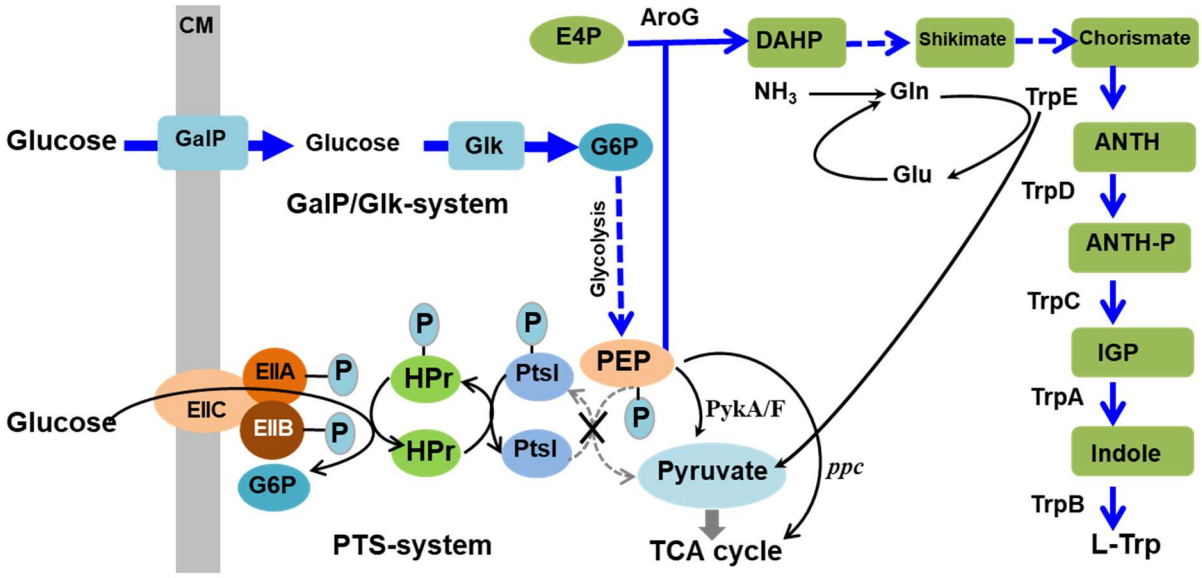

Figure 6.1: The GalP/Glk-facilitated glucose utilization system in the PTS-negative Trp-producing *E. coli* strain. To achieve a PTS-negative *E. coli* strain, one of the two cytoplasmic phosphoryl transfer proteins, PtsI, also known as EI, was first disputed, and then the GalP/Glk-facilitated glucose uptake system was integrated into the PtsI-inactivated strain. The resultant GalP/Glk-dependent strain makes the metabolic flux forced into the Trp pathway. To support the cell growth, one of the essential precursors of the TCA cycle, pyruvate, can be synthesized from the PykA/F-assisted bypass pathway and also from the TrpE-facilitated bypass pathway.

6.3 CGSS-facilitated promoter engineering of GalP/Glk-dependent glucose utilization pathway

To alleviate the decreased glucose utilization rate and growth retardation caused by the dysfunctional of GalP/Glk enzymes, our previously developed CGSS approach was applied to engineer the promoters of *galP* and *glk*, resulting in combinatorial promoter mutants with reliable modulated expression (Fig. 6.2).

In bacteria, the promoter initiates the transcription of a particular gene. It consists of two short sequence elements, approximately 10 (-10 element) and 35 (-35 element) nucleotides upstream from the transcription start site (Sharan et al., 2007). Mutations in the promoter sequences, especially at the -35 and -10 elements, could provide the constitutive promoters with a broad spectrum of strength, which is critical for protein engineering (Alper et al., 2005). Thus, it was proposed to apply the CGSS approach to replace the wild-type promoter (*ptac-galp-pJ23119-glk*) of the G028 with the promoter variants. For this purpose, the *glk* gene of the strain S028TSΔ*ptsI*/pCas9 was replaced by an antibiotic resistance gene Cm^R using the target gRNA plasmid p*glkN20-*Cm^R (Table 3.2) and the CRISPR/Cas9 technique, resulting in the strain S028TSΔ*ptsI*Δ*glk*::Cm^R/pCas9 (G028Cm). The insertion of the Cm^R gene is intended to offer the sgRNA target sequence for CRISPR/Cas9 application in further promoter engineering. As expected, the strainG028Cm was not able to grow in the M9 medium (data not shown). After

the characterization of the strain G028Cm, a mutagenesis library of the promoters was constructed (Fig. 6.2). In detail, N20 sequence of the *glk* gene in the plasmid p*glkN20*-GG^{WT} was replaced by the N20 sequence of the *Cm*^{R} gene, resulting in the plasmid p*Cm*N20-GG^{WT}. Using the p*Cm*N20-GG^{WT} DNA plasmid as a template, a random mutagenesis sequence was inserted into the -35 and -10 elements of the p*tac* and pJ23119 promoters with the synthesized mutation primers *glk-out-FR*/*XhoI-galP-MT* and *glk-out-RF*/*galP-XbaI-MT* (Table 3.3), respectively, generating the plasmid library p*Cm*N20-GG^{MT}. The promoter variants were then integrated into the chromosome of the G028 strain using the CRISPR/Cas9 technique (Fig. 6.2). Finally, the mutants were screened and selected by using the complementary growth-coupled screening and Trp biosensor-based *in vivo* characterization (Fig. 6.2).

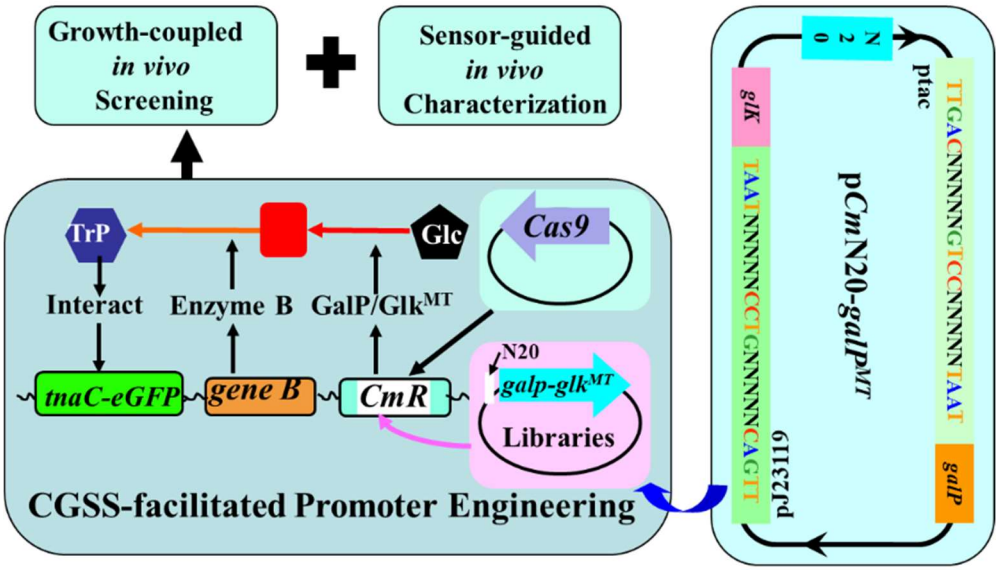

Figure 6.2: CGSS-facilitated engineering of the promoter variants for genes *galP/glk* and growth-coupled and sensor-guided *in vivo* screening of candidates. The left part illustrates the construction of the plasmids for the promoter variants of *galP/glk* genes, p*CmN20-galP*^{MT}. To construct the promoter variants, the random mutagenesis sequences (NNN) were introduced into the -35 element and -10 element of the p*tac* promoter of the *galP* gene and the p*J23119* promoter of the *glk* gene. The right part illustrates how the CGSS approach facilitates the integration of the promoter variants of *galP/glk* genes into the chromosome of the growth-deficient galP/glk-dependent strain, and this is followed by the growth complementation-assisted *in vivo* screening and Trp sensor (p*tac-tnaC-eGFP*)-guided *in vivo* characterization.

After 48 hours of cultivation, approximately 400 mutants with different sizes and various strengths of fluorescence intensity were generated on the M9 agar plate (data not shown). After confirming the phenotype of those mutants, a total of 84 promoter mutants with relatively bigger colony size and higher fluorescence signal were selected and re-checked in the M9 medium with a 96 deep well plate. A total of six samples in the H row (N1-N6 wells) were used as the positive control of strain S028 with a functional PTS, and the other six samples in the H row (N7-N12 wells) were used as the negative control of strain G028 (Figs. 6.3A and B). The growth rate and the fluorescence intensity (MFU) of all mutants are presented in heat maps

(Figs. 6.3A and B). As shown in Figs. 6.3A and B, 12 of the mutants with a higher growth (e.g., over 1.25) were coupled with a stronger fluorescence intensity (colored in green in Fig. 6.3B). Although other mutants, such as mutants in the F row (N11-N12 wells) in Fig. 6.3A exhibited a higher growth (around 1.0), their fluorescence intensities (Fig. 6.3B) were not detectable due to fact that accumulation of Trp was inconspicuous

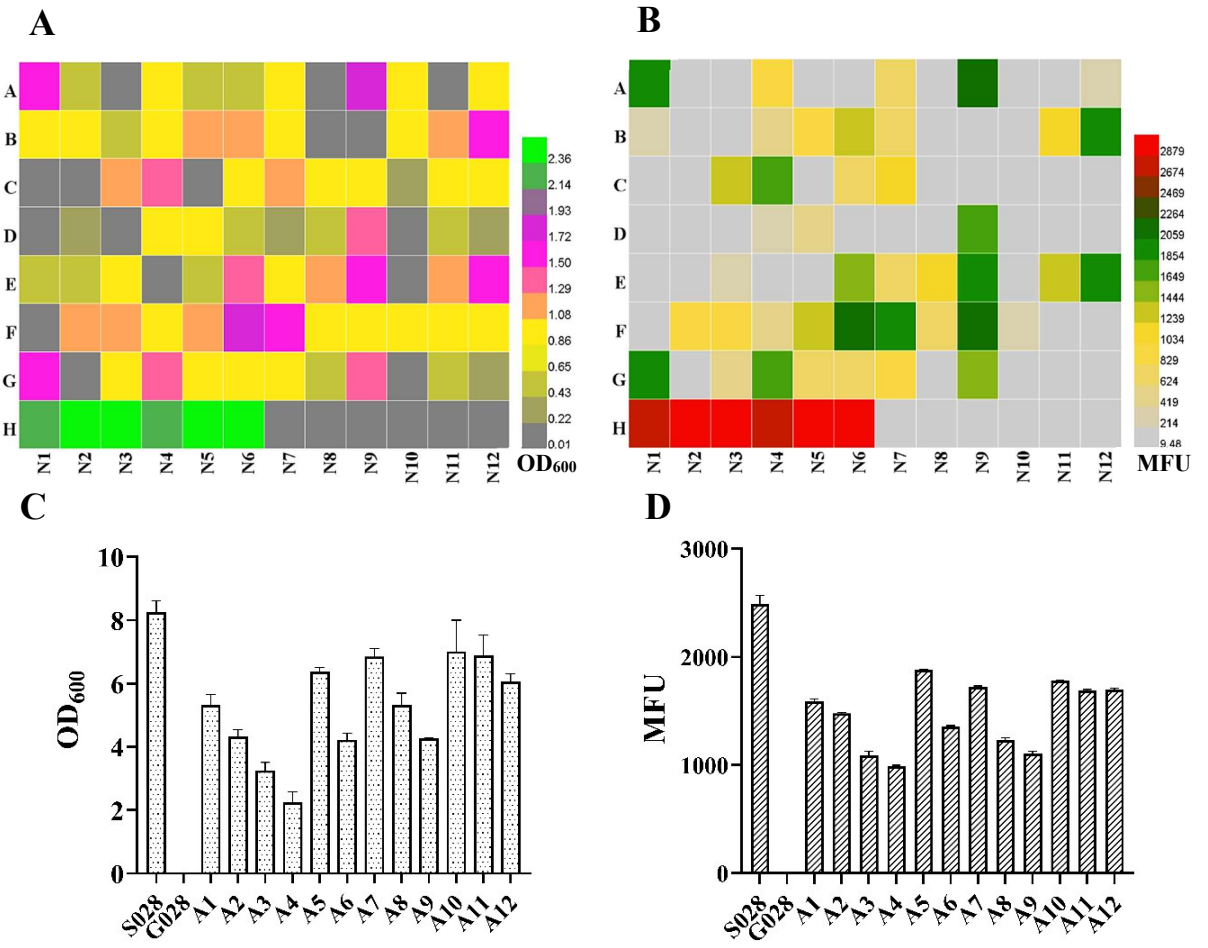

Figure 6.3: Heat maps of cell growth (OD_{600}) (A) and fluorescence intensity (MFU) (B) of selected *E. coli* strains with mutated promoters for the enzymes GalP/Glk. A total of six samples in the H row (N1-N6 wells) are positive controls (strain S028 with a functional PTS). The other six samples in the H row (N7-N12 wells) are negative controls (strain G028 with a defective PTS). Measurements of OD_{600} (C) and fluorescence intensity (D) of the S028 and selected mutants A1-12 from the second round of screening. The cells were cultured with FM-III medium in shake flasks. All the data are from two independent biological samples.

These results indicated the advantage of the growth and biosensor-coupled *in vivo* screening and characterization approach compared to the dialogue-oriented approaches, e.g., cell growth complementation-assisted library screening.

Batch-fermentations of the 12 mutants with the positive (S028) and negative (G028) controls were then carried out in F-III medium in 50 mL conical tubes. As shown in Figs. 6.3C and D, consistent with the previous results, the negative control strain G028 was not able to grow in

the M9 medium, and therefore, the cell growth and MFU was not detectable. Batch culture of all the mutants resulted in a higher growth and fluorescence intensity than those obtained with negative control (Figs. 6.3C and D). The efficiency of the CGSS approach for generating positive mutants was proved to be remarkable. Among these mutants, four of them, A5, A7, A10, and A11, showed a more significant advantage of growth compared with the others (Figs. 6.3C). However, the growth and MFU of these mutants were still lower than that of the positive control (S028). These results suggested that these mutants should be further optimized in order to achieve a better Trp production.

6.4 Batch mode laboratory adaptive evolution of the GalP/Glk-dependent *E. coli* strain

To improve the performance of GalP/Glk-dependent mutants further, mutants (A1-12) with a relatively high MFU were subjected to in a batch mode of adaptive evolution. Specifically, after cultivation for 4 days at 30°C, 10μL of culture was sub-cultured into 3 mL fresh M9 medium in 50 mL shake flasks. After five rounds of adaptive evolution, the resulting mutants were plated on M9 agar plate for further characterization.

A total of 84 mutants exhibiting a higher fluorescence intensity and a bigger colony size were selected and inoculated into a 96 deep-well plate. After 15 hours of cultivation, their cell growth rate and fluorescence density were measured. As shown in Figs. 6.4A and B, 10 of the mutants restored their growth to the same level as the strain S028 (in red, Fig. 6.4A), and these mutants also obtained a similar fluorescence intensity to the strain S028 (in red, Fig. 6.4B). After confirmation of the phenotype, promoter sequences of the genes *galP* and *glk* in these mutants were examined. The sequencing results showed that there were almost 6 types of mutations (Fig. 6.4C). Among them, the mutant pairs B1 and B8, B2 and B10, B3 and B5, B7 and B9 had each the same mutations (Fig. 6.4C). Thus, the strains B1, B2, B3, B7, B4, and B6 were selected for characterization by batch fermentation in shake flasks. Batch fermentation of strain B3 resulted in a higher production of Trp than that achieved with other strains cultured under the same conditions (data not shown). Therefore, the strain B3 together with the strain S028, was adopted for further evaluation by fed-batch fermentation.

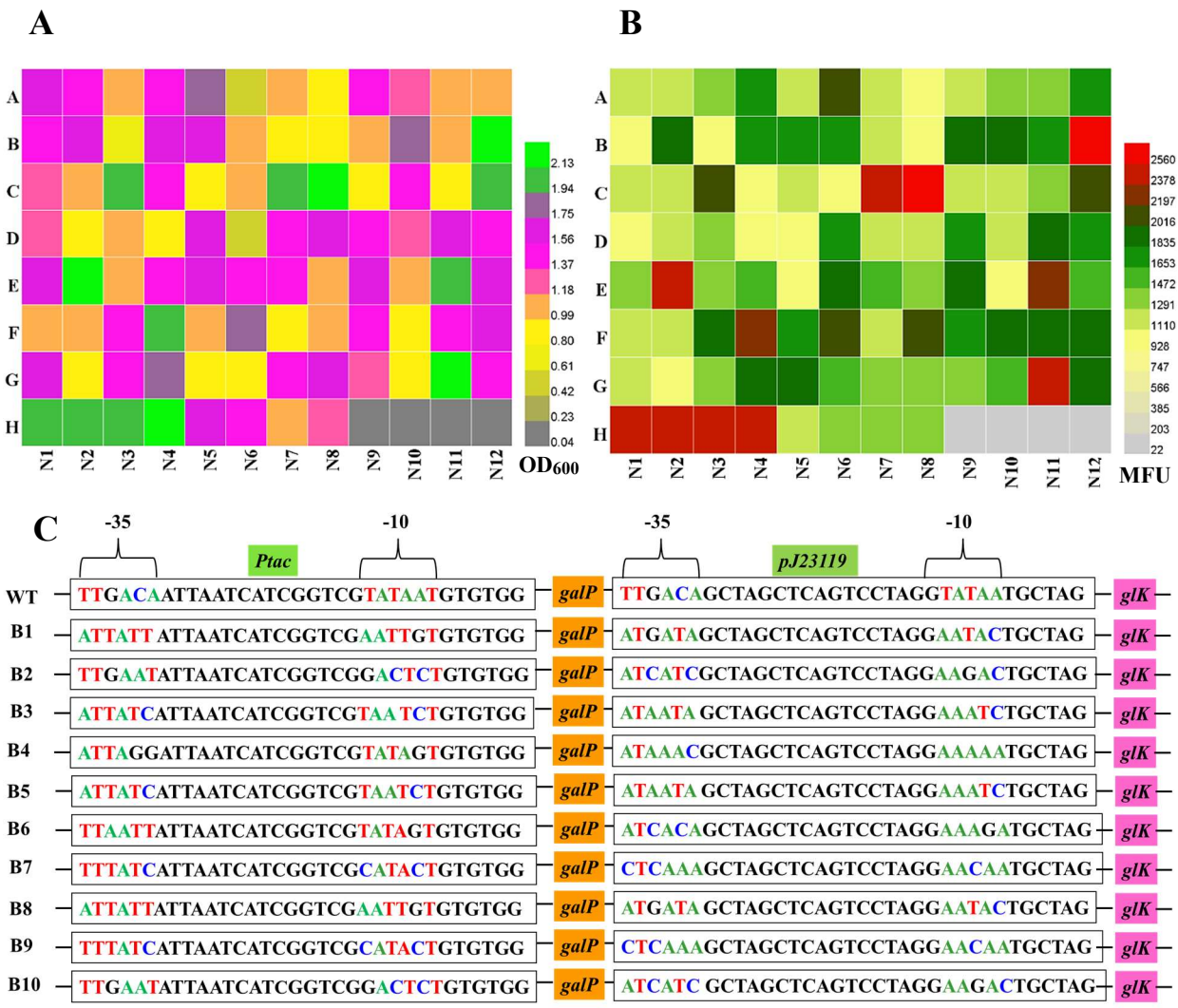

Figure 6.4: Heat maps of cell growth (OD_{600}) (A) and fluorescence density (MFU) (B) of the selected mutants after batch mode adaptive evolution. A total of four samples in the row H (N1-N4 wells) are positive control (PTS-dependent strain S028). The samples N5-N8 and N9-N12 in row H are negative control (GalP/Glk^{MT}-dependent strain G028a) and negative control (PTS-defective strain G028), respectively. (C) The mutation sequence of the promoter variants regarding the p*tac* promoter of the *galP* gene and the p*J23119* promoter of the *glk* gene.

As depicted in Fig. 6.5, fed-batch fermentation of the strain B3 suffered growth retardation severely before the stationary phase (within 30 hours, Fig. 6.5A). This result indicates that the central glucose utilization PTS system is preferable for supporting the growth of *E. coli*. Surprisingly, similar productions of Trp were observed in both strains B3 and S028 within the same period (Fig. 6.5B), indicating that with the less biomass formation and glucose consumption (Fig. 6.5C), the yield of Trp (Fig. 6.5D) and the specific formation rate of Trp (Fig. 6.5E) of strain B3 were higher than that of the positive control. However, after the mid-exponential phase, the strain B3 gradually regained its ability to utilize glucose to the level of S028 (Figs. 6.5A and B). The average biomass of the strain B3 during the stationary phase was determined to be 28.46±2.31 gDCW/L, which was approx. 25 % less than that of the strain S028 (Fig. 6.5A). Considering that the strain B3 still underwent the higher glucose uptake rate

during the stationary phase (4.10 g/L/h from 40 to 50h) but produced a similar amount of Trp as the strain S028 (Fig. 6.5B), this resulted in the reduction of the Trp yield of the strain B3 (e.g., 0.14 g/g vs. 0.15 g/g at 50h, Fig. 6.5D). According to the previous reports, the increase in glucose consumption rate is because of the sensory mutation introduced into the GalP and Glk enzymes or the deletion of GalR repressor during the adaptive evolution (Aguilar et al., 2018). Therefore, the genes *galP*, *glk*, and *galR* were sequenced. The results revealed that no mutation was generated in the *galP*, *glk*, and *galR* genes. These results demonstrated that other potential glucose uptake pathways might be activated instead of the GalP/Glk pathway, e.g., the galactose transporter MglBAC and the maltose MalABC transporter system (Alva et al., 2020).

