Tuning for light and more

Engineering phototrophy and membrane proteins in Escherichia coli

Nicolaas J.H.P. Claassens

Tuning for light and more

Engineering phototrophy and membrane proteins in *Escherichia coli*

Nicolaas J.H.P. Claassens

Thesis committee

Promotors
Prof. Dr John van der Oost
Personal chair at the Laboratory of Microbiology
Wageningen University & Research

Prof. Dr Willem M. de Vos
Professor of Microbiology
Wageningen University & Research

Co-promotor
Prof. Dr Vitor A.P. Martins dos Santos
Professor of Systems and Synthetic Biology
Wageningen University & Research

Other members
Prof. Dr René H. Wijffels, Wageningen University & Research
Prof. Dr Klaas J. Hellingwerf, University of Amsterdam
Dr Jan-Willem de Gier, Stockholm University, Sweden
Dr Arren Bar-Even, Max Planck Institute Potsdam-Golm, Germany

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences).

Tuning for light and more

Engineering phototrophy and membrane proteins in *Escherichia coli*

Nicolaas J.H.P. Claassens

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 17 March 2017
at 1.30 p.m. in the Aula.

Nicolaas J.H.P. Claassens
Tuning for light and more - Engineering phototrophy and membrane proteins in *Escherichia coli*, 328 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2017)
With references, with summaries in English and Dutch

ISBN: 978-94-6343-092-0
DOI: 10.18174/405559

Table of contents

Chapter 1

General introduction and Thesis outline

General introduction

Phototrophy supporting life and biobased production

Sunlight is the main primary energy source for our planet, and has been the major supplier of energy for biological systems for billions of years. Biological photosystems can convert the energy of sunlight into biochemical energy carriers (e.g. NADPH and ATP) to support the growth of many light-harvesting, phototrophic organisms. Phototrophy literally means 'feeding on light'. Some phototrophic microbes, the so-called photo*hetero*trophic organisms, can use light as an energy source, but also require additional organic substrates for growth. On the other hand, many phototrophic organisms are photo*auto*trophs, implying that they can perform photosynthesis, i.e. grow by using light as their sole energy source and CO_2 as their only carbon substrate. In nature a wide variety of photosynthetic organisms can be found, including bacteria such as cyanobacteria, purple and green photosynthetic bacteria, as well as eukaryotes, including algae and plants (Hohmann-Marriott and Blankenship, 2012).

Photosynthetic organisms play a key role as primary producers in many ecosystems, and plants are the basis for food and feed production. Photosynthesis is also seen as a promising process to solve one of our societies' grandest challenges: the urgent need to replace the ongoing fossil-resource-based production of chemicals and fuels by more environmental-friendly alternatives. These alternatives should be sustainable, implying their use does not lead to high net CO_2 emissions and fossil resource depletion. Many alternative, so-called 'biobased' platforms for the production of chemicals and fuels based on photosynthesis have been proposed, some of which are being developed on industrial scale. A common biobased solution for the production of chemicals and fuels is the conversion of plant biomass into products through microbial fermentation. In these so-called first generation biobased approaches, plant-derived substrates, such as glucose and starch, are being converted into several chemical and fuel compounds by heterotrophic microorganisms (Aro, 2016). As a second-generation process, the non-edible fraction of plant biomass, such as lignocellulose, is being explored as a substrate for product formation by heterotrophic microorganisms (Liao

et al., 2016); this type of biomass would decrease the competition with food and feed production. In addition, efforts are ongoing to completely get rid of plant-based biomass as a substrate. Therefore as the third-generation of biobased production, photosynthetic microorganisms, such as microalgae and cyanobacteria, are being considered for direct conversion of (sun)light, CO_2 and H_2O into biomass and product compounds (Wijffels et al., 2013). This third generation would allow for eliminating the need to grow, harvest and process plants, which would potentially allow for more direct and efficient production of chemicals and fuel compounds from light, CO_2 and H_2O.

Unfortunately, however, photosynthetic organisms, including plants, microalgae and cyanobacteria, are quite inefficient in converting the available energy of sunlight into biochemical energy carriers (Conrado et al., 2013; Zhu et al., 2008, 2010). This inefficiency is a major challenge and concern for developing efficient, cost-effective, and sustainable production platforms. Hence, there is a major incentive to look into developing more efficient photosynthetic production platforms, for which developments in the field of synthetic biology are encouraging. Engineering more efficient synthetic photosynthetic cell factories is a promising but also challenging task towards realizing fourth-generation biobased production platforms (Aro, 2016). Towards this goal, in this thesis we explore the engineering of simple photosystems using novel systems and synthetic biology tools.

The inefficiencies of biological 'light reactions'

Theoretical maximum solar-to-biomass energy conversion efficiencies of plants are around 6% (**Figure 1**), however, in practice maximum efficiencies reached are only about 4% and generally close to 1% (Zhu et al., 2008, 2010). Also for microalgae, maximum energy conversion efficiencies of only ~3% are reached (Wijffels and Barbosa, 2010). A major part of efficiency loss in photosynthesis is because 50% of the available solar light energy cannot be used by the typical oxygenic photosystems found in all plants, microalgae and cyanobacteria (Blankenship et al., 2011; Zhu et al., 2008, 2010). Their ubiquitous oxygenic photosystems only use light of 400-740 nm. However, ~50% of the solar energy is available in wavelengths longer than ~740 nm. The ubiquitous

oxygenic photosystems cannot use these longer, lower energy wavelengths to excite electrons in their reaction centers, leaving this part of solar energy unused. On top of this, many photosynthetic organisms, reflect part of the green light (~500-600 nm), for example for plants resulting in the loss of another ~5% of the available solar energy (Zhu et al., 2008, 2010). Furthermore, the maximum energy per photon that can be used for excitation in typical oxygenic photosystems is that of a red photon (typically 680-700 nm), so all additional energy in shorter wavelengths is lost. The latter is referred to as the photochemical inefficiency, which results in another 7% loss of available solar energy. Another ~11% of available energy in sunlight is lost dictated by thermodynamics. In summary, at least ~74% of the available solar energy is lost in oxygenic photosystems, mainly due to their incapability to absorb infrared light (**Figure 1**).

Apart from the mentioned losses in biological photosystems, also referred to as 'light reactions', large energy lossces occur in the 'dark reactions'; these include reactions such as CO_2 fixation pathways, which will not be discussed in detail here. However, the inefficiency of the ubiquitous Calvin-Benson Cycle for CO_2 fixation and promising alternative solutions hereto are discussed in **Chapter 3** and **Chapter 8.** For the production of specific chemical compounds by photosynthetic cells, another major loss often occurs as only a part of the metabolic fluxes is leading to the product of interest; in most cases, a high amount of biochemical energy and fixed carbon ends up in undesired biomass and by-products. Metabolic engineering of production pathways has in some cases been powerful to redirect metabolism towards higher product synthesis, as illustrated by a range of products that were generated through metabolic engineering for heterotrophic hosts (Liao et al., 2016). In recent years, similar metabolic engineering approaches have been applied for some photosynthetic hosts as well, mainly in cyanobacteria, where they have led to more efficient production of several compounds (Savakis and Hellingwerf, 2015).

Still the major energy losses occur in the upstream 'light reactions', hence they are a window of opportunity for improving photosynthetic efficiency (Blankenship et al., 2011; Ort et al., 2015). Apart from the ubiquitous oxygenic photosystems, a broad diversity of alternative photosystems is found in prokaryotes, providing inspiration and options to tackle the aforementioned

issues of photosynthetic inefficiencies. The emerging field of synthetic biology will be instrumental towards engineering of more efficient photosynthesis (Aro, 2016), as will also be explored further in this thesis.

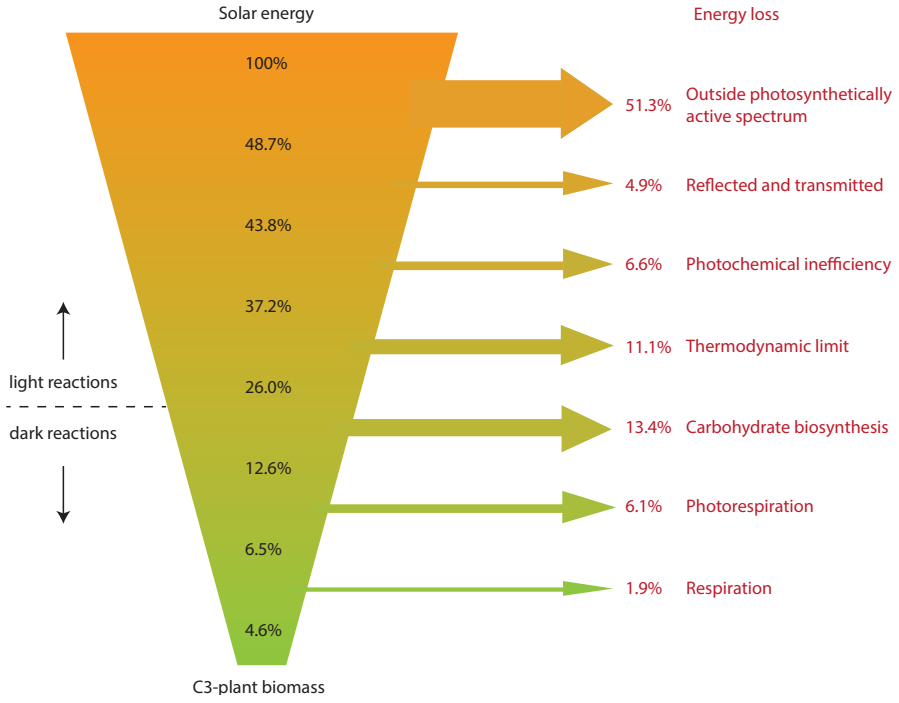


Figure 1. Maximum theoretical solar-to-biomass energy conversion efficiencies of oxygenic photosynthesis, including energy losses in the 'light' and 'dark' reactions, depicted for a C3-plant, representative for most oxygenic photosynthetic organisms. Adapted from (Zhu et al., 2010).

Inspiration from alternative photosystems in nature

Apart from the oxygenic photosystems of plants, algae and cyanobacteria, a wide diversity of other photosystem types can be found in the prokaryotic world (**Table 1, Figure 2**) (Hohmann-Marriott and Blankenship, 2012). Oxygenic photosynthesis is driven by a tandem of photosystem II (PSII) and photosystem I (PSI), both having reactions centers with chlorophyll pigments. Chlorophyll pigments in the core of the reaction center are being exciting with the energy of an absorbed photon. The absorption of photons by PSII can be used to energize electrons derived from water, hereby splitting water in protons and oxygen. Electrons derived from water have a very positive redox potential, and

their energy level will be excited by tandem operation of PSII and PSI and some redox carriers. Finally, in non-cyclic electron transport electrons will end up in redox carriers with a very negative redox potential (e.g. NADPH or ferredoxin). Part of the energy derived from the absorbed photons is also converted in electron transport chains to pump out protons resulting in a proton motive force (PMF), which can be used to drive ATP synthesis. The pigments used by plants, including chlorophylls and additional carotenoid pigments in light-harvesting antennae, can only harvest light of wavelengths in the range of 400-700 nm. Chlorophyll *d* and *f* variants present in cyanobacteria allows them to harvest available light a bit more efficiently, by expanding the light-harvesting range up to ~740 nm (Blankenship et al., 2011; Croce and van Amerongen, 2014).

Apart from oxygenic photosystems, a wide variety of anoxygenic photosystems has been found in bacteria. These anoxygenic photosystems harbor bacteriochlorophyll-pigmented reaction centers, which can also absorb longer wavelengths, even up to ~1100 nm (Blankenship et al., 2011; Ort et al., 2015). In addition, these photosystems do not use water as an electron donor, hence they do not split water and do not generate oxygen. Instead, anoxygenic phototrophs use higher energy potential electron donors, such as hydrogen, sulfur, hydrogen sulfide, or organic molecules. The fact that these alternative electron donors are generally less abundant than water, implies that these anoxygenic phototrophs are less widespread in nature. However, the ability of some anoxygenic photosystems to use almost the complete spectrum of available sunlight is a promising feature. In recent years several studies have suggested thermodynamically feasible proposals for improving oxygenic photosystems through replacing their PSI by an anoxygenic photosystem that can harvest longer wavelengths (Blankenship et al., 2011; Chen and Blankenship, 2011; Ort et al., 2015). Such engineering proposals could drastically expand the potential spectrum of light to be harvested. However such proposals are highly challenging to implement, for example given the relatively high number of proteins, electron carriers and pigments involved in anoxygenic photosystems.

A simpler type of photosystems are proton-pumping rhodopsins (PPRs) (**Figure 2**), which consist of one transmembrane protein, which is covalently bound to a single retinal pigment molecule. By the absorption of a photon, the

retinal pigment and the protein undergo conformational changes, eventually leading to the extrusion of a proton from the cell against the electrochemical gradient. The light-generated proton motive force is a form of chemical energy that can be used to regenerate ATP and facilitate energy-demanding transport across the membrane. These simple photosystems are present in a wide range of archaea, bacteria and also in some eukaryotes (Ernst et al., 2014). Unlike the more complex oxygenic and anoxygenic photosystems, rhodopsins are not capable to excite electrons, implying that PPR systems cannot solely support autotrophic photosynthesis, as therefore external electron donors are required. However, PPR photosystems have for example been observed to supply additional ATP for the survival of (photo)heterotrophic organisms which can encounter low energy conditions, in oligotrophic environments, such as in the oceans (DeLong and Beja, 2010).

So far, a PPR photosystem has been identified in only one of the cyanobacteria, the primitive thylakoid-less cyanobacterium *Gloeobacter violaceus*. Like all cyanobacteria, this organism also possesses PSII and PSI, and it has been hypothesized that the typical absorbance of this rhodopsin in the green spectrum (around ~540 nm) may complement the somewhat lower absorbance of PSII and PSI in the green light spectrum (Choi et al., 2014). However, it has to be noted that *G. violaceus* possesses for green light absorption also phycobilisomes, which are antennae complexes present in cyanobacteria but absent in plants and algae. Therefore the exact physiological role of the PPR system in *Gloeobacter violaceus* remains unelucidated so far. However, given the typical green absorption spectrum of PPR photosystems it is suggested in this thesis (**Chapter 2**) and by some others (Chen et al., 2016) that the transplantation of PPRs into oxygenic phototrophic bacteria, and on the longer term into algae and plants, may fill up their green gap in light-harvesting.

More challenging and promising would be to apply PPRs to absorb the mostly unused spectrum above 700 nm. Even though no rhodopsins have yet been found in nature that use wavelengths above 600 nm, this is thermodynamically feasible. Far beyond 600 nm, photons with far infrared wavelengths above 1100 nm contain still more than enough energy per photon to power the transport of one proton against a typical cellular electrochemical gradient. It is an outstanding question if PPRs or other light-driven proton-pumps exist in nature that can use these wavelengths. Given the typical absorption spectra

of the rhodopsins, in the 400-600 nm range, reaching these wavelengths may seem far-fetched. However, attempts have been made to red-shift the absorption of actively pumping rhodopsins above 600 nm. Towards this end, protein engineering approaches as well as introducing synthetic retinal analogues have been attempted in recent years, with some encouraging results, though often large wavelength shifts compromise proton-pumping rates (Engqvist et al., 2015; Ganapathy et al., 2015; Kim et al., 2008; Liu and Asato, 2003). A very promising step forward was published very recently, for the first time reconstitution of PPRs with a retinal analog resulted in proton-pumping in the near infrared region (Ganapathy et al., 2017). In this work one tested retinal analog reconstituted in a proteorhodopsin mutant even gave an absorption maximum of 740 nm, while retaining pumping capacity, which opens new possiblities for PPR photosystems.

Despite the potential of absorbing longer wavelengths, it has to be noted that the efficiency of PPRs (pumping only one proton per absorbed photon) is substantially lower than the efficiency of reaction center photosystems (pumping 2-3 protons per photon and having the capability of exciting electrons as well). However, for engineering photosystems, PPRs have the advantage over reaction center photosystems that they are more simple, requiring less genes. For the production of PPRs only about 6 genes are needed, while for reaction center photosystems at least ~30 genes are required (Bryant and Frigaard, 2006). Hence, PPR photosystems are more simple systems to explore and study, in particular with respect to their transplantation to heterologous hosts.

This thesis explores the potential of transplanting photosystems using *Gloeobacter* rhodopsin (GR) as a main PPR model (**Chapter 4**). GR is a special PPR variant, as it is one of the few PPRs identified to date that has a second light-harvesting pigment, i.e. the keto-carotenoid echinenone. The keto-group of the carotenoid non-covalently interacts with the GR protein, leading to a dual-pigment rhodopsin photosystem with a broad absorption spectrum, extending from about 400 to 600 nm (Balashov et al., 2010). Multiple pigments and light-harvesting complexes are a typical feature of more complex of reaction center photosystems. Therefore, lessons learnt from transplanting dual-pigment biosynthesis for GR, as performed in this thesis, can be useful towards successfully transplanting more complex, multi-pigment photosystems later on.

Table 1. Main types of photosystems found in nature, based on a detailed review (Hohmann-Marriott and Blankenship, 2012).

Type of photosystem	*Electron source*	*Wavelength absorbance*	*Pigments*	*Organisms*
proton-pumping rhodopsins	none – only generate an electrochemical gradient	~ 400 – 600 nm in nature, but thermodynamically not necessarily limited to this range	retinal, (+ for some a keto-carotenoid antennae pigment)	many heterotrophic archaea and bacteria, some autotrophs and eukaryotes
anoxygenic reaction center systems	H_2S, S, H_2, or organic compounds	~ 400 -1100 nm	bacteriochlorophylls, carotenoids	e.g. purple sulfur and non-sulfur, phototrophic bacteria, green sulfur phototrophic bacteria
oxygenic reaction center photosystems	water	400 -~ 740 nm, with generally decreased absorbance between 500-600 nm (green)	chlorophylls phycobilliproteins, light-harvesting complexes, carotenoids	e.g. cyanobacteria, algae and plants

Early attempts for modifying and transplanting photosystems

In the past decade, often with the motivation to improve photosynthesis, a few attempts have been made to modify photosystems in phototrophs. Also attempts have been made to transplant photosystems and related pigment production to non-native, sometimes heterotrophic, hosts. These endeavors were fueled by the increased capabilities for genetic engineering and the emerging tools and ideas from the field of synthetic biology (Cameron et al., 2014). Also the increased availability of (meta)genomic information on photosystems and photosynthetic organisms has boosted these endeavors. In **Table 2,** a concise overview is provided of relevant attempts for modifying and transplanting photosystems and their pigments in a range of hosts.

