Towards stable cyanobacterial cell factories

Du, W.

Publication date
2018
Document Version
Other version
License
Other

6 General discussion: Challenges for the application of synthetic biology in the synthesis of commodity products by cyanobacteria via 'direct conversion'

Cyanobacterial direct conversion of CO_2 to several commodity chemicals has been recognized as a potential contributor to support the much needed sustainable development of human societies. However, the feasibility of this "green conversion" hinders on our ability to overcome the hurdles presented by the natural evolvability of microbes. The latter may result in the genetic instability of engineered cyanobacterial strains leading to impaired productivity. This challenge is general to any 'cell factory' approach in which the cells grow for multiple generations, and based on several studies carried out in different microbial hosts, we could identify that three distinct strategies have been proposed to tackle it. These are (i) to reduce microbial evolvability by decreasing the native mutation rate; (ii) to align product formation with cell growth/fitness; and paradoxically, (iii) to efficiently re-allocate cellular resources to product formation by uncoupling it from growth. The implementation of either of these strategies requires an advanced synthetic biology toolkit. Here, we review the existing methods available for cyanobacteria and identify areas of focus in which specific developments are still needed. Furthermore, we discuss how potentially stabilizing strategies may be used in combination leading to further increases of productivity while ensuring the stability of the cyanobacterial-based direct conversion process.

This chapter has been submitted as:

Du, W., Caicedo Burbano, S. B., Hellingwerf, K. J., Branco Dos Santos, F. (2017) Challenges for the application of synthetic biology in the synthesis of commodity products by cyanobacteria via 'direct conversion'.

6.1 Introduction

The problem of the ever-increasing energy consumption of human societies, and its negative impact on global climate trends, makes it urgent to develop sustainable and environmentally friendly production processes. To this point, shifting society's dependence away from fossilized deposits towards renewable biomass resources is generally recognized as an important contributor [194]. By using energy obtained from (sun)light through photosynthesis, the dominant anthropogenic greenhouse gas, CO_2, can be utilized as a carbon source and stored in the biopolymers of which a cell is composed (proteins, nucleic acids, cell wall, neutral and phospholipids, *etc.*). Microbial molecular biology has imbued us with the ability to convert such biopolymers into different chemical compounds suitable for diverse applications. Notably, they can be used as biofuel and bioplastic, whose combustion will lead to CO_2 release, which can again be further recycled in a closed-cycle. By such means, biomass resource utilization can be regarded not only as a direct capture of the renewable energy present in sunlight, but additionally, it guarantees carbon-neutrality (*i.e.* zero net carbon emission) [195].

For compounds derived from microbial fermentation of plant biomass (*e.g.* plant crop residues, lignocellulosic feedstock, algae biomass, *etc.*) one can consider this an indirect approach, as these processes typically require first the accumulation of biopolymers via oxygenic photosynthesis in plants/algae, from which substrates can then be extracted and prepared, before they are finally converted, mostly via fermentation, into the target products [196]. In contrast, the direct-approach, enables the transformation of CO_2 into target chemicals, without the need of first accumulating and extracting biopolymers [54]. This is generally accomplished by the introduction of heterologous pathway(s) for the production of a specific compound in a cyanobacterium, using the native metabolite(s) of the host as a substrate for the heterologous production pathway. Additional gene knockouts, to channel more metabolic flux through the production pathway, may be introduced as well [197]. The overall efficiency of the direct-conversion approach is thereby increased in comparison to indirect approaches, since the latter ultimately relies on fewer steps, while all of them have their inherent losses [198]. As such, cyanobacteria, photosynthetic prokaryotes that show relatively fast (compared to *e.g.* plants) photoautotrophic growth and are easily genetically accessible, serve as an attractive microbial host for chemical production directly from CO_2. Such photo-driven cyanobacterial "direct-conversion" (elsewhere termed the "Photanol approach" [199]), has been extensively explored for the production of a variety of commodity products during the past decade [72].

Despite the tremendous progress achieved in cyanobacterial "direct-conversion", development of such green cell factories that meet the economic requirements on a commercial scale is still challenging [4,200]. One of these challenges is to achieve high product titers [201]. To reach these depends mainly on sustained microbial productivity during relatively long process operation. Yet, the longer cultivations also imply higher numbers of cell divisions, and therefore an increasing number of opportunities to acquire (a) genetic lesion(s). The evolvability of a biological system dictates that the fitter (*i.e.* faster growing) individuals of the population will ultimately be selected to dominate the culture. Cells that acquire (a) mutation(s) that hamper productivity are in most cases relieved to a certain extent from the burden associated with product formation. When growth continues, the latter type of mutant cells will generally start to out-compete the slower growing producing cells. This means that with time the population will become dominated by non-producing, faster growing, mutant cells. As it will be discussed later, especially during longer cultivations, this can be a very real problem, which certainly presents an enormous challenge to the commercialization of cyanobacterial cell factories.

Such genetic instability of production strains is not unique to cyanobacteria; it is a much more widespread phenomenon in biotechnology. At its basis is the conflict between the human desire for sustained high productivity and the microbial struggle and evolvability towards maximal fitness for growth. However, if this issue of genetic instability is not addressed within the field of cyanobacteria, it will most likely compromise the economic viability of the much-needed processes based on the application of cyanobacterial cell factories. This chapter will focus on genetic instability, primarily in cyanobacterial cell factories and on ways to minimize its consequences. We will start by listing all the instances in which instability of cyanobacterial production strains has been reported in literature, followed by a review of our current understanding of the mechanisms that may lead to mutations. Solutions towards the development of stable cell-factory based systems will then be discussed, while evaluating their benefits and downsides. The available synthetic biology toolkit and cultivation methodologies, essential for the implementation of the discussed solutions, will also be briefly covered. And finally, we will share our perspective on how all these developments may shape the field in the years to come, and may result in stable, highly productive cyanobacterial cell factories.

6.2 Genetic instability of cyanobacterial production strains

Genetic instability of production strains has been reported for at least one representative of all three of the most common representatives of the cyanobacterial species used for 'direct conversion', *i.e. Synechocystis* sp.