Overall, increasing in the glucose uptake rate of the strain B3 did not significantly improve the yield of Trp, but it showed benefits in other aspects such as the accumulation of intermediates. As seen in Figs. 6.5G and H, lower accumulation of extracellular SA and DSA was observed in the strain B3 during the entire fermentation process. It was believed that lower accumulation of intermediates indicates an increase in Trp production, especially during the stationary phase. However, the pyruvate production increased dramatically to 2.95 g/L after the exponential phase (Fig. 6.5K), which is approximately five-fold higher than that obtained with strain S028 and indicates the potential importance of this intermediates in supporting its growth. As displayed in Fig. 6.1, several pathways are involved in pyruvate synthesis, such as the PtsI-assisted PTS system, PykA/F-assisted bypass pathway, and the TrpE-facilitated bypass pathway (Fig. 6.1). We hypothesize that the strain B3 was evolved to activate the pathways of pyruvate synthesis for supporting its growth instead of Trp production. Therefore, a more efficient evolutionary method should be established for further development of the GalP/Glk-dependent *E. coli* strain.

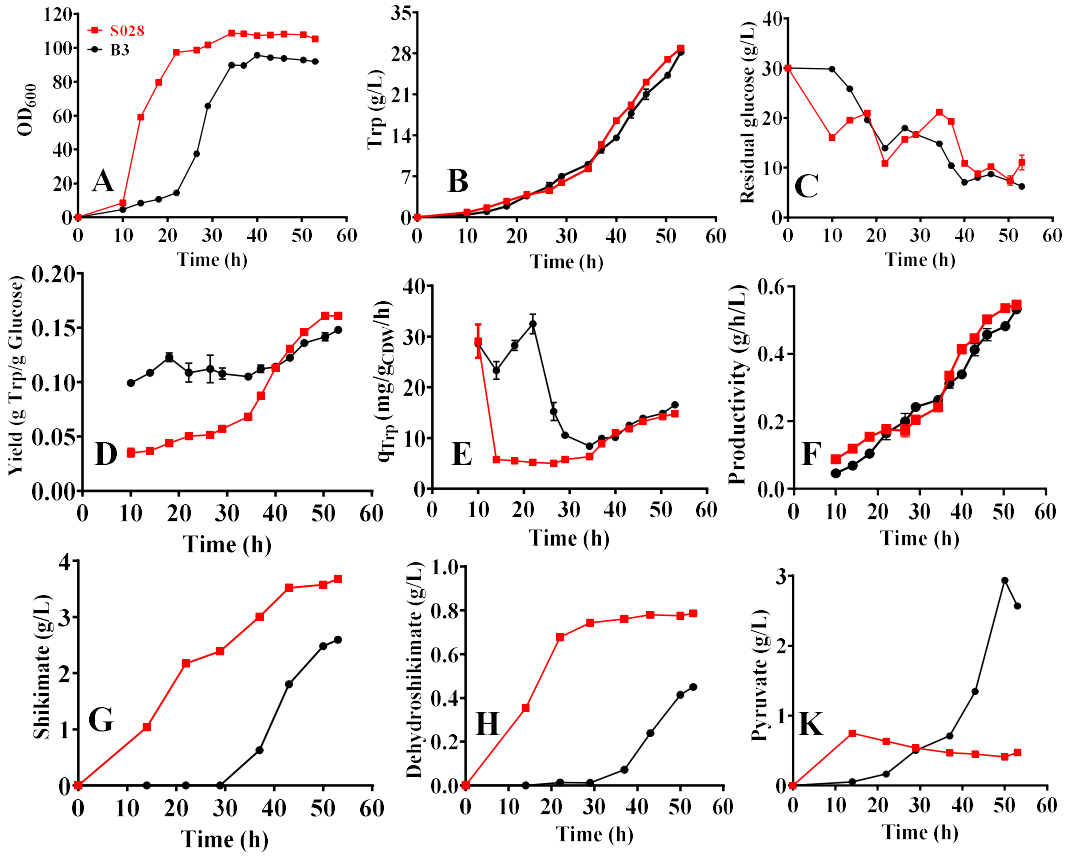

Figure 6.5: Fed-batch fermentation results of the strains S028 (red and square) and B3 (black and circle). (A) Cell growth; (B) Trp production; (C) Residual glucose; (D) Overall yield; (E) Formation rate of Trp (q_{Trp}); (F) Productivity. Accumulation of the intermediates shikimate (G), dehydroshikimate (H), and pyruvate (K). The initial inoculated OD_{600} was 0.45.

6.5 Auto-CGSS-facilitated continuous evolution of GalP/Glk-dependent Trp-producing strain

To obtain a better GalP/Glk-dependent Trp producer with a higher product yield, the strain G028 was optimized by integrating the CRISPR/Cas9-facilitated *in vivo* mutagenesis with an automatic continuous evolutionary system (Fig. 2.8 and Fig. 6.6).

Based on a continuous evolutionary system (Fig. 6.6) established by Dr. Chengwei Ma of our group, further optimization and updating were conducted by integration with the CGSS-facilitated *in vivo* continuous mutagenesis system. The efficiency of this system was demonstrated through development of the GalP/Glk-dependent glucose utilization strain. As illustrated in Fig. 6.2 and Fig. 2.8, the promoter library integrated by (p*Cm*N20-GG^{MT}) was integrated into the chromosome of the strain S028TSΔ*ptsI*Δ*glk*::*Cm*^{R}/pCas9-pJC184 (G028JC) in the same way as outlined in the aforementioned CGSS approach. The resulting cells were inoculated into a vessel containing a synthetic medium with corresponding antibiotics. The growth of the promoter mutants was monitored in real-time by a cell density meter. As soon as

the OD_{600} reached a set point of 0.8, the cultivation system was operated in an auxostat mode, in which the fresh culture supplemented with antibiotics and 10% w/v L-arabinose was supplied via a dedicated inlet, and the depleted medium and cells were pumped out to a waste container (Fig. 6.6). The feeding of L-arabinose was aimed to induce the expression of mutators in the mutagenesis plasmid (pJC184) to conduct genome-context random mutagenesis and thus to accelerate the mutation rate during continuous evolution. Moreover, the flow rates of fresh medium (inlet) and waste (outlet) were dynamically and automatically controlled in response to the real-time monitored cell density so that a constant cell density at around 0.8 was maintained. The real-time fluorescence intensity monitored by the flow cytometer was used as an indicator for intracellular concentration of Trp.

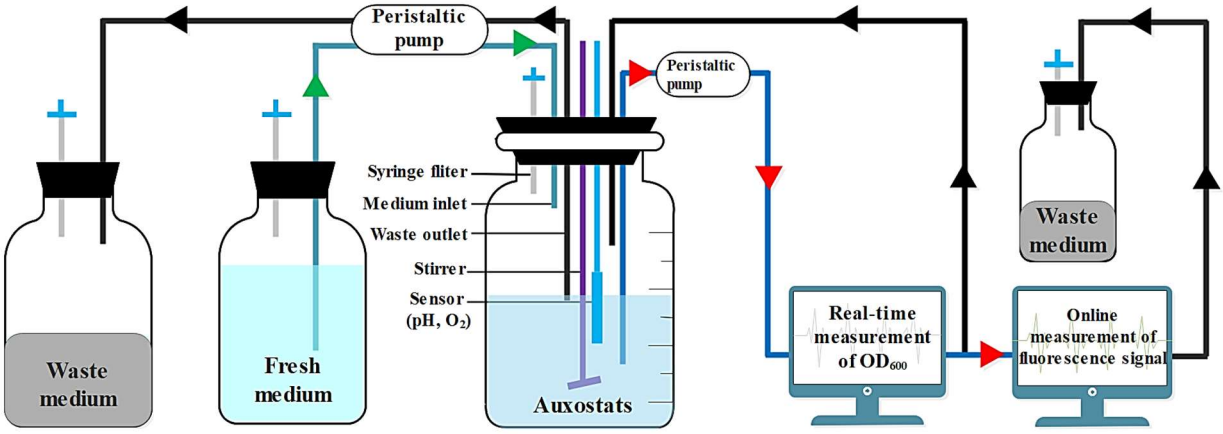

Figure 6.6: Improvement of GalP/Glk-dependent *E. coli* in an automatic continuous evolutionary system. During the continuous evolution, the flow rates of the fresh medium inlet and the waste culture outlet are dynamically regulated by peristaltic pumps in response to the real-time measured optical density of the culture. The real-time fluorescence intensity monitored by the flow cytometer is used as an indicator for intracellular concentration of Trp.

For real-time *in vivo* characterization of the mutants, the evolution vessel was directly connected to a cell density meter (Ultrospec 10, Biochrom) and a flow cytometer (CytoFLEX, Beckman Coulter) for real-time monitoring the cell growth and fluorescence intensity, respectively. As observed in Fig. 6.7A, after almost 45h of cultivation, the OD_{600} reached 0.8. Simultaneously, the cultivation system was switched to an auxostat mode. Moreover, Fig. 6.7B showed that, as the evolution continued up to 126h, the fluorescence intensity reached to the maximum level (~11,446 MFU) and then decreased to a stable level (~8,000-9,000 MFU). The possible reason might be that the PTS-disrupted cells were primarily evolved to support its growth rather than Trp production, and thus the cells with the advantage in growth might become dominant in the late evolutionary phase (~306-378h, Fig. 6.7A).

Subsequently, six samples from the first phase (labeled with triangles in Fig. 6.7B) were selected and their fluorescence intensity and Trp production were characterized via batch

fermentation in shake flasks. The best-characterized mutant G028c was then forced to the second round of continuous evolution for further improvement of its performances. In this round, the OD_{600} of the new mutants reached 0.8 after only 20h of cultivation and the fluorescence intensity increased to the maximum level (~13,541 MFU) in a short period (~390-438h), but encountered the same issue as the first phase, in which the fluorescence intensity decreased significantly in the final stage (Fig. 6.7B). Therefore, four samples from the second phase (labeled with squares in Fig. 6.7B) were further characterized. To this end, the samples from different time points were plated on the synthetic agar medium and grown for 30 hours. Afterwards, a total of 20 single colonies from each sample were selected and together with the positive controls (S028TS/pCas9, B3, and G028c) cultivated in a 96 deep-well plate with F-III medium. After 10 hours of cultivation, their growth and fluorescence intensity were measured.

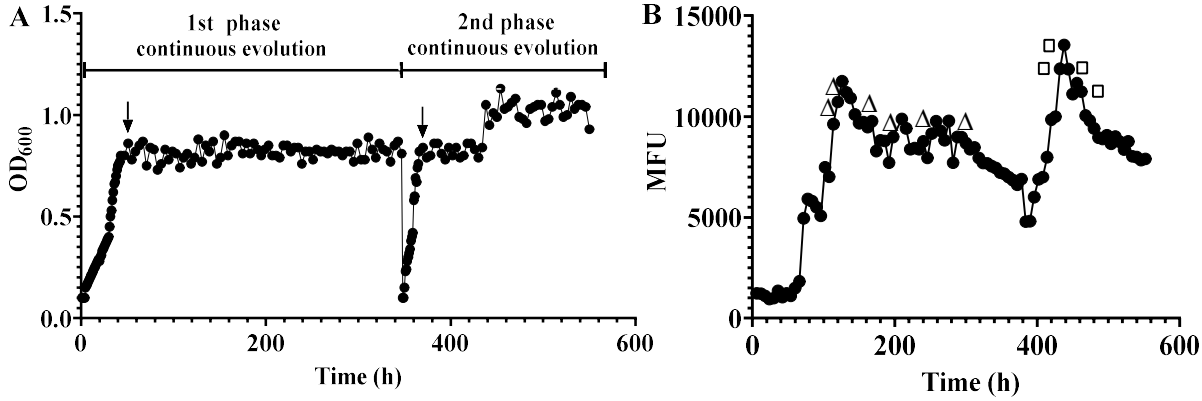

Figure 6.7: Real-time measurements of fluorescence intensity (A) and OD_{600} (B) of the mutants during the whole process of continuous evolution. The process was performed in two stages, from which the best-characterized mutant from the first phase (0-380h) was forced to the second phase of continuous evolution (381-550h) for further development of the stain. the data points in (A) marked with triangles or squares indicate that the samples at those time points were selected for the characterization. The data points in (B) labeled with arrows indicate the time points for induction of 0.1% w/v L-arabinose.

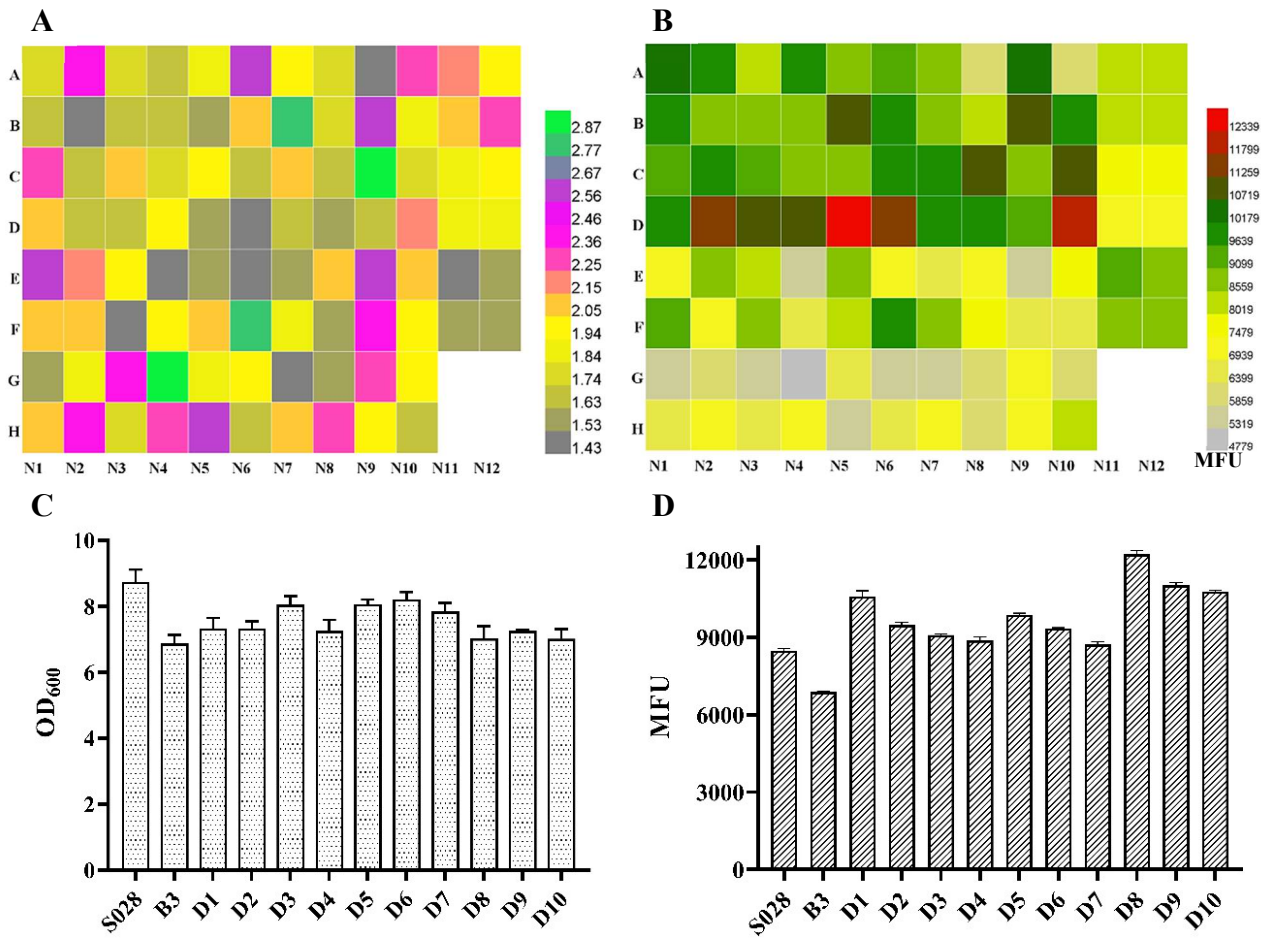

Figure 6.8: Heat maps of cell growth (OD_{600}) (A) and fluorescence intensity (MFU) (B) of the selected mutants from the second phase of continuous evolution. A total of twenty samples in rows A/B (N1-N10 wells), rows C/D (N1-N10 wells), rows E/F (N1-N10 wells), and rows G/H (N1-N10 wells) are presented as single colonies from four different time points. A total of four samples in rows A/B (N11-N12 wells), rows C/D (N11-N12 wells), and rows E/F (N11-N12 wells) are presented as the positive controls: S028, B3 and G028c, respectively. Measurements of the growth (C) and the fluorescence intensity (D) of S028, B3 and mutants D1-10. The cells were cultured with FM-III fermentation in shake flasks. All the data are from two independent biological samples.

It was found that, after 10 hours of cultivation, the average growth rate of the mutants sampled at the time point of 462h (G/H: H1-H10) was higher than those sampled at the early time point (C/D: H1-H10), but the average fluorescence intensity (MFU) was much lower (6480.3±872.7 vs. 10230.7±1136.6) (Figs. 6.8A and B). The results indicated that the PTS-defective cells are preferentially evolved to synthesize the fundamental molecules to support its growth rather than Trp synthesis. Moreover, in comparison to the strain G028c (E/F: N11-N12) and the reference strain S028TS/pCas9 (A/B: N1-N12), a total of 10 mutants shared a higher fluorescence intensity (over 10000, Fig. 6.7B). They were again subjected to characterization by cultivation in FM-III medium in shake flasks for hours. As observed in Figs. 6.8C and D, batch fermentation of all those mutants resulted in a comparable growth rate but with a higher fluorescence intensity than that of the positive strain (S028 and B3). Obviously, the strain D8

showed the best performance in Trp production and therefore it was selected as the candidate for further evaluation by fed-batch fermentation.

6.6 Characterization of the strain D8 in fed-batch fermentation

To examine the performance of the promising mutant D8 for the biosynthesis of Trp, fed-batch fermentation was carried out in bioreactors. In the meanwhile, the fermentation with strain S028 was used as the reference. The plasmid pJC184 was removed from the strain to avoid undesired mutations in subsequent experiment.

As observed in Figs. 6.9A and B, the strain G028 (Table 3.1), in which the chromosome was integrated by a tandem gene circuit (*ptac-galp-pJ23119-glk*) but without adaptive evolution, was not able to completely utilize the carbon source glucose to support its growth (9.12±0.04 at 58h). Nevertheless, as expected, a significant improvement in the glucose uptake rate of D8 (Fig. 6.9B) resulted in a substantial formation of biomass (Fig. 6.9A and Table 6.1). It was noteworthy that the growth of D8 was remarkably impaired compared to that of the reference strain S028 (Fig. 6.9A and Table 6.1). The biomass formation of D8 (22.89±0.20 g/L) was 32.02% less than that of S028 (33.67±0.10 g/L, Table 1). However, D8 resulted in a similar production of Trp as that obtained with strain S028 at the end of fermentation. Reasonably, the overall specific production rate of Trp (q_{Trp}) of strain D8 was significantly higher than that of S028, especially in the stationary phase (Fig. 6.9D). As shown in Table 6.1, the overall specific production rate was increased by 45% (25.28 vs. 17.38 mg/g $_{DCW}$/h) calculated at the end of the fermentation. Moreover, the average production yield of D8 was calculated to be 0.16 g/g during the stationary phase (40-67h, Fig. 6.9E), which is 13.10% and 22.56% greater than that achieved by S028 (0.15 g/g, Fig. 6.9E) and B3 (0.13 g/g, Fig. 6.5D), respectively. The improvements in qTrp and production yield demonstrated that the evolved D8 mutant has much favorable traits for Trp production.

It is noted that the yield of strain D8 was significantly higher than that of S028 (0.16 g/g vs. 0.15 g/g, $P < 0.001$), but the formation of Trp was completely congruent for both strains. In this context, the concentration of two intermediates (SA and DSA) from the Trp synthesis pathway was determined by HPLC to explore the potential mechanisms behind. Figs. 6.10G and H indicate that throughout 30 to 47h, the trends of SA and DSA formation in D8 were similar to those in S028. It was also found that the Trp production of the strain D8 was slightly low (Fig. 6.9C) because of its insufficient glucose conversion rate (Fig. 6.9E). After 50h of cultivation, with the improvement of glucose conversion yield, more intermediates in D8 were converted into the Trp synthesis pathway (Figs. 6.9G and H). This resulted in an increase of Trp

production (38.77±0.02 g/L), and consequently reached the same level as strain S028 (Table 6.1). Therefore, it was attractive to mine potential mutations that led to an improvement in the Trp yield of the strain D8 by comparative genomic DNA analysis.

Table 6.1: L-tryptophan fed-batch fermentation parameters of the strains S028 and D8.

strains	OD_{600}	DCW (g/L)	GlcC* (g/L)	Trp (g/L)	q_{Trp} (mg/gDCW/h)	Yield (g Trp/g Glc.)	Vp (g/L/h)
S028	96.20±0.28	33.67±0.10	261.33±7.35	39.20±0.78	17.38±0.40	0.15	0.59±0.01
D8	65.40±0.57	22.89±0.20	240.81±0.18	38.77±0.02	25.28±0.71	0.16	0.58±0.01

Fed-batch fermentations were performed in a highly instrumented and automated 4-paralleled 1.5 L bioreactors system DASGIP for 67 hours; the initial glucose concentration was 25 g/L; the initial inoculation OD_{600} was 0.45. Results are the means ± standard deviations in two independent measurements. * GlcC is the calculated cumulative consumption per reactor volume.