The most simple and relatively successful modifications actually consist of removing parts of photosystems. It has been demonstrated that by truncating light-harvesting antennae complexes of photosystems, light can be more equally spread throughout dense cell cultures, sometimes improving overall culture biomass yields (Beckmann et al., 2009; Kirst et al., 2014). In addition, very recently, plant photosynthesis in tobacco plants has been substantially improved by modifying three genes, for an accelerated response to shading

conditions. This results in more efficient use of lower light intensities during shading and increases crop yields up to 20% (Kromdijk et al., 2016)

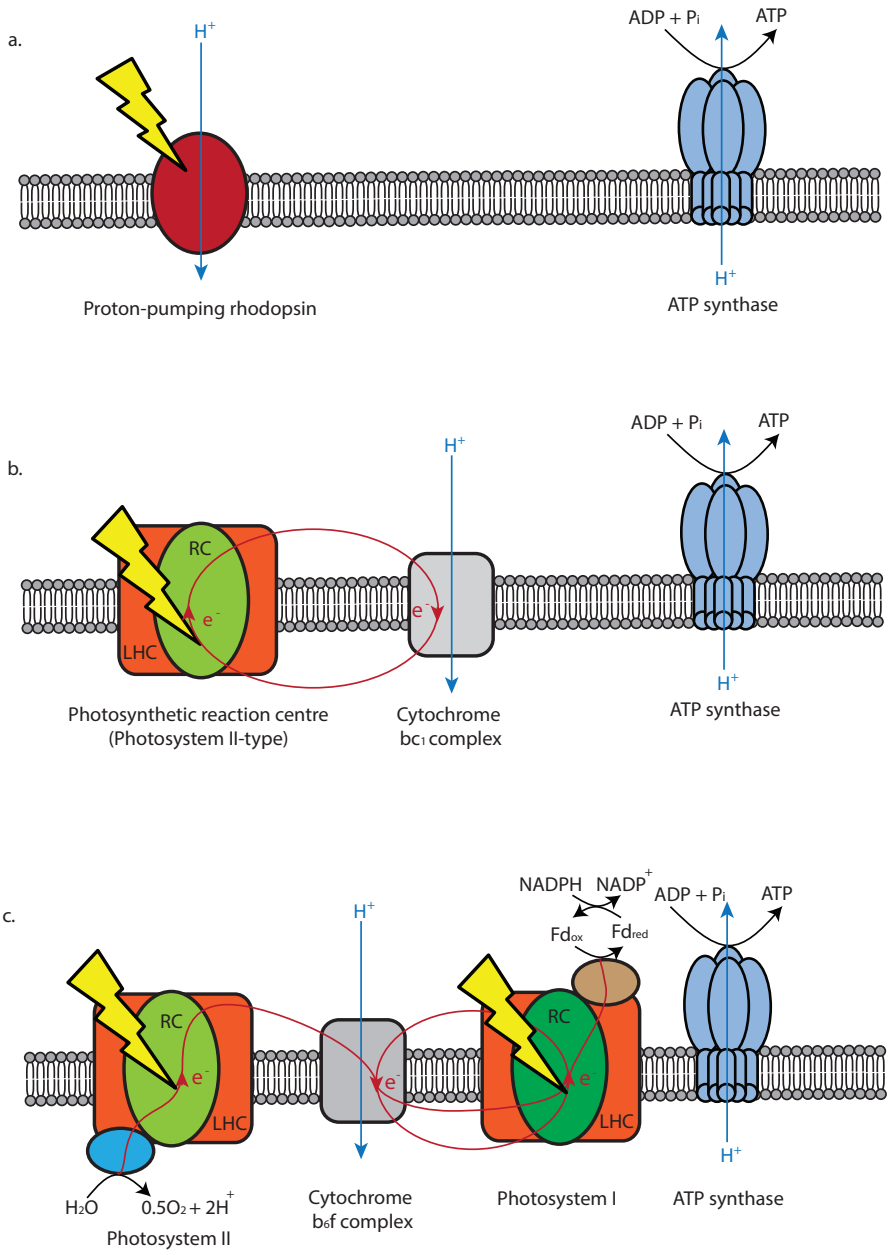


Figure 2. Schematic picture of main types of natural photosystems. a| Proton-pumping rhodopsin photosystem; b| anoxygenic photosystem, as for example found in purple anoxygenic bacteria; c| oxygenic photosystem, as for example found in cyanobacteria and algae. Red arrows indicate electron flows, blue arrows indicate proton flows. LHC: light-harvesting complex; RC: reaction centre.

Complete transplantations of complete photosystems have so far only been achieved for some bacterial PPR photosystems, and specifically proteorhodopsin. These PPRs systems, sometimes including retinal pigment biosynthesis, have been transplanted to both heterotrophic and phototrophic model hosts (**Table 2**). The physiological impact of these transplanted PPRs on the cell, including growth and production has so far been very limited, which will be further discussed in **Chapter 2**.

Transplantation of pigment biosynthesis pathways, often without the transplantation of related photosystem proteins, has been achieved for several pigments. Carotenoid biosynthesis has been extensively metabolically engineered in *E. coli* and some other hosts, also given the relevance of these compounds for human health. For transplanting the biosynthesis pathways of chlorophylls and bacteriochlorophylls some partly successful attempts have been made (Hitchcock et al., 2016; Ipekoğlu et al., 2016; Tikh et al., 2014a).

In general, stoichiometrically balanced production of multiple pigments, often from pathways involving phototoxic intermediates and products, or shared precursors molecules, will be a challenging task to realize. Another factor futher complicating transplantation is the fact that many crucial enzymes in pigment biosynthesis pathways and especially photosystem proteins are often integrated-membrane proteins (Johnson and Schmidt-Dannert, 2008). The efficient heterologous production of integrated-membrane proteins has generally been much more challenging compared to that of soluble proteins (Schlegel et al., 2014).

In this thesis, an exploration is described of the transplantation of a branched pigment biosynthesis pathway and the main focus is on producing photosystem membrane proteins. In this thesis PPRs as simple model photosystems are transplanted into the model heterotrophic host *Escherichia coli*. The proteobacterium *E. coli*, lacking any photosystems, has been applied in molecular biology for many decades as the standard bacterial host, and also recent tools developed in synthetic and systems biology are most advanced for this bacterium. Therefore we chose *E. coli* as a chassis to explore new avenues for transplanting rhodopsin photosystems, even though photosynthetic hosts may eventually be better suited systems to realize improved photosynthesis.

Table 2. Overview of examples of engineering and transplantation of photosystems and pigment biosynthesis in diverse hosts.

Host organism	*Native photosystems*	*Modifications or transplantations performed*
Nicotiana tabacum	oxygenic photosystem	- accelerated photosystem shading response (Kromdijk et al., 2016)
Chlamydomonas reinhardtii (green microalgae)	oxygenic photosystem	- truncated photosystems (Beckmann et al., 2009) - refactoring and introduction of non-native Photosystem II core (Gimpel et al., 2016)
Synechocystis sp. PCC6803 (cyanobacterium)	oxygenic photosystem	- truncated photosystem (Kirst et al., 2014) - transplantation of proteorhodopsin (Chen et al., 2016)
Rhodobacter sphaeroides (purple non-sulfur bacterium)	anoxygenic photosystem	- chlorophyll *a* biosynthesis + chlorophyll-protein complex (Hitchcock et al., 2016; Ipekoğlu et al., 2016) - transplantation of Yellow Fluorescent Protein as an artificial light-harvesting antennae (Grayson et al., 2017) - transplantation of proteorhodopsin (Tikh et al., 2014b)
Escherichia coli (heterotrophic model bacterium)	none	- proteorhodopsin + retinal biosynthesis (Kim et al., 2012b; Martinez et al., 2007; Walter et al., 2007; Wang et al., 2015) - carotenoid biosynthesis, reviewed in (Ye and Bhatia, 2012) - partial bacteriochlorophyll biosynthesis (Tikh et al., 2014a)
Shewanella oneidensis (heterotrophic bacterium with external electron exchange capabilities)	none	- proteorhodopsin (Johnson et al., 2010)
Saccharomyces cerevisiae (heterotrophic, model yeast)	none	- carotenoid biosynthesis, reviewed in (Ye and Bhatia, 2012)

Available tools for engineering phototrophy

Synthetic biology is an emerging engineering discipline in the field of biology (Cameron et al., 2014), using some other fields, including genetic engineering and systems biology, as a solid foundation. The synthetic biology approaches and tools are anticipated to become highly relevant for engineering more efficient photosystems.

Photosystems are already complex systems on their own, and especially when aiming at engineering complete autotrophic cell factories, a systems and synthetic biology approach is required. Autotrophic cell factories require a proper integration of photosystems, other energy-regenerating systems,

CO_2 fixation pathways and production pathways. To design and analyze such systems, system biology modelling-based approaches, such as Flux Balance Analysis (FBA) in metabolic models (Orth et al., 2010) and thermodynamic analyses (Flamholz et al., 2012; Noor et al., 2013, 2014) can be strong tools, as employed in **Chapter 3.**

To facilitate the experimental implementation of transplanting and improving photosystems, synthetic biology tools will be instrumental. In recent years, many tools have been developed and deployed for transplantation and refactoring of biological systems (Smanski et al., 2016). The extended synthetic biology toolbox has allowed for the successful transplantation and optimization of several systems in *E. coli*, such as nitrogen fixation (Smanski et al., 2014), light-switches (Schmidl et al., 2014), external electron uptake (Goldbeck et al., 2013) and very recently CO_2 fixation by the Calvin Cycle (Antonovsky et al., 2016). Tools that allowed these successful transplantations include developments in DNA synthesis (Kosuri and Church, 2014), codon usage algorithms (Elena et al., 2014) (**Chapter 5, 6**), DNA assembly (Casini et al., 2015), laboratory evolution (Portnoy et al., 2011) and notably novel tools for the fine-tuning of gene expression (**Table 3**). Fine-tuning is highly relevant for complex (branched) pathways and membrane proteins, which are required for the transplantation of the PPR photosystems performed in this thesis (**Chapter 4,7**).

An overview of important systems for recombinant protein production and fine-tuning thereof is provided in **Table 3**, focusing on systems for bacterial expression, especially for *E. coli*. Some of these tools rely on tunable expression by inducible promoters or constitutive promoter libraries (Blazeck and Alper, 2013). However, promoter engineering has some limitations as it cannot tackle the large influence on expression by the secondary structures of the mRNA transcript, especially around the translational start site (5'untranslated region(UTR):coding sequence junction) (Kudla et al., 2009a; Mutalik et al., 2013a). Furthermore, promoter engineering does not allow for differential tuning of multiple genes within one operon, while operons are convenient genetic units for transplanting larger (photo)systems in bacteria.

Other tools focus on tuning at the translational level, specifically by modifying Ribosome Binding Sites (RBSs) and their sequence contexts. N-terminal fusion

tags, with favorable 5'UTR:coding sequence contexts for translation initation, have been instrumental to increase the expression of several difficult-to-express proteins (Young et al., 2012). Although limitedly explored, libraries of N-terminal tags resulting in different translation inititation stenghts may allow for fine-tuning of expression (Goltermann et al., 2011). As a potential disadvantage N-terminal fusion peptides may compromise protein structures and functionality.

Alternatively, the structure of mRNAs has been modified with large libraries of hairpins at the 5'UTR and 3'UTR to vary the expression of transcripts in an operon (Pfleger et al., 2006). In addition, optimization at the translational level has been performed by approaches involving random mutations (Cheong et al., 2015; Mirzadeh et al., 2015). These random approaches result in large libraries and related screening efforts, especially when simultaneously optimizing multiple genes (Jeschek et al., 2016).

In this thesis, the focus is on rational, smaller library approaches for translational tuning, in particular using the mechanism of translational coupling. This natural genetic mechanism can occur if the RBS upstream of a coding region is in close proximity to the stop codon (either within or shortly after the coding region) of the preceding gene in the operon (Levin-Karp et al., 2013). Hereby the translation of the downstream gene is coupled to that of the upstream gene, as ribosomes can directly 'jump' to the next RBS and continue translation. As a good example of synthetic biology, this natural coupling mechanism has been redesigned, standardized and optimized for precisely varying protein production levels in *E. coli*, in so-called Bicistronic Design elements (BCDs) (Mutalik et al., 2013b). Tests of these BCD elements are described in this thesis for tuning membrane protein production for photosystems and other diverse types of membrane proteins (**Chapter 4 and Chapter 7**). In addition, rational reduced RBS libraries are employed, which are generated with the novel RedLibs algorithm (Jeschek et al., 2016). The RedLibs algorithm is based on RBS Calculator, a biophysical algorithm predicting translation initiation rates (Farasat et al., 2014; Salis, 2011; Salis et al., 2009). This RedLibs algorithm will be assessed for fine-tuning branched pigment pathway biosynthesis (**Chapter 4**).

Table 3. Overview of available systems for improving and tuning recombinant protein expression in bacteria. Most tools described are developed for *E. coli*. The given expression tools are assessed for several features relevant for engineering and tuning for photosystem transplantation and other synthetic biology projects.

Systems	***Examples***	***Fine-tuning expression***	***Decrease mRNA secondary structures around RBS***	***Differential expression in operons***	***Required efforts for screening variants***
inducible, well-titratable promoters	P_{RhaBAD} (Giacalone et al., 2006), *XylS-Pm* (Balzer et al., 2013)	+	-	-	+
constitutive synthetic promoter libraries	BIOFAB library (Mutalik et al., 2013b) and other *E. coli* promoter libraries (Davis et al., 2011; Hal Alper, Curt Fischer, Elke Nevoigt, 2005)	+	-	-	+
N-terminal fusion tags	SUMO, TrxA, GST, MBP and others, reviewed before (Young et al., 2012) or random libraries (Goltermann et al., 2011)	+/-	+	+	+
RNA hairpin libraries	TIGRs (Pfleger et al., 2006)	+	+/-	+	+/-
random synonymous mutations of first codons after start codon	synonymous mutations in first 2 codons (Mirzadeh et al., 2015), 3 codons (Nørholm et al., 2013) or 10 codons (Cheong et al., 2015)	+	+	+	-
random mutations in spacer between RBS and start codon	randomized spacers in *E. coli* (Mirzadeh et al., 2015) and *Bacillus subtilus* (Liebeton et al., 2014)	+	+	+	-
rational RBS engineering and libraries	RBS Calculator (Farasat et al., 2014; Salis, 2011; Salis et al., 2009) and similar calculators reviewed in (Reeve et al., 2014) RedLibs (Jeschek et al., 2016), EMOPEC (Bonde et al., 2016)	+	+	+	+
translational coupling elements	BIOFAB BCDs (Mutalik et al., 2013b) and some other examples (Makoff and Smallwood, 1990; Marino et al., 2015)	+	+	+/-	+

Thesis outline

The overall aim of this thesis is to explore the heterologous production of (photosystem) membrane proteins and pigment biosynthesis pathways. This is primarily approached by the transplantation of PPR photosystems into the heterotrophic host *Escherichia coli*. *In silico,* systems biology tools are used to explore the potential of PPR photosystems. Experimental synthetic biology tools are explored for improved production of PPR photosystems and their pigment biosynthesis pathways. In addition to improving PPR production, tools will be explored for improving the production of other membrane protein as well.

Firstly, in **Chapter 2,** a detailed review on PPR photosystems and their potential is provided**.** Already demonstrated examples of heterologous expression of PPRs in several hosts and their physiological impact are discussed. Based on these, some suggestions are made to improve and further exploit the these light-harvesting systems for biotechnology.

In **Chapter 3,** an *in silico* exploration is made as a basis for the engineering of autotrophy in *E. coli,* including PPR photosystems for ATP regeneration. To allow for autotrophy in *E. coli* also heterologous CO_2 fixation pathways and electron donor regeneration mechanisms are needed. To evaluate the potential and bottlenecks of complete photo-electro-autotrophic system designs, an integrated *in silico* analyses is performed. To this end, constraint-based metabolic modeling of *E. coli* is combined with novel approaches to assess thermodynamic constraints of pathways, and with tools to predict the kinetics and protein burdens of all required heterologous components.

The production of *Gloeobacter violaceus* rhodopsin (GR) and *Thermus thermophilus* rhodopsin (TR) photosystems in *E. coli* is optimized in **Chapter 4**. The production of the GR and TR protein is optimized by tuning with constitutive promoters, and especially by standardized translational coupling elements. This chapter also describes the heterologous expression of the branched pigment pathways in *E.coli* for both retinal and echinenone synthesis. Both are derived from β-carotene, and their balanced, stoichiometric production is a major challenge, for which optimization is attempted by rational, small degenerate

(RedLibs) RBS libraries. The broadened, blue-shifted absorption spectrum of GR and TR by *in vivo* synthesized echinenone is explored, as is the potential for ATP regeneration in *E. coli* by GR.

Apart from transcriptional rates and translation initiation rates, multiple other factors play a role for high, functional production of membrane proteins and other proteins. A multitude of these factors is related to the synonymous codon usage in gene sequences. In **Chapter 5** an overview is provided on how synonymous codon usage can influence both production and folding of proteins. Many explanations and mechanisms have been suggested for preferential synonymous codon usage in certain genes, organisms and growth conditions. Although the rules for predicting the optimal synonymous codon usage for heterologous protein production are far from being understood, some codon usage algorithms have been developed hereto.

In **Chapter 6,** an experimental exploration is made for two different codon usage algorithms for membrane protein production, namely codon harmonization and codon optimization. Commonly applied commercial codon optimization for heterologous protein production mainly aim to include frequent codons. Codon harmonization also includes more rare *E. coli* codons at the same locations where such rare codons are found in the wild-type gene for the native host, assuming a role of those rare codons in protein production or folding. In this chapter the performance of both algorithms is compared, also compared to the wild-type codon sequence variants, for the heterologous production of six different membrane proteins in *E. coli*. The tested proteins include several PPRs. The influence of the codon usage variants is assessed in combination with transcriptional fine-tuning in the well-established *E. coli* LEMO21(DE3) system.

The successful method for fine-tuning of membrane protein production from Chapter 4 is further explored in **Chapter 7**. Unlike previously developed membrane protein production tuning methods for *E. coli*, such as *E. coli* LEMO21(DE3), this method relies, on a non-inducible, constitutive promoter. Fine-tuning of expression for high membrane protein production is achieved by varying the translation initiation. Hereto, a standardized library of translational coupling elements, so-called Bicistronic Design elements (BCDs) are employed. This method is tested for the production of two PPRs and the

model membrane protein YidC. Furthermore, production from this method is compared to heterologous production of the same proteins in the *E. coli* LEMO21(DE3) system.

In **Chapter 8** a broad overview and perspectives are provided on the potential of microbial autotrophs for the production of value-added compounds from CO_2. Both photoautotrophic and chemolithoautotrophic production platforms are discussed, and recent progress in improving their efficiency and production potential is highlighted. Also novel *in silico* and experimental approaches, including some applied in this thesis, to engineer autotrophic systems are described. Future avenues are discussed for realizing more efficient autotrophic production platforms, including an integrated synthetic biology workflow that relies both on rational engineering and laboratory evolution approaches.

In **Chapter 9** a summary of this thesis is provided. Finally, in **Chapter 10** a general discussion of the results of this thesis is provided. Results on PPR photosystems engineering and improved membrane protein production are discussed in a wider scope, including results from recent literature. Moreover, future avenues for engineering of PPRs, other photosystems and increased membrane protein production are discussed.

Chapter 2

Potential of proton-pumping rhodopsins: engineering photosystems into microorganisms

Nico J. Claassens*[1], Michael Volpers[2,3,4], Vitor. A. P. Martins dos Santos [2, 4], John van der Oost[1*], and Willem M. de Vos[1,5]

*first author
[1] Laboratory of Microbiology, Wageningen University, Wageningen, The Netherlands
[2] Laboratory of Systems & Synthetic Biology, Wageningen University, Wageningen, The Netherlands
[3] Center for Biological Systems Analysis, University of Freiburg, Freiburg, Germany
[4] LifeGlimmer GmbH, Berlin, Germany
[5] Department of Bacteriology and Immunology, Helsinki University, Helsinki, Finland

Chapter adapted from publication:
Trends in biotechnology (2013) 31:633-642 DOI:10.1016/j.tibtech.2013.08.006

Abstract

A wide range of proton-pumping rhodopsins (PPRs) have been discovered in recent years. Using a synthetic biology approach, PPR photosystems with different features can be easily introduced in non-photosynthetic hosts. PPRs can provide hosts with the ability to harvest light and drive the sustainable production of biochemicals or biofuels. PPRs use light energy to generate an outward proton flux, and the resulting proton motive force can subsequently power cellular processes. Recently, the introduction of PPRs in microbial production hosts has successfully led to light-driven biotechnological conversions. In this review, we discuss relevant features of natural PPRs, evaluate reported biotechnological applications of microbial production hosts equipped with PPRs, and provide an outlook on future developments.

Keywords
Proton-pumping rhodopsins; synthetic phototrophy; synthetic biology; microbial biofuel and biochemical production.