PCC6803 (hereafter, *Synechocystis*), *Synechococcus elongatus* PCC7942 and *Synechococcus* sp. PCC7002. Also, it has been observed for the production of many different compounds (see Table 6.1 for a complete list) [135,71] . DNA sequence analysis of revertant cells, with hampered productivity compared to the original engineered strains, has been performed in a few reports [81–83]. It is important to note that in all of the few cases in which revertant cells were characterized, the sequencing efforts were limited to the heterologous cassette and immediately flanking regions. Consequently, mutations in other chromosomal loci would implicitly be overlooked. Nonetheless, suppressor mutations in revertant cells were often found in the coding region of the introduced gene(s) encoding the heterologous production pathway. They were varied in nature ranging from point mutations, to small and large insertions and deletions. Many of these mutations affect an active-site residue, or lead to a truncated gene product through the introduction of a stop codon in the target gene [185]. In both cases, the functionality of the enzyme(s) is severely impaired, and product formation will be minimized (if not totally abolished). This instability of engineered cyanobacterial strains has been described as "an elephant in the room", because it is "*important, obvious, yet largely ignored*" [71]. A better understanding of the burden imposed on cells by product formation is of paramount importance, before effective remedies can be proposed and tested.

Acquisition of a mutation can help a microbe to decrease the burden of product formation, and consequently, to re-gain fitness (*i.e.* often faster growth)[202]. However, a closer look at the production process in a bacterial cell factory reveals that there are several cellular processes that may be affected by the expression of a heterologous production pathway. To name but a few: *(i)* the transcriptional or translational level with competition for RNA polymerases or ribosomes, respectively [203]; *(ii)* the cellular transport capacity due to potential membrane occupancy constraints [204]; *(iii)* the availability of substrates (*e.g.* competition for intracellular metabolites) [205]; and *(iv)* the ability to maintain homeostatic levels of metabolites, which may become toxic if accumulating at too high levels [206]. Many of these burdens are generally intertwined in the experiments carried out so far. For example, whenever higher expression levels of the enzymes constituting the production pathway were obtained, the burden on transcription, translation and the requirement for the initial substrate of the heterologous pathway also increased. This makes it very difficult to dissect how each of these contributes to the overall genetic instability of a production strain, thus preventing us from getting a clearer picture of which burden weighs heavier during the cultivation, and thus, could be the focus for development of remedies.

Table 6.1 A list of publications related to cyanobacterial genetic instability

strain	product	brief description	reference
Synechocystis sp. PCC6803	ethylene	inactivation of the *efe* gene resulted from duplications at the sequence CTATG, leading to truncated peptides.	81
	lactic acid	a duplication of ~160 bp in the *sth* gene, which generated premature stop codons	60
	lactic acid	nucleotide insertions or deletions and point mutations in *ldh* gene, leading to a truncated protein and defected predicted active sites, respectively	185
Synechococcus elongatus PCC7942	isopropanol	missense point mutation or IS10 insertion in the *thl* or *atoAD* genes	82
	ethylene	duplicated sequence of GATGG within the *efe* gene causing a frame shift, resulting in truncated protein	84
Synechococcus sp. PCC7002	mannitol	single base deletion in the *mtlD* gene causing a frame-shift, resulting in truncated protein.	83
Certain cyanobacteria	ethanol	point mutations, insertions/deletions, mobile genetic elements in ethanologenic DNA cassette	207

In a recent study from our group (**Chapter 2**), we made an attempt to dissect the influence of the different burdens for a lactate producing *Synechocystis* [185]. We did this by supplementing an engineered lactate producing *Synechocystis* derivate [55], with a non-metabolized analogue (2,5-anhydro-mannitol, AHM) of a known activator of the introduced L-lactate dehydrogenase (L-LDH, from *Lactococcus lactis sp. cremoris*). Without changing the expression levels of *L-ldh,* we were able to enhance L-LDH enzyme activity (both *in vitro* and *in vivo*), ultimately leading to an increase of the lactate production rate by over 2-fold. We also observed that lactate productivity of the culture dropped dramatically at later growth-phases, from which revertant cells with diverse mutations in the *L-ldh* cassette could be isolated. All the identified mutations resulted in limited or no lactate production. Because adding this allosteric effector only modulates the activity of the target enzyme (and not its expression level), we could associate the increased phenotypic instability observed with the burden of deviating the fixed carbon from biomass mainly to increased amounts of product formation. This result is disconcerting as it suggests that production instability will be a problem irrespective of the nature of the product. It also suggests that efforts to devise solutions to prevent instability of production strains should focus more on trying to minimize the conflict between the

microbial interest (*i.e.* growth rate or fitness) and the biotechnological priority (*i.e.* high productivity).

6.3 Mechanisms causing genetic instability

Since strain instability results directly from the occurrence of genetic mutations, a better understanding of how these changes in the genome occur is of paramount importance [134]. Although not much research on this has been carried out directly on cyanobacteria, eubacteria do share the basic replication machinery, whose (mal)functioning is at the core of the emergence of genetic alterations. Hence, below we summarize what is known about the mechanism of generation of mutations in model organisms like *Escherichia coli* (hereafter, *E.coli*)*, Bacillus subtilis and Pseudomonas aeruginosa*.

There are six known common sources of genetic variation in bacteria, which may lead to the generation of genetic diversity in microbial offspring (Figure 6.1).