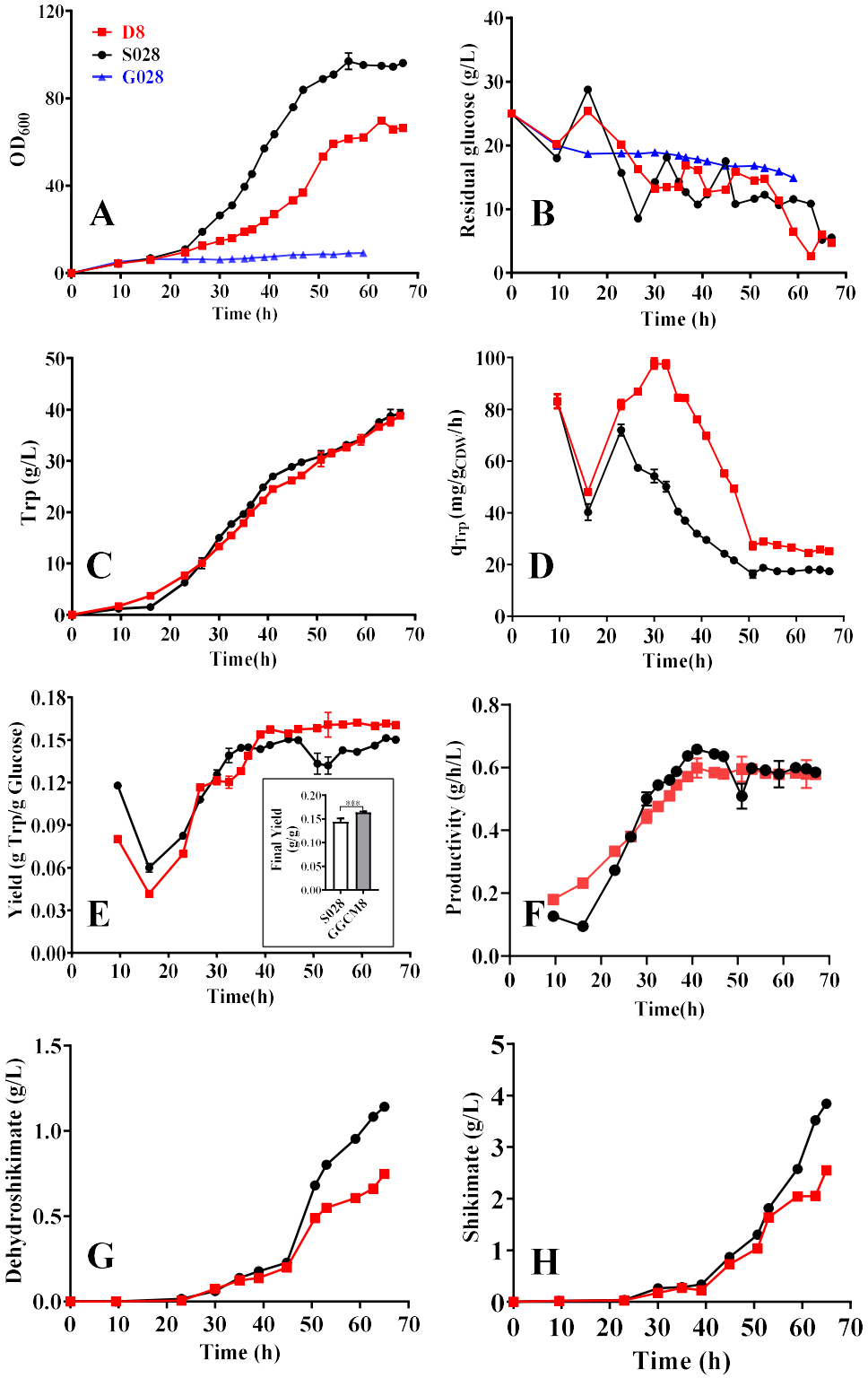

Figure 6.9: Fed-batch fermentation results of the strains S028 (black and circle), G028 (blue and triangle), and D8 (red and square). (A) cell growth; (B) residual glucose; (C) Trp production (D) specific production rate of Trp Trp production (q_{Trp}); (E) overall yield of Trp and the yield, '***' designates $P < 0.001$; (F) Productivity. Accumulation of the intermediates dehydroshikimate (DSA) (G) and shikimate (SA) (H).

6.7 Inactivation of Mlc repressor to activate the expression of PTS genes

In order to elucidate potential mutations that resulted in the behaviors of D8, the whole genome of D8 was sequenced by using the Next-Generation Sequencing (NGS) approach. With the genomic information of *E. coli* W3110 as a control, pairwise genome alignment of the *E. coli* W3110 and D8 was conducted and the results revealed that a total of 15 genes in the D8 strain were modified and 5 genes were deleted (Appendix A: Table A1). Among them, three genes are of great interest because of their functions in glucose utilization and cell growth.

In particular, one of the genes that are directly related to PTS system is found to be deleted in the mutant D8: *mlc* (also known as *dgsA*, EcoCyc ID: G6852). Mlc is a dual transcriptional regulator that controls the expression of a number of genes encoding enzymes of the *E. coli* phosphotransferase (PTS) and phosphoenolpyruvate (PEP) systems (Plumbridge, 2001; S.Y. Kim et al., 1999). In particular, Mlc represses the expression of two genes *ptsG* and *manXYZ* which encode respectively the two transporters EIICBGlc and EIIABCD^{Man} involved in the uptake of glucose (Plumbridge, 2001; Plumbridge and Kolb, 1993). In *E. coli*, PtsG-facilitated PTS is the major glucose transporter (Fig. 6.1). As shown in Figs. 6.10A and B, in the absence of glucose, PtsG (EIICB^{Glc}) is predominately phosphorylated and Mlc is bound to its operators; in the presence of glucose, the dephosphorylated-form of PtsG interacts directly with Mlc and induces the transcription of Mlc-regulated genes (*ptsGH*) by displacing Mlc from its target sequences. As well as being the major uptake system for mannose, the *manXYZ* operon encodes three proteins EIIABCD^{Man} to form the enzyme II of the mannose PTS, the latter also transports glucose efficiently (Fig. 6.10C) (Plumbridge, 1998). Plumbridge found that a mutation in the *mlc* gene resulted in a threefold activation of *manX* expression (Plumbridge, 1998). It was assumed that the deletion of *mlc* gene in D8 could also result in derepression of *manX* and *ptsG* expression (Figs. 6.10.B and C). As shown in Fig. 6.1 and Fig. 6.10B, an activation of *manX* and *ptsG* expression also leads to the accumulation of glucose-6-phosphate, and the accumulated Glc-6-P can be subsequently converted to PEP by the Glycolysis pathway. This indicates that inactivation of Mlc repressor could be one of the important strategies to lead to increased availability of PEP, and it also explains why the fed-batch fermentation of D8 resulted in a higher Trp yield (Table 6.1). However, further experiments are required to confirm this hypothesis.

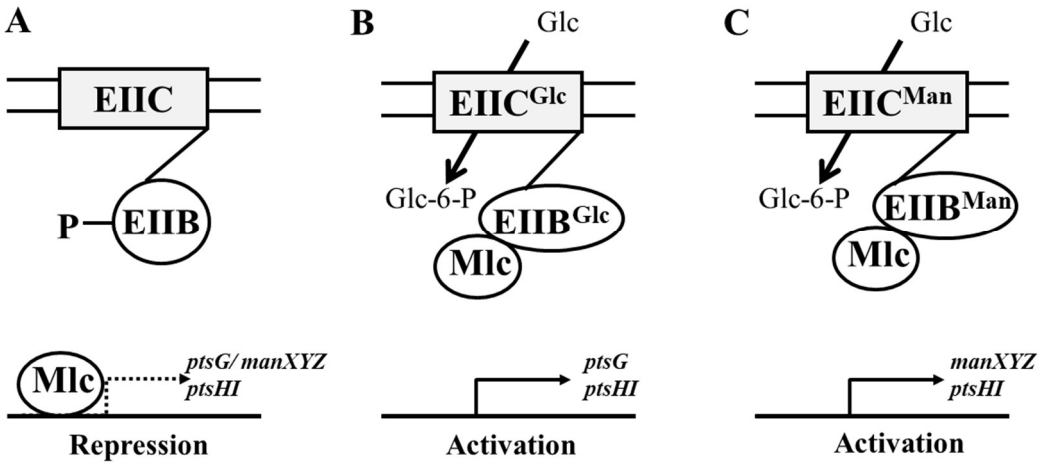

Figure 6.10: Models for PtsG and ManXYZ regulations. PtsG (EIICBGlc) and ManXYZ (EIIABCD^{Man}) are shown as a membrane-anchored EIIC domain and a soluble EIIB domain. (A) In the absence of glucose, PtsG or ManXYZ is predominately phosphorylated and Mlc is bound to its operators. In the presence of glucose, EIIB^{Glc} or EIIB^{Man} becomes dephosphorylated, Mlc binds to the EIIB domain and Mlc controlled genes are activated (adapted from J Plumbridge (2001)).

6.8 Structure-based analysis of the enzyme variant of GalP for glucose permeation

The *galP* gene in D8 was found to have three mutations of Lys137Thr, Pro195Ala, and Phe204Leu compared with that in the strain W3110 (Appendix A: Table A1). The promoter sequence of the strain D8 also coincided with that of the strain B3 (Fig. 6.4C). In this regard, it was attractive to explore how these mutations affect the structure of the GalP protein and its transport capability.

The prototypical H^+/Galactose symporter GalP, a specialized membrane channel for the transport of sugars into and out of cells, are members of the Major Facilitator Superfamily (MFS) (Zheng et al., 2010). To date, many of the available crystal structures have been determined for the MFS, e.g., symporter STP1 form *Arabidopsis thaliana* (PDB: 6H7D), xylose transporter XylE from *E. coli* (PDB: 4JA3), and glucose/H+ symporter $GlcP_{Se}$ from *Staphylococcus epidermidis* (PDB: 4LDS), but the crystal structure for GalP symporter is not yet available. Therefore, a homology model of GalP was carried out using Modeller version 9.24 (Webb and Sali, 2016) with the crystal structure of *A. thaliana* STP10 as the structural template (Paulsen et al., 2019). The model was based on the STP10 rather than the $GlcP_{Se}$ structure that has the highest alignment identity (38.95%), because the crystal structure of STP10 in complex with glucose is available. Afterward, docking of flexible ligand glucose to the GalP-symporter was performed with AutoDock Vina (Trott and Olson, 2010), which is integrated into Chimera (Pettersen et al., 2004). The homology model of GalP and the result of docking were presented in Fig. 6.11.

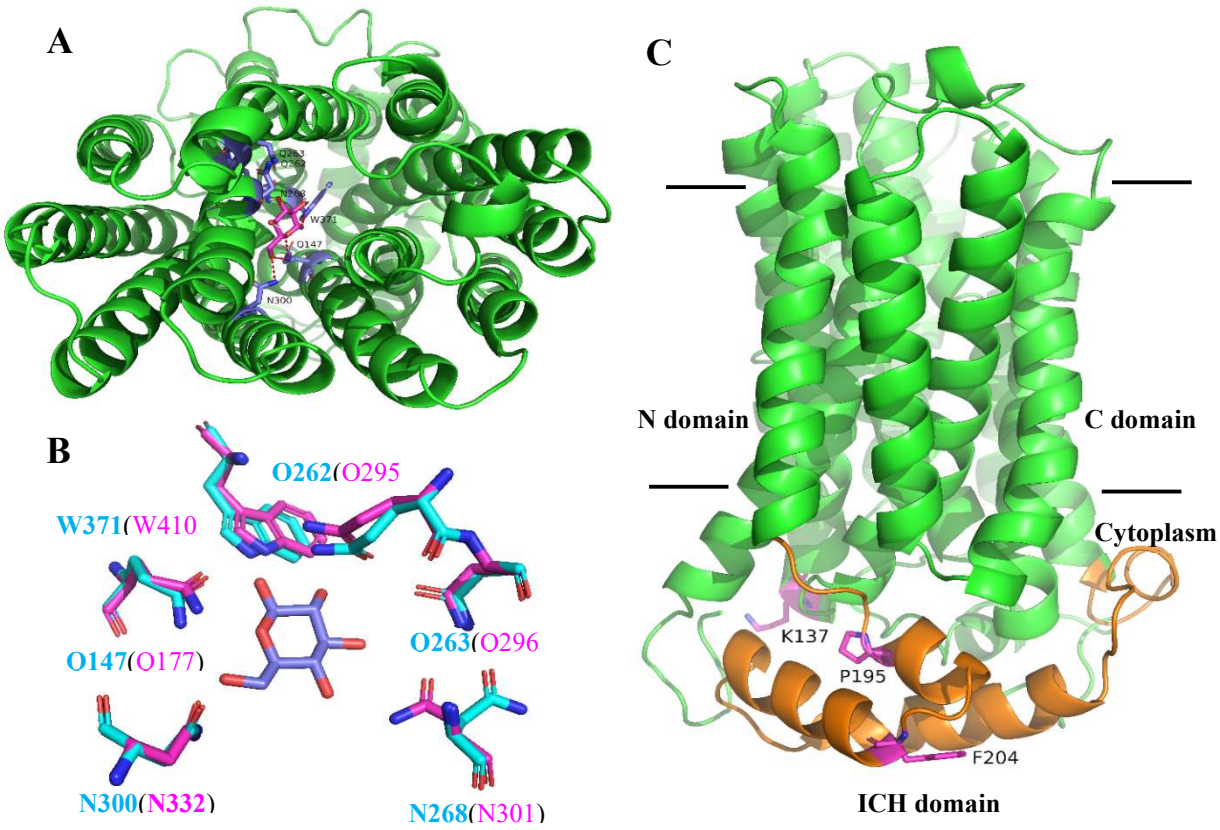

Figure 6.11: A 3D model of GalP was generated based on the crystal structure of *Arabidopsis thaliana* symporter STP10 (PDB: 6H7D). (A) The structure (viewed from the periplasmic face) represents an outward-facing conformation in complex with D-glucose. The glucose-binding sites were predicted by the molecular docking, in which the binding residues are shown in the sticks. (B) Comparison of the glucose binding sites of GalP (blue) and STP10 (pink). The binding residues in blue are presented to the modeled structure of GalP, and the corresponding residues in STP1 are shown in pink. (C) Side view of modeled GalP structure in a ribbon representation. In this view, the intracellular helical bundle (ICH) domain is shown in orange, where the mutated residues in GalP are presented in pink sticks.

As depicted in Figs. 6.11A and C, the outward-facing conformation of the modeled GalP symporter consists of the N- and C-terminal domains, each domain with six transmembrane helices, and both domains are connected by a long loop, namely the intracellular helical bundle (ICH) domain (Fig. 6.11C, in orange). According to early studies, although it has two short helical segments, the ICH domain is an extended and flexible structure that determines the change between inward and outward conformations (Iancu et al., 2013; Paulsen et al., 2019). Fig. 6.11A shows that glucose could be docked into the homology model of GalP at the potential binding site, and as shown in Fig. 6.11B, the predicted glucose binding-sites were almost 100% conserved in the template structure (STP10, in pink). These results indicated that similar glucose binding sites exist in GalP.

In the homology model of GalP, the glucose binding sites are located at the interior hydrophilic cavity surrounded by the N- and C-domains, and as shown in Fig. 6.11C, the mutation site T137 is located at the Helix V in N-domain. As observed in Fig. 6.12A, the backbone ε-ammonium

group (NH_3^+) of wild-type K137 is bound to the carboxyl group of residue P2, while the location of mutated residue T137 is out of H-bonding distance of the P2 residue (Fig. 6.12B). It is assumed that although the mutated residue Thr137 is far from the predicated glucose binding sites, the hydrophilic amino acid Thr may have an indirect effect on the hydrophilic cavity, e.g., an increase in the affinity of glucose-binding.

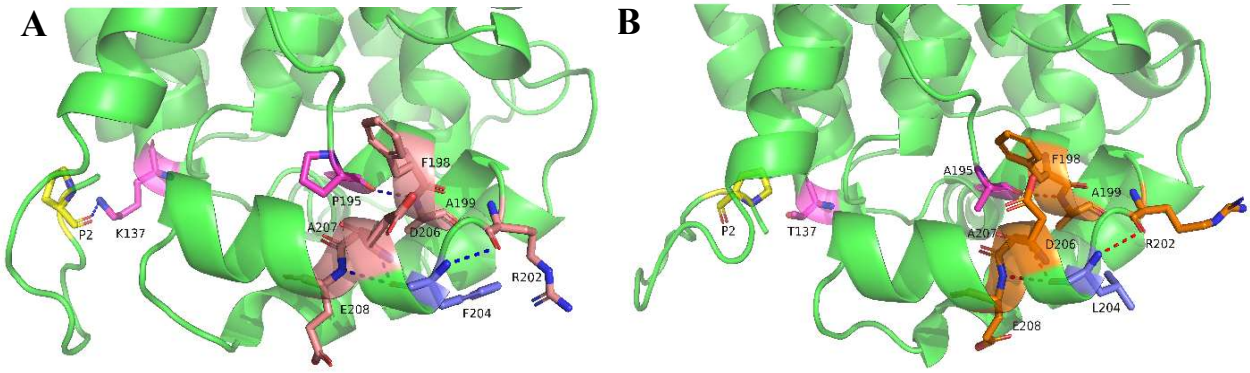

Figure 6.12: Possible interactions between the mutated residues and its adjacent parts. (A) For the predicted wild-type residue K137, its backbone ε-ammonium group (NH_3^+) is attached to the carboxyl group of the residue P2, while (B) the mutated residue T137 is out of the H-binding distance of the P2 residue. In comparison to the wild type, the other two mutated A195 and L204 residues are still bound to the adjacent residues to form the flexible ICH domain.

Moreover, Fig. 6.11C shows that the other two mutated residues Ala195 and Leu204 are located in the ICH domain. Compared to the wild-type model of GalP, there is no significant change of conformation of the helical bundle in the mutated model of GalP (Fig. 6.12B), and this indicates that the flexible conformation of ICH domain is necessary for the stabilization of hydrophilic cavity. Looking into the mutated residues, the long-chain Pro at the position of 195 was mutated to Ala and the aromatic acid Phe at the position of 204 was changed to a branched-chain amino acid Leu, which is supposed to result in the increase of the flexibility of helical bundle and lower the energetic barrier of the transition between the inward and outward conformations. Collectively, the mutated residues Thr137, Ala195, and Leu204 are supposed to be conducive to the flexibility of the ICH domain and the affinity of glucose-binding, and further experiments are required to confirm this hypothesis.

6.9 Decease of cell growth by interruption of the pentose phosphate pathway

As a key enzyme in the pentose phosphate pathway (PP pathway), the $NADP^+$-dependent glucose-6-phosphate dehydrogenase (Zwf) was found to be mutated in D8, in which the residue E165 was replaced by a stop codon. PP pathway is a metabolic pathway parallel to glycolysis, and it is mainly used for generation of ribose 5-phosphate (R5P). R5P is mainly used in the synthesis of nucleotides and nucleic acids, and the deficiency of R5P therefore will lead to the insufficiency of nucleotides and nucleic acids synthesis. The process of DNA duplication and

cell division will consequently be influenced. This might partially explain why D8 exhibited a severe growth deficiency.

6.10 Conclusions

In vivo continuous mutagenesis, as a method for the directed evolution of biomolecules, has been extensively explored for enhancing enzymatic activity, the generation of novel functional enzymes, and the optimization of valued compounds-producing strains. In this study, the CGSS-facilitated *in vivo* continuous mutagenesis (CMCGSS) was proved to be able to evolve the GalP/Glk-dependent strain to utilize the glucose and increase glucose conversion yield. To investigate the potential Trp yield in GalP/Glk-dependent glucose utilization strain, the promoter mutants S028Δ*ptsI*::GG^{MT} were generated by deletion of the gene *ptsI* and integration of GalP and Glk promoter variants. The strains S028Δ*ptsI*::GG^{MT} were first forced to the batch adaptive evolution, and one of the candidates B3 was identified. However, the fed-batch cultivation of the strain B3 exhibited a similar pattern of Trp yield and production to that of the strain S028. In this regard, a novel and scalable CGSS-facilitated *in vivo* continuous mutagenesis system with real-time measurement of cell growth and online monitoring of fluorescence intensity was established for continuous evolution of the mutants S028Δ*ptsI*::GG^{MT}. As expected, one best candidate D8 was identified and the fed-batch fermentation results indicated that the average yield of strain D8 was higher than that of strain B3 by 22.56% (0.16 g/g vs. 0.13 g/g) during the stationary phase and the specific Trp production rate of D8 was increased to 26.64±0.21 mg/gDCW/h. These results demonstrated that *in vivo* continuous mutagenesis is more efficient compared to the traditional adaptive evolution and indicated the utility of an *in vivo* mutagenesis approach.

Chapter 7

7. Engineering microbial cell factories: Assembling of selected enzyme variants in Trp-producing strains for Trp overproduction

7.1 Introduction

As demonstrated in Chapter 5, engineering of one of the DAHP synthase isozymes, phenylalanine-sensitive AroG enzyme, is one of the crucial strategies for efficient biosynthesis of Trp. Using the CGSS approach, an AroG variant, AroG$^{D6G-D7A}$, was screened and it was proved to be more powerful than the originally used variant AroG^{S180F} in terms of the Phe resistance (Fig. 5.6) *in vitro* and facilitating Trp biosynthesis *in vivo* (Table 5.2). As presented in Table 5.2, using AroG$^{D6G-D7A}$ enabled the strain S028G (also named as S028GM1) to produce 24.03±1.02 g/L of Trp in 37 hours, which is 38.50 % higher than that produced by the strain S028 (17.35±1.16 g/L) where the variant AroG^{S180F} was used. As shown in Fig. 5.6, the specific activity of AroG$^{D6G-D7A}$ was notably higher than that of AroG^{S180F} under identical assay conditions. These led us to believe that the increased activity of DAHP synthase resulted in the increased Trp production. It subsequently raised the hypothesis: an improved Trp production could be done by further increasing the activity of DAHP synthase in the strain S028G. Although the mutation D6G-D7A in AroG did not make it completely resistant to Phe, it showed very strong resistance to Phe: approximately only 40% of the enzymatic activity at the presence of an extremely high concentration of Phe (20 mM Phe, Fig. 5.6). Since the intercellular concentration of Phe of the strain S028G might not be extremely high (<20 mM), it was reasonable to believe that introducing another copy of *aroG*$^{D6G-D7A}$ could significantly increase the activity availability of DAHP synthase in the current strain S028G. In this regard, another copy of *aroG*$^{D6G-D7A}$ was introduced to the strain S028G resulting in the strain S028GGMT.