Proton-pumping rhodopsins as simple photosystems for synthetic biology

Photosynthetic microorganisms are promising hosts for the light-driven, sustainable production of relevant compounds, such as biochemicals and biofuels (Wang et al., 2012). To synthesize these products from carbon dioxide, photosynthetic microorganisms harvest light through their photosystems (**Box 1**). Photosynthetic microorganisms, such as cyanobacteria and algae, naturally harbor photosystems. However, most of these photosynthetic microorganisms cannot yet be genetically engineered efficiently, and they are not well developed as production hosts (Wang et al., 2012). As an alternative, it has recently been suggested that non-photosynthetic production hosts with well-established genetic systems can be converted to phototrophs by introducing photosystems (Johnson and Schmidt-Dannert, 2008; Maurino and Weber, 2013). This could lead to a diverse set of biotechnological production hosts that can harvest light, and use its energy to drive the production of chemicals of interest. Moreover, photosystems may also be engineered into natural photosynthetic microorganisms as additional photosystems to increase their photosynthetic efficiency (Walter et al., 2010).

Of the natural photosystems, proton-pumping rhodopsins (PPRs) are the most amenable to introduction into non-photosynthetic production hosts, as only a small number of genes has to be introduced (**Box 1**). Apart from the gene that encodes PPR protein, approximately 5 genes are needed for the retinal biosynthesis (Martinez et al., 2007). Disadvantages of PPRs are that they cannot generate reducing equivalents, so they cannot support complete autotrophy. Furthermore they are less efficient at generating a proton motive force (PMF) from light than chlorophyll-pigmented photosystems. However, even the most simple chlorophyll-pigmented photosystem needs at least 30 unique genes to be expressed (Bryant and Frigaard, 2006). Hence, we consider the simpler PPRs as the most attractive photosystems to introduce in a host at the current state of synthetic biology.

Some first successful attempts to engineer PPRs into microorganisms for biotechnological applications have been reported, such as increased hydrogen production from light in *E. coli* by co-expression of a hydrogenase and a PPR

(Kim et al., 2012b). Another example is increased production of electrical current from light in *Shewanella oneidensis* engineered with a PPR (Johnson et al., 2010).

Another successful application of PPRs is as optogenetic tools in neuroscience. When PPRs are expressed in neurons, their activity can silenced by light (Chow et al., 2010; Han et al., 2011). Some insights gained by applying PPRs in neuroscience can also be beneficial for the application of PPRs in microorganisms and vice versa. However, as some excellent reviews have recently covered this topic (Fenno et al., 2011; Mattis et al., 2012; Zhang et al., 2011), we here restrict ourselves to microbial applications.

Many different types of PPR sequences have been discovered over the last few years, mainly in metagenomic libraries and newly sequenced genomes (**Box 2**). This natural diversity in PPRs provides an attractive basis for different applications. In this review we will first discuss features of discovered PPRs, especially their relevance for engineering applications. Specific attention will be given to the quantitative features of PPRs, as these are important for biotechnological applications. Furthermore, we will address applications of PPRs in non-photosynthetic and photosynthetic microbial hosts.

Box 1. Proton-pumping rhodopsins compared to other photosystems

The center of light conversion in each biological photosystem is the photoconverter, a pigmented protein complex that converts the light energy to chemical energy. Two main types of photoconverters can be found in nature. The first main type of photoconverters is the chlorophyll-pigmented reaction centers, which use the light energy to transfer electrons to a higher energy level. The second main type of photoconverters are the retinal-pigmented proton-pumping rhodopsins (PPRs), which pump protons against a proton gradient (Hohmann-Marriott and Blankenship, 2012).

The chlorophyll-pigmented reaction centers are found in oxygenic photosystems of photosynthetic organisms, such as cyanobacteria, algae and plants. Furthermore the chlorophyll-pigmented reaction centers are found in anoxygenic photosystems in some photosynthetic and phototrophic bacteria, such as green sulfur bacteria or purple bacteria. In addition to the reaction center, these photosystems consist of antenna complexes and electron transport proteins (Hohmann-Marriott and Blankenship, 2012). Most chlorophyll-pigmented photosystems can both pump protons and shuttle electrons to for example NADPH to drive metabolic conversions. The proton-pumping activity can generate a proton motive force (PMF) which for example can drive ATP synthesis by the membrane-bound ATPases. By producing both ATP and reducing equivalents, the chlorophyll-pigmented photosystems can facilitate CO2 fixation pathways, so they can support photosynthesis.

The second type of photoconverters, the retinal-pigmented PPRs are found in PPR photosystems. These photosystems mostly only consist of the retinal-pigmented PPR photoconverter itself. As an exception, some xanthorhodopsin PPRs have an additional antenna pigment (**Box 2**). PPR photosystems are mostly found in heterotrophic archaea, bacteria and eukarya (Hohmann-Marriott and Blankenship, 2012). However, PPRs have also been found as additional photosystems in an alga and a cyanobacterium (Imasheva et al., 2009; Tsunoda et al., 2006). The protein of PPR photosystems is a 7-helix transmembrane polypeptide (~ 250-300 amino acids) that is associated with the retinal pigment (**Figure 1**). This pigment is isomerized from the all-trans state to the 13-cis state after absorption of a photon. This photo-isomerization of the retinal results in conformational changes of the PPR protein, which facilitate the transport of one proton across the cytoplasmic membrane (Brown and Jung, 2006). Chlorophyll-pigmented photosystems can pump more than one proton per absorbed photon, so PPR photosystems are less efficient in generating ATP from photons (Bryant and Frigaard, 2006). Furthermore, PPRs photosystems cannot generate reducing equivalents, which is needed for CO2 fixation.

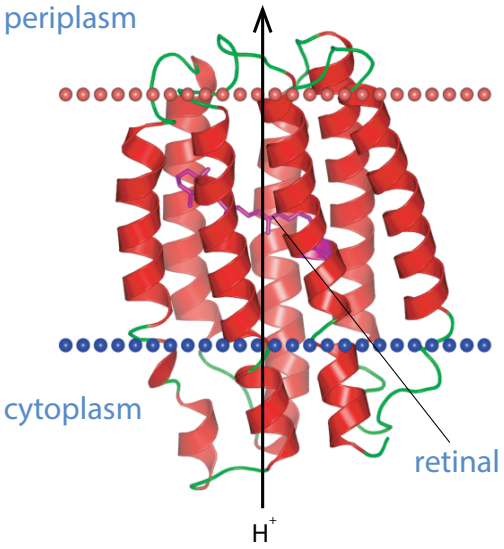


Figure 1. Protein structure of a PPR (proteorhodopsin), illustrated with an outward proton flux, adapted with permission from http://opm.phar.umich.edu/protein.php?pdbid=2l6x.

Features of proton-pumping rhodopsins relevant for engineering

Using currently available tools, in principle any natural PPR gene sequence can be cloned either in an expression vector or in the host chromosome. When desired, PPR genes can be synthesized and codon optimized for a host to increase expression efficiency. In addition, the production levels of PPR protein and pigment biosynthesis enzymes can be fine-tuned by using constitutive or regulated promoters. Before introducing the PPR in a host, the first step is to select an appropriate PPR from nature based on several features. Here the following features will be discussed: (i) functional production of PPR protein, (ii) functional production of pigment biosynthesis pathways, (iii) abundance in the host membrane, (iv) proton-pumping rate, and (v) light-harvesting spectrum.

Functional expression of PPR proteins

The first PPR that was heterologously expressed in *E. coli*, correctly folded for retinal pigment binding, and showed proton pumping was a bacterial proteorhodopsin from a marine uncultured gamma-proteobacterium (Béjà et al., 2000). Several proteorhodopsins have been engineered in several bacterial hosts: *E. coli* (Kim et al., 2012b; Martinez et al., 2007; Walter et al., 2007), *S.*

oneidensis (Johnson et al., 2010), and *Pseudomonas putida* (J.D. Buck, PhD thesis, MIT, 2012). All three hosts belong to the class of gamma-proteobacteria; a class from which many proteorhodopsin genes originate (Atamna-Ismaeel et al., 2008). Engineering hosts with PPRs from similar origin most likely increases the chance of functional expression. However, several archaeal PPRs have also been functionally expressed in *E. coli*, a phylogenetically distant bacterial host (Feng et al., 2006; Fu et al., 2010; Kamo et al., 2006; Zhou et al., 2009a). On the contrary, the first discovered archaeal PPR, bacteriorhodopsin from *H. salinarum*, thus far has not been functionally expressed in *E. coli* (Karnik et al., 1990). In summary, functional, heterologous expression of PPRs is unpredictable and should be tested by trial and error.

A specific group of PPRs that is potentially relevant for biotechnological applications are the xanthorhodopsins, as some PPR proteins in this group bind an antenna pigment in addition to retinal (**Box 2**). The antenna-binding xanthorhodopsin from *G. violaceus* has been expressed functionally in *E. coli* (Balashov et al., 2010; Imasheva et al., 2009; Miranda et al., 2009). Recently, xanthorhodopsins were found in thermophiles, the thermophilic xanthorhodopsins may be interesting for introduction in thermophilic production hosts. The *T. thermophilus* PPR has already been functionally expressed in the mesophile *E. coli* (Tsukamoto et al., 2013a).

In addition to prokaryotic hosts, microbial, eukaryotic hosts such as yeast and filamentous fungi are useful production hosts for a large range of biotechnological applications. Some archaeal and eukaryotic PPRs have been functionally expressed in yeast membranes, for example in *Schizosaccharomyces pombe* and *P. pastoris* (Hildebrandt et al., 1993; Janke et al., 2013). In *S. pombe*, the archaeal PPR bacteriorhodopsin was expressed in the cytoplasmic membrane, and by changing the leader signal sequence of the gene, it was functionally expressed as well into the mitochondrial membrane (Hoffmann et al., 1994). Changing the signal sequence may also be applied to direct PPRs to another membrane, such as the thylakoid membrane of cyanobacteria.

In conclusion, many PPRs have been found in nature in diverse hosts that live in a wide range of conditions. Hence, suitable candidate PPR sequences are available for engineering into many hosts. However, functional expression of PPRs has so far only been reported in a few hosts.

Box 2. Natural diversity in proton-pumping rhodopsins

PPRs appear to be very 'cosmopolitan' proteins, and it has been suggested that this simple photosystem is transferred by horizontal gene transfer among all domains of life: archaea, bacteria, eukarya (**Figure 2**) (Frigaard et al., 2006; Sharma et al., 2006). Most PPRs were found in aquatic environments. However, terrestrial bacterial communities isolated from leaf surfaces were found to contain PPR sequences and have PPR absorption spectra similar to PPRs from aquatic environments (Atamna-Ismaeel et al., 2012).

In the early 1970s the first PPR was discovered in the halophilic archaeum *Halobacterium salinarum* (Oesterhelt and Stoeckenius, 1971). This PPR was called bacteriorhodopsin, as its discovery was before the discovery of the archaea as a distinct prokaryotic domain apart from bacteria. Three decades later, the first bacterial PPR was found in the uncultivated marine gamma-proteobacterium EBAC31A08 in a Sargasso sea metagenome and called proteorhodopsin (Béjà et al., 2000). Many more bacterial PPRs have since been discovered, mainly in marine metagenomic samples where PPR sequences are abundant (Fuhrman et al., 2008). The first eukaryotic PPR was found in the genome of the fungus *Leptosphaeria maculans* (Waschuk et al., 2005), followed by the discovery of PPRs in eukaryotic algae and dinoflaggelates (Slamovits et al., 2011; Tsunoda et al., 2006).

Another remarkable type of PPR was found in the bacteria *Gloeobacter violaceus* and *Salinibacter ruber*, their PPR can bind an antenna pigment in addition to the retinal pigment (Balashov et al., 2005; Imasheva et al., 2009). Their antenna pigments are keto-carotenoids and broaden the absorption spectrum. These antenna-binding PPRs were recently grouped as subgroup 1 in a larger group of xanthorhodopsins, based on sequence similarity (**Figure 2**) (Vollmers et al., 2013). Subgroup 1 also contains xanthorhodopsin sequence from thermophiles, such as *Thermus thermophilus*. The PPR from *T. thermophilus* was characterized very recently, and was found to pump protons fast, especially at high temperatures around 75°C. However, antenna-binding has not been verified yet for the *T. thermophilus* xanthorhodopsin (Tsukamoto et al., 2013a) . The PPRs in subgroup 2 of xanthorhodopsins, found for example in psychrophilic bacteria and dinoflaggelates, probably do not bind antenna pigments (Vollmers et al., 2013).

Many of the microbial rhodopsins found in nature do not pump protons (Klare et al., 2008). However, all microbial rhodopsins bind a retinal molecule that can absorb light. Other light-stimulated functions of microbial rhodopsins are light sensing, chloride-pumping, sodium pumping, or passive cation transport (Inoue et al., 2013; Klare et al., 2008). PPRs can sometimes be identified among other microbial rhodopsin sequences, as all the PPRs share four crucial residues involved in proton pumping. However, having these residues does not guarantee proton pumping (Brown and Jung, 2006).

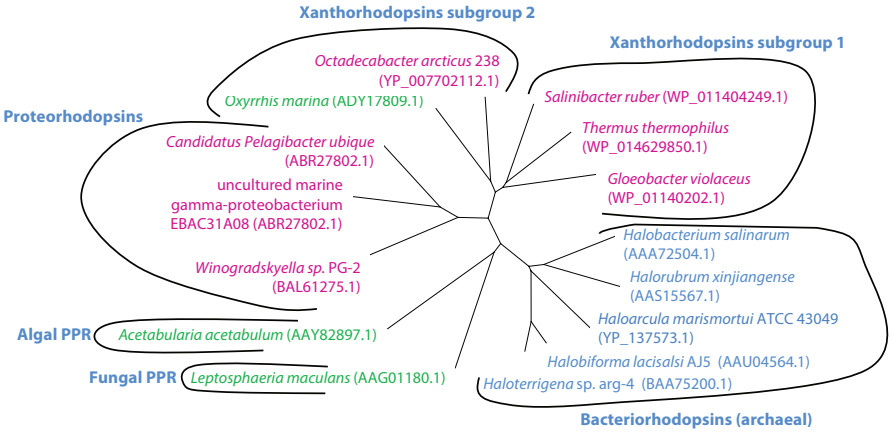


Figure 2. Unrooted phylogenetic tree of amino acid sequences of PPRs that are discussed or referred to in this review. A multiple sequence alignment is made by ClustalX 2.1. (Thompson et al., 1997), and the tree is drawn by TreeView. Colors of the species names indicate the domains of life: archaea, blue; bacteria, magenta; eukarya, green. For each PPR the NCBI accession number is given between parentheses.

Functional expression of pigment biosynthesis pathways

For potential industrial applications of PPRs it is desirable to produce, in addition to PPR protein, the pathway for biosynthesis of the retinal and eventually antenna pigments. Otherwise, the need to add pigments to the medium makes the industrial process more expensive.

Retinal can be biosynthesized from the ubiquitous isoprenoid intermediate isopentenyl diphosphate (IPP), which is for example an intermediate in the quinone synthesis (Lange et al., 2000). The biosynthesis of retinal from IPP

requires seven additional enzymes. However, in many bacteria, such as *E. coli,* only five additional genes need to be expressed for retinal biosynthesis, as these bacteria already synthesize the more downstream intermediate farnesyl diphosphate (FPP) (Martinez et al., 2007). Of the additional genes, four are for the conversion of FPP to β-carotene; these genes can be found in *crtEBIY* gene clusters. The remaining gene encodes a β-carotene oxygenase for the cleavage of β-carotene into two retinal molecules. Because of this oxygenase, retinal biosynthesis requires molecular oxygen, so aerobic, or at least, microaerobic conditions (Peck et al., 2001).

Complete synthesis of retinal in *E. coli* was demonstrated by expressing *crtEBIY* genes and genes for bacterial (*blh)* or eukaryotic β-carotene oxygenases (Kim et al., 2012b; Lintig et al., 2000; Martinez et al., 2007). Retinal pathway sequences are available from the PPR gene clusters, which are found in various species (McCarren and DeLong, 2007; Sabehi et al., 2005). Hence, suitable retinal pathways are probably available for introduction in all types of production hosts, though it might be that some pathway intermediates are toxic in specific hosts.

In addition to retinal synthesis, antenna pigments of xanthorhodopsins can potentially be biosynthesized in the production host (see **Chapter 4**).

Abundance in the host membrane

The number of functionally expressed PPRs in the host membrane of each cell influences the proton flux that can be generated. The number of PPRs per cell has been spectroscopically quantified in a few studies only, mostly for PPRs in their natural hosts (**Table 1a**). In the standard engineering host *E. coli*, several PPRs have been heterologously expressed, but no quantification of PPRs per cell was reported so far, in **Chapter 4** a quantification of PPR production in *E. coli* is provided.

Expressing elevated numbers of PPRs per cell also means higher metabolic costs of synthesis. These metabolic costs of PPR synthesis together with the metabolic benefit of proton pumping by PPRs were analyzed for natural hosts; this analysis demonstrated that a high number of PPRs per cell is required to have a physiologically relevant net energy benefit from PPRs (Kirchman and

Hanson, 2013). Likewise, to provide the engineered host with high metabolic benefits, it is probably important to increase PPRs per cell by overexpression. Increasing the PPRs per cell is probably beneficial as long as all PPRs can absorb sufficient light and all PPRs are functionally incorporated into the membrane. PPRs that are not incorporated in the membrane cannot pump protons and are likely to be rapidly degraded, as was observed for overexpressed bacteriorhodopsin in the cytoplasm of *E. coli* (Karnik et al., 1987).

The number of functional PPRs is thus limited by the available membrane. An option to increase the number of PPRs could be to increase the membrane surface of the engineering hosts by inducing membrane invaginations (Johnson and Schmidt-Dannert, 2008).

In addition to available membrane, also the density of PPRs in the membrane of an engineered host should be considered. A high density of PPRs (> 100,000 per cell) seems possible in the specialized purple membrane of *Halobacterium salinarum* harboring bacteriorhodopsin (Wagner and Hope, 1976). In these purple membranes, bacteriorhodopsins occur as trimers. For an archaeal PPR that was heterologously expressed in *E. coli* a monomer configuration was reported (Feng et al., 2006). Hypothetically, this is related to the lack of specific archaeal cardiolipins in *E. coli*, needed for stabilization of archaeal PPRs trimers in the membrane (Corcelli et al., 2002). For proteorhodopsin a natural host, hexamers or pentamers configurations were found and for xanthorhodopsin from *G. violaceus* trimers were observed (Klyszejko et al., 2008; Tsukamoto et al., 2013b). For this xanthorhodopsin, the trimer configuration also seems present when heterologously expressed *E. coli* (Tsukamoto et al., 2013b). Summarizing, there is potential to engineer higher PPR densities in the host membrane by aiming for oligomerization, but this only relevant when the PPRs are not yet in a dense, oligomer configuration.

A third option to increase PPR numbers could be to lower the expression of certain other membrane proteins, creating more available membrane surface for PPRs. Some membrane proteins, such as specific transporters, could be redundant in specific biotechnological applications.

Proton-pumping rate

Besides the number of PPRs per cell, the proton-pumping rate per PPR determines the generated proton flux. In a few studies, proton-pumping rates per PPR are reported (**Table 1b**). Although the published values of the different PPR types vary substantially, it is unfortunately hard to these compare proton-pumping rates because different methods and conditions have been used.

Basically, the maximal proton-pumping rate is determined by two biophysical characteristics: the length of the photocycle of a PPR and the quantum yield. Once the photocycle is complete after the absorption of a photon, the PPR is ready to absorb and pump another proton. However, not a lot and consistent photocycle lengths are available for most PPRs. The quantum efficiency is the part of all the absorbed photons that results in a proton being exported. Determining quantum efficiency is rather complicated and different values have been reported for the well-studied bacteriorhodopsin from *H. salinarum*, but its quantum efficiency is generally assumed to be approximately 60% (Birge, 1990; Kirchman and Hanson, 2013).

In addition, to biophysical factors that determine the maximal proton-pumping rate, environmental factors as light intensity and extracellular pH determine the actual proton-pumping rate. The light absorption is dependent on the cross-section of the PPR and the spectrum; absorption spectra are discussed in the next section.