I. The most basic form is DNA replication errors. For *E. coli*, this error rate per base pair is about 10^{-10} per generation. This means approximately one mutation in the genome during 2000 cell divisions [134]. However, because of the very high number of cells in a relevant population [208], even after one division, approximately half a million cells will carry a mutation in only 1 mL of *E. coli* culture at an optical density (OD_{600}) of 1. If such a single-base change causes an amino acid substitution, it may disable/inactivate the encoded protein [209].
II. The second type of error is a result of homologous recombination. These errors generally occur between two neighboring long direct repeats (homologous regions, HRs), where DNA sequences between repeats may either be deleted or duplicated, depending on how these HRs recombine. The repetitive use of biological sequences, such as promoters and terminators, may increase the frequency of occurrence of such recombination events. These DNA arrangements will impair the functionality of an (over)expressed synthetic construct. Their frequency is difficult to predict, as it is very much sequence dependent [210].
III. The third type of error is indels caused by DNA replication slippage on simple sequence repeats (SSRs) [211]. During DNA replication, DNA polymerase may temporarily stall [212], then realign the nascent strand to the template strand. Misalignments would easily happen at sites with SSRs, leading to either a deletion or duplication of the sequences between SSRs.
IV. The fourth type of error is caused by mobile genetic elements, like insertion sequences, transposons and lysogenic phages. These are

DNA sequences which are able to integrate into a new site of their resident genome via the process of transposition [213]. Transposition can result in not only gene insertional inactivation, but also in DNA deletions, duplications, and inversions [214].

V. For genes (over)expressed on a self-replicative plasmid(s), variability of plasmid copy number affects the expression level of the target gene(s), and thereby could vary productivity. Furthermore, plasmid loss has been identified as a key factor contributing to limited recombinant gene expression [215], especially for high-copy number plasmids, which form a large cellular metabolic burden.

VI. Last but not least, genetic errors could also be introduced through error-prone DNA replication/repair during growth-limited stress conditions, such as starvation, hypoxia, antibiotic or other stresses [216]. This is particularly relevant because in an industrial setup, especially under long-term cultivation, and/or "toxic" byproduct accumulation, such stress conditions are indeed very common. Genetic errors would occasionally generate fitter mutants, and their selection could affect strain composition and target product formation.

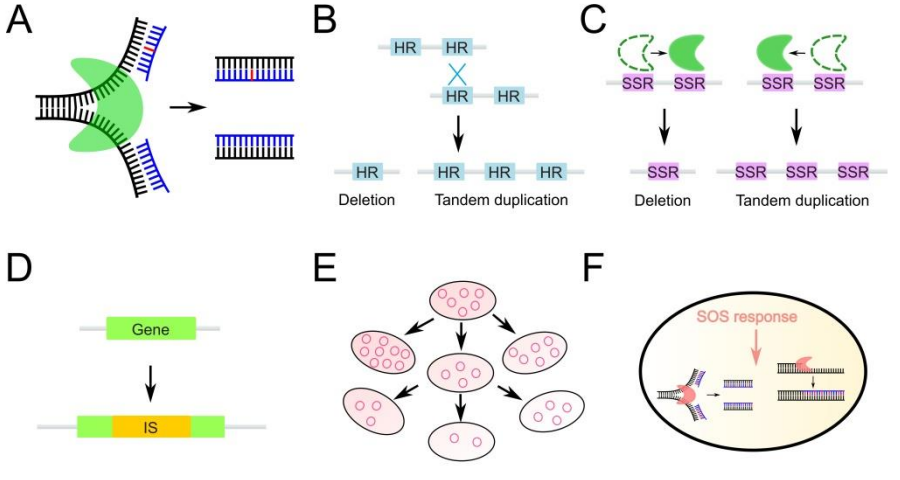

Figure 6.1 Schematic view of six typical sources of genetic instability in prokaryotes. A) DNA replication errors; B) homologous recombination errors; C) DNA replication slippage errors; D) mobile genetic elements errors; E) plasmid loss; F) stress-responsive error-prone DNA replication/repair.

In addition to the listed sources of genetic error, instability can also be caused during DNA methylation at specific sequences [217], a common process in restriction/modification systems. Because DNA methylation causes mismatches for base pairing, those methylated bases are mutational hotspots

[218]. Essentially, strain instability is mainly due to genome instability shaped by microbial evolvability via Darwinian selection [133]. The rate at which this underlying genetic variation can emerge to its manifestation as strain instability depends on many factors (*e.g.* how often non-producing revertant cells occur in the population; the cultivation mode; and how revertant cells perform in comparison to the producing parent). A deep understanding of the population dynamics of such suppressor mutations (and its clear visualization) equips us with the ability to try to predict this process, and distinguish the contribution of each variable involved.

6.4 Population dynamics of suppressor mutants

To unravel the population dynamics when revertant cells appear, in different modes of cultivation, we simulated the development in time of the percentage of revertant cells within a population, using a dynamical growth model. Simulations were performed for both batch and continuous (*e.g.* turbidostat [96]) cultivation. The gradual increase in the fraction of revertant cells within the population was calculated based on the difference in growth rate between revertant cells and a product-forming *Synechocystis* reference strain (Figure 6.2).

For batch cultivation, we assumed the initial OD_{730} to be 0.05 and the final value 20. A conversion factor of OD_{730} to number of cells of 1×10^8 cells $OD_{730}^{-1} mL^{-1}$ [164] was used. Verhulst logistic equation (Eq. 1) was used to model growth in such a batch culture [219] as follows:

$$\frac{dN}{dt} = \mu N\left(1 - \frac{N}{K}\right) \qquad (1)$$

in which, *N* represents the *Synechocystis* population size; μ defines the growth rate (h^{-1}), t represents time in hours, while *K* stands for the maximal cell density (*i.e.* the final OD_{730} of 20), which is assumed to be reached due to nutrient limitations. To simplify the simulation, we consider the percentage of revertant cells to be present in the pre-culture as 0.0001%, 0.001%, 0.01%, and 0.1% of the total number of cells. The growth rate of the producing cells is set to be 0.02 h^{-1}, while revertant cells are assumed to grow with a μ ranging from 0.04 to 0.08 h^{-1} [185]. For the simulations of turbidostat conditions, the thresholds were set to be 0.45 and 0.5 (OD_{730}) [164].