Results showed that S028G accumulated notable intermediates SA and DSA (Fig. 5.7), it was assumed that the strain S028GGMT would accumulate them even worse. Nevertheless, the heterogeneous expression of *An*TrpC^{R378F} would be a solution to that issue. As illustrated in Chapter 4, either additionally expressing *An*TrpC^{WT} or *An*TrpC^{R378F}, especially the later, made

the strain produced much fewer intermediates SA and DSA. In addition, it promoted Trp yield by 26.4 % compared to that of the parent strain S028 (0.19 g/g vs. 0.15 g/g, Table 4.2). Therefore, the strain S028AARF as well as S028AAWT was constructed and characterized in this study.

Moreover, Trp yield could be enhanced by other strategies, i.e., enhancing the supply of precursors such as E4P and PEP (Fig. 2.1). As revealed in Fig. 2.1 and Fig. 6.1, the availability of PEP for Trp biosynthesis could be achieved by replacing PEP-dependent sugar transport system, namely PTS system, with the PEP-independent GalP/Glk-facilitated glucose uptake (Carmona et al., 2015; Chen et al., 2018; Lu et al., 2012). Indeed, fermentation results of strain D8 which obtained by applying that strategy showed an increased Trp yield than that of the PTS-positive strain (Fig. 6.9). In a microbial cell factory, a high glucose conversion yield is an essential parameter for cost-effective production. In this regard, A Trp-producing microbial cell factory was designed for a high flux towards the biosynthesis of Trp by integrating the candidate modules as mentioned above. To this end, the strain D8AA was constructed and characterized by fed-batch fermentation in this work.

7.2 Trp production improved by increasing the activity of DAHP synthase and knocking out the *fruR* gene

Increasing the activity of DAHP synthase was hypothesized to be an effective strategy for improving the Trp production in the strain S028G. Therefore, another copy of $aroG^{D6GD7A}$ was introduced into this strain at the loci of the *fruR* gene with the following reasons.

Fructose repressor (FruR), also known as catabolite repressor activator (Cra), is a dual transcriptional regulator that modulates the direction of carbon flow through different central pathways of energy metabolism. Particularly in the presence of glucose, the intracellular fructose 1-phosphate interacts with the FruR repressor to prevent its binding to the target operons, thereby activating the gluconeogenic pathway and the Krebs cycle of energy metabolism (Ramseier et al., 1995; Saier Jr and Ramseier, 1996). It was found that the deletion of the *fruR* gene led to an increase in chorismate production (Liu et al., 2016).

The replacement of the *fruR* gene in the strain S028G by the gene variant $aroG^{D6G\text{-}D7A}$ was done by using the CRISPR/Cas9 technique. To this end, the target plasmid p*fruR*N20-$aroG^{GA}$ was first constructed based on the plasmid p*Cm*N20-$aroG^{GA}$ (Table 3.2), whereby the *Cm*N20 sequence was modified to *fruR*N20 sequence and the homology recombination arms were also altered accordingly (Fig. 2.7). The final plasmid p*fruR*N20-$aroG^{GA}$ was then transferred into

the competent cells of S028GM1/pCas9 for targeted genome editing, resulting in the strain S028GGMT. The comparison of the performance of the strain S028GGMT and the reference strain S028G were carried out with fed-batch fermentations.

As shown in Fig. 7.1 and Table 7.1, S028GGMT showed good reproducibility of Trp formation, specific production rate, and production yield. In particular, S028GGMT exhibited the same Trp production pattern as the reference strain before the mid-log phase (0-25h). After that, S028GGMT produced Trp more efficiently than the reference strain. As a result, 42.95±0.91 g/L of Trp was achieved by S028GGMT in 65h, which is almost 18.85% higher than that achieved by S028G (36.04±1.17 g/L, Table 7.1). As a comparable cell density was obtained for both S028G (30.21±1.10 gDCW/L) and S028GGMT (26.67±0.10 gDCW/L) (Fig. 7.1A and Table 7.1), it was reasonable that the specific formation rate (24.77±0.23 mg/gDCW/h, Table 7.1) obtained by S028GGMT is higher than that (18.35±0.48 mg/gDCW/h) obtained by S028G. Moreover, the production yield of Trp (0.16 g/g, Table 7.1) for both strains is almost equal, indicates the Trp pathway after AroG required optimization.

The results also revealed that the average specific production rate (q_{trp}) and the average Trp yield of the strain S028GGMT were similar to those obtained with S028G during the stationary phase (Figs. 7.1E and F). These results also coincided with the previous result that the variant $AroG^{D6G-D7A}$ could neither promote the specific production rate nor the yield in strain S028G (Fig. 5.7), indicating the native Trp enzymes could not further streamline the metabolic flux. As illustrated in Fig. 2.1, enhancing the catalytic efficiency of *trp* operon could be done by many ways, including removing other metabolic controls (e.g., feed-forward inhibition) in the Trp biosynthetic pathway, engineering of the Trp exporter (YddG), and increasing the availability of precursors (e.g., L-serine). Interestingly, the productivity of Trp of S028GGMT was greater than that of S028G until the end of fermentation, and this difference was more noticeable during the stationary phase (Fig. 7.1D). For example, the average productivity of Trp of S028GGMT was calculated to be 0.72±0.05 g/L/h from 35h to 62 h, which is 38.46% higher than that of S028G (0.52±0.03 g/L/h) (Fig. 7.1D). Therefore, the increased Trp productivity of S028GGMT could be one of the main factors for the improvement of Trp production because of the enhanced enzyme activity of $AroG^{D6G-D7A}$.

As shown in Figs. 7.1G and H, the trends in the formation of SA and DSA revealed that the concentrations of SA and DSA in S028GGMT were higher than that of reference strain throughout the fermentation process. In addition to the improved Trp production, this suggested S028GGMT was able to rewire metabolic flux into the chorismite pathway more efficiently

than the reference strain S028G. In one of the early reports, metabolomics analysis showed that the deletion of *fruR* significantly enhances the metabolic flow through glycolysis, the pentose phosphate pathway and the TCA cycle, increasing levels of key intermediates (e.g., SA and DSA) and substrates (e.g., L-glutamine and L-serine) for Trp biosynthesis (Liu et al., 2016). In this regard, to what extent the introduction of an extra enzyme variant AroG$^{D6G-D7A}$ resulted in an increase of the metabolic flux towards chorismate synthesis seems to not easy to conclude. However, the flux could not be very efficiently converted into the Trp pathway might due to the low catalytic efficiency of the *trp* operon and the complicated regulations (e.g., feed-forward inhibition), thus led to the accumulation of SA and DSA (Figs. 7.1G and H). This also indicated why the variant AroG$^{D6G-D7A}$ could not promote Trp yield. It reasonably led us to overexpress the *An*TrpC^{R378F} variant.

Table 7.1: Comparison of the performances of *E. coli* S028G derivative strains in fed-batch fermenter cultures

strains	OD_{600}	DCW (g/L)	GlcC* (g/L)	Trp (g/L)	q_{Trp} (mg/gDCW/h)	Yield (g Trp/g Glc.)	Vp (g/L/h)
S028G	86.31±3.14	30.21±1.10	192.73±10.47	36.04±1.17	18.35±0.48	0.16	0.55±0.01
S028GGMT	76.20±0.28	26.67±0.10	275.32±11.42	42.95±1.26	24.77±0.23	0.16	0.66±0.01
S028AAWT	86.30±0.14	30.21±0.05	246.26±2.89	46.05±0.38	23.45±0.23	0.19	0.71±0.01
S028AARF	75.50±1.27	26.43±0.44	276.70±3.11	51.19±0.40	29.80±0.74	0.19	0.79±0.01

Fed-batch fermentations were done in a highly instrumented and automated 4-paralleled 1.5 L bioreactors system DASGIP for 65 hours; the initial glucose concentration was 36 g/L; the initial inoculation OD_{600} was 0.45. Experiments were performed in two biologically independent experiments. * GlcC is the calculated cumulative consumption per reactor volume; q_{Trp}, specific production rate of Trp; Yield, Trp production vs. glucose consumed; Vp, volumetric productivity.

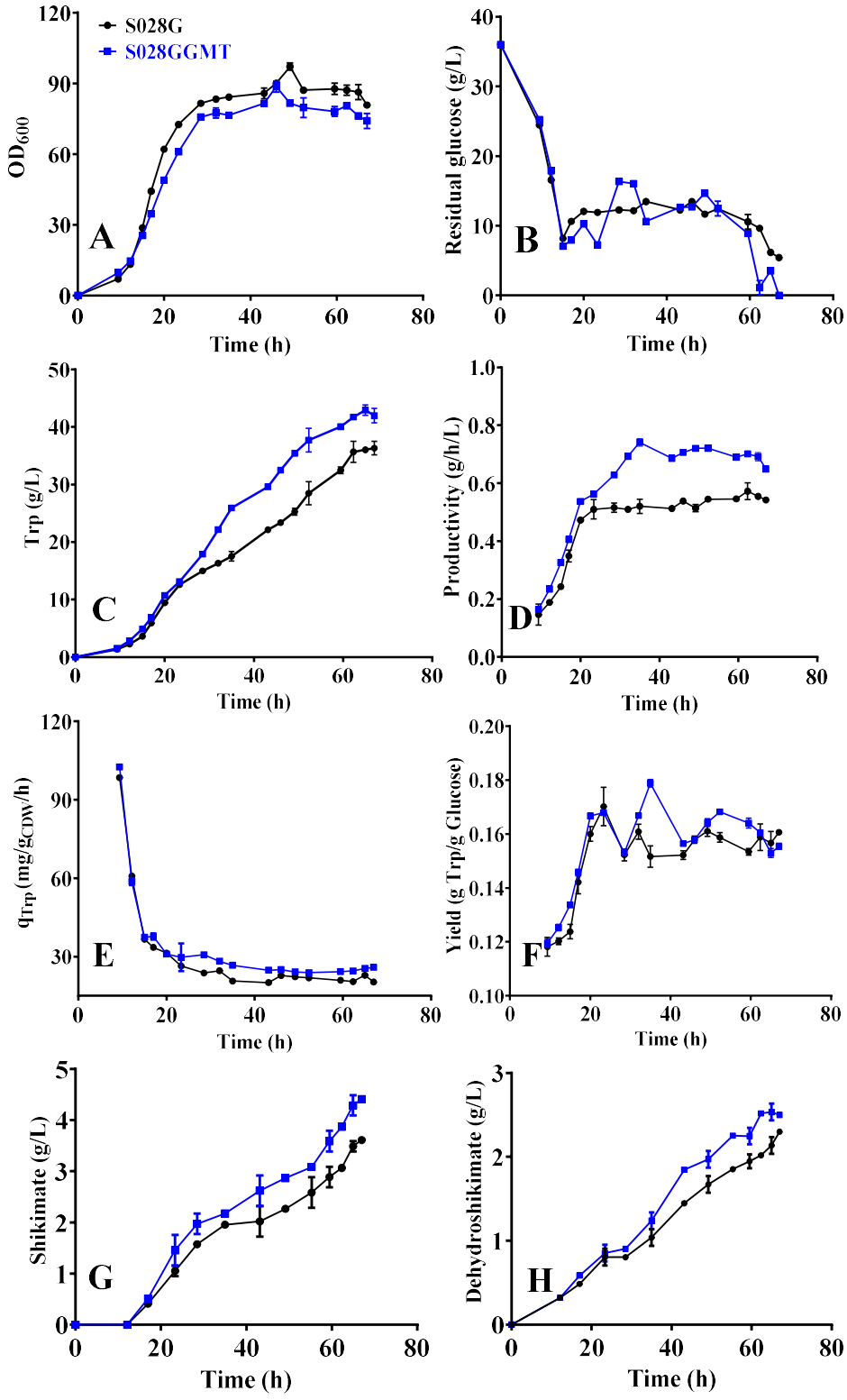

Figure 7.1 Fed-batch fermentation of the strains S028G (black and circle) and S028GGM (blue and square). (A) Cell growth; (B) Residual glucose concentration (g/L); (C) Trp production (g/L); (D) Overall productivity (g/L/h), (E) Formation rate of Trp (q_{Trp}, mg/g DCW/h), (F) Yield of Trp (g Trp/ g glucose) and Accumulation of the intermediates shikimate (G) and dehydroshikimate (H). The inoculated OD_{600} was 0.45.

7.3 Overexpression of *An*TrpC^{R378F} in strain S028GGMTΔ*trpR*

Subsequently, the gene *AntrpC* was integrated into the chromosome of the strain S028GGMT to enhance the flux in the direction of Trp biosynthesis. To this end, the all-in-one plasmids p*Cas9-AntrpC*^{WT} and p*Cas9-AntrpC*^{R378F} (Table 2.3) were constructed and used for the chromosome integration of *AntrpC*^{WT} and *AntrpC*^{R378F} in the strain S028GGMT, respectively. The resulting strains S028AAWT and S028AARF, together with the strain S028GGMT, were then carried out by fed-batch fermentation to evaluate their performances.

As seen in Fig. 7.2A, the trends of growth of the strains S028AAWT and S028AARF were similar to that of the strain S028GGMT. Moreover, a high biomass formation could provide a high Trp formation titer during the stationary phase (Figs. 7.2C and E). Remarkably, S028AARF produced 51.19±0.40 g/L of Trp (Table 7.1), which is 19.19% and 13.04% higher than those obtained by the strains S028GGMT (42.95±1.26 g/L, Table 7.1) and S028AAWT (46.05± 0.38 g/L, Table 7.1), respectively. It suggested that the metabolic flux for the biosynthesis of Trp was more streamlined after additionally overexpressing the *AntrpC*^{WT} or *AntrpC*^{R378} and the deletion of *trpR* gene in the strain S028GGMT. Recently, many attempts have been reported to develop a Trp-overproducing strain by defined genetic modifications (Chen and Zeng, 2017; Chen et al., 2018). Among these engineered strains, one of the strains S028 could produce around 40 g/L of Trp within 65 hours (Chen and Zeng, 2017). To our best of knowledge, the Trp producing capability of the strain S028AARF is the best one among those rationally designed strains reported so far.

Additionally, the specific formation rate of Trp in the strain S028AARF (29.80±0.74 mg/gDCW/h, Table 7.1) showed an advantage over those of the strains S028GGMT (24.77±0.23 mg/gDCW/h, Table 7.1) and S028AAWT (23.45±0.23 mg/gDCW/h, Table 7.1). Also, it was notable that Trp yield for both strains S028AAWT and S028AARF increased to around 0.19 g/g at 65h (Fig. 7.2F and Table 7.1), whereas, it was 0.16 g/g for the reference strain S028GGMT. These results obviously showed overexpressing the *AntrpC*^{WT} or *AntrpC*^{R378} in the strain S028GGMT improved the Trp yield very significantly (by approximately 20%). This result coincided with the previous result that the overexpression of the enzyme *An*TrpC had a beneficial effect on the yield of Trp (Fig. 4.8). As expected, less accumulation of intermediates SA and DSA in the strain S028AARF (Figs. 7.2.G and H). It reflected that integrating the variant *An*TrpC^{R378F} made the *trp* operon more effective. The performance of S028AARF in terms of the increased Trp production titer, rate, yield, and the decreased concentration of SA and DSA makes it more attractive for its industrial application.

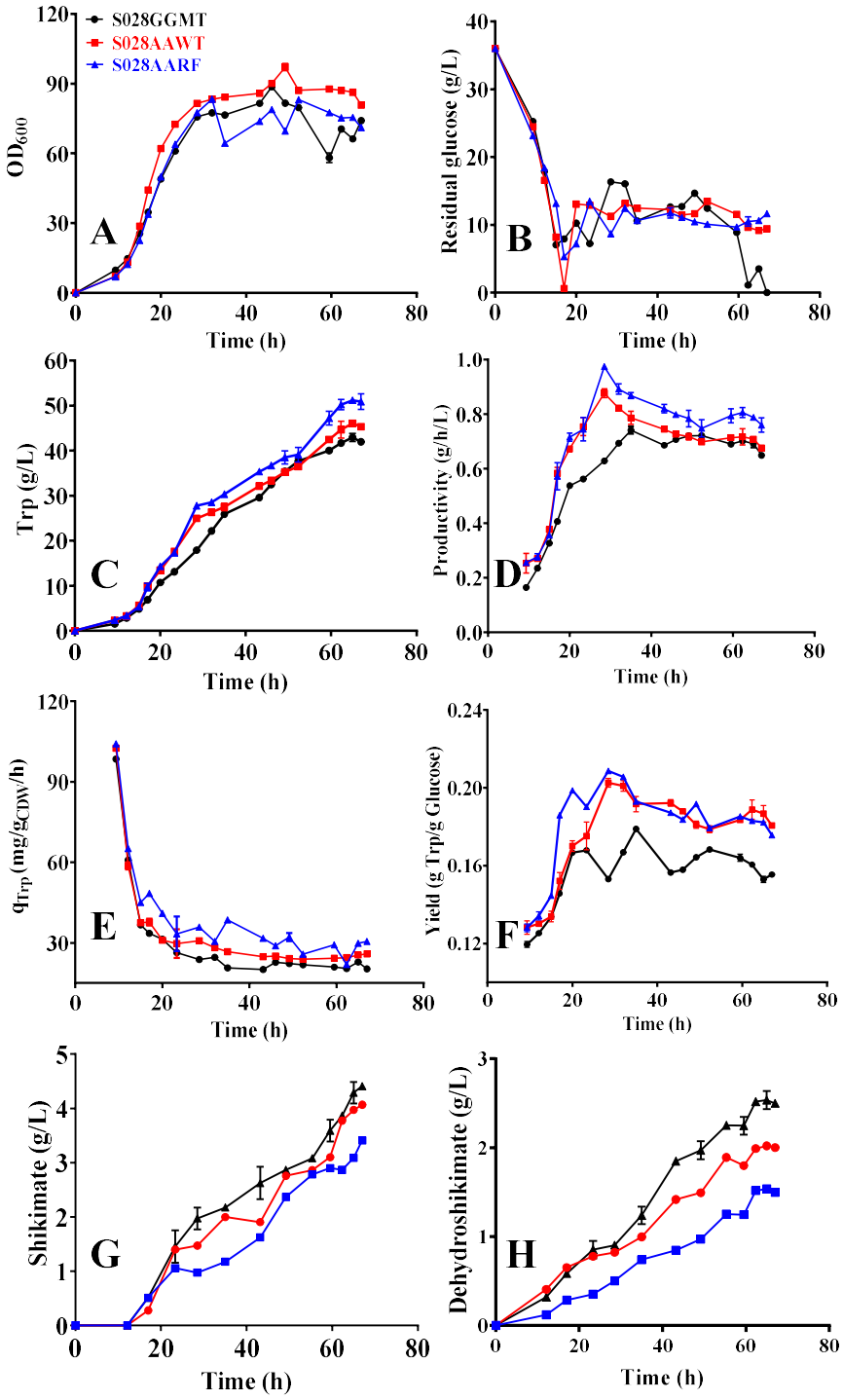

Figure 7.2 Fed-batch fermentation of the strains S028GGMT (black and circle), S028AAWT (red and square), and S028AARF (blue and triangle). (A) Cell growth; (B) Glucose concentration; (C) Trp production; (D) Overall productivity, (E) Formation rate of Trp (q_{Trp}), and (F) Production yield of Trp. Accumulation of the intermediates shikimate (G) and dehydroshikimate (H). The fed-batch fermentation was performed in a 1.5-L bioreactor. The initial OD_{600} was 0.45.

7.4 Integration of AroG$^{D6G-D7A}$ and *An*TrpC^{R378F} in strain D8

As observed above, additional expression of AroG$^{D6G-D7A}$ and *An*TrpC^{R378F} can lead to a better performance (as shown in Figs. 4.6E and 5.6C), since they are able to enhance the chorismate

pathway and the *trp* operon, respectively. In addition, to increase the availabilities of precursors has been proved to be a common and effective strategy for strain improvement in terms of production titer, yield, and productivity (Carmona et al., 2015; Chen et al., 2018; Lu et al., 2012). Indeed, the engineered strain D8 showed an increase of Trp yield (Table 6.1) after inactivation of the PEP-dependent PTS because of an increased availability of PEP. To further enhance the Trp production of the strain D8, $aroG^{D6G\text{-}D7A}$ and $AntrpC^{R378F}$ were subsequently integrated into its chromosome, resulting in a strain designated as D8AA. Thereafter, a comparison of the strain D8AA and the strain S028AARF was carried out by fed-batch fermentations.

For a biological replicate, two single clones that designated as D8AA-1 and D8AA-2 were picked out to evaluate their performances. Surprisingly, these clones showed strong variations in growth and Trp production (Figs. 7.3A and B). As seen from Fig. 7.3B, D8AA-2 had severe growth retardation. Its growth began to restore after 30h of retardation, and reached a final biomass concentration of 22.33±0.10 gDCW/L (Table 7.2). Because of the delayed exponential growth phase, D8AA-2 only produced 17.13±0.70 g/L of Trp. Furthermore, the Trp yield on glucose obtained by D8AA-2 was very low (0.10 g/g, Table 7.2). Unlike D8AA-2, the growth of D8AA-1 reached up to the stationary phase (60.12±0.74, Table 7.2) in 30h of cultivation without obvious retardation. D8AA-1 was able to reach a final Trp production of 41.49±0.53 g/L with a yield of 0.20 on glucose in 63h (Table 7.2).