The dependence of proton-pumping rate on light intensity has been demonstrated for proteorhodopsin (Friedrich et al., 2002; Walter et al., 2007). This dependence follows roughly Michaelis-Menten kinetics, with light intensity as 'substrate concentration'. Light saturation occurs only at very high light intensities, as the K_m value is about three times solar light intensity (Walter et al., 2007). An artificially high light intensity is needed to approach the maximal proton-pumping rate of these PPRs.

Furthermore, extracellular pH influences proton pumping by changing the protonation state of the primary proton-accepting aspartic acid residue (Asp). This Asp should be deprotonated to allow for proton pumping. In some PPR types the proton-accepting Asp has a high pK_a value, such as

in proteorhodopsin (pKa 7.5) (Brown and Jung, 2006). Proton pumping is inhibited around and below this pK_a value. Other PPRs have lower pK_a values for this Asp, such as xanthorhodopsin from *Gloeobacter violaceus* (pK_a 4.8) or archaeal PPR from *Haloterrigena turkmenica* (pKa 2.2) (Tsukamoto et al., 2013b). The pK_a of the proton acceptor residue seems to be related to the interaction of this Asp with other nearby residues (Brown, 2013; Brown and Jung, 2006). Thus, for engineering hosts that grow at neutral or lower pH values, it is preferable to introduce PPRs with a lower pK_a for the primary proton acceptor residue.

Light-harvesting spectrum

PPRs can absorb photons from a limited window of the light spectrum. PPRs with only a retinal pigment have a single absorption peak; these peaks are in the blue-green part of the light spectrum (~ 490 – 570 nm) (Fuhrman et al., 2008). Multiple absorption peaks and shoulders are found in the spectrum of antenna-binding xanthorhodopsins, which makes their total absorption spectrum broader (**Figure 3**) (Balashov et al., 2005, 2010). Having PPRs with a broad absorption spectrum is useful for biotechnological applications, because these PPRs can more efficiently absorb photons from the solar spectrum. Though the antenna-binding PPRs have a broader absorption than PPRs with only a retinal pigment, all known PPRs absorb light between 400 and 600 nm only. The solar spectrum on the earth surface starts around 300 nm and extents into the far infrared wavelengths beyond 1200 nm (Blankenship et al., 2011). Especially, absorbing the longer wavelengths (> 700 nm) is an interesting strategy to increase photosynthetic efficiencies (Blankenship et al., 2011). The aquatic environment of most PPRs found probably explains why most PPRs absorb only green and blue light, as only these wavelengths penetrate relatively deep into the water (Béja et al., 2001).

Photons of longer wavelengths still have an energy content that is much higher than the thermodynamic limit to pump protons. For example, photons of 1100 nm have an energy content of 1.270 eV; the thermodynamic limit to pump protons is 0.200 eV or lower, because the electrochemical gradient in most bacteria is lower than 0.200 V (Kashket, 1985). Hence, proton pumping by absorbing photons from longer wavelengths seems thermodynamically feasible for PPRs. However, absorption of these wavelengths by PPRs has not yet been found in nature. If PPRs using these wavelengths exist in nature,

they will most likely be found in terrestrial environments or potentially near hydrothermal vents in the deep sea where infrared radiation is emitted.

Besides searching in nature for PPRs absorbing longer wavelengths, PPRs might be engineered to absorb these longer wavelengths. Absorption of PPRs was slightly shifted to longer wavelengths by mutating residues close to the retinal binding pocket, or at other locations in the PPR protein (Kim et al., 2008; Yoshitsugu et al., 2009). However, mutated proteorhodopsins absorbing at slightly longer wavelengths appeared to pump protons slower than the wild-type proteorhodopsin (Kim et al., 2008). Another approach to change the absorption spectrum is the use of retinal analogs as alternative pigments. The analog azulenic retinal gave absorption up to 830 nm in bacteriorhodopsin, however, this retinal analog most likely does not support proton pumping (Liu and Asato, 2003).

The light spectrum of known PPRs seems somewhat limiting for efficient use of the solar light spectrum. As an alternative, artificial light sources with a narrower emission spectrum overlapping PPR absorption could be considered, as was for example demonstrated in powering PPR engineered *S. oneidensis* by green light-emitting diodes (LEDs) (Johnson et al., 2010).

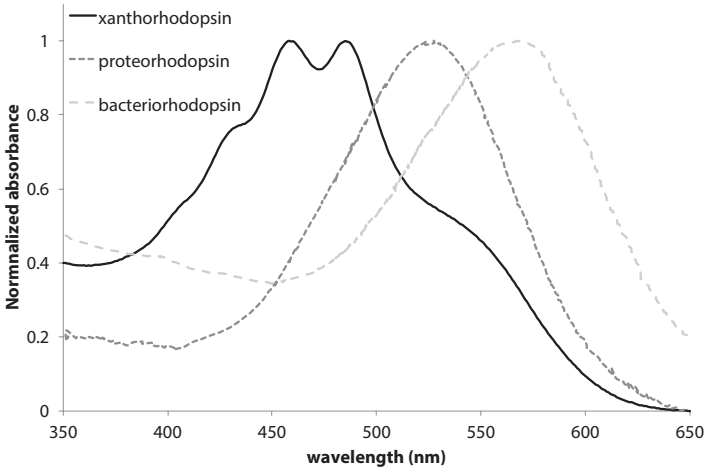


Figure 3. Absorption spectra of different types of PPRs. Bacteriorhodopsin from *H. salinarum* in purple membrane (blue), proteorhodopsin from gamma-proteobacterium EBAC31A08 in reconstituted *E. coli* membranes (red) and xanthorhodopsin from *G. violaceus* in reconstituted *E. coli* membranes (green), redrawn from (Balashov et al., 2010; Béja et al., 2001; Oesterhelt and Stoeckenius, 1971) resp. The highest absorption peaks were normalized to 1, hence only the broadness and peak wavelengths of the spectrum can be compared.

Potential of proton-pumping rhodopsins in powering cell physiology

PPRs can generate a PMF by transducing light energy. Hence, microorganisms engineered with PPRs can potentially use light to drive PMF-dependent cellular processes (**Figure 4**). We note that introduction of extra PMF generation could potentially disrupt the cellular bioenergetics and might require reregulation of the complex bioenergetics in some cases. However, some examples from literature have already shown that PMF-driven processes can be boosted relatively easily after introducing a PPR, and that growth rates or product yields can be increased. Several PMF-driven processes including demonstrated and potential applications will be discussed here, starting with the major PMF-driven process: ATP synthesis.

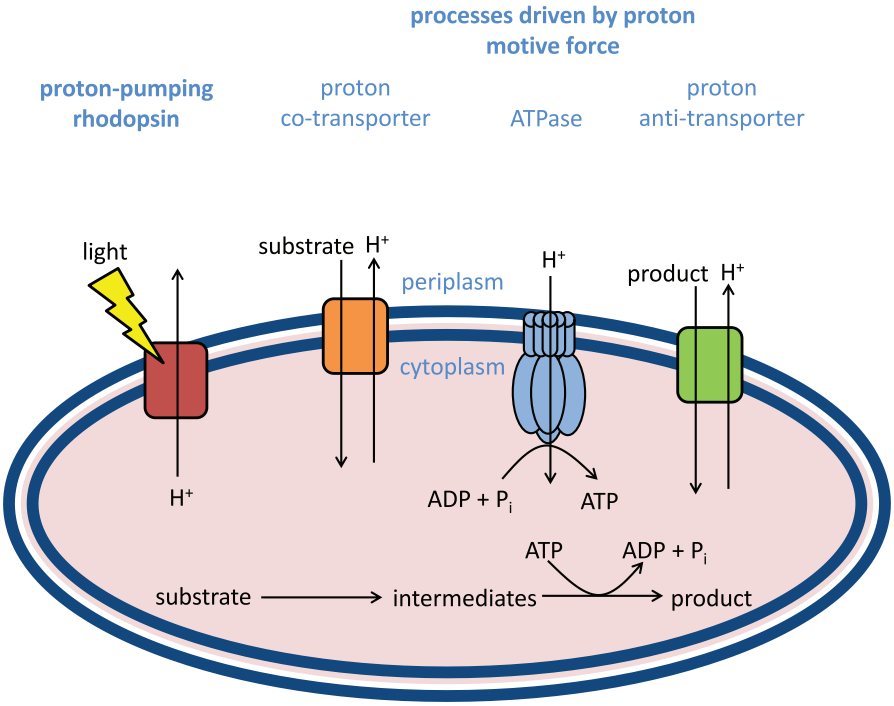


Figure 4. Schematic PPR-harboring cell, with different PMF-driven processes. The cell is depicted with a bilayer inner membrane harboring a proton-pumping rhodopsin, relevant processes driven by a proton motive force, and a conceptual production pathway.

Driving ATPase: powering ATP consuming processes

The PMF for ATP synthesis in membrane-bound ATPases is typically generated by respiratory pathways using oxygen or other electron acceptors, a process generally referred to as oxidative phosphorylation. Alternatively, when PPRs generate a PMF to support ATP synthesis this is called photophosphorylation. One of the natural roles of PPRs is hypothesized to be phosphorylation in heterotrophic microorganisms experiencing substrate or oxygen limitations (Bickel-Sandkötter et al., 1996; Fuhrman et al., 2008). In some natural PPR hosts, it was observed that light inhibits respiration, so these species can partly switch from oxidative phosphorylation to photophosphorylation (Boichenko et al., 2006).

In *E. coli* in which a proteorhodopsin was introduced, increased ATP synthesis was demonstrated in light and in the absence of oxidative phosphorylation using succinate (Martinez et al., 2007). The increased cellular ATP level from photophosphorylation was measured, but this ATP increase was about four times lower than that from oxidative phosphorylation (Martinez et al., 2007). For further quantitative comparison of PPR proton fluxes with respiratory proton fluxes, we simulated standard aerobic growth on glucose (10 mM/gDW/h) of *E. coli* in a genome-scale metabolic reconstruction by performing a flux balance analysis (Orth et al., 2010, 2011). In this simulation, the predicted respiratory proton flux through the ATPase is approximately 22 times higher than our estimate for a feasible PPR proton flux (**Table 1c, Table 2**). When this estimated PPR proton flux is included in the simulation of aerobic growth of *E. coli*, a growth rate increase of only 2% is predicted (**Table 2**). Correspondingly, *E. coli* engineered with PPRs did not exhibit increased growth, while *P. putida* engineered with a PPR did increase aerobic growth rates in light, but only at low carbon substrate concentrations (J.D. Buck, PhD thesis, MIT, 2012).

Summarizing, photophosphorylation by PPRs can be a relevant process, but at least in the model engineering host *E. coli*, to a lesser extent than oxidative phosphorylation. During conditions without oxidative phosphorylation, for example anaerobic production processes, the relevance of photophosphorylation for the engineered cell may increase.

Table 1. Overview of quantitative PPR features: a| reported numbers of PPRs per cell, and b| reported proton-pumping rates per PPR, and c| an estimate for a feasible PPR proton flux per cell based on reported values.

a

PPRs/cell	*PPR type*	*Organism*	*Method*	*Reference*
10^4	proteorhodopsin	*P. ubique*	Laser flash induced spectroscopy	(Giovannoni et al., 2005)
$2.4*10^4$	proteorhodopsin	Uncultivated gamma-proteobacterium EBAC31A08	Laser flash induced spectroscopy	(Béja et al., 2001)
$4*10^4$	proteorhodopsin	*S. oneidensis* MR-1 (engineered)	Spectroscopy	(Johnson et al., 2010)
$5.22*10^4$ +/- $3.07*10^4$	proteorhodopsin	*Winogradskyella* sp. PG-2	Spectroscopy	(Yoshizawa et al., 2012)

b

Proton- pumping rate (H^+/PPR/min)	*PPR type*	*Organism*	*Method*	*Reference*
204	bacteriorhodopsin	*H. salinarum*	Liposomes	(Mogi et al., 1988)
124 +/- 73	proteorhodopsin	*Winogradskyella* sp. PG-2	Cells	(Yoshizawa et al., 2012)
300[1]	xanthorhodopsin (gloeobacter rhodopsin)	*E. coli* (engineered)	Cells	(Ganapathy et al., 2015)

c

Assumed PPRs per cell[2]	*Assumed proton-pumping rate per PPR (H^+/PPR/min)*[3]	*Estimated proton-pumping rate per cell*	
		(H^+/cell/min)	*(mmol H^+/gDW/h)*[4]
10^5	300	$6.0*10^7$	10

[1] For this measurement, gloeobacter rhodopsin was reconstituted with retinal, but without its antenna pigment, echinenone.

[2] A ~ 100% higher number of PPRs per cell than the highest value reported in literature for natural hosts is assumed to be feasible by PPR overexpression in e.g. *E. coli*.

[3] A 50% higher proton flux per PPR than reported in literature is assumed to be feasible at high light intensities.

[4] A cell dry weight of $3*10^{-13}$ gDW/cell is assumed for *E. coli* (http://bionumbers.hms.harvard.edu//bionumber.aspx?id=103904&ver=16)

Increased ATP synthesis from photophosphorylation could increase yields of ATP-limited production pathways. However, many industrially relevant production pathways are fermentations, which have no external electron acceptor. These fermentation pathways are actually favored by low ATP production in the cell, so extra ATP synthesis from photophosphorylation is likely decreasing fermentation yields (Weusthuis et al., 2011). On the contrary, extra ATP synthesis could be relevant for biotechnological processes producing

more reduced products. Pathways to reduced products are consuming ATP, and are mostly performed aerobically to have sufficient oxidative phosphorylation (Weusthuis et al., 2011). These pathways could be performed anaerobically if production organisms are engineered with PPRs that provide sufficient photophosphorylation. Anaerobic production can be beneficial because it eliminates both substrate loss to respiration and the need for aeration of the bioreactors (Weusthuis et al., 2011).

Table 2. Simulated growth rates and proton fluxes through ATPases for *E. coli* MG1655 and *Synechocystis* sp. PCC6803 both with and without PPRs, simulated by flux balance analysis of genome-scale metabolic reconstructions.

Model simulation	*Growth rate (h^{-1})*	*Proton flux ATPase (mmol/gDW/h)*	*Simulation conditions*	*Model reference*
E. coli wild-type	0.98	223	-Standard aerobic growth -Glucose substrate (10 mmol/gDW/h)	(Orth et al., 2011)
E. coli with PPRs[1]	1.0	228		
Synechocystis wild-type	0.052	50.2	- Autotrophic growth - Light limited (30 mmol photons/gDW/h) - Standard carbon dioxide uptake (3.7 mmol/gDW/h)	(Nogales et al., 2013)
Synechocystis with PPRs[2]	0.057	55.2		

[1] An estimated PPR proton flux of 10 mmol protons/gDW/h (Table 1c) is added to the genome-scale metabolic reconstruction from cytoplasm to periplasm.
[2] An estimated PPR proton flux of 10 mmol protons/gDW/h (Table 1c) is added the genome-scale metabolic reconstruction from cytoplasm to thylakoid.

PMF-driven transport of substrates and products

The PMF is also used to power solute transport of many nutrients and substrates across the membrane. However, many of these solutes are micronutrients for which the fluxes are generally low, and it is probably negligibly beneficial to boost those by PPRs. However, larger solute fluxes that could be boosted are fluxes that include carbon substrate import and product export.

Some carbon substrates are actively taken up into the cell by proton symport, which can be powered by PPRs. This has indeed been demonstrated for lactate uptake in *S. oneidensis*; introduction of a PPR in this species resulted in up to 100% more electrical current generation from lactate in light conditions (Johnson et al., 2010).

Boosting solute export across the membrane by introducing a PPR has also been demonstrated. After introducing a PPR and a PMF-driven ethidium transporter in *E. coli*, export of the cation ethidium was increased while growing in light (Kamo et al., 2006). In conclusion, PPRs could potentially be applied to power PMF-driven transport of substrates or products during the production of biochemicals or biofuels.

Other PMF-limited processes

In addition to ATP synthesis and transport, the PMF generated from a PPR photosystem may also boost other biotechnologically relevant pathways. As an example, when introducing a PPR system together with an oxygen-tolerant hydrogenase in *E. coli*, the hydrogen production in light could be increased by 30% (Kim et al., 2012b). It was hypothesized that the light-boosted PMF makes more protons available for the hydrogenase.

Concluding, in principle all production pathways that are somehow PMF-limited can be potentially improved by introducing a PPR, but few examples have been demonstrated so far.

Potential of proton-pumping rhodopsins in powering photosynthetic microorganisms

Natural photosynthetic bacteria harvest light with photosystems and a range of pigments. These photosystems cannot absorb the complete spectrum of available sunlight. The absorption spectrum of these photosystems can be potentially complemented by introducing additional photosystems in photosynthetic bacteria, thereby increasing their photosynthetic efficiency (Blankenship et al., 2011; Walter et al., 2010). Green photosynthetic microorganisms have a lower absorption in the green spectrum, a so-called 'green gap'. PPRs can potentially complement absorption in these organisms as most PPRs absorb in the green spectrum (Blankenship et al., 2011; Larkum et al., 2012; Walter et al., 2010). In cyanobacteria this 'green gap' is already partly filled by additional pigments, but in green algae there is a 'larger gap', so a clear opportunity to improve light absorption in the green spectrum (Larkum et al., 2012). PPRs could potentially be engineered as additional photosystem into the chloroplast of green algae.

A larger potential for improving the photosynthetic efficiency of cyanobacteria and green algae is to increase absorption at wavelengths above 700 nm, where their native photosystems absorb no light at all (Blankenship et al., 2011; Larkum et al., 2012). However, absorption in this range is not yet found for PPRs, as discussed above.

An important aspect for introducing PPRs into photosynthetic microorganisms is that proton-pumping efficiency of the native chlorophyll-based photosystems is generally higher than that of PPRs (Box 1). Hence, introducing a PPR into photosynthetic organisms will probably only increase total proton-pumping fluxes in the cell, if the PPRs do not compete directly for photons with native photosystems. However, when an increased total proton flux can be generated in photosynthetic cells by introducing PPRs, this could potentially support an increased growth rate or production yield. We simulated the growth rate of a cyanobacterium engineered with a PPR. Therefore we added our estimated, feasible PPR proton flux in the genome-scale metabolic reconstruction of the cyanobacterium *Synechocystis* sp. PCC6803 and performed flux balance analyses (Nogales et al., 2013). As a result, the PPR proton flux in light-limited conditions increased the simulated growth rate of *Synechocystis* sp. PCC6803 by 10% (Table 2).

Moreover, in cyanobacteria and green algae the ATPases are normally found in the thylakoid membranes. Hence, the PPRs can probably best be expressed in these membranes of photosynthetic microorganism in such orientation that they pump protons to generate a PMF for the ATPases.

There is one primitive cyanobacterium known that has no thylakoid membrane and instead has its photosystems and ATPases located in the cytoplasmic membrane: *G. violaceus* (Nakamura et al., 2003). Noteworthy, this is the only currently known cyanobacterium having a PPR gene. The fact that a PPR system has been included to a cyanobacterium by natural evolution may indicate that introducing PPRs in cyanobacteria or green algae having thylakoids is also worth a try.

Concluding remarks and perspectives

Only a few biotechnological applications of introducing PPRs into relevant production hosts have been reported so far. In addition, several potential applications have been discussed in this review. Generated proton fluxes from introduced PPR photosystems have the potential to increase yields of pathways that are limited by ATP or other PMF-driven processes. PPR-boosted processes form a new opportunity for metabolic engineering.