Both the batch and the turbidostat cultivation mode show a similar population dynamics of the revertant *Synechocystis* cells, in the sense that ultimately, given sufficient relative abundance or time, respectively, they will dominate the population at the end of the experiment. As expected, the faster the revertant cells grow, and the more revertant cells are present in the pre-culture, the less

time it takes for revertant cells to become the dominant fraction of the whole population. Compared with batch cultivation, the turbidostat mode tends to favor revertant cells to take over the whole population faster, when all other parameters (*e.g.* growth rates, initial fraction of revertant cells, ***etc***…) remain constant. This is because in the turbidostat mode, no time is wasted by cells that cannot grow because they reached stationary phase. When simulating the population dynamics of cells in the batch mode, even with an initial fraction of 0.0001% of revertant cells in the pre-culture, these cells will exceed 50% of the whole population after only 250 hours of cultivation if their growth rate is four times higher than the growth rate of the product-forming cells. If growth rate differences would be even bigger, which is likely to occur in highly productive strains with a very high burden of product formation [185], even less time would be needed. Since in most instances revertant cells will not synthesize product, the total (*i.e.* biomass-specific) productivity of the culture would drop proportionally with the fraction of producing cells present.

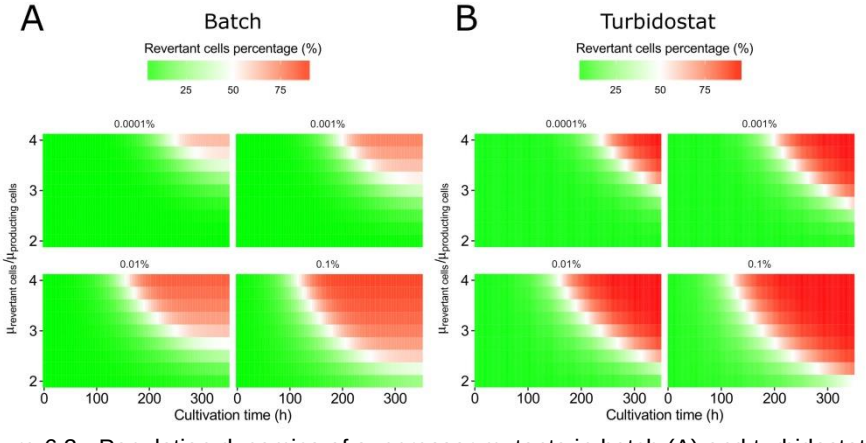

Figure 6.2. Population dynamics of suppressor mutants in batch (A) and turbidostat (B) cultivation mode. The restricted cell growth in batch cultivation mode was simulated based on Verhulst logistic equation [219], while in turbidostat mode respective dilution (10% of the total volume) was considered when a threshold (OD_{730} = 0.5) was reached. The growth rate of producing cells was set to be 0.02 h^{-1}, while the revertant cells ranged from 0.04 to 0.08 h^{-1}. For simulation of both cultivation modes, a start OD_{730} of 0.05 with a conversion factor of 1 × 10^8 cells $OD_{730}^{-1}mL^{-1}$. The percentage numbers on each subplot indicate the portion of reventant cells present in the pre-culture.

The conclusion from the simulations based on the simple phenomenological model presented here, shows that irrespective of the cultivation mode used, the presence of non-producing revertant cells in the population will have a severe negative effect on productivity. The development and implementation of (novel)

strain engineering strategies that aim at preventing suppressor mutants of gaining a growth-rate advantage should therefore urgently be sought after.

6.5 How to obtain more stable production strains

Different strategies have been proposed to obtain producing strains that are stable over an extended period of time. Because productivity is impaired directly by mutations in the heterologous product-forming pathway, enhancing microbial genome stability appears to be the first and a straightforward one. Secondly, production pathway(s) could be selected and designed such that when incorporated into the host's metabolic network, product formation becomes mandatory or beneficial for growth of the cells. Hence, natural selection will stabilize product formation in such strains. Thirdly, product-forming cells grow slower than the corresponding revertant, due to loss of the burden of product formation in the latter. If the size of this burden can be genetically manipulated, producing cells will be less easily outcompeted by those revertants. Consequently, productivity will more likely be maintained. Below, we provide an extensive review of the state-of-the-art for these three strategies.

6.5.1 Decreasing the frequency of occurrence of spontaneous mutations

Enhancing microbial genome stability with the aim of preventing strain phenotypic change, could theoretically be achieved by decreasing the host mutation rate. Because different microbial hosts have different genotypes, choosing the proper host can be very useful to avoid specific sources of genetic instability. For instance, DH5α and XL-1 Blue are two *E.coli* strains, which have been widely employed for molecular cloning experiments because they both lack *recA*, a major component of the homologous-recombination machinery of the cell. Gene inactivation mediated through homologous recombination can thereby largely be prevented [220]. Theoretically, all the mobile elements from a genome can be deleted. Indeed, such efforts have been made for *E.coli* MG1655. The resulting strain, MSD42, has been shown to be completely devoid of IS activity [221]. As predicted, MSD42 displayed better performance for maintenance of unstable genetic constructs, as compared to other *E.coli* cloning hosts [222]. Stress-induced DNA replication/repair leads to error-prone DNA replication, presumably to generate genetically diversified offspring, to increase the chances of survival. Different types of error-prone DNA replication, induced by different stress conditions, are each mediated through variants of DNA polymerase. Deletion of these error-prone polymerases will result in a reduction of the mutation rate under stress conditions in the host [223].

Besides pursuing minimized host mutation rates, another approach would be to improve the genetic stability of the heterologous synthetic constructs. As mentioned above, constructs with long repeats and SSRs have a high risk of undergoing mutation, through homologous recombination and DNA replication slippage, respectively. A web tool, Evolutionary Failure Mode (EFM) Calculator, has been recently developed to computationally detect such motifs in a designed DNA sequence [224]. By altering the codon usage of open-reading frames, or by swapping out promoters and/or terminators by alternatives of equal strength, the presence of sequences prone to mutation could then be efficiently prevented. Host DNA also undergoes frequent chemical alteration during chromosomal replication and repair, because of the possibility of mispairing. A typical example is the DNA methylation due to the activity of restriction-modification systems, like the Dcm and Dam enzymes in *E.coli* [225]. Compared to other bases, methylated bases in these two types of sites experience 3- to 5-fold elevated chances of undergoing mutation [218]. Deletion of these restriction-modification systems can eliminate the corresponding mutational hotspots.