Table 7.2: Comparison of the strains S028AARF and D8AA for L-Trp production in fed-batch fermentations

strains	**OD_{600}**	**DCW (g/L)**	**GlcC* (g/L)**	**Trp (g/L)**	**q_{Trp} (mg/gDCW/h)**	**Yield (g Trp/g Glc.)**	**Vp (g/L/h)**
S028AARF	92.20±0.28	32.27±0.10	260.92±1.21	48.27±0.29	23.74±0.07	0.19	0.77±0.01
D8AA-1	66.20±0.84	23.17±0.29	212.77±0.89	41.49±0.03	28.83±0.35	0.20	0.67±0.01
D8AA-2	63.80±0.28	22.33±0.10	174.80±0.21	17.13±0.70	12.17±0.44	0.10	0.27±0.01

Fed-batch fermentations were performed in 1.5 L bioreactors at 37°C and pH 6.8 for 63 hours; The initial glucose concentration was 35 g/L; The initial inoculation OD_{600} was 0.45. The data represent the means ± SD from three measurements. * GlcC is the calculated cumulative consumption per reactor volume.

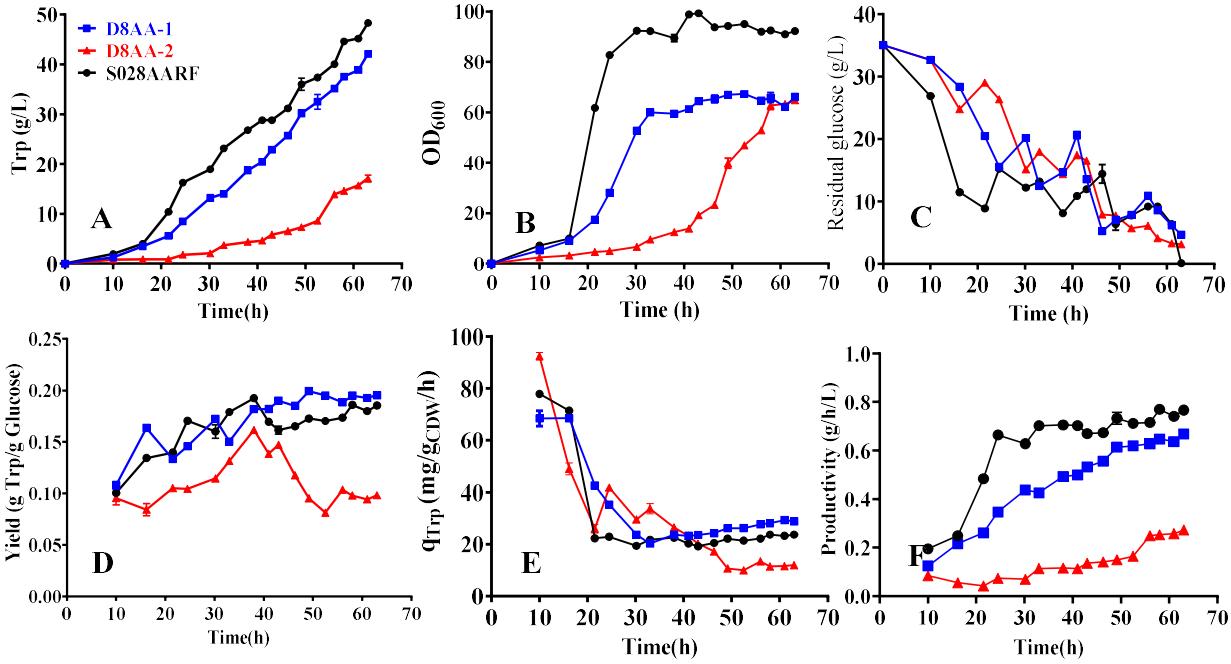

Figure 7.3 Fed-batch fermentations of the strains S028AARF (black and circle) and D8AA (red and square). (A) Cell growth; (B) Glucose concentration; (C) Trp production; (D) Specific formation rate of Trp (q_{Trp}); (E) Production yield of Trp; and (F) Overall productivity.

To further validate the variations of D8AA, three new single clones were selected for the second batch of fermentation. Indeed, the stronger variations in growth and Trp formation were also found in clones D8AA-3, -4, and -5 (Fig. 7.4). Among them, similar trends of growth and Trp formation were observed for D8AA-2, -3, and -5 (Fig. 7.4), and similar results were observed for D8AA-1 and -4 (Fig. 7.4). The cultivation conditions for these clones were identical in all the two repeating fermentations. Moreover, to confirm the unity of the D8AA clones, three single clones from the Cyro-stocks of D8AA-1 and -3 were selected and forced to batch fermentation in shake flasks. The results showed that all progeny clones from D8AA-1 exhibited similar growth trends to the parental strain D8AA-1, while all clones from D8AA-3 still suffered severe growth retardation (data not shown). The significant variations in growth and Trp production of D8AA clones were considered to be due to unknown mutations generated on the PTS-negative strain D8 regardless of the cultivation conditions. So far, the reason for the variations is not clear with our knowledge. Therefore, to characterize the genetic basis of D8AA clones exhibiting the variations in growth and Trp production, the genome sequences of the D8AA-1 and D8AA-3 clones are to be determined and compared.

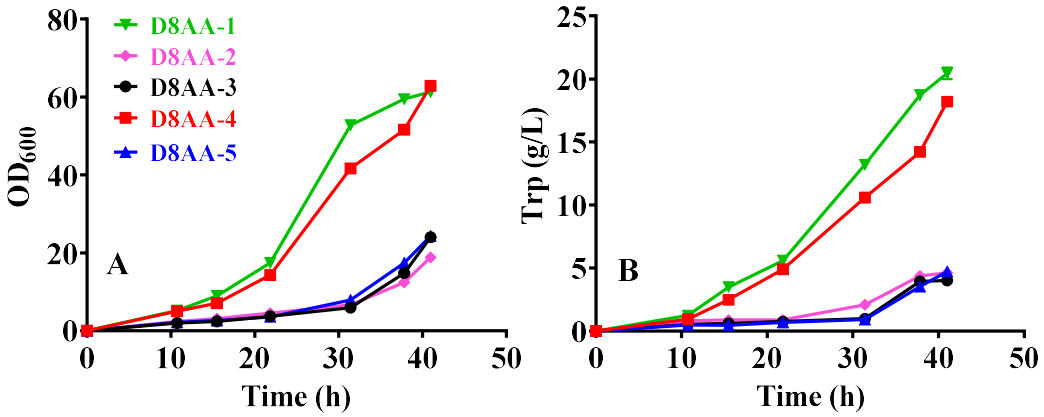

Figure 7.4 Cell growth (A) and Trp production (B) of the strain D8AA during fermentations in bioreactor. D8AA-1 and -2, single colonies 1 and 2 were selected for the first fermentation; D8AA-3, -4, and -5, single colonies 3, 4, and 5 were selected for the second fermentation. The initial glucose was 35 g/Land the initial OD600 = 0.45.

Compared with the parent strain D8, Trp concentration (41.49±0.03 g/L, Table 7.2) and yield (0.20 g/g, Table 7.2) achieved by D8AA-1 were both significant improved (by 7.03% and 21%, respectively). These results again demonstrated that additional expression of *AnTrpC*^{R378F} and *AroG*$^{D6G\text{-}D7A}$ facilities the biosynthesis of Trp, making it a powerful strategy for the improvement of Trp-producing strain. Also, expression of the enzyme variants in the strain D8 contributed to the glucose conversion yield. The results showed that Trp yield of D8AA-1 (0.20 g/g, Table 7.2) is remarkably higher than that of S028AARF (0.19 g/g, Table 7.2). It was consistent with the previous result that activation of GalP/Glk-facilitated glucose uptake system in PTS-negative strain could be beneficial for the Trp yield (Fig. 6.9E). Although the Trp production of D8AA-1 exhibited a lower level than that of S028AARF (41.49±0.03 g/L vs. 48.27±0.29 g/L, Table 7.2), as expected, the q_{Trp} of D8AA-1 (28.83±0.35 mg/gDCW/h) is 1.22-fold higher than that of S028AARF (23.74±0.07 mg/gDCW/h, Table 7.2). Since both strains contained the enzyme variants *An*TrpC^{R378F} and AroG$^{D6G\text{-}D7A}$, an improved q_{Trp} could be mainly resulted from the increased availability of PEP.

The cause of a lower Trp production is the lower biomass formation level of D8AA-1. Biomass achieved in the fed-batch cultivation of D8AA-1 was 23.17±0.29 gDCW/L (Table 7.2), which was 28.20% lower than that of the strain S028AARF (32.27±0.10 gDCW/L, Table 7.2). As discussed in Chapters 7.2 and 7.3, to some extent the biomass formation level is one of the critical parameters for the efficient formation of Trp in terms of the production and productivity (Fig. 7.1A and Fig. 7.2A). Thus, further development of D8AA could be identifying and removing the inhibition factor(s) involved in biomass formation.

7.5 Conclusions

Enzyme engineering has been widely used for Trp-producing strain improvement towards a higher titer, rate, and yield (Chen, 2017; Ger et al., 1994; Ikeda, 2006b; Carmona et al., 2015). In the previous chapters, two enzyme variants $AroG^{D6G-D7A}$ and $AnTrpC^{R378F}$ with desired functional characteristics were successfully selected using the growth-coupled screening and sensor-guided *in vivo* characterization. They were proved to be a benefit for improving Trp production effectively but separately. In this chapter, a genetically stable and Trp-producing microbial cell factory S028AARF was successfully developed by the integrations of $AroG^{D6G-D7A}$ and $AnTrpC^{R378F}$. The engineered strain S028AARF exhibited the advanced production capacity to the previously reported strains. Specifically, the strain S028AARF produced 51.19±0.40 g/L of Trp with the overall productivity of 0.79±0.01g/L/h, the specific production rate of 29.80±0.74 g/g DCW/h, and the formation yield of 0.19 g/g (Table 7.1). Considering the strain D8, which has an inactivated PTS system and an engineered GalP/Glk-facilitated glucose uptake system, had a much higher glucose conversion yield than the parent strain. Both of $aroG^{D6G-D7A}$ and $AntrpC^{R378F}$ were also reasonably integrated into the chromosome of D8 for developing the strain D8AA. Surprisingly, fed-batch cultivations of D8AA clones showed strong variations in growth and Trp production (Figs. 7.4A and B). So far, the reason for the variations is not clear with our knowledge. Nevertheless, one of the clones D8AA-1 exhibited a Trp yield as high as 0.20 g/g (vs. 0.19 g/g with S028AARF), representing the highest Trp yield reported in the literature so far and making it attractive for industrial-scale Trp bioproduction.

Chapter 8

8. Summary and perspectives

The main objective of this work was to design and construct Trp producing *E. coli* strain(s) with more robust growth and more efficiency in Trp production and yield by the engineering of key enzymes involved in the Trp pathway. To effectively *in vivo* screen and characterize enzyme variants for the above purpose, three approaches for library screening and characterization were developed and examined for engineering enzyme variants with desired performances and characteristics. Subsequently, integrative metabolic engineering was performed to assemble the enzyme variants into a previously constructed Trp-producing strain S028 (Chen, 2017). The main results are summarized below.

8.1 Summary of this thesis

8.1.1 Plasmid-based growth-coupled screening and sensor-guided *in vivo* characterization of TrpC variants

In a previous study from our group (Chen et al., 2018), the bifunctional enzyme TrpC of *E. coli* (*Ec*TrpC) was discovered to be feed-forward inhibited by anthranilate. Furthermore, TrpC of *A. niger* (AnTrpC) was found to be feed-forward activated by anthranilate. Engineering of TrpC to alleviate the inhibition or further increase the activation is thus one of the critical issues to develop a more efficient Trp biosynthetic pathway. In Chapter 4, an efficient approach was proposed which combines the complementary cell-auxotrophy coupled screening and biosensor-driven *in vivo* characterization. The effectiveness of this approach was demonstrated for the engineering of enzymes *Ec*TrpC and *An*TrpC.

Starting from the strain S028, a Trp-auxotrophic strain S093 was designed, in which the cell growth is linked directly to the catalytic efficiency of TrpC variants, and the intracellular concentration of Trp is monitored by the fluorescence intensity via a Trp sensor. Subsequently, the necessity of Trp for the cell growth was tested by cultivating the strain S093 in the M9 medium. It was found that S093 was unable to grow in synthetic medium, while the auxotrophy was alleviated by supplementation of the lost enzymatic function. Moreover, the growth and the fluorescence intensity of S093 cells were found to correlate strongly with the extracellular

Trp concentration in a range from 0 to 4 mM (Fig. 4.1D). These results revealed that the developed platform offered a higher selectivity and a more robust sensitivity screening threshold for the engineering of TrpC.

After testing the sensitivity and selectivity of the screening platform, one set of *EctrpC* gene variant libraries was constructed by codon saturation mutagenesis on the rational target residues (S58-P59-S60-K61, Fig. 4.2A). Using the PGSS approach, a highly efficient anthranilate-resistant candidate, *Ec*TrpC$^{S58Q\text{-}P59V\text{-}S60F\text{-}K61Q}$ (*Ec*TrpC^{QVFQ}), was successfully identified. *In vitro* enzyme assay also indicated that only 52% of its activity was lost under 3 mM of anthranilate. In contrast, the activity of the wild-type enzyme was inhibited entirely under the same condition (Fig. 4.3B). Also, the catalytic efficiency (kcat/km) of *Ec*TrpC^{QVFQ} was similar to that of the wild-type enzyme (Table 3.1). To some extent, it suggested that the anthranilate-resistant mutant *Ec*TrpC^{QVFQ} is a better candidate for enhancing strain tolerance against a high concentration of anthranilate when the strain is applied to the industrial scale. After testing the feasibility of the PGSS approach, it was subsequently employed to develop another anthranilate-activated enzyme *An*TrpC. In the part of this work, the catalytic site Arg378 of *An*TrpC was selected for codon saturation mutagenesis, and then the resulting mutants were also screened and evaluated using the same procedure as above. One of the mutants, *An*TrpC^{R378F}, was identified that showed increased Trp formation. *In vitro* enzyme assay also revealed that the variant *An*TrpC^{R378F} had more potential of the activation effect. Subsequently, the performances of the strain *An*TrpC^{R378F} were compared with that of the reference strain *An*TrpC^{WT} and the parental strain S028 by carrying out fed-batch fermentation in bioreactors. As presented in Fig. 4.8, fed-batch fermentation of *An*TrpC^{R378F} enabled the production of 35.36±0.81 g/L of Trp, which is 13.52% higher than that produced by *An*TrpC^{WT} (31.15±1.01 g/L, Table 4.2) at 51 h, indicating that the variant *An*TrpC^{R378F} is more efficient for the Trp biosynthetic pathway.

8.1.2 CRISPR/Cas9-facilitated engineering with growth-coupled screening and sensor-guided *in vivo* characterization of AroG variants

One of the isozymes of DAHP synthase, AroG enzyme, is feedback inhibited by the phenylalanine in *E. coli*. Engineering of the enzyme AroG to eliminate the feedback inhibition is another critical issue to develop a more efficient Trp biosynthetic pathway. In Chapter 5, another approach was developed which combines the integration of gene variants directly into the chromosome of the host cell with *in vivo* screening and characterization of the enzyme

variants. To demonstrate this method, the optimization of the enzyme AroG in the Trp pathway was implemented as an example.

Starting from the strain S028 (harboring a reference AroG variant, $AroG^{S180F}$), an aromatic amino acids-auxotrophic strain was first developed by removal of all the endogenous DAHP synthases. As expected, the auxotrophic strain was not able to grow in the M9 medium even when the wild-type enzyme AroG was reintroduced, but the auxotrophic strain was able to restore its growth by reintroducing the reference variant $AroG^{S180F}$ (Fig. 5.4). Consequently, the growth of the auxotrophic strain was directly linked to the activity of the introduced AroG variants. In principle, an engineered enzyme variant AroG with a higher activity should lead to a faster accumulation of Trp, which in turn stimulates the expression of a report gene regulated by a Trp biosensor.

To demonstrate the utility of this auxotrophic strain to achieve AroG variants with higher resistance and catalytic efficiency, a mutation library of the *aroG* gene was first generated. To this end, the saturation mutagenesis was adopted to the Phe-binding sites D6 and D7 of AroG protein and then introduced them into the target plasmid pCm-$aroG^{fbr}$. The *aroG* gene variants were then integrated into the chromosome of the auxotrophic strain by the technique of CRISPR/Cas9. After several rounds of screening and characterization, one of the mutants, $AroG^{D6G\text{-}D7A}$, was identified. *In vitro* enzyme assay showed that the variant $AroG^{D6G\text{-}D7A}$ exhibited a higher specific activity at the presence of 10 mM Phe compared to other variants and the reference variant $AroG^{S180F}$ (Fig. 5.6). Thereafter, the replacement of $AroG^{S180F}$ with the best variant ($AroG^{D6G\text{-}D7A}$) in a Trp-producing strain S028 was performed, resulting in the strain S028GM1. The strain was then carried out together with the reference strain S028 by fed-batch fermentation in bioreactors to examine the impact of the variant $AroG^{D6G\text{-}D7A}$ on Trp production. Fig. 5.7 shows that the strain S028GM1 was able to remarkably increase the Trp production by 38.5% compared to the strain S028 at 37h (24.03±1.02 g/L vs. 17.35±1.16 g/L). These results demonstrated that the variant $AroG^{D6G\text{-}D7A}$ is more efficient for the chorismate pathway and demonstrates the high efficiency of the CGSS approach.

8.1.3 CGSS-facilitated optimization of GalP/Glk-dependent glucose utilization system

A high glucose conversion yield is an essential parameter for cost-effective production. It was assumed that activation of galactose permease/glucose kinase (GalP/Glk) in a PTS-negative strain could provide an alternative glucose utilization system with an increased Trp yield. Indeed, the maximum theoretical Trp yield in a GalP/Glk-dependent strain was calculated to be 0.45 g/g, which is approximately twice that of a PTS-positive strain (0.23 g/g, Fig. 1.3).

To investigate the potential achievable Trp yield in the GalP/Glk-dependent glucose utilization strain, one of the phosphoryl transfer proteins, PtsI protein (EI, Fig. 6.1), was first removed from the strain S028TS/pCas9. Afterward, a tandem gene circuit *ptac-galp-pJ23119-glk* was integrated into the chromosome of the PTS-negative strain, resulting in the strain S028Δ*ptsI*::GG^{WT} (G028). However, it was revealed that after 70 hours of cultivation, G028 suffered severe growth retardation (data not shown). One conceivable explanation was that the co-expression of *galP* and *glk* genes was disturbed or unbalanced since the transcription of both genes *galP* and *glk* were controlled under two stronger promoters, p*tac* and p*J23119*, respectively. Therefore, using the CRISPR/Cas9 technique, the promoter variants were integrated into the p*tac* and p*J23119* promoters to realize the modulation expression of *galP* and *glk* genes, resulting in the promoter mutants S028Δ*ptsI*::GG^{MT} (G028a). The resulting mutants were further subjected to batch adaptive evolution, and one promising candidate B3 was obtained, and it proved to be advantageous for the Trp yield in shake flask cultivation (data not shown). However, in a subsequent test in fed-batch fermentation, it was found that the yield (0.13 g/g in 50h, Fig. 6.5) and production titer (28.17±0.05 g/L in 50h, Fig. 6.5) of the strain B3 exhibited similar values to those of the reference Trp-producing strain S028. One conceivable explanation is that the PTS-negative strain is to synthesize intermediates for supporting their growth rather than Trp synthesis. In this regard, a continuous evolution of the GalP/Glk-dependent glucose utilization strain with the aid of a continuous mutagenesis method was conducted to evolve the strain with the desired phenotype and function.

In the part of this work, a novel and scalable continuous evolution approach with real-time measurement of cell growth and online monitoring of fluorescence intensity was established for the adaptive evolution of the GalP/Glk-dependent glucose utilization strain (Fig 6.6). Unlike conventional discrete evolution methods, this approach enables a broader and genome-wide random and continuous mutagenesis of strains. In this regard, the promoter library (p*Cm*N20-GG^{MT}) was integrated into the chromosome of S028TSΔ*ptsI*Δ*glk*::*Cm*^{R} /pCas9-pJC184 (G028JC) strain in the same way as described in Chapter 6.3. The resulting promoter mutants were then inoculated in the cultivation vessel for further adaptive evolution (Fig. 6.6). The growth of the promoter mutants was monitored in real-time (every 10 min). As soon as the OD_{600} reached to approximately 0.8, the system was operated in an auxostat mode. For real-time *in vivo* characterization of the mutants, the cultivation vessel was connected directly to a cell density meter and a flow cytometer to monitor the cell growth and fluorescence intensity in real-time, respectively. In the scheme for continuous evolution, the inlet flow rate of the fresh medium and the outlet flow rate of waste culture were dynamically regulated via a peristaltic

pump in response to the real-time monitored cell density so that a constant cell density at around 0.8 was maintained. In this case, the mutants with desired properties were to be enriched in the cultivation vessel, but the undesired mutants were to be excluded from the vessel (Fig. 6.6). As expected, a new promising candidate D8 was identified. Subsequently, fed-batch fermentation of the strain D8 found that the Trp yield of D8 showed 22.56% higher than that of strain B3 (0.16 g/g vs. 0.13 g/g, Fig. 6.9). Also, the specific Trp production rate of D8 was simultaneously increased to 26.64±0.21 mg/g DCW/h (Fig. 6.9). These results demonstrated that *in vivo* continuous mutagenesis method is superior to the conventional discrete method and that the developed approach for *in vivo* mutagenesis approach is highly efficient.