PPRs can also contribute to a recent development in synthetic biology: introducing autotrophy into heterotrophic organisms. These endeavors often have a goal to overcome inefficiencies of natural photoautotrophy, such as improving the inefficiency of CO_2-fixation. The most common pathway for photosynthesis is the Calvin Cycle, in which the CO_2-fixing RuBisCO is a notably slow enzyme and has competing oxygenase activity. The Calvin Cycle with an improved RuBisCO mutant was engineered in *E. coli* (Parikh et al., 2006). Recently, more efficient CO_2 fixation pathways from extremophiles were engineered in the heterotrophs *E. coli* and *Pyrococcus furiosis* (Bonacci et al., 2012; Keller et al., 2013; Mattozzi et al., 2013). Integration of such CO_2-fixation pathways together with pathways for the uptake of non-carbon electron donors has been suggested; these electron donors can for example be cathodes of bioelectrochemical systems and hydrogen (Hawkins et al., 2013; Lovley and Nevin, 2013; Rabaey et al., 2011). However, much ATP is needed to support efficient reduction of CO_2 by these electron donors (Fast and Papoutsakis, 2012; Rabaey et al., 2011). So, integration of CO_2 fixation processes and electron donor uptake pathways has a potential bottleneck of ATP-limitation. This ATP bottleneck can potentially be relieved by integrating PPRs into this new synthetic biology concept (further explored by *in silico* analysis in **Chapter 3**).

Acknowledgements

N.J.C is supported by Wageningen University by the IP/OP program Systems Biology. V.A.P.M.d.S acknowledges financial support by the EU through the project Microme (FP7-grant 222886-2). W.M.d.V. acknowledges financial support by the Netherlands Organization for Scientific Research (NWO) through a Spinoza grant.

Chapter 3

Integrated *in silico* analysis of pathway designs for synthetic photo-electro-autotrophy

Michael Volpers[1,2*], Nico J. Claassens[3*], Elad Noor[4], John van der Oost[3], Willem M. de Vos[3,5], Servé W. M. Kengen[3], and Vitor A.P. Martins dos Santos[1,2]

*contributed equally
[1]Laboratory of Systems and Synthetic Biology, Wageningen University, Dreijenplein 10, 6703 HB, Wageningen, The Netherlands
[2]LifeGlimmer GmbH, Markelstr. 39a, 12136, Berlin, Germany
[3]Laboratory of Microbiology, Wageningen University, Dreijenplein 10, 6703 HB, Wageningen, The Netherlands
[4]Institute of Molecular Systems Biology, ETH Zürich, Auguste-Piccard-Hof 1, 8093, Zürich, Switzerland
[5]Department of Bacteriology and Immunology, Helsinki University, Haartmaninkatu 3, 00014, Helsinki, Finland

Chapter adapted from publication:
PLoSONE (2016) DOI:10.1371/journal.pone.0157851

Abstract

The strong advances in synthetic biology enable the engineering of novel functions and complex biological features in unprecedented ways, such as implementing synthetic autotrophic metabolism into heterotrophic hosts. A key challenge for the sustainable production of fuels and chemicals entails the engineering of synthetic autotrophic organisms that can effectively and efficiently fix carbon dioxide by using sustainable energy sources. This challenge involves the integration of carbon fixation and energy uptake systems. A variety of carbon fixation pathways and several types of photosystems and other energy uptake systems can be chosen and, potentially, modularly combined to design synthetic autotrophic metabolism. Prior to implementation, these designs can be evaluated by the combination of several computational pathway analysis techniques. Here we present a systematic, integrated *in silico* analysis of photo-electro-autotrophic pathway designs, consisting of natural and synthetic carbon fixation pathways, a proton-pumping rhodopsin photosystem for ATP regeneration and an electron uptake pathway. We integrated Flux Balance Analysis of the heterotrophic chassis *Escherichia coli* with kinetic pathway analysis and thermodynamic pathway analysis (Max-min Driving Force). The photo-electro-autotrophic designs are predicted to have a limited potential for anaerobic, autotrophic growth of *E. coli*, given the relatively low ATP regenerating capacity of the proton pumping rhodopsin photosystems and the high ATP maintenance of *E. coli*. If these factors can be tackled, our analysis indicates the highest growth potential for the natural reductive tricarboxylic acid cycle and the synthetic pyruvate synthase – pyruvate carboxylate -glyoxylate bicycle. Both carbon fixation cycles are very ATP efficient, while maintaining fast kinetics, which also results in relatively low estimated protein costs for these pathways. Furthermore, the synthetic bicycles are highly thermodynamic favorable under conditions analyzed. However, the most important challenge identified for improving photo-electro-autotrophic growth is increasing the proton-pumping rate of the rhodopsin photosystems, allowing for higher ATP regeneration. Alternatively, other designs of autotrophy may be considered, therefore the herein presented integrated modeling approach allows synthetic biologists to evaluate and compare complex pathway designs before experimental implementation.

Keywords
Photo-electro-autotrophy; CO_2 fixation; integrated *in silico* analysis; constraint-based metabolic modelling; thermodynamic analysis

Introduction

One of the current grand societal and technological challenges is to establish sustainable production processes for chemicals and fuels. The current biotechnological production is based on a series of relatively inefficient steps, including plant photosynthesis, harvesting and transport of biomass from fields to a biorefinery, pre-treatment of crude biomass and subsequent microbial fermentation (Hawkins et al., 2013). A potentially more efficient production process is offered by autotrophic microorganisms directly converting carbon dioxide and sustainable energy, such as light and electricity, into products. However, many autotrophic microorganisms are not genetically amenable or for other reasons not suitable for such industrial applications. Major advances in synthetic biology start to allow engineering of complex features related to autotrophy into heterotrophic chassis microorganisms, such as *Escherichia coli*. Recently, features such as (partial) carbon fixation pathways (Bonacci et al., 2012; Ducat and Silver, 2012; Gong et al., 2015; Guadalupe-Medina et al., 2013; Keller et al., 2013; Mainguet et al., 2013; Mattozzi et al., 2013), assimilation pathways of one-carbon compounds (Müller et al., 2015; Siegel et al., 2015), photosystems (Claassens et al., 2013; Wang et al., 2015) and direct electron transfer (Goldbeck et al., 2013; Jensen et al., 2010) have been successfully introduced in *E. coli* and some other heterotrophic hosts. However, complete synthetic autotrophy engineered in a heterotrophic host has not yet been demonstrated.

Completely autotrophic, synthetic microorganisms require integration of subsystems for both carbon fixation and energy uptake to regenerate electron donors and ATP. Those subsystems need to be integrated into properly evaluated designs before going into the challenging, time-consuming and expensive process of experimental implementation. Here we present an *in silico* analysis of different designs for so-called anaerobic photo-electro-autotrophy in *E. coli*. These designs consist of both uptake pathways for electron donors and photosystems to harvest light energy (**Figure 1**). Electron donor uptake can be achieved through uptake pathways for e.g. low potential electrons from a cathode, formate or hydrogen for regenerating intracellular electron donors, known as microbial electrosynthesis. Most pathways for electrosynthesis function best anaerobically, however, this mostly results

in an ATP limitation (Fast and Papoutsakis, 2012; Hawkins et al., 2013; Lovley and Nevin, 2013; Pandit and Mahadevan, 2011; Rabaey and Rozendal, 2010; Rabaey et al., 2011). Therefore we included in our designs a proton-pumping rhodopsin (PPR), which is a photosystem that can generate a proton motive force and hence regenerate ATP under anaerobic conditions (Claassens et al., 2013). These proton-pumping rhodopsin photosystems are less complex to engineer by synthetic biology tools than the common water-splitting reaction-center photosystems (Claassens et al., 2013). However, contrary to reaction-center photosystems, proton-pumping rhodopsins cannot regenerate electron donors, so for autotrophic growth they need to be complemented with electron donor uptake mechanisms. These electron uptake mechanisms and photosystems could be integrated with a carbon fixation pathway that can be chosen from a variety of (anaerobic) synthetic and natural pathways (Bar-Even et al., 2010).

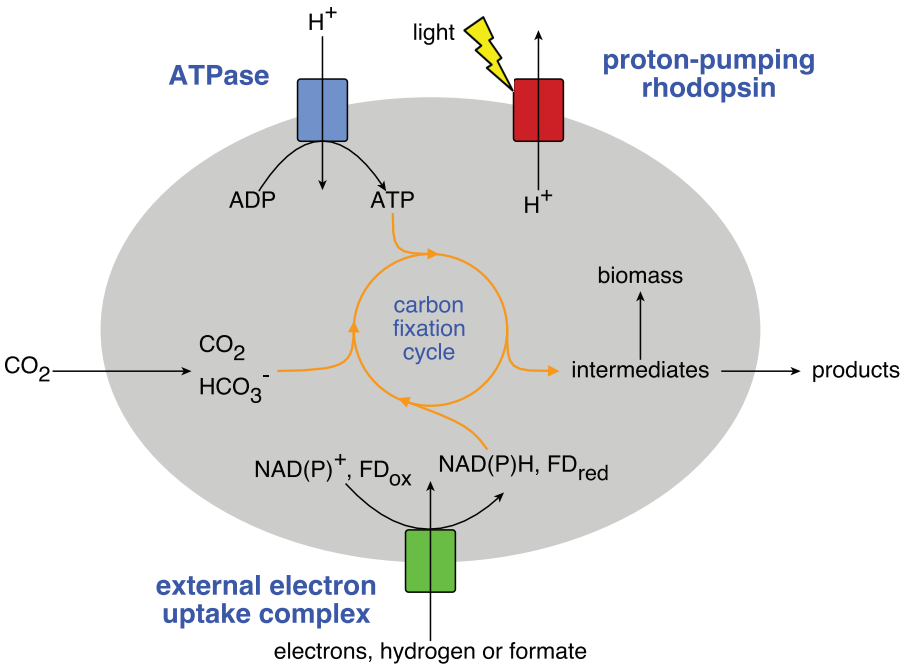


Figure 1. General principle of the photo-electro-autotrophic *E. coli*. Key components are a carbon fixation cycle, a mechanism to take up external electrons and the proton-pumping rhodopsins photosystem.

Here we investigate the feasibility of the different designs for synthetic photo-electro-autotrophy by using an integrated *in silico* analysis approach. Several *in silico* tools enabled the analysis of key performance factors such as growth rates, pathway thermodynamics, pathway kinetics and systems protein burden. We used the Max-min Driving Force (MDF) to determine the thermodynamic feasibility of pathways (Flamholz et al., 2012; Noor et al., 2012, 2013, 2014). We used Flux Balance Analysis (FBA) (Orth et al., 2010) and Flux Variability Analysis (FVA) (Mahadevan and Schilling, 2003) to predict growth yields and integrated these data with enzyme and pathway kinetic parameters (Bar-Even et al., 2010) to predict the protein burden on the cell of the components of the photo-electro-autotrophic designs. Some of these tools have been applied before to analyze carbon fixation pathways and other integrated autotrophic designs (Bar-Even et al., 2010, 2012a, 2012b, 2013; Boyle and Silver, 2012; Fast and Papoutsakis, 2012; Pandit and Mahadevan, 2011). Here we combine all of these state-of-the-art *in silico* analysis tools with input of available experimental data to analyze novel photo-electro-autotrophic designs.

The integrated *in silico* analysis allowed us to compare the feasibility of different designs and identify key bottlenecks for realizing photo-electro-autotrophy in e.g. *E. coli*. Photo-electro-autotrophic growth in *E. coli* under anaerobic conditions, seems drastically limited by ATP. The assumed ATP regeneration through the rhodopsin photosystems is too limited to achieve high growth rates. Improving rhodopsin photosystems pumping rates and/or levels, or alternative ATP regeneration systems have to be considered before successful implementation of the envisioned photo-electro-autotrophic designs in *E. coli*. Given the limitations of rhodopsin photosystems, reasonably high photo-electro-autotrophic growth rates above 0.1 h^{-1} are only feasible when assuming that the Non-Growth-Associated Maintenance (NGAM), of *E. coli* can be drastically decreased to ~1 mmol ATP/gCDW/h^{-}. Furthermore, ATP-efficient carbon fixation cycles, such as the natural reductive TCA (rTCA) cycle and synthetic pyruvate synthase – pyruvate carboxylase – glyoxylate (PyrS-PyrC-Glx) bicycle, result in higher predicted growth rates, while giving relatively low protein costs. Generally, the herein presented, integrated *in silico* approach allows synthetic biologists and metabolic engineers to better evaluate and compare complex designs for e.g. autotrophy before experimental implementation.

Methods

Autotrophy subsystems for the designs

Carbon fixation pathways and reactions

Carbon fixation is a central part of any autotrophic system. From the six known natural pathways for carbon fixation and a large theoretical repertoire of synthetic pathways for carbon fixation we analyzed six pathways, which are listed in Table 1 with their most important characteristics. Firstly we included the naturally dominant, oxygen-tolerant (i) Calvin cycle (**Figure S1**). However, for the photo-electro-autotrophic designs we focused on anaerobic conditions. This enabled us to include carbon fixation pathways containing oxygen-sensitive ferredoxin-oxidoreductase enzymes, which is advantageous, as these pathways are often both kinetically fast and ATP-efficient (Bar-Even et al., 2010). Therefore we selected the natural, anaerobic (ii) rTCA cycle (**Figure S2**) (Berg, 2011; Berg et al., 2010).

Furthermore we selected the natural (iii) 3-hydroxypropionate – 4-hydroxybutyrate (3HP-4HB) cycle (**Figure S3**); in our analyses, in contrast to the natural aerobic situation for this cycle, the primary product acetyl-CoA is assimilated into the central metabolism using oxygen-sensitive pyruvate synthase. This anaerobic alternative is a more ATP-efficient alternative than the natural assimilation via succinyl-CoA (Estelmann et al., 2011). Even more importantly, as in *E. coli* the succinate dehydrogenase generates ubiquinol, which can be only be regenerated to ubiquinone with a high potential electron acceptor, such as oxygen, which is not present in envisioned anaerobic conditions. Furthermore, we note that the 3-HP-4HB cycle we analyzed is the crenarchaeal version, and not the recently elucidated, more ATP-efficient, thermodynamically less favorable, thaumarchaeal verion (Konneke et al., 2014). While analyzing the 3HP-4HB cycle in *E. coli* we identified a more promising alternative cycle, being the natural, anaerobic (iv) – dicarboxylate-4-hydroxybutyrate (DC-4HB) cycle (**Figure S4**).

In addition, we selected two synthetic pathways from an extensive overview of identified synthetic carbon fixation pathways from a study by the Milo group (Bar-Even et al., 2010). From these we selected ferredoxin-oxidoreductase

containing pathways with both low ATP requirements and favorable kinetics and thermodynamics (Bar-Even et al., 2010), for which experimental implementation in *E. coli* is ongoing (Mainguet et al., 2013): the (v) PyrS-PyrC-Glx and (vi) pyruvate synthase – phosphoenol carboxylase - glyoxylate (PyrS-PEPC-Glx) bicycles (**Figure 2**). The latter pathway was proposed some decades ago to be a natural cycle, named reductive dicarboxylic acid cycle (Ivanovsky et al., 1993), but its presence in nature could so far not be confirmed (Bar-Even et al., 2010). We did not pursue the analysis of the natural 3-hydroxypropionate bicycle – for which engineering has been attempted in *E. coli* (Mattozzi et al., 2013) – because it requires, similar to the natural 3HP-4HB pathway, oxygen or another electron acceptor for the regeneration of the high potential electron acceptor of the succinate dehydrogenase. The natural carbon fixation with the lowest ATP consumption known, the Wood-Ljungdahl pathway (Berg, 2011; Berg et al., 2010), was also not included, as this complex pathway is probably hard to engineer into *E. coli*, as it has complex metallo-chemistry (Ducat and Silver, 2012).

Carbon uptake was assumed to occur by passive diffusion of CO_2 over the cell membranes, which is already implemented as a reaction in the applied *E. coli* core model. Alternatively, one may consider to increase carbon uptake rate by expressing heterologous bicarbonate transporters. However, such a bicarbonate transporter may not be required in *E. coli*. It was demonstrated that heterologous expression of a bicarbonate transporter in *E. coli* did not further increase CO_2 fixation rates in an *E. coli* strain engineered with Calvin cycle enzymes (Gong et al., 2015).

Table 1. General characteristics of the analyzed carbon fixation pathways.

Carbon fixation pathway	*Origin pathway*	*Oxygen sensitive enzymes*	*Electron donors*	*Carbon species*	*Total number of enzymes in cycle**	*References, figures*
Calvin cycle (reductive pentose phosphate pathway)	natural e.g. plants, cyanobacteria	no	NADPH	CO_2	12	(Berg, 2011), Fig. S1
rTCA cycle (reductive tricarboxylic acid cycle)	natural many anaerobic bacteria	yes**	$ferredoxin_{red}$ NADPH NADH	CO_2	8	(Berg, 2011), Fig. S2
3HP-4HB cycle (3-hydroxypropionate - 4-hydroxybutyrate cycle)	natural, *aerobic crenarchaeota*	no***	NADPH	HCO_3^-	13	(Berg, 2011), Fig. S3
DC-4HB cycle (dicarboxylate-4-hydroxybutyrate cycle)	natural, *anaerobic crenarchaeota*	yes	$ferredoxin_{red}$ NADPH	CO_2, HCO_3^-	14	(Berg, 2011) Fig. S4
PyrS-PyrC-Glx bicycle (pyruvate synthase-pyruvate carboxylase-glyoxylate bicycle)	synthetic	yes	$ferredoxin_{red}$ NADH	CO_2, HCO_3^-	10	(Bar-Even et al., 2010, 2012c), Fig. 2
PyrS-PEPC-Glx bicycle (pyruvate synthase-phosphoenol pyruvate carboxylase-glyoxylate bicycle)	synthetic	yes	$ferredoxin_{red}$ NADH	CO_2, HCO_3^-	11	(Bar-Even et al., 2010, 2012c), Fig. 2

*the number of enzymes includes all enzymes which are part of the CO_2 fixation cycles, note that for the synthetic cycles we also included the number of enzymes for the required glyoxylate assimilation cycle, which consists of 5 additional enzymes

**some natural bacteria harboring the rTCA cycle are microaerobic, they have some mechanism to protect the ferredoxin-oxidoreductases from oxygen or have oxygen-tolerant variants of those (Berg, 2011)

***the natural 3HP-4HB cycle is aerobic, however the variant modelled in this paper contains pyruvate synthase, an oxygen sensitive ferredoxin-oxidoreductase

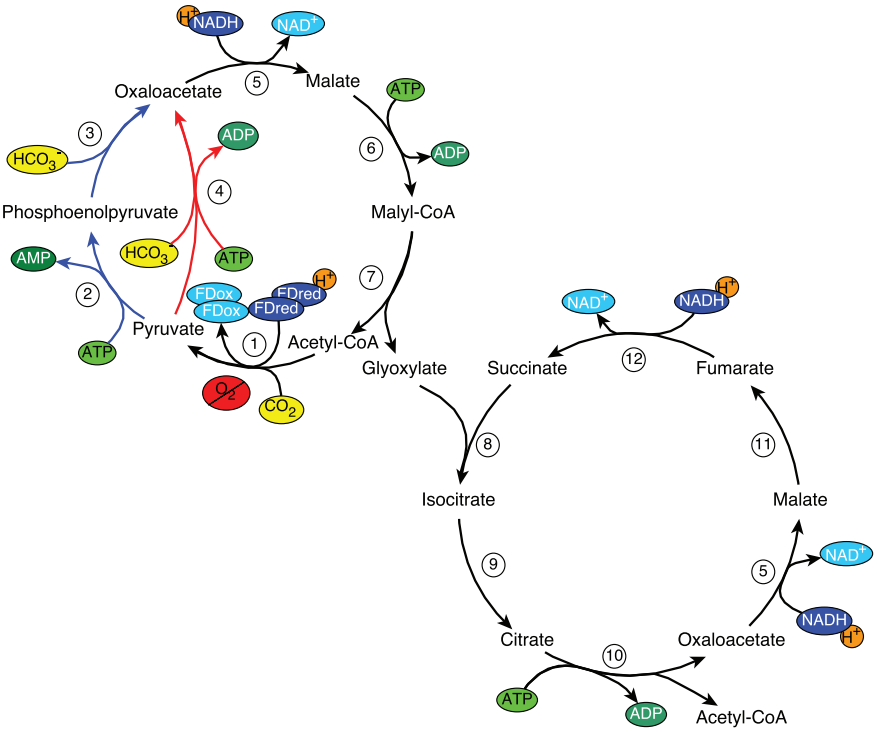


Figure 2. Synthetic carbon fixation pathways: PyrS-PyrC-Glx bicycle (red and black) and the PyrS-PEPC-Glx bicycle (blue and black). Enzymes involved in the bicycles: 1) pyruvate synthase (oxygen sensitive), 2) pyruvate water dikinase, 3) phosphoenolpyruvate carboxylase, 4) pyruvate carboxylase, 5) malate dehydrogenase, 6) malate thiokinase, 7) malyl-CoA lyase, 8) isocitrate lyase, 9) aconitate hydratase, 10) ATP citrate lyase, 11) fumarate hydratase and 12) fumarate reductase.