The enormous size of the average microbial population works strongly against our ability to reduce the number of mutations to a level at which it can be neglected. For *Synechocystis* in an industrial setup (*i.e.* 10.000 L and 2 $gDW.L^{-1}$) the total number of cells can easily reach 10^{16}. In such a large population, even if the mutations induced by mobile elements and restriction-modification systems are prevented and heterologous sequences have been designed optimally, it is hard to fully prevent the appearance of revertant cells. Some mutations (*e.g.* DNA replication errors) can never be completely avoided, not even if the most reliable proof-reading DNA polymerase would be used. Hence, preventing genetic instability only through elimination of mechanisms that may lead to mutation is somehow limited, as it may be seen as a strategy to only delay the problem (rather than to eliminate it). More permanent solutions will require novel strategies.

6.5.2 On the relation between growth and product formation

In literature, it is often stated that ‘product formation is coupled to growth (of the producing organism)’. Growth, however, is the result of a complex interplay of a myriad of (bio)chemical reactions plus physicochemical interactions. The simplest way to relate growth of a bacterium to its underlying biochemistry is to split this biochemistry up into the sum of reactions that form its catabolism (to generate free energy from the available substrates) and anabolism (the sum of the reactions in which new building blocks are synthesized, from which in the end new cells will emerge). Product formation can be coupled to either of these two sets of processes. Examples are lactic acid production in lactic acid

bacteria, and the production of acetate in *Synechocystis* [226], respectively. Products that are produced as a (by)product of anabolism often can be re-channeled into intermediary metabolism. However, by knock-out mutagenesis of the enzymes involved in this assimilation, a stoichiometric coupling between anabolism and product formation can be achieved [226].

Significantly, the degree of coupling between catabolism and anabolism can vary a lot between different classes of organisms [227], and between physiological conditions (*e.g.* through the uncoupling activity of weak organic acids). Hence the degree of coupling between growth (particularly: catabolism) and product formation can also vary. However, not only coupling between catabolism and anabolism itself is variable, such variability can also occur in the coupling of product formation to either anabolism or catabolism itself. The latter depends on the biochemical redundancy of the metabolic network of the cell. To give an example: In homolactic fermentations the production of lactic acid is strictly and stoichiometrically coupled to catabolism (*i.e.* not necessarily to growth; the latter depends on the degree of coupling between anabolism and catabolism: During strong uncoupling by weak acids, the production of lactic acid may continue long after cells have stopped growing, as long as sufficient sugar is available (see also above)). In heterofermentative (lactic acid) bacteria the energy-generating catabolic flux of sugar can to a variable extent be converted into lactic acid, and into any of a series of alternative products, like ethanol, hydrogen, acetic acid, *etc.* The regulation of carbon flux over these different end products is usually very complex and mostly not resolved at the molecular level. Therefore, in such examples formation of any specific product may show any degree of coupling to catabolism/growth. Similarly, divergent metabolic routes may present themselves for product formation linked to anabolism: Metabolic intermediates like fumaric and acetic acid are liberated in the pathways that lead to the synthesis of specific amino acids. The cellular content of a specific amino acid can, however, vary and also alternative routes of supply of the amino acid(s) may be available.

In conclusion, in selected cases product formation in bacteria may stoichiometrically be coupled to either anabolism or catabolism, but this cannot be taken for granted and needs to be investigated physiologically for every individual cell factory/product combination. Furthermore, often this coupling will be quite sub-stoichiometric. However, when this (strict) coupling exists, the intrinsic evolvability of living organisms can be exploited to select for faster growing variants, and hence for variants that then will provide higher (volumetric) productivity.

In engineered cell factories, *e.g.* when a heterologous pathway for biofuel production is engineered into a cyanobacterium, the new pathway is coupled to

neither catabolism, nor anabolism, but is rather inserted and operating orthologous to these two parts of the cellular metabolic network. The same holds for the landmark achievement in synthetic biology: artemisinin synthesis in *Escherichia coli* [228]. In such situations the orthologous pathway directly competes with anabolism for substrate(s), and may thus form a metabolic burden for the cell factory, with the consequence that it will be out-competed by spontaneously originating loss-of-function mutants.

In fermentative cell factories the generally low ATP-yield of catabolism, and the requirement that overall metabolism must have a closed redox balance, may supply additional mechanisms to stabilize/favor product-forming organisms over revertants. For instance, *E.coli* has been genetically engineered to enhance the anaerobic operation of the oxidative tricarboxylic acid (TCA) cycle, such that more of the reducing power generated is directed to compensate for redox equivalents (*i.e.* NAD(P)H) required for the production of a very reduced compound, 1,4-butanediol [229].

In selected cases metabolism can be re-engineered such that growth becomes fully dependent on the synthesis of a specific (by)product. For instance, by blocking all the metabolic routes from α-ketoglutarate to succinate, to disrupt TCA cycle activity in *E.coli*, introduction of a proline-4-hydroxylase, which can mediate the same conversion, allows the cells to resume growth. Because proline-4-hydroxylase also converts proline to hydroxyproline, the production of this latter compound can thereby be coupled to cell growth [230].

6.5.3 Computational tools for the design of strains with product formation coupled to anabolism or catabolism

Computational tools have played an important role to confirm or identify candidate compounds of which the production is coupled stoichiometrically to cell growth. These tools are based on the analysis of microbial genome-scale metabolic models (GSM), constructed for *in silico* simulation of phenotypic characteristics of the organism under study (*e.g.* impact of a gene deletion). The most widely-used tools for simulation use Flux Balance Analysis (FBA) [168] and Flux Variability analysis (FVA) [169], to predict the adjustment of the distribution of the metabolic fluxes after a certain pathway has been modified. Accordingly, several computation algorithms have been developed [152,231,232], which greatly facilitate the analysis of coupling of metabolite production, like *e.g.* OptKnock [152]. For the production of a selected target compound, OptKnock suggests which gene(s) have to be deleted, such that flux towards product formation is maximized [152]. Initial tests showed that the computational results obtained with OptKnock agreed well with the experimental results available from the literature [152], which highlights the usefulness of this computational

tool. However, if one allows more than two genes to be deleted *in silico* by OptKnock, the number of gene combinations that have to be analyzed rapidly increases. Computational costs increase exponentially with the number of genes to be deleted, and this limits the possibilities for further simulation. To circumvent this problem, new algorithms, such as OptGene [233], OptFlux [234], and FastPros [235] have been developed. These latter tools aim for more efficient *in silico* simulation with less computational costs. This is realized using an iterative approach, instead of the comprehensive one used in OptKnock.