8.1.4 Engineering microbial cell factories: Assembling of selected enzyme variants in Trp-producing strains

In Chapter 7, iterative metabolic engineering was applied to integrate the gene variants (*aroG*$^{D6G\text{-}D7A}$ and *AntrpC*^{R378F}) into the chromosome of previous Trp-producing strain S028G or D8 to develop a Trp-overproducing microbial cell factory. A comparison of the performance of strain D8AA to strain S028AARF was performed in the part of this work.

Using one of the previous Trp-producing strains S028G as starting strain, the gene variants *aroG*$^{D6G\text{-}D7A}$ and *AntrpC*^{R378F} were subsequently integrated into the chromosome of the host strain to develop a Trp-hyperproducing strain S028AARF. The performance of S028AARF was further evaluated in well-controlled fed-batch fermentation. Remarkably, that after 65 hours of cultivation, S028AARF was able to produce a very high Trp production (51.19±0.40 g/L, Table 7.1), which is approximately 19.19% higher than that of S028GGMT (42.95±0.26 g/L) and about 13.04% more than that of S028AAWT (46.05±0.38 g/L). Moreover, the Trp yield of S028AARF was raised to 0.19 g/g at 65h (Fig. 7.2F), which is almost 20% higher than that of S028G. To date, many attempts have recently been reported to develop a Trp overproducing strain through defined genetic modifications or directed evolution (Chen and Zeng, 2017; Chen et al., 2018; Panichkin et al., 2016). Among these engineered strains, one of the Trp-producing strains S028 was reported to produce around 40 g/L of Trp within 65 hours of cultivation (Chen and Zeng, 2017). Thus, the engineered strain S028AARF exhibited the merits of comparable production capacity to the previously reported strains.

In addition to the biosynthetic pathway itself, to increase the availabilities of precursors (eg., PEP, E4P, and L-Ser) has been proved to be a common and effective strategy for improving production titer (Carmona et al., 2015; Chen et al., 2018; Lu et al., 2012). Indeed, a previous study on the strain D8 showed an increase of the Trp yield due to the availability of PEP (Table

6.1). To further enhance the Trp yield and Trp formation, *aroG*$^{D6G-D7A}$ and *AntrpC*^{R378F} were also subsequently integrated into the chromosome of strain D8, resulting in the strain D8AA. Thereafter, a comparison of the performance of the strains D8AA and S028AARF was carried out using fed-batch fermentation. Surprisingly, fed-batch cultivations of D8AA clones showed strong variations in growth and Trp production (Figs. 7.4A and B). Since the cultivation conditions for D8AA clones were identical in all fermentations, the significant variations of D8AA clones were suspected due to unknown mutations generated on the PTS-negative strain D8 regardless of the cultivation conditions. So far, the reason for the variations is not clear with our knowledge. Therefore, to characterize the genetic basis of D8AA clones exhibiting the variations in growth and Trp production, the genome sequences of the D8AA-1 and D8AA-3 clones are to be determined and compared. Nevertheless, fed-batch cultivation of clone D8AA-1 exhibited a higher level of Trp yield (0.20 g/g, Table 7.2) than S028AARF (0.19 g/g, Table 7.2); however, final titer of Trp formation in D8AA-1 was less than that in S028AARF (41.49±0.03 g/L vs. 48.27±0.29 g/L at 63h). Reasonably, the Trp yield achieved in D8AA-1 is higher than that reported in literature so far, making it attractive for industrial-scale Trp bioproduction.

8.2 Outlook for future work

8.2.1 Further prospects and optimizations of the CGSS approach

As an efficient approach for *in vivo* screening and characterization of enzyme variants, CGSS approach facilitates the integration of gene variants into the chromosome of the host microorganism and examines the performance of enzyme variants under intracellular environments.

The targeted microbial genome editing has recently played a prominent role in protein engineering. For instance, the multiplex automated genomic engineering (MAGE) approach could simultaneously target many genes for genome engineering (Wang et al., 2009). However, it suffered low editing efficiency and a lack of traceability. Owing to the simplicity and high efficiency of the technique of CRISPR/Cas9, multiplex genome editing has become possible for engineering of various industrial microbes. On the basis of MAGE, several approaches have emerged for efficient construction of large-scale and trackable libraries through combination with CRISPR/Cas9 (Reisch and Prather, 2015; Ronda et al., 2016). For instance, CRISPR enabled trackable genome engineering (CREATE) technique was developed with the capacity of highly efficient CRISPR editing, DNA barcoding, and multiplexed rational design, and it has been applied in site-directed saturation mutagenesis of target enzymes, reconstruction of

adaptive laboratory evolution, and identification of stress tolerance genes (Garst et al., 2017). Based on CREATE, an iterative CREATE strategy (iCREATE) was adopted for engineering of the utilizing pathway of sugar mixture to achieve a tolerance against conventional hydrolysate inhibitors in *E. coli* (Liu et al., 2018). Collectively, the capacity of multiplex genome editing involved in these approaches can be integrated into the CGSS approach to achieve a broad range of targets that require a simultaneous modulation of multiple genes.

Nevertheless, CRISPR/Cas9-facilitated dsDNA break repair induces cellular stress responses and leads to increased off-target mutagenesis rates. Insufficient understanding of sgRNA sequence design may reduce the efficiency of Cas9 protein-facilitated cleavage at the select loci. Therefore, due to the restricted efficiency of transformation and recombination, the CGSS approach can only be applied in less than two site mutations for protein engineering. Typically, the number of theoretical mutation colonies for random mutagenesis of two residues should be more than 380 in order to cover all possibilities, when 95% reliable coverage is required. However, Fig. 5.5B shows that the maximum amount of enzyme mutants generated by using agar plate-based screening was approximately 400. If the number of mutations continues to increase, the number of colonies generated on the agar plate may not cover all theoretical possibilities. Ideally, the aforementioned drawback can be optimized by enhancing the efficiency of transformation and recombination, by increasing the gene variant library in genome context, or by linking the CGSS approach with the naturally modified mutators such as MP plasmid (pJC184, Badran and Liu, 2015a), MutaT7 (Moore et al., 2018), and polymerases DNAP *Pol*3 M (Halperin et al., 2018) (Table 2.2). In this regard, combining the CGSS approach with the naturally modified mutators makes generation of the random mutagenesis within the specific, well-defined DNA regions without construction of gene variants *in vitro* and followed by transformation of gene variants.

8.2.2 Further improvements of continuous evolution approach

In Chapter 6, a CGSS-facilitated continuous automatic adaptive evolution platform was developed and successfully demonstrated in evolving the GalP/GlK-dependent glucose utilization strain.

As reported in Fig. 6.9, fed-batch fermentation of the strain D8 resulted in an increase of Trp yield (0.16 g/g at 65h), which is 13.10% higher than that of the strain S028. However, obviously, it is almost 2.5-fold as low as the maximum theoretical yield (0.45 g/g, Fig. 1.3B). Due to the fact that no evolutionary pressure was imposed on the Trp biosynthetic pathway during continuous evolution, the PTS-negative cell was forced to enhance its growth rather than the

Trp synthesis. In this regard, additional evolutionary pressures such as the auxotrophy for Trp synthesis are to be imposed during the continuous evolution of the strain GG^{MT}. In practice, the target enzymes in Trp pathway, such as AroCK enzymes or TrpEDCBA operon, can be considered to be knocked out from the GG^{MT} strain, resulting in a double-auxotrophic strain (Trp- and PTS-auxotrophic strain). In this regard, only the GalP/Glk-facilitated glucose utilization system regains its ability and the corresponding mutant contains the key enzyme variants with desired performances; the mutants will remain dominant during the continuous evolution. Correspondingly, the inlet flow rate of the fresh medium and outlet flow rate of the waste culture are dynamically regulated in response to the fluorescence intensity instead of cell density, so that only the fluorescence intensity of candidates over a threshold remains dominant in the cultivation vessel, whereby the threshold can be manually adjusted in accordance with the fluorescence intensity of the mutants.

Although precise control of the cultivation parameters (e.g., pH, temperature, and nutrient conditions) and the characterization methods (e.g., optical density and fluorescence intensity) have been achieved by continuous evolution, single-phase continuous cultivation systems are susceptible to microbial contamination (Faassen and Hitzmann, 2015). Microdroplet-assisted continuous cultivation can effectively address this problem by compartmentalization, i.e., by encapsulating microbial cells in droplets on the scale of microliters or even nanoliters (Bowman and Alper, 2019; Zeng et al., 2020). For instance, a microbial microdroplet culture system (MMC) has been developed by Jian et al., (2019) for automated high-throughput cultivation and adaptive evolution of microorganisms (Jian et al., 2019). Moreover, cultivation of microbes is necessary for microbial sorting, identification, and screening. Therefore, development of sorting approaches is demand for identification of mutants with desired performance during continuous evolution, i.e., the fluorescence-activated cell sorting (FACS) or microfluidic fluorescence-activated droplet sorting (mFADS). To date, the compartmentalization assays in droplets make the microfluidic flow sorting system the most flexible and widely used screening platform. In practice, individual mutants are compartmentalized into emulsion droplets, and the mutants can be sorted by dielectrophoresis in a fluorescence-activated manner (as in FACS).

8.2.3 Further developments of the Trp-producing strain D8AA

As discussed in section 7.4, inefficient Trp production of D8AA-1 may be caused by insufficient formation of biomass. Indeed, a sufficient specific formation rate of Trp was observed during the production phase and thus it is expected that Trp production of the strain S028AARF could be enhanced by generating more biomass (Fig. 7.2). Therefore, it is

promising to further develop D8AA-1 to generate sufficient biomass for an increased productivity of Trp.

As presented in Fig. 6.1, disruption of the PTS system in *E. coli* aims to block the conversion of PEP to pyruvate and switch PEP to the Trp pathway. Since pyruvate is an essential precursor for many intermediates and for energy metabolism in the TCA cycle. Thus, activation of the enzymes involved in pyruvate formation is one of the promising strategies for supporting the growth of PTS-negative strain. These enzymes include pyruvate kinases encoded by *pykA* and *pykF*, anthranilate synthase encoded by TrpE, and carboxylase encoded by *ppc* (Fig. 6.1 and Fig. 1.2). In theory, overexpression of these enzymes in the PTS-negative strain can to some extent contribute to recovery of cell growth to the same level as the PTS-positive strain. However, with this strategy, the precursor of PEP will be mainly used for the energy metabolism to support cell growth rather than Trp biosynthesis (Fig. 1.3). Therefore, there is a need to establish a dynamic regulation at the key node of PEP in a PTS-negative strain. With this dynamic regulation, enzymes like PykAF could be activated during the growth phase; subsequently, PykAF could be inactivated and another enzyme involved in the Trp pathway such as AroG is switched on once the cell enters into the stationary phase. In this case, PEP is mainly utilized for Trp biosynthesis during the stationary phase. So far, various well-characterized models for stationary phase-responsive regulation of gene expression have been developed and applied for microbial production, e.g., stationary phase sigma factor RpoS (Hengge-Aronis, 1993), quorum sensing (QS) (Miller and Bassler, 2001), and growth/stationary phase-dependent promoter and degradation degron (Gao et al., 2019). Specifically, Gao et al., (2019) designed a dynamic control circuit using a growth/stationary phase-dependent promoters and degrons for shikimic acid production. It was found that uncoupling cell growth from the production phase has significant potential for shikimic acid production. Thus, applying these stationary phase-responsive regulators at the node of PEP for dynamically controlled expression of enzymes PykAF and AroG in PTS-negative strain is a promising strategy for microbial production of Trp.

9. References

Aguilar, C., Martínez-Batallar, G., Flores, N., Moreno-Avitia, F., Encarnación, S., Escalante, A., and Bolívar, F. (2018). Analysis of differentially upregulated proteins in *ptsHIcrr*- and *rppH*- mutants in *Escherichia coli* during an adaptive laboratory evolution experiment. Appl Microbiol Biotechnol *102*, 10193-10208.

Aiba, S., Imanaka, T., and Tsunekawa, H. (1980). Enhancement of tryptophan production by *Escherichia coli* as an application of genetic engineering. Biotechnol Lett *2*, 525-530.

Albanese, A.A., HIGGONS, R.A., Hyde, G.M., and Orto, L. (1956). Lysine and tryptophan content of proteins and their utilization for human growth. AM J Clin Nutr *4*, 161-168.

Alper, H., Fischer, C., Nevoigt, E., and Stephanopoulos, G. (2005). Tuning genetic control through promoter engineering. PNAS *102*, 12678-12683.

Alper, H., Moxley, J., Nevoigt, E., Fink, G.R., and Stephanopoulos, G. (2006). Engineering yeast transcription machinery for improved ethanol tolerance and production. Science *314*, 1565-1568.

Alva, A., Sabido-Ramos, A., Escalante, A., and Bolívar, F. (2020). New insights into transport capability of sugars and its impact on growth from novel mutants of *Escherichia coli*. Appl Microbiol Biotechnol, 1-17.

Amann, E., Ochs, B., and Abel, K.-J. (1988). Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene *69*, 301-315.

Amrein, B.A., Steffen-Munsberg, F., Szeler, I., Purg, M., Kulkarni, Y., and Kamerlin, S.C.L. (2017). CADEE: Computer-aided directed evolution of enzymes. IUCrJ *4*, 50-64.

Arnold, F.H., and Georgiou, G. (2003). Directed enzyme evolution: screening and selection methods (Springer Science & Business Media).

Arzumanyan, G.A., Gabriel, K.N., Ravikumar, A., Javanpour, A.A., and Liu, C.C. (2018). Mutually orthogonal DNA replication systems *in vivo*. ACS Synth Biol *7*, 1722-1729.

Atsumi, S., and Liao, J.C. (2008). Directed evolution of *Methanococcus jannaschii* citramalate synthase for biosynthesis of 1-propanol and 1-butanol by *Escherichia coli*. Appl Environ Microbiol *74*, 7802-7808.

Azuma, S., Tsunekawa, H., Okabe, M., Okamoto, R., and Aiba, S. (1993). Hyper-production of L-trytophan *via* fermentation with crystallization. Appl Microbiol Biotechol *39*, 471-476.

Badawy, A.A. (2017). Kynurenine pathway of tryptophan metabolism: regulatory and functional aspects. International Journal of Tryptophan Research *10*, 1178646917691938.

Badran, A.H., Guzov, V.M., Huai, Q., Kemp, M.M., Vishwanath, P., Kain, W., Nance, A.M., Evdokimov, A., Moshiri, F., and Turner, K.H. (2016). Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. Nature *533*, 58-63.

Badran, A.H., and Liu, D.R. (2015a). Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra. Nature Commun *6*, 8425.

Badran, A.H., and Liu, D.R. (2015b). *In vivo* continuous directed evolution. Curr Opin Chem Biol *24*, 1-10.

Bae, S., Park, J., and Kim, J.-S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics *30*, 1473-1475.

Báez, J.L., Bolívar, F., and Gosset, G. (2001). Determination of 3-deoxy-D-arabino-heptulosonate 7-phosphate productivity and yield from glucose in *Escherichia coli* devoid of the glucose phosphotransferase transport system. Biotechnol Bioeng *73*, 530-535.

Baret, J.-C., Miller, O.J., Taly, V., Ryckelynck, M., El-Harrak, A., Frenz, L., Rick, C., Samuels, M.L., Hutchison, J.B., and Agresti, J.J. (2009). Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity. Lab on a Chip *9*, 1850-1858.

Bergquist, P.L., Reeves, R.A., and Gibbs, M.D. (2005). Degenerate oligonucleotide gene shuffling (DOGS) and random drift mutagenesis (RNDM): two complementary techniques for enzyme evolution. Biomol Eng *22*, 63-72.

Binder, S., Siedler, S., Marienhagen, J., Bott, M., and Eggeling, L. (2013). Recombineering in *Corynebacterium glutamicum* combined with optical nanosensors: a general strategy for fast producer strain generation. Nucleic Acids Res *41*, 6360-6369.

Bischoff, L., Berninghausen, O., and Beckmann, R. (2014). Molecular basis for the ribosome functioning as an L-tryptophan sensor. Cell Rep *9*, 469-475.

Boville, C.E., Romney, D.K., Almhjell, P.J., Sieben, M., and Arnold, F.H. (2018). Improved synthesis of 4-cyanotryptophan and other tryptophan analogues in aqueous solvent using variants of TrpB from *Thermotoga maritima*. J ORG CHEM *83*(14), 7447-7452.

Bowman, E.K., and Alper, H.S. (2019). Microdroplet-assisted screening of biomolecule production for metabolic engineering applications. Trends Biotechnol 38(7), 701-714.

Bridges, B.A., and Woodgate, R. (1985). Mutagenic repair in *Escherichia coli*: products of the *recA* gene and of the *umuD* and *umuC* genes act at different steps in UV-induced mutagenesis. PNAS *82*, 4193-4197.

Brödel, A.K., Isalan, M., and Jaramillo, A. (2018). Engineering of biomolecules by bacteriophage directed evolution. Curr Opin Biotechnol *51*, 32-38.

Brödel, A.K., Jaramillo, A., and Isalan, M. (2016). Engineering orthogonal dual transcription factors for multi-input synthetic promoters. Nature Commun *7*, 13858.

Caligiuri, M., and Bauerle, R. (1991). Identification of amino acid residues involved in feedback regulation of the anthranilate synthase complex from *Salmonella typhimurium*. Evidence for an amino-terminal regulatory site. J Biol Chem *266*, 8328-8335.

Calvo, J.M., and Matthews, R.G. (1994). The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. Microbiol Mol Biol Rev *58*, 466-490.

Carlson, J.C., Badran, A.H., Guggiana-Nilo, D.A., and Liu, D.R. (2014). Negative selection and stringency modulation in phage-assisted continuous evolution. Nat Chem Biol *10*, 216-222.

Carmona, S.B., Moreno, F., Bolívar, F., Gosset, G., and Escalante, A. (2015). Inactivation of the PTS as a strategy to engineer the production of aromatic metabolites in *Escherichia coli*. J Mol Microbiol Biotechnol *25*, 195-208.

Chan, E.-C., Tsai, H.-L., Chen, S.-L., and Mou, D.-G. (1993). Amplification of the tryptophan operon gene in *Escherichia coli* chromosome to increase L-tryptophan biosynthesis. Appl Microbiol Biotechnol *40*, 301-305.

Chandran, S.S., Yi, J., Draths, K., Daeniken, R.v., Weber, W., and Frost, J. (2003). Phosphoenolpyruvate availability and the biosynthesis of shikimic acid. Biotechnol Prog *19*, 808-814.

Chen, C.-Y., Georgiev, I., Anderson, A.C., and Donald, B.R. (2009). Computational structure-based redesign of enzyme activity. PNAS *106*, 3764-3769.

Chen, L. (2017). Rational metabolic engineering and systematic analysis of *Escherichia coli* for L-tryptophan bioproduction (Technische Universität Hamburg-Harburg).

Chen, L., Chen, M., Ma, C., and Zeng, A.-P. (2018). Discovery of feed-forward regulation in L-tryptophan biosynthesis and its use in metabolic engineering of *E. coli* for efficient tryptophan bioproduction. Metab Eng *47*, 434-444.

Chen, L., and Zeng, A.-P. (2017). Rational design and metabolic analysis of *Escherichia coli* for effective production of L-tryptophan at high concentration. Appl Microbiol Biotechnol *101*, 559-568.

Chen, Y., Liu, Y., Ding, D., Cong, L., and Zhang, D. (2018). Rational design and analysis of an *Escherichia coli* strain for high-efficiency tryptophan production. J Ind Microbiol Biotechnol *45*, 357-367.

Chen, Z., Rappert, S., Sun, J., and Zeng, A.-P. (2011). Integrating molecular dynamics and co-evolutionary analysis for reliable target prediction and deregulation of the allosteric inhibition of aspartokinase for amino acid production. J Biotechnol *154*, 248-254.

Cho, S., Shin, J., and Cho, B.-K. (2018). Applications of CRISPR/Cas system to bacterial metabolic engineering. Int J Mol *19*, 1089.

Christians, F.C., Scapozza, L., Crameri, A., Folkers, G., and Stemmer, W.P. (1999). Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. Nat Biotechnol *17*, 259-264.

Coulet, P.R., and Blum, L.J. (2019). Biosensor principles and applications. CRC Press. ISBN 0-8247-8546-0.

d'Oelsnitz, S., and Ellington, A. (2018). Continuous directed evolution for strain and protein engineering. Curr Opin Biotechnol *53*, 158-163.

Dell, K., and Frost, J. (1993). Identification and removal of impediments to biocatalytic synthesis of aromatics from D-glucose: rate-limiting enzymes in the common pathway of aromatic amino acid biosynthesis. J Am Chem Soc *115*, 11581-11589.

Di Mascio, P., Kaiser, S., and Sies, H. (1989). Lycopene as the most efficient biological carotenoid singlet oxygen quencher. ARCH Biochem Biophys *274*, 532-538.

Díaz-Quiroz, D.C., Cardona-Félix, C.S., Viveros-Ceballos, J.L., Reyes-González, M.A., Bolívar, F., Ordoñez, M., and Escalante, A. (2018). Synthesis, biological activity and molecular modelling studies of shikimic acid derivatives as inhibitors of the shikimate dehydrogenase enzyme of *Escherichia coli*. J Enzyme Inhib Med Chem *33*, 397-404.

Dickinson, B.C., Packer, M.S., Badran, A.H., and Liu, D.R. (2014). A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. Nature Commun *5*, 5352.

Dietrich, J.A., McKee, A.E., and Keasling, J.D. (2010). High-throughput metabolic engineering: advances in small-molecule screening and selection. Annu Rev Biochem *79*, 563-590.

Ding, R., Liu, L., Chen, X., Cui, Z., Zhang, A., Ren, D., and Zhang, L. (2014). Introduction of two mutations into AroG increases phenylalanine production in *Escherichia coli*. Biotechnol Lett *36*, 2103-2108.