The reactions for all six analyzed carbon fixation pathways are listed in Table S1. These reactions were used for FBA and FVA in the *E. coli* core model and further computational pathway analyses. The carbon dioxide uptake flux was not constrained for FBA and FVA.

Proton-pumping rhodopsin reaction

The PPR photosystem absorbs photons to pump protons from the cytoplasm to the periplasm. However, we use the *E. coli* core model for our computational analysis (see section Flux Balance and Variability analysis below and **Supplementary Text S1**), which contains no periplasmic compartment. Due to this we modelled the PPRs by a transport reaction for cytoplasmic protons (H^+_c) to extracellular protons (H^+_e):

$$1H^+_c \rightarrow 1\ H^+_e$$

Light or photons were not included as a factor in the model; the constraint of the availability of photons we implicitly incorporated by the constraints we put on the PPR reaction. The range of constraints analyzed for this reaction was based on the range of PPR proton pumping values available from literature: 0-10 mmol H^+/gCDW/h (Claassens et al., 2013; Ganapathy et al., 2015). However, to explore the potential of further optimizing the PPR proton pumping we extended the analyzed range up to 50 mmol H^+/gCDW/h.

Electron-uptake reactions

To enable carbon fixation, electron donors need to be regenerated continuously. This cannot be done by the proton-pumping rhodopsin photosystems. Hence, those electron donors need to be regenerated through external electron uptake mechanisms. Several external electron donors are promising for electro-autotrophy, e.g. low potential electrons, formate or hydrogen. For practical application, these electron donors basically all require a pathway that couples them to the required electron donors for carbon fixation: NADPH, NADH and/or reduced ferredoxin. Therefore the electron uptake reaction was generalized for FBA and FVA by adding an exchange reaction for extracellular electrons ($e_{\bar{e}}$) and subsequent reactions that use these electrons to provide electron donors, similar to the previously published simplified FBA reaction for electrosynthesis (Pandit and Mahadevan, 2011).

NADPH regeneration:

$$1\ NADPH^+ + 2e^-_e + 1H^+_c \rightarrow NADPH$$

We added no reaction to regenerate NADH, since both the *E. coli* core and the genome-scale model have an NAD transhydrogenase, which transfers electrons from NADPH to NADH (Sauer et al., 2004).

$Ferredoxin_{red}$ regeneration was implemented in the model as well:

$$1\ Ferredoxin_{ox} + 1e^-_e \rightarrow 1Ferredoxin_{red}$$

It has to be noted, that unlike $NAD(P)^+$, the reduction of ferredoxin by H_2 quickly becomes unfavorable even at moderately low H_2 concentrations or $ferredoxin_{red}$:$ferredoxin_{ox}$ ratios. However, recently several flavin-based electron bifurcating enzymes (Buckel and Thauer, 2013; Schuchmann and

Müller, 2014; Wang et al., 2013) have been described that couple endergonic reduction of ferredoxin with exergonic reduction of NAD(P)H. Given that the ferredoxin-requiring carbon fixation pathways analyzed, require ferredoxin as well as NAD(P)H, such systems would be suitable for overall energy-neutral, simultaneous regeneration of both types of electron donors. If this option turns out not to be practically feasible, proton motive force-driven enzymes may be employed for ferredoxin reduction, however this would add energetic costs, which are not accounted for in our analyses here.

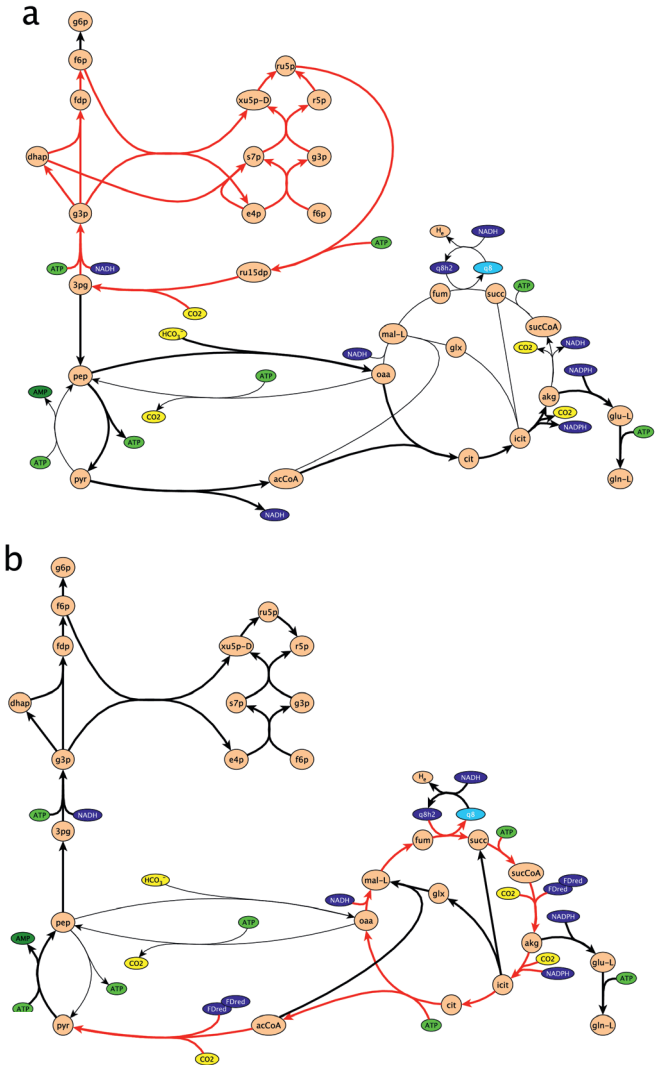
The uptake of external electron donors and subsequent electron donation reactions were not constrained for FBA and FVA. However, simulated fluxes for optimal growth could be compared to scarce literature data on feasible specific activities for electron uptake mechanisms.

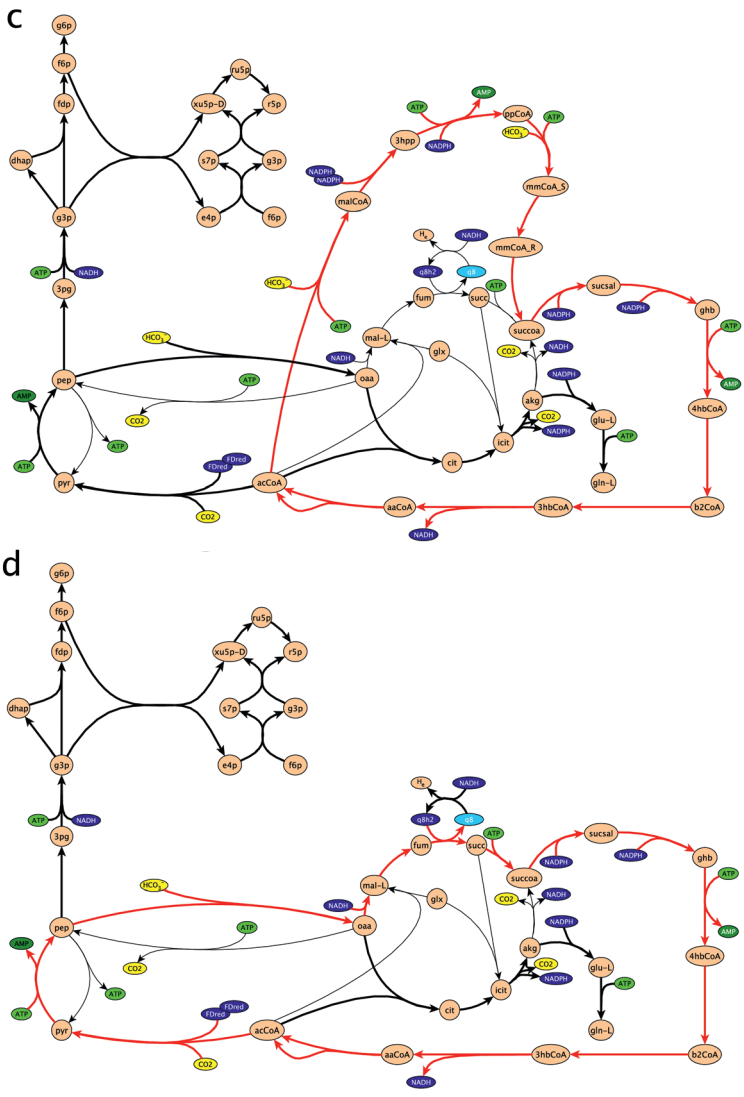
Computational methods of analysis

Flux Balance and Variability Analysis

We used the *E. coli* core model (Orth et al., 2009) for FBA (Orth et al., 2010) and FVA (Mahadevan and Schilling, 2003) of photo-electro-autotrophic designs in *E. coli*. We decided to use the core model instead of the genome-scale model of *E. coli*, as it is easier to analyze, for example for assessing total ATP consumption. Furthermore, the core model is more general and not overly specific to *E. coli;* this may facilitate more general conclusions that are more likely to apply to other promising chassis organisms having similar central metabolism (Bar-Even et al., 2013). Even though all carbon fixation cycles directly link to intermediates in the core metabolism (**Figure 3**), we verified that the full model gave similar predictions as the core model, to rule out biased results by using the core model (**Supplementary Text S1 and Supplementary Table S2**). The reactions listed in Table S1 were added to the model, including the missing reactions for the carbon fixation pathways, reactions to provide electron donors and a reaction representing the PPRs. The constraints used for the FBA and FVA are included in Table S1 and S2. We performed FBA and FVA with both the standard Non Growth Associated Maintenance (NGAM) for *E. coli of* 8.39 mmol ATP/gCDW/h (Feist et al., 2007) and a lower NGAM estimate of 1.00 mmol ATP/gCDW/h. To check for possible alternative flux distributions for the optimal growth solution, we also performed Flux Variability Analysis (FVA).

This is mainly interesting for the calculation of the protein burden to the cell and the ATP consumption analysis, since in these methods the flux through a certain reaction and the flux distribution, respectively, play a crucial role. For FBA and FVA we used Python 2.7 and the cobrapy toolbox (Ebrahim et al., 2013) with GLPK 4.48 as the solver for the linear optimization. We optimized for growth utilizing the default biomass objective function in the *E. coli* core model (Orth et al., 2009) (**Supplementary Table S1 and S2**).




c
g6p
f6p
fdp
dhap
g3p
3pg
pep
pyr
ru5p
xu5p-D
r5p
s7p
g3p
e4p
f6p
3hpp
ppCoA
mmCoA_S
malCoA
mmCoA_R
fum
succ
succoa
mal-L
glx
oaa
cit
icit
akg
glu-L
gln-L
sucsal
ghb
4hbCoA
acCoA
aaCoA
3hbCoA
b2CoA
ATP
NADH
NADPH
AMP
CO2
HCO3
H2
q8h2
q8
FDred
fDred
d
g6p
f6p
fdp
dhap
g3p
3pg
pep
pyr
ru5p
xu5p-D
r5p
s7p
g3p
e4p
f6p
fum
succ
succoa
mal-L
glx
oaa
cit
icit
akg
glu-L
gln-L
sucsal
ghb
4hbCoA
acCoA
aaCoA
3hbCoA
b2CoA
ATP
NADH
NADPH
AMP
CO2
HCO3
H2
q8h2
q8
FDred
fDred

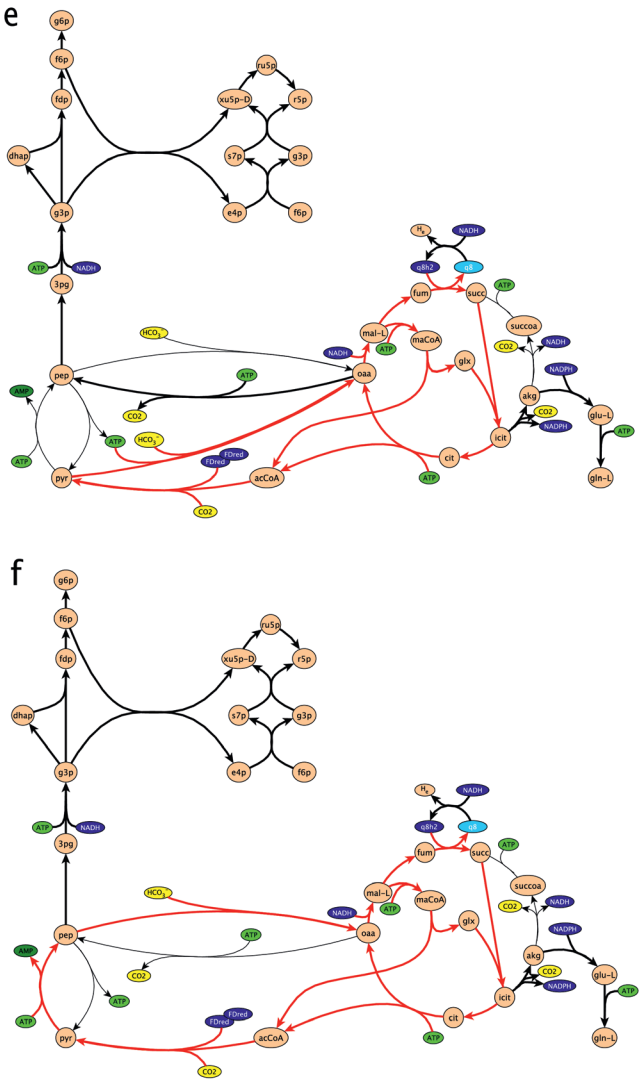


Figure 3.Integration of **different carbon fixation pathways in the *E. coli* core metabolism.** a) Calvin cycle, b) rTCA cycle, c) 3HP-4HB cycle, d) DC-4HB cycle, e) PyrS-PyrC-Glx bicycle, f) PyrS-PEPC-Glx bicycle. Reactions that are part of carbon fixation pathways (red arrows), reactions carrying flux (bold red arrows) and not carrying flux (thin arrows) in the FBA simulations are indicated. ATP (green), electron donors (blue) and CO_2/HCO_3^- (yellow) are color coded. Reactions of some non-branching pathways were lumped into one reaction for better visualization. For more detailed maps of the carbon fixation pathways see **Figure 2 and Supplementary figures S1-S4.**

ATP consumption analysis of carbon fixation

We quantified the ATP consumption of the carbon fixation pathways to produce biomass. This was done by calculating the number of ATPs consumed for each precursor in the *E. coli* core model biomass equation. All consumed ATPs were calculated, starting from inorganic carbon to the biomass precursors through the pathways taken in the solution from the FBA. We took into account two ATPs consumed for AMP-producing reactions. We calculated the absolute ATP consumption per mole biomass precursor. From this we calculated the total ATP consumption per mole biomass produced, by taking into account the stoichiometry of the precursors in the biomass objective function and the Growth Associated Maintenance (GAM) of the biomass reaction.

Thermodynamic analysis of carbon fixation

For analyzing the thermodynamics of the carbon fixation pathways, Gibbs free energies of reactions ($\Delta_r G'^0$) were acquired by the Component Contribution method, implemented in the eQuilibrator software (Flamholz et al., 2012; Noor et al., 2012, 2013). The Gibbs free energy of a reaction ($\Delta_r G'$) under physiological conditions depends on the metabolite concentrations of both substrates and products. These metabolite concentrations can be varied to optimize the thermodynamic profile of the carbon fixation pathways. In this analysis we focus on the reaction's thermodynamic driving force, that is the negative change of Gibbs free energy ($-\Delta_r G'$) (Noor et al., 2014). A pathway is only thermodynamically feasible if the thermodynamic driving force of each reaction is positive. The thermodynamic profile optimization was done by varying the metabolite concentrations within physiological constraints to maximize the minimal driving force in a pathway. After optimization we found the thermodynamic bottleneck reaction of a pathway, which is the reaction with the minimal thermodynamic driving force. This value for the maximized, minimal thermodynamic driving force of a pathway is further referred to as the Max-min Driving Force (MDF) (Noor et al., 2014). This MDF determines the magnitude of the thermodynamic bottleneck of a given pathway, i.e. a higher MDF indicates a less severe thermodynamic bottleneck. The MDF analysis done in here is an improved implementation of the original MDF analysis (Noor et al., 2014). This implementation of MDF applied here and elsewhere (Gerosa et al., 2015), also incorporates the standard errors in Gibbs free energy estimates made by the Component Contribution method (**Supplementary text 2**).

For the thermodynamic profile optimization, we fixed some co-factor concentrations and where possible ratios of co-factors, if not stated otherwise. For other metabolite concentrations we chose the range to be between 1 µM and 10 mM, representing a physiological range (Noor et al., 2014). The concentration of the two inorganic carbon species consumed in different carbon fixation pathways was fixed: [CO_2 (aq)]=0.3 mM, [HCO_3^-]=7 mM. These concentrations are based on pH 7.5 and equilibrium with a gas feed phase containing 1% (v/v) CO_2, which is a feasible CO_2 concentration for an industrial set-up. For the following co-factors concentrations and ratios were fixed: [orthophosphate]=10 mM, [pyrophosphate]=1 mM, [CoA]=1 mM, [ATP]/[ADP]=10, [ADP]/[AMP]=1, [NADPH]/[$NADP^+$]=10, [NADH]/[NAD^+]=0.1 as in (Noor et al., 2014). We calculated all the Gibbs free energies for pH 7.5 and an ionic strength of 0.2 M (Noor et al., 2014). For [$ferredoxin_{red}$]/[$ferredoxin_{ox}$] we assumed a ratio of 10, contrary to the ratio of 1 assumed before (Noor et al., 2014). This ratio can be generally maintained in cells by flavin-based electron bifurcating enzymes as discussed above. Furthermore, we assumed that the PPRs together with the operation of ATP synthase can maintain the [ATP]/[ADP] ratio.

Most concentrations and ratios chosen, are based on the characteristics of *E. coli* cytoplasm (Bennett et al., 2009; Noor et al., 2014). However, in real life conditions and literature a wide range of co-factor concentrations and ratios can be found. Hence, we also performed a sensitivity analysis for the carbon species concentrations, ratios of electron donor and ATP/ADP. We varied the ratios given above one at a time as follows; [$ferredoxin_{red}$]/[$ferredoxin_{ox}$]: 0.5 – 50, [ATP]/[ADP]: 5 - 20, [NADPH]/[$NADP^+$]: 0.5 – 50 and [NADH]/[NAD^+]: 0.1 – 10. Additionally, we computed the MDF for a varying range of [CO_2 (aq)]: 0.01 – 3 mM and [HCO_3^-]: 0.25 - 70 mM. We optimized the thermodynamic profile using Python 2.7 and a module for sequential least squares programming from the package scipy.optimize (fmin_slsqp).

Kinetic analysis of carbon fixation pathways

To compare the kinetics of the carbon fixation pathways, we used two methods. The first method, Flux-Force Efficacy (FFE), is related to the thermodynamic driving force of the thermodynamic bottleneck reaction in a pathway, which is quantified by the MDF, as discussed in the section above (Noor et al., 2014).