To have more powerful computational algorithms available, capable of target compound identification for production coupled to anabolism, we have recently developed an algorithm to 'Find Reactions Usable In Tapping Side-products' – FRUITS [226] (**Chapter 4**). This tool analyzes existing GSMs for the identification of side-products of anabolism to which the production can be coupled by deletion of their respective assimilation/re-utilization pathway(s). If such metabolites can be identified, the corresponding genetic intervention(s) would enable the production of these metabolites, strictly coupled to biomass synthesis. As the deletion approach implies very low chances of genetic reversion, production systems based on such mutant would be expected to be genetically stable. Even more, the evolutionary potential of the cells that would lead to faster-growing cells, would increase the rate of product formation as well. Based on a GSM of *Synechocystis* [18,154] a total of 9 compounds have been confirmed to be coupled to *in silico* growth. This result has been obtained with a knockout of maximally 4 genes. Out of these 9 candidates, acetate [226] (**Chapter 4**) and fumarate [236](**Chapter 5**) were chosen for experimental validation. The model-guided engineered strains show an obligatory coupling between compound production (*i.e.* acetate and fumarate) and photoautotrophic growth, as predicted. Furthermore, stable productivity was confirmed by performing prolonged turbidostat cultivations [226].

It has to be kept in mind that the approach using FRUITS generally may suffer from some inherent limitations. Firstly, the number of compounds which can be produced using this approach is somewhat limited. That is because these compounds need to fulfill specific criteria, such as being a side-products, and no reaction can be selected for knock-out that prohibits intracellular redox-balancing. Secondly, the flux towards compounds synthesized as a side product of anabolism will generally not be very high, which may limit the production of large quantities. And thirdly, the growth/anabolism-coupled approach relies on cellular growth – and for batch cultivation, at stationary phase productivity will halt. For this reason, it is worthwhile to also explore product formation in the stationary phase, where cell growth has stopped, but a large population of cells is potentially available for product formation.

6.5.4 Uncoupling product formation from growth of the cells

Rather than trying to couple product formation to cell growth, in batch culturing systems, one may also try to deliberately uncouple product formation from growth by bringing the genes, encoding a product-forming pathway, under inducible (transcriptional) control. These latter genes would then be induced only after growth has halted, for instance because cells have reached the stationary phase. This approach prevents direct competition for resources between growth and product formation and hence lowers the risk that the producing cells will be overgrown by non-producing mutant derivatives. However, for this approach it is very important to clearly state which type of metabolism one is considering in a specific example. The three relevant types of metabolism in this respect are: phototrophic, respiratory and fermentative metabolism. In batch cultures of respiratory and fermentative metabolism, in contrast to phototrophic, the thermodynamic driving force for growth rapidly dissipates itself at the time when cells reach stationary phase. This is different in phototrophy: As long as (the) light is on, all metabolic pathways in a phototroph (*i.e.* also the orthogonal ones) receive enough free energy to allow a significant metabolic flux. For respiratory and fermentative metabolism this is only true when specific precautions are made in the medium composition, so that stationary phase is induced by mineral limitation and not by exhaustion of the source of free energy and carbon for growth and product formation of/by the cells. Accordingly, the genes encoding the product-forming enzymes would be repressed during exponential growth, to minimize the differences in growth rate between producing cells and spontaneously appearing non-producing revertant cells, so that producing cells would not be outcompeted. During stationary phase, in which (net) cell growth has ended, the biological burden can be increased to maximal levels, provided enough carbon and free energy is available for this (see above). Under these circumstances, revertant cells cannot out-compete the producer cells any more, due to the limited growth of both types of cells in the stationary phase. Alternatively, target gene(s) expression for product formation can be initiated in the late-exponential growth phase.

In the early days of microbial physiology this approach of uncoupling product formation from growth of the cells was studied extensively in chemostat systems, in the absence of any genetic engineering. It was shown that the rate of product formation in selected chemotrophic bacteria [237], when surplus carbon- and energy source is available (so-called ‘overflow metabolism’), depends strongly on the chemical nature of the mineral limiting growth rate of the cells. Presumably, this is due to the mechanisms of metabolic regulation operative in the specific cell factory selected. The selection of the optimal

nutrient for uncoupled production, that will drive the batch culture into stationary phase, presumably can be selected on the basis of the knowledge derived from such 'overflow experiments'. It is relevant to note that in phototrophs, metabolism generally is strictly regulated, often by mechanisms based on redox-regulation. This causes metabolism to halt when growth stops. It will be of interest to find out whether or not in phototrophs conditions can be found that will allow 'overflow metabolism'.

Generalist microbes contain a large number of regulation systems, to control a myriad of reactions under a variety of extra- and intracellular conditions [238]. Among them, inducible promoters and transcription factor–based genetic regulation systems [239] play important roles. Inducible promoters can be used to precisely switch on/off the expression of (a) target gene(s) at pre-set time points. For instance, if we aim to start up product after (exponential) growth of the cells for improved strain phenotypic stability, a simple approach would be to add an inducer that allows target gene expression only after the exponential phase. The inducible promoter should be tightly regulated, with no/little gene expression during the exponential phase, and drive high-level expression of target gene(s) afterwards. For cyanobacteria, only a few truly inducible promoters have been reported, including ones induced by physical factors, like temperature [240] and light [241–243], and by chemical signals, *e.g.* IPTG [36], anhydrotetracycline [244], and any of several metal-ions [245–247]. Among them, the P*nrsB* promoter, induced by nickel ions, has been tested as one of the tightly-regulated promoters [28,248], that can be applied to selectively initiate product formation after passage of the exponential phase. Beyond these several strictly regulatory promoter systems are available from chemotrophic bacteria [249–251].