Dodge, T.C., and Gerstner, J.M. (2002). Optimization of the glucose feed rate profile for the production of tryptophan from recombinant *E. coli*. J Chem Technol Biotechnol *77*, 1238-1245.

Doroshenko, V., Airich, L., Vitushkina, M., Kolokolova, A., Livshits, V., and Mashko, S. (2007). YddG from *Escherichia coli* promotes export of aromatic amino acids. FEMS Microbiol Lett *275*, 312-318.

Doukyu, N., and Iida, S. (2020). Production of styrene oxide from styrene by a recombinant *Escherichia coli* with enhanced AcrAB-TolC efflux pump level in an aqueous-organic solvent two-phase system. Biosci Biotechnol Biochem 1-8.

Du, D., Wang, Z., James, N.R., Voss, J.E., Klimont, E., Ohene-Agyei, T., Venter, H., Chiu, W., and Luisi, B.F. (2014). Structure of the AcrAB-TolC multidrug efflux pump. Nature *509*, 512-515.

Eggeling, L., Bott, M., and Marienhagen, J. (2015). Novel screening methods-biosensors. Curr Opin Biotechnol *35*, 30-36.

Faassen, S.M., and Hitzmann, B. (2015). Fluorescence spectroscopy and chemometric modeling for bioprocess monitoring. Sensors *15*, 10271-10291.

Fan, L.-Q., Li, M.-W., Qiu, Y.-j., Chen, Q.-m., Jiang, S.-J., Shang, Y.-J., and Zhao, L.-M. (2018). Increasing thermal stability of glutamate decarboxylase from *Escherichia coli* by site-directed saturation mutagenesis and its application in GABA production. J Biotechnol *278*, 1-9.

Fang, M., Wang, T., Zhang, C., Bai, J., Zheng, X., Zhao, X., Lou, C., and Xing, X.-H. (2016). Intermediate-sensor assisted push-pull strategy and its application in heterologous deoxyviolacein production in *Escherichia coli*. Metab Eng *33*, 41-51.

Finney-Manchester, S.P., and Maheshri, N. (2013). Harnessing mutagenic homologous recombination for targeted mutagenesis *in vivo* by TaGTEAM. Nucleic Acids Res *41*, e99-e99.

Floras, N., Xiao, J., Berry, A., Bolivar, F., and Valle, F. (1996). Pathway engineering for the production of aromatic compounds in *Escherichia coli*. Nat Biotechnol *14*, 620-623.

Flores, S., Gosset, G., Flores, N.d., De Graaf, A., and Bolívar, F. (2002). Analysis of carbon metabolism in *Escherichia coli* strains with an inactive phosphotransferase system by ^{13}C labeling and NMR spectroscopy. Metab Eng *4*, 124-137.

Fujii, R., Kitaoka, M., and Hayashi, K. (2004). One-step random mutagenesis by error-prone rolling circle amplification. Nucleic Acids Res *32*, e145-e145.

Fujiwara, R., Noda, S., Tanaka, T., and Kondo, A. (2020). Metabolic engineering of *Escherichia coli* for shikimate pathway derivative production from glucose-xylose co-substrate. Nature Commun *11*, 1-12.

Gao, C., Hou, J., Xu, P., Guo, L., Chen, X., Hu, G., ... & Liu, L. (2019). Programmable biomolecular switches for rewiring flux in Escherichia coli. Nat Commun, *10*(1), 1-12.

Garst, A.D., Bassalo, M.C., Pines, G., Lynch, S.A., Halweg-Edwards, A.L., Liu, R., Liang, L., Wang, Z., Zeitoun, R., and Alexander, W.G. (2017). Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. Nature Biotechnol *35*, 48.

Ger, Y.-M., Chen, S.-L., Chiang, H.-J., and Shiuan, D. (1994). A single Ser-180 mutation desensitizes feedback inhibition of the phenylalanine-sensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthetase in *Escherichia coli*. J Biochemistry *116*, 986-990.

Ghosh, S., and Banerjee, U.C. (2015). Generation of *aroE* overexpression mutant of *Bacillus megaterium* for the production of shikimic acid. Microb Cell Fact *14*, 69.

Gibbs, M.D., Nevalainen, K.H., and Bergquist, P.L. (2001). Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling. Gene *271*, 13-20.

Gosset, G., Yong-Xiao, J., and Berry, A. (1996). A direct comparison of approaches for increasing carbon flow to aromatic biosynthesis in *Escherichia coli*. J. Ind. Microbiol *17*, 47-52.

Gu, P., Yang, F., Kang, J., Wang, Q., and Qi, Q. (2012). One-step of tryptophan attenuator inactivation and promoter swapping to improve the production of L-tryptophan in *Escherichia coli*. Microb Cell Fact *11*, 30.

Guo, X.-W., Zhang, Y., Li, L.-L., Guan, X.-Y., Guo, J., Wu, D.-G., Chen, Y.-F., and Xiao, D.-G. (2018a). Improved xylose tolerance and 2, 3-butanediol production of *Klebsiella pneumoniae* by directed evolution of *rpoD* and the mechanisms revealed by transcriptomics. Biotechnol Biofuels *11*, 307.

Guo, X., Chavez, A., Tung, A., Chan, Y., Kaas, C., Yin, Y., Cecchi, R., Garnier, S.L., Kelsic, E.D., and Schubert, M. (2018b). High-throughput creation and functional profiling of DNA sequence variant libraries using CRISPR–Cas9 in yeast. Nature Biotechnol *36*, 540.

Guo, Z.-p., Khoomrung, S., Nielsen, J., and Olsson, L. (2018c). Changes in lipid metabolism convey acid tolerance in *Saccharomyces cerevisiae*. Biotechnol Biofuels *11*, 1-15.

Hagino, H., and Nakayama, K. (1975). Regulatory properties of anthranilate synthetase from *Corynebacterium glutamicum*. Agric Biol Chem *39*, 323-330.

Hall, B.G. (1981). Changes in the substrate specificities of an enzyme during directed evolution of new functions. Biochemistry *20*, 4042-4049.

Halperin, S.O., Tou, C.J., Wong, E.B., Modavi, C., Schaffer, D.V., and Dueber, J.E. (2018). CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. Nature *560*, 248-252.

Hardiman, E., Gibbs, M., Reeves, R., and Bergquist, P. (2010). Directed evolution of a thermophilic *β*-glucosidase for cellulosic bioethanol production. Appl Biochem Biotechnol *161*, 301-312.

Heipieper, H.J., Weber, F.J., Sikkema, J., Keweloh, H., and de Bont, J.A. (1994). Mechanisms of resistance of whole cells to toxic organic solvents. Trends Biotechnol *12*, 409-415.

Hendricks, C.L., Ross, J.R., Pichersky, E., Noel, J.P., and Zhou, Z.S. (2004). An enzyme-coupled colorimetric assay for *S*-adenosylmethionine-dependent methyltransferases. Anal Biochem *326*, 100-105.

Hengge-Aronis, R. (1993). Survival of hunger and stress: the role of *rpoS* in early stationary phase gene regulation in *E. coli*. *Cell*, *72*(2), 165-168.

Henkin, T.M., and Yanofsky, C. (2002). Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. Bioessays *24*, 700-707.

Herry, D., and Dunican, L. (1993). Cloning of the *trp* gene cluster from a tryptophan-hyperproducing strain of *Corynebacterium glutamicum*: identification of a mutation in the *trp* leader sequence. Appl Environ Microbiol *59*, 791-799.

Hess, G.T., Tycko, J., Yao, D., and Bassik, M.C. (2017). Methods and applications of CRISPR-mediated base editing in eukaryotic genomes. Mol Cell *68*, 26-43.

Hong, S.H., Lee, J., and Wood, T.K. (2010). Engineering global regulator Hha of *Escherichia coli* to control biofilm dispersal. Microb Biotechnol *3*, 717-728.

Hu, C., Jiang, P., Xu, J., Wu, Y., and Huang, W. (2003). Mutation analysis of the feedback inhibition site of phenylalanine-sensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase of *Escherichia coli*. J. Basic Microbiol. *43*, 399-406.

Hubbard, B.P., Badran, A.H., Zuris, J.A., Guilinger, J.P., Davis, K.M., Chen, L., Tsai, S.Q., Sander, J.D., Joung, J.K., and Liu, D.R. (2015). Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. Nat Methods *12*, 939-942.

Huccetogullari, D., Luo, Z.W., and Lee, S.Y. (2019). Metabolic engineering of microorganisms for production of aromatic compounds. Microb Cell Fact *18*, 41.

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*. Biotechnol Biofuels *11*, 103.

Iancu, C.V., Zamoon, J., Woo, S.B., Aleshin, A., and Choe, J.-y. (2013). Crystal structure of a glucose/H^+ symporter and its mechanism of action. PNAS *110*, 17862-17867.

Ikeda, M. (2003). Amino acid production processes. In Microbial production of l-amino acids (Springer), pp. 1-35.

Ikeda, M. (2006). Towards bacterial strains overproducing L-tryptophan and other aromatics by metabolic engineering. Appl Microbiol Biotechnol *69*, 615.

Ikeda, M., Nakanishi, K., Kino, K., and Katsumata, R. (1994). Fermentative production of tryptophan by a stable recombinant strain of *Corynebacterium glutamicum* with a modified serine-biosynthetic pathway. Biosci Biotechnol Biochem *58*, 674-678.

Jakočiūnas, T., Pedersen, L.E., Lis, A.V., Jensen, M.K., and Keasling, J.D. (2018). CasPER, a method for directed evolution in genomic contexts using mutagenesis and CRISPR/Cas9. Metab Eng *48*, 288-296.

Jian, X., Guo, X., Wang, J., Tan, Z.L., Xing, X.-h., Wang, L., and Zhang, C. (2019). Microbial microdroplet culture system (MMC): an integrated platform for automated, high-throughput microbial cultivation and adaptive evolution. Biotechnol Bioeng. doi: 10.1002/bit.27327

Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., and Yang, S. (2015). Multigene editing in the *Escherichia coli* genome *via* the CRISPR-Cas9 system. Appl Environ Microbiol *81*, 2506-2514.

Ju, J., Misono, H., and Ohnishi, K. (2005). Directed evolution of bacterial alanine racemases with higher expression level. J Biosci Bioeng *100*, 246-254.

Jůzlová, P., Martinkova, L., and Křen, V. (1996). Secondary metabolites of the *fungus Monascus*: A review. J Ind Microbiol *16*, 163-170.

Kikuchi, Y., Tsujimoto, K., and Kurahashi, O. (1997). Mutational analysis of the feedback sites of phenylalanine-sensitive 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase of *Escherichia coli*. Appl Environ Microbiol *63*, 761-762.

Kim, S.W., and Keasling, J. (2001). Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. Biotechnol Bioeng *72*, 408-415.

Kirschner, K., Szadkowski, H., Jardetzky, T. S., & Hager, V. (1987). Phosphoribosylanthranilate isomerase-indoleglycerol-phosphate synthase from *Escherichia coli*. In Methods in enzymology (Vol. 142, pp. 386-397). Academic Press.

Kleinstiver, B.P., Sousa, A.A., Walton, R.T., Tak, Y.E., Hsu, J.Y., Clement, K., Welch, M.M., Horng, J.E., Malagon-Lopez, J., and Scarfò, I. (2019). Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. Nat Biotechnol *37*, 276-282.

Klig, L.S., Carey, J., and Yanofsky, C. (1988). *trp* Repressor interactions with the *trparoH* and *trpR* operators: Comparison of repressor binding *in vitro* and repression *in vivo*. J Mol Biol *202*, 769-777.

Klingenberg, M. (1974). Nicotinamide-adenine dinucleotides (NAD, NADP, NADH, NADPH): Spectrophotometric and fluorimetric methods. In Methods of enzymatic analysis (pp. 2045-2072). Academic Press.

Kolkman, J.A., and Stemmer, W.P. (2001). Directed evolution of proteins by exon shuffling. Nat Biotechnol *19*, 423-428.

Korendovych, I. V. (2018). Rational and semirational protein design. In Protein Engineering (pp. 15-23). Humana Press, New York, NY.

Koresawa, M., and Okabe, T. (2004). High-throughput screening with quantitation of ATP consumption: a universal non-radioisotope, homogeneous assay for protein kinase. Assay Drug Dev Technol *2*, 153-160.

Kumamaru, T., Suenaga, H., Mitsuoka, M., Watanabe, T., and Furukawa, K. (1998). Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. Nat Biotechnol *16*, 663-666.

Lai, Y.P., Huang, J., Wang, L.F., Li, J., and Wu, Z.R. (2004). A new approach to random mutagenesis *in vitro*. Biotechnol Bioeng *86*, 622-627.

Lawley, B., and Pittard, A. (1994). Regulation of *aroL* expression by TyrR protein and Trp repressor in *Escherichia coli* K-12. J Bacteriol *176*, 6921-6930.

Lee, J.-H., and Wendisch, V.F. (2017). Production of amino acids-genetic and metabolic engineering approaches. Bioresour Technol *245*, 1575-1587.

Leemhuis, H., Kelly, R.M., and Dijkhuizen, L. (2009). Directed evolution of enzymes: library screening strategies. IUBMB life *61*, 222-228.

Li, M., Liu, C., Yang, J., Nian, R., Xian, M., Li, F., and Zhang, H. (2020). Common problems associated with the microbial productions of aromatic compounds and corresponding metabolic engineering strategies. Biotechnol Adv, 107548.

Lin, S., Meng, X., Jiang, J., Pang, D., Jones, G., OuYang, H., and Ren, L. (2012). Site-directed mutagenesis and overexpression of *aroG* gene of *Escherichia coli* K-12. Int J Biol Macromol *51*, 915-919.

Liu, L., Duan, X., and Wu, J. (2016). Modulating the direction of carbon flow in *Escherichia coli* to improve L-tryptophan production by inactivating the global regulator FruR. J Biotechnol *231*, 141-148.

Liu, R., Liang, L., Garst, A.D., Choudhury, A., i Nogué, V.S., Beckham, G.T., and Gill, R.T. (2018). Directed combinatorial mutagenesis of *Escherichia coli* for complex phenotype engineering. Metab Eng *47*, 10-20.

Lu, J., Tang, J., Liu, Y., Zhu, X., Zhang, T., and Zhang, X. (2012). Combinatorial modulation of *galP* and *glk* gene expression for improved alternative glucose utilization. Appl Microbiol Biotechnol *93*, 2455-2462.

Lutz, S. (2010). Beyond directed evolution—semi-rational protein engineering and design. Curr Opin Biotechnol *21*, 734-743.

Man, T.-K., Pease, A.J., and Winkler, M.E. (1997). Maximization of transcription of the serC (pdxF)-aroA multifunctional operon by antagonistic effects of the cyclic AMP (cAMP) receptor protein-cAMP complex and Lrp global regulators of *Escherichia coli* K-12. J Bacteriol *179*, 3458-3469.

Maresca, M., Erler, A., Fu, J., Friedrich, A., Zhang, Y., and Stewart, A. (2010). Single-stranded heteroduplex intermediates in λ Red homologous recombination. BMC Mol Biol *11*, 54.

Mascarenhas, D., Ashworth, D., and Chen, C.S. (1991). Deletion of *pgi* alters tryptophan biosynthesis in a genetically engineered strain of *Escherichia coli*. Appl Environ Microbiol *57*, 2995-2999.

Mateos, S.S., Sánchez, C.L., Paredes, S.D., Barriga, C., and Rodríguez, A.B. (2009). Circadian levels of serotonin in plasma and brain after oral administration of tryptophan in rats. Basic Clin Pharmacol *104*, 52-59.

McCandliss, R.J., Poling, M., and Herrmann, K. (1978). 3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase. Purification and molecular characterization of the phenylalanine-sensitive isoenzyme from *Escherichia coli*. J Biol Chem *253*, 4259-4265.

Miller, M. B., & Bassler, B. L. (2001). Quorum sensing in bacteria. Annu Rev Microbiol *55*(1), 165-199.

Molla, K.A., and Yang, Y. (2019). CRISPR/Cas-mediated base editing: technical considerations and practical applications. Trends Biotechnol 37 (10), 1121-1142.

Molnos, J., Gardiner, R., Dale, G.E., and Lange, R. (2003). A continuous coupled enzyme assay for bacterial malonyl–CoA: acyl carrier protein transacylase (FabD). Anal Biochem *319*, 171-176.

Moore, C.L., Papa III, L.J., and Shoulders, M.D. (2018). A processive protein chimera introduces mutations across defined DNA regions *in vivo*. J Am Chem Soc *140*, 11560-11564.

Moura, M., Pertusi, D., Lenzini, S., Bhan, N., Broadbelt, L.J., and Tyo, K.E. (2016). Characterizing and predicting carboxylic acid reductase activity for diversifying bioaldehyde production. Biotechnol Bioeng *113*, 944-952.

Mukhopadhyay, A. (2015). Tolerance engineering in bacteria for the production of advanced biofuels and chemicals. Trends Microbiol *23*, 498-508.

Mustafi, N., Grünberger, A., Kohlheyer, D., Bott, M., and Frunzke, J. (2012). The development and application of a single-cell biosensor for the detection of l-methionine and branched-chain amino acids. Metab Eng *14*, 449-457.

Naeem, A., James, N., Tanvir, M., Marriam, M., and Nathaniel, S. (2017). Fluorescence Activated Cell Sorting (FACS): An Advanced Cell Sorting Technique. PSM Biological Research *2*, 83-88.

Nagaraja, P., Yathirajan, H.S., and Vasantha, R.A. (2003). Highly sensitive reaction of tryptophan with p-phenylenediamine. Anal Biochem *312*, 157-161.

Neuenschwander, M., Butz, M., Heintz, C., Kast, P., and Hilvert, D. (2007). A simple selection strategy for evolving highly efficient enzymes. Nat Biotechnol *25*, 1145-1147.

Nikaido, H., and Takatsuka, Y. (2009). Mechanisms of RND multidrug efflux pumps. BBA-Proteins and Proteomics *1794*, 769-781.

Nilsson, L., Vanet, A., Vijgenboom, E., and Bosch, L. (1990). The role of FIS in trans activation of stable RNA operons of *E. coli*. The EMBO Journal *9*, 727-734.

O'Gara, J.P., and Dunican, L.K. (1995). Mutations in the *trpD* gene of *Corynebacterium glutamicum* confer 5-methyltryptophan resistance by encoding a feedback-resistant anthranilate phosphoribosyltransferase. Appl Environ Microbiol *61*, 4477-4479.

Ogino, T., Garner, C., Markley, J.L., and Herrmann, K.M. (1982). Biosynthesis of aromatic compounds: 13C NMR spectroscopy of whole *Escherichia coli* cells. PNAS *79*, 5828-5832.

Oldiges, M., Kunze, M., Degenring, D., Sprenger, G., and Takors, R.J.B.p. (2004). Stimulation, monitoring, and analysis of pathway dynamics by metabolic profiling in the aromatic amino acid pathway. *20*, 1623-1633.

Olsen, M. J., Gam, J., Iverson, B. L., & Georgiou, G. (2003). High-throughput FACS method for directed evolution of substrate specificity. In Directed enzyme evolution (pp. 329-342). Humana Press.

Packer, M.S., Rees, H.A., and Liu, D.R. (2017). Phage-assisted continuous evolution of proteases with altered substrate specificity. Nature Commun *8*, 1-11.

Panichkin, V.B., Livshits, V.A., Biryukova, I.V., and Mashko, S.V. (2016). Metabolic engineering of *Escherichia coli* for L-tryptophan production. Appl Biochem Microbiol *52*, 783-809.

Park, J., Bae, S., and Kim, J.-S. (2015). Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. Bioinformatics *31*, 4014-4016.

Patnaik, R., Louie, S., Gavrilovic, V., Perry, K., Stemmer, W.P., Ryan, C.M., and del Cardayré, S. (2002). Genome shuffling of Lactobacillus for improved acid tolerance. Nat Biotechnol *20*, 707-712.

Patthy, L. (1999). Genome evolution and the evolution of exon-shuffling—a review. Gene *238*, 103-114.

Paulsen, P.A., Custódio, T.F., and Pedersen, B.P. (2019). Crystal structure of the plant symporter STP10 illuminates sugar uptake mechanism in monosaccharide transporter superfamily. Nature Commun *10*, 1-8.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera-a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Pfeifer, G.P., You, Y.-H., and Besaratinia, A. (2005). Mutations induced by ultraviolet light. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *571*, 19-31.

Plumbridge, J. (1995). Co-ordinated regulation of amino sugar biosynthesis and degradation: the NagC repressor acts as both an activator and a repressor for the transcription of the *glmUS* operon and requires two separated NagC binding sites. The EMBO journal *14*, 3958-3965.

Postma, P. (1996). Phos-phoenolpyruvate: carbohydrate phosphotransferase systems. *Escherichia coli* and *Salmonella*: cellular and molecular biology, 206-216.

Postma, P.W., Lengeler, J.W., and Jacobson, G.R. (1993). Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. Microbiol Mol Biol Rev *57*, 543-594.

Pritchard, L., Corne, D., Kell, D., Rowland, J., and Winson, M. (2005). A general model of error-prone PCR. J Theor Biol *234*, 497-509.

Pu, J., Zinkus-Boltz, J., and Dickinson, B.C. (2017). Evolution of a split RNA polymerase as a versatile biosensor platform. Nat Chem Biol *13*, 432-438.

Qi, Y., Liu, H., Chen, X., and Liu, L. (2019). Engineering microbial membranes to increase stress tolerance of industrial strains. Metab Eng *53*, 24-34.

Ramos, J.L., Duque, E., Gallegos, M.-T., Godoy, P., Ramos-Gonzalez, M.I., Rojas, A., Terán, W., and Segura, A. (2002). Mechanisms of solvent tolerance in gram-negative bacteria. Annual Reviews in Microbiology *56*, 743-768.