The FFE represents the forward flux as a fraction of the total flux through the bottleneck reaction. The FFE has values between 0 and 1; a value close to 1 is preferred as this means that the backward flux through the thermodynamic bottleneck reaction is negligible. The FFE for bottleneck reactions is calculated according to:

$$\mathrm{FFE} = \frac{-1+\exp\left(\frac{-\Delta_r G'}{RT}\right)}{1+\exp\left(\frac{-\Delta_r G'}{RT}\right)} \quad (1)$$

R is the gas constant (8.314 J/mol/K), *T* is the temperature (310.15 K) and $-\Delta_r G'$ is the MDF of the pathway.

The second method for estimating kinetics is the Pathway Specific Activity (PSA) (Bar-Even et al., 2010). PSA is defined as the upper limit flux (µmol pathway product/mg/min) carried by 1 mg of total protein in the pathway. PSA is calculated according to:

$$\mathrm{PSA} = 1 \Big/ \sum_{i=1}^{m} \frac{w_i}{v_i} \quad (2)$$

With *m*, w_i and v_i as the number of enzymes in the pathway, their stoichiometric coefficients in the pathway and the specific activities of those enzymes, respectively. Most of the specific activities of the enzymes were taken from (Bar-Even et al., 2010). The specific activities of some of the reactions in the carbon fixation pathways were not available before, due to a lack of experimental data especially for the ferredoxin-oxidoreductase enzymes (Bar-Even et al., 2010). However, recently some new experimental data became available for those enzymes in BRENDA (Schomburg et al., 2013). So, we estimated their specific activity by taking the mean, after discarding the lower 50% and upper 10% of the specific enzyme activities found in BRENDA. These values were discarded to limit the inclusion of unnaturally low or high activities, because of experimental errors and non-representative *in vitro* conditions, as done in (Bar-Even et al., 2010). All enzymes and their specific activities are listed in Table S1. In contrast to (Bar-Even et al., 2010), we assumed CO_2 and HCO_3^- concentrations to be saturated, similarly to all other concentrations. This assumption allowed to have the most fair comparison between pathways, since the K_m values for the carbon species for the ferredoxin-oxidoreductase enzymes are not available. Furthermore, in a potential industrial application a gas feed with high CO_2 concentrations close to saturation is feasible. As the pathway end product for

PSA we chose pyruvate, since this is a central metabolite, and an important precursor for biosynthesis of many products of commercial interest, further it is an important product of several analyzed carbon fixation pathways.

Protein burden of photo-electro-autotrophic system

We propose the Pathway Protein Burden (PPB) (g protein/gCDW) as a method to estimate the fraction of cellular dry weight that needs to be dedicated to carbon fixation enzymes to produce pyruvate from CO_2 for the growth rate predicted by FBA. It is defined as:

$$\mathrm{PPB} = F_{CO2}/(n^{*}\mathrm{PSA}) \tag{3}$$

where F_{CO2} (mmol/gCDW/h) is the CO_2 uptake flux from the FBA for the simulated growth rates, PSA is the Pathway Specific Activity (µmol/mg protein/h) and n is the number of CO_2/HCO_3^- molecules needed to produce one molecule of pathway end product. In our analysis this pathway end product is pyruvate, (n = 3).

Since PSA is an estimate for the fastest kinetics of a carbon fixation pathway, the PPB represents a minimal estimate for the protein burden of the carbon fixation pathway to support the predicted CO_2 uptake flux to produce pyruvate.

For electron uptake mechanisms and PPRs, which are mostly single proteins or protein complexes, similarly an estimate for the Protein Burden (PB) (g protein/gCDW) of these systems can be made as follows:

$$\mathrm{PB} = F/(n^{*}SA) \tag{4}$$

Where F (mmol/gCDW/h) is the electron uptake flux or proton-pumping flux for the simulated growth conditions, SA (µmol/mg protein/h) is the specific activity of the related protein, i.e. the electron uptake enzyme or PPR. n is a stochiometric constant for example for converting the number of electrons per hydrogen or formate molecule (n = 2).

Table 2. *In silico* results for all analyzed carbon fixation pathways and overview of number of enzymes natively present, engineered or not yet engineered in *E. coli* (see also **Supplementary text S3**). Results shown are based on a high flux of PPR proton pumping of 50 mmol/gCDW/h.

	Flux Balance Analysis (FBA)		*Thermodynamics*	*Stoichiometry*	*Kinetics*		*Pathway Protein Burden*		*Expression challenges*		
	growth rate (hr^{-1})						*g protein /gCDW*		*Number of enzymes*		
Carbon fixation pathway	*low maintenance (1.00 mmol ATP/gCDW/h)*	*high maintenance (8.39 mmol ATP/gCDW/h)*	*Max-Min Driving Force (MDF) (kJ/mol)*	*ATP required (mmol/ gCDW)*	*Flux-Force Efficacy (FFE)*	*Pathway Specific Activity (PSA) (µmol pyruvate/mg pathway protein/min)*	*low maintenance (1.00 mmol ATP/gCDW/h)*	*high maintenance (8.39 mmol ATP/gCDW/h)*	*non-native in E. coli*		*native in E. coli*
									engineered	*not engineered*	
Calvin cycle	0.083	0.030	4.61	238	0.73	0.86	1.3%	0.5%	2	1	10
rTCA cycle	0.209	0.071	5.66	124	0.82	1.97	1.5%	0.5%	1	1	6
3HP-4HB cycle	0.083	0.030	9.93	227	0.96	0.51	2.3%	0.8%	6	2	5
DC-4HB cycle	0.110	0.039	9.27	203	0.95	0.54	2.8%	1.0%	3	1	10
PyrS-PyrC-Glx bicycle	0.164	0.058	7.11	144	0.89	1.22	1.9%	0,6%	4	0	6
PyrS-PEPC-Glx bicycle	0.130	0.047	7.12	175	0.89	1.10	1.6%	0.6%	3	0	8

Results and discussion

FBA and FVA of photo-electro-autotrophic designs

Predicted growth and fluxes depend on carbon fixation cycles

We used FBA to predict photo-electro-autotrophic growth rates for different carbon fixation cycles (**Table 2**) and to investigate which parts of the central metabolism are possibly used for this growth (**Figure 3**). Different carbon fixation cycles resulted in different growth rates, which could be related to their ATP requirements, as will be discussed later. Furthermore, we found by FBA that the added reactions for a certain carbon fixation cycle were not always necessarily taken for an optimal growth solution. When we simulated the 3HP-4HB cycle by FBA it only utilized a part of the cycle, i.e. from succinyl-CoA to acetyl-CoA. The subsequent conversion from acetyl-CoA to succinyl-CoA was replaced by a more ATP efficient route involving some of the other core model reactions of *E. coli* and the added pyruvate synthase. The cycle taken, turned out to be an alternative natural cycle, the DC-4HB cycle, found in anaerobic archaea (Huber et al., 2008), which we then included for further analyses (**Supplementary figure S4**).

Additionally, we analyzed the complete, 3HP-4HB cycle as well, by knocking out succinyl-CoA ligase reaction and thus forcing FBA to use the 3HP-4HB route including pyruvate synthase, which resulted in a lower predicted growth rate than for the DC-4HB cycle. Furthermore, we observed that the FBA carbon fixation simulations resulted in formation of some products, e.g. acetate or formate. These by-products were formed to regenerate some of the ATP used for the carbon fixation, either directly by substrate level phosphorylation or indirectly by generating additional proton motive force via proton symport in the by-product efflux reaction. The by-product formation was driven by additional uptake of electrons, since electron donors were needed to produce those by-products from CO_2. This led to additional flux for electron uptake, other than for biomass production. In an industrial setting, where photo-electro-autotrophic *E. coli* strains would possibly produce a specific product, separating undesired by-products complicates downstream purification of desired products, reducing the economic advantage of the whole process. Hence, we excluded certain reactions to block the generation of all by-

products (Table S1). Whereas this can be easily done in the metabolic model, in an experimental set-up this may require several gene knockouts.

On top of FBA, FVA was performed to analyze if the fluxes and flux distributions found by FBA could vary for the same optimal growth solutions. FVA showed that there is only very little variability in flux distributions to achieve the optimal growth rate given the constraints. Only some futile cycles and side-reactions showed some variability (Table S2). The overall flux distribution did not vary and, not unexpectedly, the unconstrained uptake rates of CO_2 and electrons did not vary. The FVA demonstrated that all relevant fluxes which we further apply to estimate ATP consumption and protein burdens gave no variation ranges for the optimal growth scenario, hereby we justified taking the fluxes resulting from FBA for further analysis.

ATP regeneration limits photo-electro-autotrophic growth

We found by FBA that PPR proton-pumping flux and, consequently, ATP regeneration severely limited autotrophic growth rates. We varied the PPR flux between 0 and 50 mmol H^+/gCDW/h; autotrophic growth was predicted for all carbon fixation pathways only above a threshold PPR flux of 34.0 mmol H^+/gCDW/h (**Supplementary figure S5**). At 50 mmol H^+/gCDW/h growths rates varied from 0.030 h^{-1} (Calvin cycle and 3HP-4HB cycle) to 0.071 h^{-1} (rTCA cycle). This PPR flux threshold is caused by the fact that the non-growth associated ATP requirement for maintenance (NGAM) needs to be overcome before ATP is available for growth. *E. coli* normally has a NGAM flux of 8.39 mmol ATP/gCDW/h. This NGAM value is based on standard heterotrophic growth conditions and includes ATP consumption for other processes than anabolic reactions, e.g. membrane leakage (Feist et al., 2007).

The NGAM value is normally determined in chemostat experiments. However, for a completely new growth phenotype of an engineered, and potentially laboratory evolved, autotrophic *E. coli,* a reliable estimate for the maintenance term is not available and it may turn out to be higher or lower (Bar-Even et al., 2013). The burden of heterologous expression of autotrophic components may increase the NGAM, which would make the designed photo-electro-autotrophic designs infeasible. Therefore proper estimates of the ATP burden of heterologous expression (Ceroni et al., 2015), of for example for autotrophy

pathways, and methods to minimize this burden are of key importance. On the other hand, nature demonstrates that lower NGAM values than the standard *E. coli* value are feasible, as for several organisms lower NGAM values have been determined. For example, for the strictly anearobic *Shewanella oneidensis* and *Geobacter metallireducens*, which both can transfer external electrons directly, NGAM values were determined to be 1.03 and 0.81 mmol ATP/gCDW/h, respectively (Feist et al., 2014; Pinchuk et al., 2010). The anoxygenic photosynthetic *Rhodobacter sphaeroides* has a NGAM of 8.4 ATP/gCDW/h during photoheterotrophic growth at high light intensities, but during light-limited photoheterotrophic growth this can decrease down to 0.5 ATP/gCDW/h (Imam et al., 2015).

Based on this naturally feasible lower range of NGAM values, we repeated our FBA simulations with a lower NGAM of 1.00 mmol ATP/gCWD/h. With the low NGAM, the PPR flux threshold above which growth is feasible decreased to 4.0 mmol H^+/gCDW/h (**Supplementary figure S6**). Consequently, the maximum growth rates at a PPR flux of 50 mmol H^+/gCDW/h increased substantially compared to the high NGAM scenario (Table 2). These maximum growth rates varied from 0.083 h^{-1} (Calvin cycle and 3HP-4HB cycle) to 0.209 h^{-1} (rTCA cycle). However, when we considered a somewhat more realistic proton flux for the PPRs of 20.0 mmol H^+/gCDW/h, for the ATP-efficient synthetic PyrS-PyrC-Glx bicycle, a growth rate of 0.057 h^{-1} seemed feasible. Such a growth rate would be a high growth rate for autotrophs, considering that most natural autotrophic organisms have growth rates below 0.1 h^{-1}. For instance, the photosynthetic model cyanobacterium *Synechocystis* PCC6803 has a maximum photoautotrophic growth rate of 0.085 h^{-1} (Boyle and Morgan, 2011; Shastri and Morgan, 2005). However, a photo-electro-autotrophic *E. coli* may only achieve growth rates in this range if several criteria are met: a low NGAM, a high PPR flux and an ATP-efficient CO_2 fixation cycle.

As PPR proton export was generally limiting growth, we analyzed proton-pumping reactions in the FBA in more detail. The PPR mainly created the proton efflux from the cytoplasm. Furthermore, the pathways that included fumarate reductase (rTCA, DC-4HB, PyrS-PyrC-Glx and PyrS-PEPC-Glx) had an additional efflux of protons from ubiquinol recovery (**Supplementary figure S7**). For most organisms harboring the rTCA, the natural electron donor for

fumarate reductase is not known, though in *Hydrogenobacter* it is reported to be NADH (Berg, 2011). Alternatively, in *E. coli* and consequently in our analyses, a native ubiquinol-dependent fumarate reductase could be employed for the rTCA cycle, DC-HB cycle and the synthetic bicycles. The reduction of resulting ubiquinone via an NADH dehydrogenase created an additional export of protons. The majority of available proton motive force resulting from proton efflux, was imported by the ATP synthase to regenerate ATP. The influx through the ATP synthase was even larger than the total efflux of PPRs and ubiquinol-dependent reductases. This gap could be explained by additional proton uptake from the external medium, which we could relate to the formation of water in the carbon fixation pathways (**Supplementary figure S7**).

In addition to the PPR proton export flux we analyzed external electron uptake flux and CO_2 uptake flux for all scenarios. The maximum electron uptake varied from 14.4 mmol e^-/gCDW/h for the cycles with the slowest predicted growth (3HP-4HB and Calvin cycle) to 35.7 mmol e^-/gCDW/h for the fastest predicted growth (rTCA cycle) (**Supplementary figure S8**). Accordingly, the CO_2 uptake ranged from 3.5 mmol CO_2/gCDW/h (both, 3HP-4HB and Calvin cycle)) to 8.7 mmol CO_2/gCDW/h (rTCA cycle) (**Supplementary figure S8**). The feasibility of all these fluxes is related to the kinetic activities of related enzymes and the resulting protein burdens, which is discussed in later. Summarizing the electron and CO_2 uptake flux practically limit photo-electro-autotrophic growth to a lesser extent than the PPR proton-pumping flux.

ATP consumption of carbon fixation pathways

By FBA we found that the PPR flux, and consequently the ATP production, was the limiting factor for the photo-electro-autotrophic growth of *E. coli*. In addition, we found considerably varying growth rates for the different carbon fixation pathways. Therefore we analyzed in more detail the ATP consumption per mole biomass formed for every carbon fixation pathway, given the reactions toward biomass precursors that carry flux according to FBA. The FVA showed that there is no variability in the reactions toward biomass precursors that carry flux, which justifies using the results from FBA as a basis for this ATP calculation.

The ATP consumption per mole biomass consists of the ATP needed for the biosynthesis of biomass precursors and the Growth Associated Maintenance

(GAM). For *E. coli* this GAM is 59.8 mmol ATP/gCDW, independent of the carbon fixation pathway utilized, whereas the ATP requirement for the biomass precursors is different for each carbon fixation pathway. Therefore, the total ATP requirement varied from 124 mmol ATP/gCDW (rTCA cycle) to 238 mmol ATP/gCDW (Calvin cycle) (**Table 2, Supplementary figure S9**). The ATP requirement per mole biomass is negatively correlated with the growth rates predicted by the FBA for the pathways (**Figure 4**). Roughly the rTCA cycle and the PyrS-PyrC-Glx bicycle had the lowest ATP requirements and gave clearly higher growth rates in the FBA than the other cycles. The 3HP-4HB cycle had a growth rate similar to the Calvin cycle, even though the ATP requirement is lower for the 3HP-4HB cycle (227 mmol ATP/gCDW). This is related to the fact that the 3HP-4HB cycle generates NADH (**Supplementary figure S3**) which needs to be regenerated to NADPH by consuming some of the exported protons (**Figure S7**).

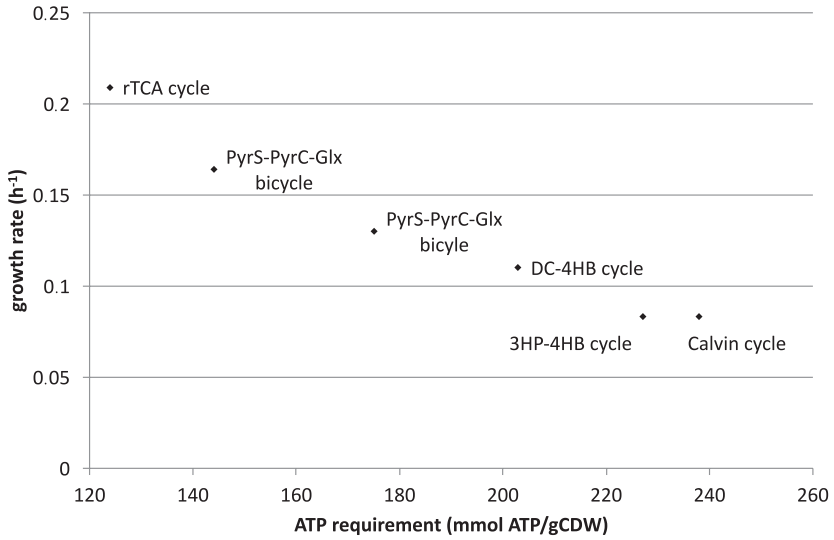


Figure 4. Photo-electro-autotrophic growth rates in relation to the ATP requirement for biomass. Growth rates were based on a PPR flux of 50.0 mmol H^+/gCDW/h and a NGAM of 1.00 mmol ATP/gCDW/h.

Thermodynamics of carbon fixation under different conditions

We determined MDF values for all carbon fixation pathways to determine their thermodynamical favorability. A negative MDF means the pathway is not feasible, and a higher MDF value indicates that a pathway is thermodynamically

more favorable (Noor et al., 2014). Our analysis showed that under standard conditions all pathways were feasible. The natural 3-HP-4HB cycle and DC-4HB cycle were thermodynamically most favorable. Also both synthetic bicycles were very favorable, the Calvin cycle and rTCA cycle were slightly less favorable (**Table 2**).

To check how sensitive the thermodynamics were if the conditions differed from the defined standard conditions we varied the pH, as well as important co-factor concentrations and ratios. From this, we observed that for a pH range between 5.0 and 9.0 all pathways were thermodynamically feasible (**Supplementary figure S10**). The DC-4HB and 3HP-4HB cycles had a very low MDF for pH values close to 5.0, however, this pH is already far off from the physiological range of *E. coli* (pH ~7.5). So thermodynamically these pathways seemed all quite robust for pH deviations around the *E. coli* optimum.

With respect to the co-factors, the sensitivity analysis showed that the ferredoxin-utilizing pathways were thermodynamically less favorable (MDF < 4.0 kJ/mol) when the $ferredoxin_{red}$/$ferredoxin_{ox}$ ratio got very low, i.e. 0.5 (**Supplementary figure S11**), in particular for the rTCA cycle. This would practically not be a problem if there is an electron uptake mechanism and sufficient external electron donors, thereby maintaining a sufficiently high ratio. However, this low-potential redox couple will require systems, such as flavin-based bifurcating enzymes to allow for the thermodynamically less favorable reduction of $ferredoxin_{ox.}$.

Low NADPH/$NADPH^+$ ratios had only a limited effect on thermodynamic favorability, however again the rTCA had the strongest decrease in MDF values when this cofactor ratio dropped (**Supplementary figure S12**). A low NADH/NAD^+ ratio had only a large impact of thermodynamic favorability of the Calvin cycle, the sole electron donor in this cycle, at the typical physiological ratio of 0.1 or lower the MDF dropped below 4 kJ/mol (**Supplementary figure S13**). It has to be noted that in natural organisms harboring the Calvin cycle, instead of NADH, NADPH is employed as the electron donor. However in our analysis we primarily included NADH as electron donor, as *E. coli* natively harbors a reversible glyceraldehyde-3-phosphate dehydrogenase that uses NADH. Given the much higher typical ratio of NADPH:NADP (10), substituting this enzyme

by a heterologous variant of using NADPH as electron donor would result in a higher MDF value of around 6 kJ/mol (**Supplementary figure S13**), as was found in the sensitivity analysis for the physiologically unrealistic NADH:NAD ratio of 10. So to improve thermodynamic favorability of the Calvin cycle the electron donor specificity of glyceraldehyde-3phosphate dehydrogenase is crucial.