The use of transcription factor–based signal-transduction systems, commonly containing a sensory input module, several control devices, and at least one output module with transcription-modulating activity, permits flux adjustment in a dynamic mode, which is preferrable [252]. This is due to the fact that nutrient concentrations and the cellular environment are changing during cultivation. Under these conditions, such regulation allows dynamic adjustment of metabolic fluxes, to better manage the trade-offs between growth and product formation and can further help to avoid build-up of undesired intermediates [253]. The use of such transcription factor–based signal transduction systems, designed and applied in a variety of micro-organisms, has significantly boosted product formation [254–259]. In general, the use of flexible inducible regulation systems can help microbes to better cope with the imposed metabolic burden, thereby enhancing microbial robustness and potentially improve strain stability. Yet, the limited pool of metabolite-responsive sensors limits their broad application. Novel approaches to identify or design the sensory part of such

regulatory circuits is therefore necessary [260–263]. For cyanobacteria, some progress has been reported already with sensors responding to oxygen [264], or the nitrogen status [265]. However, compared to other microorganisms, like *E. coli*, the choice of inducible gene expression systems in cyanobacteria is much more restricted.

6.5.5 New design strategies to uncouple product formation from growth of the cells

Novel strategies for efficient cellular resource re-allocation to stabilize and increase product formation have been reported recently. One way is to create a synthetic growth–switch, through either control of transcription, of global mRNA degradation, or of an essential anabolic pathway. For instance, Izard *et. al.* have reported a growth–switch, developed through regulating RNA polymerase synthesis, using an IPTG-induced promoter [266]. By manipulating the IPTG concentration in the medium, bacterial growth can be dynamically controlled, and this allowed the authors to achieve a two-fold increase in the product titer. Another study reported construction of a similar system with the inducible expression of the sequence-dependent endoribonuclease MazF to control global transcriptome degradation, while in the heterologous genes, necessary for product formation, the MazF recognition sites were removed through recoding of the sequence, while maintaining the amino acid sequence of their reading frames [267]. Cell growth was arrested when *mazF* was expressed. The heterologous genes remained comparatively highly expressed, so that more cellular resources were allocated to product formation, rather than to growth. A similar system was recently created by conditionally expressing the TCA cycle for itaconic acid production by *E.coli* [268]. A temperature-dependent promoter was used to selectively control the expression levels of *icd*, encoding isocitrate dehydrogenase, a key enzyme for a functional TCA cycle. A knock-down of *icd* enables a switch from growth to product formation. Such a two-stage process may significantly increase maximal product titers that can be reached.

These latter strategies are reminiscent to how a phage/virus hijacks a cell to use the host's cellular resources to rebuild itself. During this process, a virus can efficiently stop or minimize growth of its host, and manages to replicate and synthesize its own DNA and proteins. This virus infection process thus provides a good concept for efficient cellular resource re-allocation. However, for such strategies to be successful the availability of finely tunable gene expression systems is an essential prerequisite.

6.6 Identification of new inducible gene expression systems

New inducible gene expression systems, *e.g.* inducible promoters, can be identified and developed following both rational and random approaches. The rational approach commonly relies on measuring the abundancy of intracellular mRNA levels with microarrays or RNA-seq, under a range of different conditions. That information can then be used as an indication of how the different promoters behave. For instance, Berla *et al.* investigated the genes which were highly expressed in the stationary-phase of *Synechocystis* [269]. Those highly expressed genes hint to the possibility that their promoters were up-regulated in that growth phase. Because cells enter stationary phase due to a nutrient or environmental limitation, it is important to find the specific condition(s) to which a specific promoter responds. Kopf *et al.* then studied the changes in gene expression, and by inference the activity of the corresponding promoters, in *Synechocystis,* also using RNA-seq, under a range of different stress conditions [270], including environmental stress conditions like the extremes of light intensity and temperature, and nutrient limitation, *e.g.* of carbon, nitrogen, phosphorus, and iron. Through detailed analysis of those data, useful information for promoter(s) induced by each condition is obtained. Inducible promoters that may be identified this way still need to be further tested in the relevant microbial host, using a suitable reporter gene.

The random approach, through the utilization of a DNA library, has also been tested to identify new inducible promoters. This is typically achieved by first creating a library of randomly digested genomic DNA fragments, which are then used, together with a reporter gene, to characterize those fragments under different conditions. Promoters induced under a specific condition can be identified by screening the promoter library under the corresponding condition. By monitoring the signals from the library, target constructs containing a "promoter fragment" can be selected and isolated. After sequencing the target fragment, the corresponding gene and its function can also be characterized.

For screening purposes, in principle, a promoter library could be integrated either in a replicative plasmid or into the host's chromosome. For cyanobacteria, it is comparatively easier to have this library on a replicative plasmid, especially for those species, which are polyploid (*i.e.* those that contain multiple chromosome copies per cell). This is because it is time-consuming and difficult to achieve a fully segregated, chromosome-based promoter library. On the other hand, although no segregation step is needed when a replicative plasmid is used, its copy number might vary even more than the number of chromosomes per cell within a population [271]. This causes uncertainty in assays of the transcriptional strength of a promoter fragment. So far in cyanobacteria, the only promoter library that has been constructed was harbored in a replicative plasmid [272]. It would be of interest to know whether

such a library can actually also be constructed on the basis of a chromosomal location. This of course will depend on whether the cyanobacterial ploidy issues can be solved.