Ramseier, T.M., Bledig, S., Michotey, V., Feghali, R., and Jr Saier, M.H. (1995). The global regulatory protein FruR modulates the direction of carbon flow in *Escherichia coli*. Mol Microbiol *16*, 1157-1169.

Ravikumar, A., Arzumanyan, G.A., Obadi, M.K., Javanpour, A.A., and Liu, C.C. (2018). Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. Cell *175*, 1946-1957. e1913.

Rees, H.A., Komor, A.C., Yeh, W.-H., Caetano-Lopes, J., Warman, M., Edge, A.S., and Liu, D.R. (2017). Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. Nature Commun *8*, 1-10.

Reisch, C.R., and Prather, K.L. (2015). The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in Escherichia coli. Sci Rep *5*, 1-12.

Ren, Kun Xu, David Jay Segal, and Zhang, Z. (2018). strategies for the enrichment and selection of genetically modified cells. Trends Biotechnol 37 (1), 56-71.

Ren, C., Xu, K., Liu, Z., Shen, J., Han, F., Chen, Z., and Zhang, Z. (2015). Dual-reporter surrogate systems for efficient enrichment of genetically modified cells. Cell Mol Life Sci *72*, 2763-2772.

Ronda, C., Pedersen, L.E., Sommer, M.O., and Nielsen, A.T. (2016). CRMAGE: CRISPR optimized mage recombineering. Sci Rep *6*, 1-11.

Ross, W., Thompson, J.F., Newlands, J.T., and Gourse, R.L. (1990). *E. coli Fis* protein activates ribosomal RNA transcription *in vitro* and *in vivo*. The EMBO Journal *9*, 3733-3742.

Rothman, S.C., and Kirsch, J.F. (2003). How does an enzyme evolved in vitro compare to naturally occurring homologs possessing the targeted function? Tyrosine aminotransferase from aspartate aminotransferase. J Mol Biol *327*, 593-608.

Rüffer, N., Heidersdorf, U., Kretzers, I., Sprenger, G., Raeven, L., and Takors, R. (2004). Fully integrated L-phenylalanine separation and concentration using reactive-extraction with liquid-liquid centrifuges in a fed-batch process with *E. coli*. Bioprocess Biosystems Eng *26*, 239-248.

Russell, D. W., & Sambrook, J. (2001). Molecular cloning: a laboratory manual (Vol. 1, p. 112). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sabnis, N.A., Yang, H., and Romeo, T. (1995). Pleiotropic regulation of central carbohydrate metabolism in *Escherichia coli* via the gene *csrA*. J Biol Chem *270*, 29096-29104.

Sadler, J.C., Currin, A., and Kell, D.B. (2018). Ultra-high throughput functional enrichment of large monoamine oxidase (MAO-N) libraries by fluorescence activated cell sorting. Analyst *143*, 4747-4755.

Saier Jr, M.H., and Ramseier, T.M. (1996). The catabolite repressor/activator (Cra) protein of enteric bacteria. J Bacteriol *178*, 3411.

Savile, C.K., Janey, J.M., Mundorff, E.C., Moore, J.C., Tam, S., Jarvis, W.R., Colbeck, J.C., Krebber, A., Fleitz, F.J., and Brands, J. (2010). Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. Science *329*, 305-309.

Schaaper, R.M. (1993). Base selection, proofreading, and mismatch repair during DNA replication in *Escherichia coli*. J Biol Chem *268*, 23762-23765.

Schoner, R., and Herrmann, K.M. (1976). 3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase. Purification, properties, and kinetics of the tyrosine-sensitive isoenzyme from *Escherichia coli*. J Biol Chem *251*, 5440-5447.

Schuster, A., Erasimus, H., Fritah, S., Nazarov, P.V., van Dyck, E., Niclou, S.P., and Golebiewska, A. (2019). RNAi/CRISPR screens: from a pool to a valid hit. Trends Biotechnol *37*, 38-55.

Sen, S., Dasu, V.V., and Mandal, B. (2007). Developments in directed evolution for improving enzyme functions. Appl Biochem Biotechnol *143*, 212-223.

Sharan, R., Karni, S., and Felder, Y. (2007). Analysis of biological networks: transcriptional tetworks-promoter pequence analysis. Tel Aviv University, 1-5.

Singer, P. (2007). High-dose amino acid infusion preserves diuresis and improves nitrogen balance in non-oliguric acute renal failure. Wiener Klinische Wochenschrift *119*, 218-222.

Song, L., and Zeng, A.-P. (2017). Engineering 'cell robots' for parallel and highly sensitive screening of biomolecules under *in vivo* conditions. Sci Rep *7*, 1-9.

Sprenger, G. A. (2006). Aromatic amino acids. In Amino acid biosynthesis~pathways, regulation and metabolic engineering (pp. 93-127). Springer, Berlin, Heidelberg.

Stano, N.M., and Patel, S.S. (2004). T7 lysozyme represses T7 RNA polymerase transcription by destabilizing the open complex during initiation. J Biol Chem *279*, 16136-16143.

Sun, M.G., Seo, M.-H., Nim, S., Corbi-Verge, C., and Kim, P.M. (2016). Protein engineering by highly parallel screening of computationally designed variants. Sci Adv *2*, e1600692.

Sylvestre, J., Chautard, H., Cedrone, F., and Delcourt, M. (2006). Directed evolution of biocatalysts. Org Process Res Dev *10*, 562-571.

Tan, Z.L., Zheng, X., Wu, Y., Jian, X., Xing, X., and Zhang, C. (2019). *In vivo* continuous evolution of metabolic pathways for chemical production. Microb Cell Fact *18*, 82.

Tatarko, M., and Romeo, T. (2001). Disruption of a global regulatory gene to enhance central carbon flux into phenylalanine biosynthesis in *Escherichia coli*. Curr Microbiol *43*, 26-32.

Tribe, D.E., and Pittard, J. (1979). Hyperproduction of tryptophan by *Escherichia coli*: genetic manipulation of the pathways leading to tryptophan formation. Appl Environ Microbiol *38*, 181-190.

Trott, O., and Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem *31*, 455-461.

Umbarger, H.E. (1978). Amino acid biosynthesis and its regulation. Annu Rev Biochem *47*, 533-606.

Valle, F., Munoz, E., Ponce, E., Flores, N., and Bolivar, F. (1996). Basic and applied aspects of metabolic diversity: the phosphoenolpyruvate node. J Ind Microbiol *17*, 458-462.

Vallejo, D., Nikoomanzar, A., Paegel, B.M., and Chaput, J.C. (2019). Fluorescence-activated droplet sorting for single-cell directed evolution. ACS Synth Biol *8*, 1430-1440.

Van Dyk, T.K., Templeton, L.J., Cantera, K.A., Sharpe, P.L., and Sariaslani, F.S. (2004). Characterization of the *Escherichia coli* AaeAB efflux pump: a metabolic relief valve? J Bacteriol *186*, 7196-7204.

Vanhercke, T., Ampe, C., Tirry, L., and Denolf, P. (2005). Reducing mutational bias in random protein libraries. Anal Biochem *339*, 9-14.

Verma, N., Singh, A.K., and Singh, M. (2017). L-arginine biosensors: A comprehensive review. Biochem Biophys Rep *12*, 228-239.

Volkov, A.A., Shao, Z., and Arnold, F.H. (1999). Recombination and chimeragenesis by in *vitro* heteroduplex formation and *in vivo* repair. Nucleic Acids Res *27*, e18-i-e18-vi.

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

Wang, Y., Feng, S., Zhan, T., Huang, Z., Wu, G., and Liu, Z. (2013). Improving catalytic efficiency of endo-*β*-1,4-xylanase from *Geobacillus stearothermophilus* by directed evolution and H179 saturation mutagenesis. J Biotechnol *168*, 341-347.

Webb, B., and Sali, A. (2016). Comparative protein structure modeling using MODELLER. Curr Protoc Bioinformatics *54*, 5.6. 1-5.6. 37.

Wendisch, V. F. (Ed.). (2007). Amino acid biosynthesis-pathways, regulation and metabolic engineering (Vol. 5). Springer Science & Business Media.

Wendisch, V.F. (2020). Metabolic engineering advances and prospects for amino acid production. Metab Eng *58*, 17-34.

Wendisch, V.F., Jorge, J.M., Pérez-García, F., and Sgobba, E. (2016). Updates on industrial production of amino acids using *Corynebacterium glutamicum*. World J Microb Biot *32*, 105.

Whitfield, J.H., Zhang, W.H., Herde, M.K., Clifton, B.E., Radziejewski, J., Janovjak, H., Henneberger, C., and Jackson, C.J. (2015). Construction of a robust and sensitive arginine biosensor through ancestral protein reconstruction. Protein Science *24*, 1412-1422.

Williams, B.B., Van Benschoten, A.H., Cimermancic, P., Donia, M.S., Zimmermann, M., Taketani, M., Ishihara, A., Kashyap, P.C., Fraser, J.S., and Fischbach, M.A. (2014). Discovery and characterization of gut microbiota decarboxylases that can produce the neurotransmitter tryptamine. Cell host microbe *16*, 495-503.

Williams, T.C., Pretorius, I.S., and Paulsen, I.T. (2016). Synthetic evolution of metabolic productivity using biosensors. Trends Biotechnol *34*, 371-381.

Wilson, C.J., Husain, S.S., Stimson, E.R., Dangott, L.J., Miller, K.W., and Maggio, J.E. (1997). p-(4-Hydroxybenzoyl) phenylalanine: A photoreactive amino acid analog amenable to radioiodination for elucidation of peptide- protein interaction application to substance P receptor. Biochemistry *36*, 4542-4551.

Wu, S., Acevedo, J.P., and Reetz, M.T. (2010). Induced allostery in the directed evolution of an enantioselective *Baeyer Villiger* monooxygenase. PNAS *107*, 2775-2780.

Wu, W.B., Guo, X.L., Zhang, M.L., Huang, Q.G., Qi, F., and Huang, J.Z. (2018). Enhancement of l-phenylalanine production in *Escherichia coli* by heterologous expression of *Vitreoscilla hemoglobin*. Biotechnol Appl Biochem *65*, 476-483.

Wu, Z., Kan, S.J., Lewis, R.D., Wittmann, B.J., and Arnold, F.H. (2019). Machine learning-assisted directed protein evolution with combinatorial libraries. PNAS *116*, 8852-8858.

Xue, T., Chen, D., Su, Q., Yuan, X., Liu, K., Huang, L., Fang, J., Chen, J., He, W., and Chen, Y. (2019). Improved ethanol tolerance and production of *Saccharomyces cerevisiae* by global transcription machinery engineering via directed evolution of the SPT8 gene. Food Biotechnol *33*, 155-173.

Yajima, Y., Sakimoto, K., Takahashi, K., Miyao, K., Kudome, Y., and Aichi, K. (1990). L-Tryptophan-producing microorganism and production of L-tryptophan. Japan Patent Appl *2*, 182.

Yakandawala, N., Romeo, T., Friesen, A., and Madhyastha, S. (2008). Metabolic engineering of *Escherichia coli* to enhance phenylalanine production. Appl Microbiol Biotechnol *78*, 283-291.

Yamamoto, K., and Ishihama, A. (2006). Characterization of copper-inducible promoters regulated by CpxA/CpxR in *Escherichia coli*. Biosci Biotechnol Biochem *70*, 1688-1695.

Yang, H., Wolff, E., Kim, M., Diep, A., and Miller, J.H. (2004). Identification of mutator genes and mutational pathways in *Escherichia coli* using a multicopy cloning approach. Mol Microbiol *53*, 283-295.

Yang, K.K., Wu, Z., and Arnold, F.H. (2019). Machine-learning-guided directed evolution for protein engineering. Nat Methods *16*, 687-694.

Yu, D., Ellis, H.M., Lee, E.-C., Jenkins, N.A., and Copeland, N.G. (2000). An efficient recombination system for chromosome engineering in *Escherichia coli*. PNAS *97*, 5978-5983.

Yuan, L., Kurek, I., English, J., and Keenan, R. (2005). Laboratory-directed protein evolution. Microbiol Mol Biol Rev *69*, 373-392.

Zaccardi, M.J., Mannweiler, O., and Boehr, D.D. (2012). Differences in the catalytic mechanisms of mesophilic and thermophilic indole-3-glycerol phosphate synthase enzymes at their adaptive temperatures. Biochem Biophys Res Commun *418*, 324-329.

Zeng, W., Guo, L., Xu, S., Chen, J., and Zhou, J. (2020). High-throughput screening technology in industrial biotechnology. Trends Biotechnol. doi: 10.1016/j.tibtech.2020.01.001

Zhang, B., Zhou, N., Liu, Y.-M., Liu, C., Lou, C.-B., Jiang, C.-Y., and Liu, S.-J. (2015). Ribosome binding site libraries and pathway modules for shikimic acid synthesis with *Corynebacterium glutamicum*. Microb Cell Fact *14*, 71.

Zhang, H., Chong, H., Ching, C.B., Song, H., and Jiang, R. (2012). Engineering global transcription factor cyclic AMP receptor protein of *Escherichia coli* for improved 1-butanol tolerance. Appl Microbiol Biotechnol *94*, 1107-1117.

Zhang, J., Zong, W., Hong, W., Zhang, Z.-T., and Wang, Y. (2018). Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. Metab Eng *47*, 49-59.

Zhang, Y., Ma, C., Dischert, W., Soucaille, P., and Zeng, A.P. (2019). Engineering of phosphoserine aminotransferase increases the conversion of L-homoserine to 4-hydroxy-2-ketobutyrate in a glycerol-independent pathway of 1, 3-propanediol production from glucose. J Biotechnol *14*, 1900003.

Zhang, Y., Perry, K., Powell, K., Stemmer, P., and Del Cardayré, S. (2002). Evolution of *Streptomyces fradiae* by whole genome shuffling. Nature *415*, 644-646.

Zhao, C., Cheng, L., Wang, J., Shen, Z., and Chen, N. (2016a). Impact of deletion of the genes encoding acetate kinase on production of L-tryptophan by *Escherichia coli*. Ann Microbiol *66*, 261-269.

Zhao, D., Yuan, S., Xiong, B., Sun, H., Ye, L., Li, J., Zhang, X., and Bi, C. (2016b). Development of a fast and easy method for *Escherichia coli* genome editing with CRISPR/Cas9. Microb Cell Fact *15*, 205.

Zhao, H., Giver, L., Shao, Z., Affholter, J.A., and Arnold, F.H. (1998). Molecular evolution by staggered extension process (StEP) *in vitro* recombination. Nat Biotechnol *16*, 258-261.

Zhao, Z., Chen, S., Wu, D., Wu, J., and Chen, J. (2012). Effect of gene knockouts of L-tryptophan uptake system on the production of L-tryptophan in *Escherichia coli*. Process Biochem *47*, 340-344.

Zheng, H., Taraska, J., Merz, A.J., and Gonen, T. (2010). The prototypical H^+/galactose symporter GalP assembles into functional trimers. J Mol Biol *396*, 593-601.

Zhou, L.-B., and Zeng, A.-P. (2015). Exploring lysine riboswitch for metabolic flux control and improvement of L-lysine synthesis in *Corynebacterium glutamicum*. ACS Synth Biol *4*, 729-734.

Zhu, X., Zhao, D., Qiu, H., Fan, F., Man, S., Bi, C., and Zhang, X. (2017). The CRISPR/Cas9-facilitated multiplex pathway optimization (CFPO) technique and its application to improve the *Escherichia coli* xylose utilization pathway. Metab Eng *43*, 37-45.

Zurawski, G., Gunsalus, R., Brown, K., and Yanofsky, C. (1981). Structure and regulation of *aroH*, the structural gene for the tryptophan-repressible 3-deoxy-D-arabino-heptulosonic acid-7-phosphate synthetase of *Escherichia coli*. J Mol Biol *145*, 47-73.

Appendix A: Comparative genomic analysis of the strains D8 and *E. coli* W3110

Table A.1: Summary and annotation of the modified ORFs from the D8 strain in alignment with the reference strain *E. coli* W3110.

Number	Gene	Enzyme	Residue_mutate	EcoCyc ID[1]//KEGG number[2]//UniPro ID[3]	Function
1	*galP*	Galactose:H^+ symporter	Lys137Thr	EG12148//b2943//P0AEP1	Major facilitator superfamily (MFS)
2	*galP*	Galactose:H^+ symporter	Pro195Ala	EG12148//b2943//P0AEP1	Major facilitator superfamily (MFS)
3	*galP*	Galactose:H^+ symporter	Phe204Leu	EG12148//b2943//P0AEP1	Major facilitator superfamily (MFS)
4	*folD*	Bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase	Leu36Gln	EG10328//b0529//P24186	(6*R*)-5,10-methylene-5,6,7,8-tetrahydrofolate + $NADP^+$ = 5,10-methenyltetrahydrofolate + NADPH
5	*rzoD*	DLP12 prophage	Arg34Pro	G0-10436//b4510//P58041	Polypeptide; Putative prophage lysis lipoprotein RzoD
6	*aroG* *	3-deoxy-D-arabino-heptulosonate-7-phosphate synthase; Phenylalanine repressible	Ser180Phe	EG10079//b0754//P0AB91	D-erythrose 4-phosphate + H_2O + phosphoenolpyruvate = 7-phospho-2-dehydro-3-deoxy-D-arabino-heptonate + phosphate
7	*cmk*	Cytidylate kinase	Glu153Gly	EG11265//b0910//P0A6I0	ATP + CMP = ADP + CDP
8	*ycbW*	Cell division protein ZapC	Gln168Leu	G6486//b0946//P75862	Contributes to the efficiency of the cell division process
9	*ycdT*	Probable diguanylate cyclase DgcT	Val130Ala	G6532//b1025//P75908	2 GTP = cyclic di-3',5'-guanylate + 2 diphosphate
10	*hflD*	Lysogenization regulator	Trp181Arg	EG11345//b1132//P25746	Negative regulator of phage lambda lysogenization

Table A.1: Summary and annotation of the modified ORFs from the D8 strain in alignment with the reference strain *E. coli* W3110.

Number	Gene	Enzyme	Residue_mutate	EcoCyc ID[1]//KEGG number[2]//UniPro ID[3]	Function
11	*umuC*	DNA polymerase V catalytic protein	Phe287Leu	EG11056//b1184//P04152	Poorly processive, error-prone DNA polymerase involved in translesion repair
12	*trpE**	Anthranilate synthase subunit TrpE	Ser40Phe	EG11028//b1264//P00895	chorismate + L-glutamine = anthranilate + H^+ + L-glutamate + pyruvate
13	*zwf*	NADP+-dependent glucose-6-phosphate dehydrogenase	Glu165 Stop codon	EG11221//b1852//P0AC53	D-glucose 6-phosphate + NADP+ = 6-phospho-D-glucono-1,5-lactone + H^+ + NADPH
14	*fliC*	Flagellar filament structural protein	Glu115Lys	EG10321//b1923//P04949	Polymerizes to form the filaments of bacterial flagella
15	*yohC*	Putative inner membrane protein	Val33Ala	EG12016//b2135//P0AD17	Composes of five predicted transmembrane domains
16	*yeiB*	DUF418 domain-containing protein YeiB	Ser28Pro	EF11290//b2152//P25747	Involves in transport
17	*yfcO*	DUF2544 domain-containing protein YfcO	Met218Ile	G7203//b2332//P76498	Contributes to adhesion to various surfaces in specific environmental niches
18	*serA**	Phosphoglycerate dehydrogenase	Asp364Ala	EG10944//b2913//P0A9T0	(2R)-3-phosphoglycerate + NAD^+ = 3-phosphooxypyruvate + H^+ + NADH
19	*serA**	Phosphoglycerate dehydrogenase	His344Ala	EG10944//b2913//P0A9T0	(2R)-3-phosphoglycerate + NAD^+ = 3-phosphooxypyruvate + H^+ + NADH
20	*rpsL*	30S ribosomal subunit protein S12	Lys43Arg	EG10911//b3342//P0A7S3	With S4 and S5 plays an important role in translational accuracy

Table A.1: Summary and annotation of the modified ORFs from the D8 strain in alignment with the reference strain *E. coli* W3110.

Number	Gene	Enzyme	Residue_mutate	EcoCyc ID[1]//KEGG number[2]//UniPro ID[3]	Function
21	*yhaC*	uncharacterized protein YhaC	Glu134 Stop codon	EG11174//b3121//P11864	--
22	*yhaC*	uncharacterized protein YhaC	Thr216Ala	EG11174//b3121//P11864	--
23	*yibJ*	Putative uncharacterized protein YibJ	Lys35Gln	EG11766//b3595//P32109	--
24	*ykfF*	CP4-6 prophage; protein YkfF	Deletion[4]	G6124//b0249//P75677	--
25	*ypjJ*	CP4-57 prophage; DUF987 domain-containing protein YpjJ	Insertion[5]	G0-10464//b4548//P58033	--
26	*ykfH*	DUF987 domain-containing protein YkfH	Deletion	G0-10432//b1504//Q9XB42	--
27	*ybfD*	H repeat-associated putative transposase YbfD	Deletion	EG11524//b0706//P28916	Expression of ybfD is upregulated in long-term stationary phase cultures
28	*dgsA* *(mlc)*	DNA-binding transcriptional repressor Mlc	Deletion	G6852//b1594//P50456	Regulates the expression of proteins involved in phosphotransferase system for sugar uptake
29	*rrlD*	23S ribosomal RNA	Deletion	EG30080//b3275//--	Belongs to the large subunit (50S subunit) of ribosome

Note: EcoCyc ID[1], https://biocyc.org/organism-summary?object=ECOLI; KEGG number[2], https://www.genome.jp/dbget-bin/www_bget?eco; Unipro ID[3], https://ebi14.uniprot.org/uniprot; * the gene variants $aroG^{S180F}$, $serA^{H344A/N364A}$, and $trpE^{S40F}$ were generated by rational protein design (Chen, 2017); Deletion[4] or insertion[5] indicate that part of sequences of gene is deleted or inserted.