6.7 Additional challenges for cyanobacteria as efficient cell factories

As discussed above, cyanobacterial ploidy generally impedes genetic modifications when such modifications are based on a chromosomal location. *Synechocystis* has a copy number ranging from about 2 to 20, depending on cultivation conditions and growth phase [69]. Hence, when positive colonies appear on an agar plate after transformation, commonly, additional segregation steps are necessary. To achieve full chromosomal segregation, homozygous mutants are usually selected with an increasing dosage of the antibiotic. Depending on the specific gene to be modified, this process could take up to a few weeks or even more. Recently, a CRISPR-based technique has been validated to accelerate this segregation process in certain cyanobacterial species [149,273]. Yet, how to incorporate this technique for efficient genome editing and other purposes, like constructing a chromosomal promoter library, will require more work. The exact mechanism behind the (cyano)bacterial ploidy is unknown, though different hypotheses were proposed that associate an evolutionary advantage to it [274]. Hypothetically, ploidy might not only allow for low mutation rates and gene redundancy for improved evolvability under extreme conditions, but also potentially ensure a less stochastic regulatory network and global gene dosage control [275]. It can also provide additional template for repair of damaged DNA. More experimental evidence is necessary to clarify or validate these issues. Such insight would make these organisms more suitable for genome editing.

Cyanobacteria generally have a low maximal growth rate, compared to other model microbes, like *E. coli* and *Saccharomyces cerevisiae.* The low growth rate clearly impairs their ability to reach high biomass-specific and volumetric rates of production. The doubling time of *Synechocystis* in many published studies is between 12 and 16 hours, although at specific cultivation conditions, this time can be shortened to less than 6 hours [100]. For other cyanobacteria, *e.g. Synechococcus* sp. PCC7002, ~ 2.5 h doubling time has been reported under conditions with high irradiance levels [276]. Interestingly, the high rates of photosynthetic electron transfer, especially cyclic electron flow (that leads to ATP formation, rather than the combination of NADPH and ATP that is the result of electron transfer through the Z-scheme), that occur under these conditions, were assumed to be important for this. *Synechococcus* UTEX 2973 has also been reported as a very fast-growing cyanobacterium (with a doubling

time of even less than 2 h under optimal conditions). Compared to its close relative - *Synechococcus elongatus* PCC7942 which has a two-times lower maximal growth rate - surprisingly, only a small number of differences in its nucleotide sequence were found between their genomes [277,278]. These studies may help us understand what is necessary to make a cyanobacterium grow fast, to make it a better chassis for a cyanobacterial cell factory.

6.8 High throughput screening systems

High throughput screening, permitting multiple tests simultaneously, is urgently needed for a range of aspects of cyanobacteria. For instance, the above-mentioned random approach to build a promoter library can be further exploited with a high throughput screening system to efficiently characterize large numbers of individual clones. So far, the systems available have been developed essentially for chemoheterotrophic microbes. For cyanobacteria, at least the use of an additional light source is required. For microplate based screening systems, a simple modification is to put the microplate inside an illuminated incubator [185]. Simple as it looks, the microplate however needs to be manually measured in a microplate reader at target time intervals for several days. Thus, a microplate system enabling controlled automatic measurement with custom light sources would be a significant improvement. Efforts are made to devise such a setup by adding programmable illumination regimes to a multimode microplate reader [279]. Besides modifications to a microplate reader, a completely new platform has also been developed, which relies on a pipetting robot for automatic sample processing [77]. Ideally, such high throughput systems should enable the measurement of both OD, and fluorescence. Minimal liquid evaporation, efficient gas exchange and availability of a range of light sources (*i.e.* for actinic light, a measuring beam for OD_{730} and fluorescence excitation light) would further optimize the versatility of such a system.

Beyond microplates, microdroplet-based single cell analysis systems have also been developed for analysis and screening of cyanobacterial production strains. This is basically realized by first encapsulating single cells in a microdroplet, and then monitoring its product formation, usually via a fluorescence signal derived from enzymatic conversion of the product into a fluorescent product. Hence within a population of cells, it is possible to characterize the production capacity of each single cell. Highly productive strains could then be selected with a sorting system [112]. In recent studies, such microdroplet systems have been described for cyanobacteria producing lactate [280] and ethanol [75]. Because such systems allow single cell analysis within a population, they may prove very useful for library screening, like a promoter library. Furthermore, it would be very advantageous if a variety of compounds

could be detected in such a system, even more if the intracellular portion of the target product could be measured as well.

6.9 Trends towards the future

Optimized, stable, cyanobacterial cell factories for product formation should, whenever possible, combine aspects of growth-coupled and growth-uncoupled characteristics; that is, product formation must be enabled in a growth-coupled fashion during the exponential growth phase, such that revertant cells will have little chance to outcompete the producing cells. During the stationary phase, *i.e.* when cells do not grow, a growth-uncoupled system would allow continued product formation. Because "stable" here of course is always relative, a reduction of microbial evolvability by decreasing the spontaneous mutation rate is also preferred. This can be achieved by making the synthetic construct for the heterologous genes less prone to spontaneous mutation, and by reducing the microbial host's spontaneous mutation rate. For each of these strategies, the corresponding toolkit must be further developed. For instance, computational tools would help identify compounds that can be synthesized in a growth-coupled mode, while a high-throughput screening system would tremendously assist the identification and development of inducible gene expression systems. In addition, if cyanobacterial ploidy can be rationally manipulated, genetic modification would be greatly facilitated. Furthermore, a deeper understanding of what determines the growth rate of a cyanobacterium will help to design a super-efficient cyanobacterial chassis. By tackling all these challenges, the design and construction of stable and efficient cyanobacterial cell factories, which may support the sustainable development of human society, will be significantly accelerated.

6.10 Conclusions

In this chapter, we tackled one of the main challenges in enabling cyanobacteria to convert CO_2 into commodity products: sustained stable productivity. The need to obtain high and sustained productivity of a product by a microbe can be regarded as a battle against the microbial distinctive ability to evolve. Keeping in mind that microbial evolvability is inherent to life, we can either attempt to directly undermine their evolvability by force, *i.e.* by decreasing their random/spontaneous mutation rate, or try to indirectly make full use of their ability to form product, *i.e.* by aligning product formation to fitness. By efficiently manipulating cellular resource allocation, more resource can be stably channeled towards product formation. With the further development of toolkits and methods, and the implementation of production-stabilizing strategies or a combination thereof, obtaining stable and efficient cyanobacterial cell factories will be greatly facilitated.