The ATP/ADP ratio only had a small impact on thermodynamic favorability of all pathways (**Supplementary figure S14**). Another thermodynamically critical factor could be the dissolved CO_2 concentration range. However, even at a dissolved CO_2 concentration of 10 µM all cycles were still feasible, which would be the dissolved concentration range resulting from an equilibrium with ambient air CO_2 (**Supplementary figure S15**). The Calvin cycle was thermodynamically not sensitive to dissolved CO_2 concentration changes; for all other cycles the MDF dropped to lower, but still positive values for lower dissolved CO_2. Especially the rTCA got a relatively low thermodynamic favorability at the lowest CO_2 concentration. HCO_3^- concentrations did not impact the thermodynamic bottleneck reactions for any of the cycles (**Supplementary figure S16**).

Summarizing, the rTCA cycle and Calvin cycle were generally slightly less thermodynamically favorable than the other analyzed natural and synthetic cycles. Especially at less favorable ratios of electron donors these cycles had lower thermodynamic favorability for their bottleneck reactions, which may result in lower net fluxes through these reactions.

Kinetics of carbon fixation pathways

For analyzing the kinetics of the carbon fixation pathways we calculated both the FFE and PSA (**Table 2**). The 3-HP-4HB cycle, DC-4HB cycle and the synthetic bicycles all had a FFE of 0.89 or higher, which is related to their higher thermodynamic favorability as discussed above. The rTCA cycle and Calvin cycle therefore had a slightly lower FFE value (**Table 2**), however still the majority of the flux of the thermodynamic bottleneck reactions was in the forward direction. The highest PSA value was found for the rTCA cycle (1.97 µmol pyruvate/mg pathway protein/min), partly because this cycle involves the lowest number of enzymes. Both synthetic bicycles have the also relatively

high with PSA values around 1.1 µmol pyruvate/mg pathway protein/min. The other three natural cycles had the slowest kinetics calculated by PSA. In contrast to an earlier PSA analysis (Bar-Even et al., 2010), now all PSA values could be compared among the different pathways. As more specific activities for the ferredoxin:oxidoreductase enzymes were available now and included in our analysis (Schomburg et al., 2013). In summary, the synthetic PyrS-PyrC-Glx, PyrS-PEPC-Glx bicycles presented the best combination of FFE and PSA scores, whereas the rTCA cycle had a notably high PSA but a somewhat lower FFE due to its lower thermodynamic favorability. This makes the synthetic bicycles most promising from a pathway kinetics point of view.

Protein burden of photo-electro-autotrophic systems

Carbon fixation pathways

To determine the protein burden of each carbon fixation pathway, we calculated the Pathway Protein Burden (PP) (**Table 2**). For the higher NGAM scenario lower protein burdens for carbon fixation were found than for the low NGAM scenario; as for a high NGAM scenario lower growth rates were predicted, consequently lower carbon fixation fluxes and related pathway proteins were required. By assuming a very low NGAM and a high PPR flux, the more promising growth rates were found, in this scenario the smallest protein burden was found for the Calvin cycle (1.3%). This is surprising, as the carboxylase of the Calvin cycle, RubisCO (ribulose-bisphosphate carboxylase-oxygenase) is a notoriously slow catalyst, which therefore is expressed highly in many natural organisms(Bar-Even et al., 2010). This kinetic inefficiency of RubisCO is also reflected in the relatively high PSA for the Calvin cycle, however given the low growth rate feasible for this pathway, the overall protein burden for CO_2 fixation is still the lowest. Other than for the Calvin cycle, low burdens were found for the cycles that resulted in fastest growth,: rTCA cycle (1.5%), the PyrS-PEPC-Glx bicycle (1.6%) and the PyrS-PyrC-Glx bicycle (1.9%). For these cycles the extra pathway protein requirement to sustain their higher CO_2 fixation flux was compensated by their high PSA values. The other nature cycles had a higher burden up to 2.8% for the DC-4HB cycle.

Under the assumption that the total proteome of cells is about 50% of the cell dry weight (Bar-Even et al., 2013), the burden of the carbon fixation pathways

with a PPB of ~2% of the cell dry weight resulted in a proteome burden of 4% of total cell protein weight. A fraction of 4% heterologous protein expression is well below the 20% considered to be a limit beyond which cell viability is critically affected (Bar-Even et al., 2013). However, for a photo-electro-autotrophic *E. coli* additional heterologous proteins need to be expressed for PPRs and electron uptake enzymes, as will be discussed below.

Proton-pumping rhodopsin systems

The expression of proton-pumping rhodopsin systems (PPRs) imposes a significant additional protein burden to the cell. For the PPRs based on literature, a proton-pumping activity of 300 mmol H^+/mmol PPR/min may be feasible (Claassens et al., 2013; Ganapathy et al., 2015). Assuming a molecular weight for proteorhodopsin of 29 kDa (Stone et al., 2013), and hence a specific activity of 10.3 µmol H^+/mg protein/min, this would give a substantial protein burden of 8.1% at the maximum pumping rate (50 mmol H^+/gCDW/h) used in FBA. Assuming this specific activity and $3*10^{-13}$ gCDW/cell, one cell would require to have ~500,000 PPRs/cell, while so far only levels up to 100,000/cell have been measured (Claassens et al., 2013). A protein burden of 8.1% per CDW corresponds to ~ 16% of the total proteome, and as the cytoplasmic membrane contains about 20-30% of the total proteome (Bernsel and Daley, 2009),so this is a very significant burden to the membrane proteome. So to achieve a high growth by PPRs, either the pumping rates or the number of PPRs/cell has to be significantly increased. Increased membrane surface area may be required to accommodate a higher number of PPRs, analogous to the extra thylakoid membranes in which cyanobacteria accommodate their photosystems.

Electron uptake mechanisms

For estimating the protein burdens of electron uptake mechanisms specific activities of the relevant enzymes are needed. With formate as an electron donor, formate dehydrogenase is needed, for this enzyme some specific activity values are available (Tishkov and Popov, 2006). For the aerobic, engineered NADPH-dependent formate dehydrogenases the highest specific activity reported is 2.5 µmol formate/mg protein/min (Tishkov and Popov, 2006). At the maximum electron uptake predicted by FBA this would give an infeasibly high protein burden of 23.8% of cell dry weight. However, for simple aerobic NADH-dependent formate dehydrogenases values up to 10 µmol

formate/mg protein/min have been reported, reducing their burden fourfold to 6.0% of cell dry weight. This would still result in a significant burden on the proteome of ~12%. However, for the anaerobic scenario oxygen-sensitive formate dehydrogenases may even allow for higher specific activities and lower burdens(Liu and Mortenson, 1984), but not many reliable values are available yet.

For using hydrogen as an electron donor, hydrogenases are required that are $NADP^+$-dependent, ferredoxin-dependent and/or bifurcating. For those no reliable specific activities with physiological electron acceptors were measured yet. Therefore, the protein burden for hydrogen uptake systems, including flavin-based bifurcating enzymes, could not be estimated. Also for direct electron uptake mechanisms also very limited kinetic data are known, however before it was assumed that an electron uptake flux of 30.0 mmol e^-/gCDW/h is feasible in engineered *E. coli* (Pandit and Mahadevan, 2011), this is close to the maximum flux of 35.7 mmol e^-/gCDW/h that is required for our highest growth scenario (low NGAM with rTCA cycle).

The estimates here for protein burdens are all based on kinetic data from *in vitro* assays, which not necessarily correspond directly to *in vivo* conditions, but it was recently demonstrated that *in vitro* activities generally concur with *in vivo* activities (Davidi et al., 2016). Based on burden estimates, for fast photo-electro-autotrophic growth assuming a high proton-pumping rate and low NGAM a total burden on the proteome of at least 20% is estimated to be put on the organism, consisting of a burden of PPRs (~16%) and a burden of the CO_2 fixation cycle (e.g. ~4% for synthetic bicycles and rTCA). This still excludes the burden for electron uptake, which may be high as well, but is hard to estimate given limited data. To reduce the added burden of this electron uptake, kinetically fast uptake enzymes have to be found or engineered. In comparison, in the photosynthetic cyanobacterium *Synechocystis* sp. PCC6803, 26% of the proteome mass is devoted to energy metabolism (Liebermeister et al., 2014; Wegener et al., 2010). The majority is also for the photosystem (18%) and a smaller part for carbon fixation (2.9%). So generally the photosystems seem to give a relative large contribution to the proteome of a photoautotroph.

Challenges for experimental implementation of photo-electro-autotrophic systems

The analysis of the presented designs suggests that photo-electro-autotrophic growth of *E. coli* has a limited feasibility. In general, ATP limited the growth rate and the available models cannot predict what the effective ATP maintenance will be for an engineered photo-electro-autotrophic *E. coli*. Better *in silico* prediction of NGAM and strategies for minimizing the maintenance burden in engineered cells requires a lot of attention.

For experimentally implementing anaerobic photo-electro-autotrophy the limitation of growth by ATP, related to the PPR proton export flux is a key challenge. It is therefore crucial to find ways to increase the kinetic rates of the proton-pumping rhodopsin photosystems and/or to reach higher numbers of PPRs per cell without surpassing heterologous expression limits in the membrane. In a previous bioenergetic analysis of PPRs in natural photoheterotrophs it was also shown that PPRs generate a relatively low amount of metabolic energy at relatively high protein costs (Kirchman and Hanson, 2013), also confirmed experimentally by only very small growth advantages measured for microorganisms heterologously expressing rhodopsins (Chen et al., 2016; Wang et al., 2015). Therefore, research efforts could be aim to increase their energetic efficiency, e.g. by broadening the PPR light absorption and/or proton-pumping kinetics, e.g. by applying PPRs with an antennae pigment (Balashov et al., 2010; Claassens et al., 2013) or through protein engineering of PPRs (Engqvist et al., 2015). If PPRs turn out not to be able to generate sufficient proton pumping for ATP regeneration, it could be considered (first *in silico*) to heterologously express other, more complex (anaerobic) photosystems (Tikh and Schmidt-Dannert, 2013). One could potentially apply the photosystems found anoxygenic photosynthetic organisms, which are natural photo-electro-autotrophs. These microorganisms generate ATP by their anoxygenic photosystems and use sulfur, hydrogen sulfide or hydrogen as an electron source (Bryant and Frigaard, 2006). Alternatively, these anoxygenic photosynthetic microorganisms themselves may be employed as chassis for engineering photo-electro-autotrophic production. Some basic tools for genetic and metabolic engineering are available for some of these organisms, such as for the green bacterium *Chlorobaculum tepidum*(Azai et al., 2013) and

purple bacterium *Rhodobacter sphaeroides* (Jaschke et al., 2011; Tikh et al., 2014b).

Apart from applying more efficient photosystems for ATP regeneration, electron acceptors could be used to allow for oxidative phosphorylation, such as oxygen or nitrate. However, oxygen may inhibit certain electron uptake pathway enzymes and carbon fixation pathway enzymes, but there are indications that for some of the oxygen-sensitive enzymes there are natural, oxygen-tolerant alternatives (Berg, 2011). This could facilitate a (micro)aerobic scenario for engineering electro-autotrophy, where instead of photophosphorylation by PPRs, oxidative phosphorylation is used for ATP production. In such a (micro)aerobic scenario, the favorable and ATP-efficient synthetic carbon fixation pathways can still be advantageous compared to other less ATP-efficient (aerobic) pathways for potentially high product and biomass yields. However, this scenario is only advantageous if the required oxygen-protecting mechanisms of oxygen-tolerant enzymes are not too energy-demanding. Alternatively, for such (micro)aerobic conditions other synthetic non-ferredoxin-dependent, ATP-efficient pathways could be promising and should be analyzed (Bar-Even et al., 2010).

This work also has shown an extensive integrated analysis of several CO_2 fixation pathways. These results may guide choices for engineering CO_2 fixations cycles in autotrophs in general. Given their relatively low ATP costs, especially the rTCA cycle and synthetic PyrS-PyrC-Glx seem to be attractive, resulting in the highest predicted growth for anaerobic photo-electro-autotrophy. In addition to their ATP efficiency, these pathways were found to have a high kinetic activity, resulting in a limited protein burden. The synthetic PyrS-PyrC-glx cycle, may despite its slightly higher ATP consumption and lower growth rates, still be advantageous compared to the rTCA if the less favorable thermodynamics of the rTCA cycle are taken into account.

Furthermore, experimental implementation of these carbon fixation pathways in *E. coli* requires functional expression of the pathway enzymes. Therefore, we reviewed native presence of all carbon fixation pathway enzymes in *E. coli*, and if absent, we checked literature for reports of successful heterologous

expression in *E. coli* (**Supplementary Text S3**). Accordingly, we determined the number of native *E. coli* enzymes and the number of already heterologously expressed, *in vivo* functionally enzymes (**Table 2**). The analysis shows that for the synthetic PyrS-PyrC-Glx bicycle only 4 non-native enzymes are needed. All of them have already been (individually) expressed heterologously in *E. coli* recently (Mainguet et al., 2013). For the rTCA cycle only 2 enzymes have to be heterologously expressed, however running the reverse TCA cycle in *E. coli* will probably also involve reregulation of the native oxidative TCA cycle. As the PyrS-PyrC-Glx bicycle and the rTCA cycle require reduced ferredoxin, attention has to be paid to achieve heterologous expression of e.g. recently discovered flavin-based bifurcating enzymes in *E. coli*, which allow for thermodynamically favorable ferredoxin reduction.

Also other hosts than *E. coli* may be considered for engineering (photo-electro-) autotrophy, as *E. coli* for example may appear to have high ATP maintenance costs. A more detailed analysis of the potential of (photo-electro-)autotrophic designs in other hosts, as performed in this work, can be performed if a core or genome-scale metabolic model is available for that host. Promising alternative chassis organisms could be acetogens or methanogens, some of which are genetically accessible (Köpke et al., 2010; Schiel-Bengelsdorf and Dürre, 2012). Acetogens and methanogens harbor the extremely ATP-efficient Wood-Ljungdahl carbon fixation pathway and they generally have uptake mechanisms for hydrogen. However, these organisms produce large amounts of acetate or methane to generate ATP by substrate level phosphorylation; these are mostly unwanted by-products for industrial applications. This by-product formation could potentially be avoided by introduction of PPRs for photophosphorylation into these organisms. This could be analyzed *in silico* using the approach described in this paper.

In summary, the herein described integrated *in silico* analysis is an in-depth exploration of the potential of photo-electro-autotrophy to indicate bottlenecks and steer strategic choices before experimental implementation. We envision that such an integration of tools will prove to be very useful in developing and analyzing designs for synthetic autotrophy and other challenging pathway designs in synthetic biology.

Author contributions

NJC, MV, JvdO, WMdV and VAPMds conceived this work, and the detailed study design was prepared by NC and MV. MV performed the modelling with some assistance and advice from NJC and EN. NJC drafted the manuscript, with support from MV. SKM was involved in advice regarding microbial physiological aspects. EN, SKM, JvdO, WMdV, and VAPMdS critically revised and approved the manuscript.

Supplementary information

Supplementary text S1.

Comparison of FBA results of core model versus genome-scale model

Here we compare the FBA results of the core model versus the full model. The *E. coli* core model is based on the *E. coli* genome-scale metabolic model AF1260 and consists only of the core metabolism of *E. coli* [31] (including glycolysis, pentose phosphate cycle, TCA cycle, glyoxylate shunt, respiration, etc.). To rule out a bias due to the use of the core model in our analysis, we included the reactions for the PPRs, the electron donor regeneration and the carbon fixation cycles in the genome-scale AF1260 model [33] and performed the FBA for the same constraints and modifications discussed in before. For the Calvin cycle and the 3HP-4HB cycle, it was necessary to additionally allow for succinate transport out of the cell. This byproduct formation is needed to have feasible growth conditions in these two scenarios for the genome-scale model, and not for the core model. Even though the resulting growth rates differ by up to 15%, the ranking of the different carbon fixation cycles does not change (**Table A**). Remarkably, the three better performing pathways, that is the rTCA cycle, the PyrS-PyrC-GLX and the PyrS-PEPC-GLX bicycle, show a lower growth rate in the genome-scale model compared to the core model, whereas the growth rates of the other three pathways increase when using the genome-scale model. This is due to the different composition of the biomass objective function. However, the rTCA cycle and the two synthetic bicycles still show a higher growth rate and combined with the results from the thermodynamic and pathway kinetics analysis, which do not depend on the choice of the model, the conclusions based on the core model are still valid. We believe, that the advantages of the core model, that is the better accessibility and the more general conclusions, overpower the more realistic but very specific representation when using the genome-scale metabolic model.

Table A. Growth rates resulting from the FBA for the *E. coli* core model and the genome-scalemodel of *E. coli* for low and high non-growth associated maintenance.

carbon fixation pathway	*growth rate (hr^{-1})*			
	low maintenance		*high maintenance*	
	(1.00 mmol ATP/gCDW/h)		*(8.39 mmol ATP/gCDW/h)*	
	Core model	*Genome-scale model*	*Core model*	*Genome-scale model*
Calvin cycle	0.083	0.090	0.030	0.032
rTCA cycle	0.209	0.173	0.071	0.062
PyrS-PyrC-Glx bicycle	0.164	0.149	0.058	0.053
PyrS-PEPC-Glx bicycle	0.130	0.115	0.047	0.041
3HP-4HB cycle	0.083	0.095	0.030	0.034
PyrS-PEPC-4HB cycle	0.110	0.113	0.039	0.041

Supplementary text S2.

Update Max-min driving force algorithm

Before a thermodynamic-based pathway analysis technique was developed, denoted Maxmin Driving Force (MDF), which identifies those reactions within a pathway whose rates are constrained by low thermodynamic driving force (Noor et al., 2014). One of the sources of data required for performing MDF analysis, is the standard Gibbs free energies of each one of the reactions, in the specified pH and ionic strength levels (ΔrG°). In this work and earlier work (Noor et al., 2014), we obtained the Gibbs energy data from the Component Contribution method (Noor et al., 2013), which estimates the values for the vast majority of enzyme-catalyzed reactions. This method is based on multivariate linear regression and provides both a mean and standard error of each estimated value. In the original implementation of the MDF algorithm (Noor et al., 2014), we ignored the standard errors. In this and other work (Gerosa et al., 2015), we improved the implementation by relaxing the thermodynamic constraints in the MDF linear program with uncertainty slack variables, which reflect the standard error of the Component Contribution estimates. This new implementation is now available freely on GitHub at https://github.com/eladnoor/component-contribution, under "scripts/max min driving force.py".

Sampling from a Multivariate Gaussian

Consider a D-dimensional random variable with a Gaussian distribution of mean μ and covariance Σ. If we have a 1-dimensional random Gaussian sampler, we

can sample from the multivariate distribution by sampling D times and then stretching and rotating the vector according to Σ. Specifically, we define the square root of the covariance matrix as

$$\sqrt{\Sigma} = U \cdot \sqrt{\Sigma} \cdot U^{T} \quad (1)$$

where *S* is a diagonal real matrix and *U* is unitary which are given by the Singular Value Decomposition (SVD) of the covariance matrix, i.e. Σ = U · S · U > (note that Σ is Hermitian and thus diagonalizable with real eigenvalues).

If $\forall i$: yi ∼ N(0, 1) and we define $z \equiv \mu + y \cdot \sqrt{\Sigma}$ then

$$z \sim N(\mu, \Sigma) \quad (2)$$

Linear Programming with Multivariate Gaussian

The same approach can be applied for setting hard linear constraints on a the original variable in the context of linear programming. We define the auxiliary variable $y \in [-1, 1]^D$ replace the random variable by the expression $\mu + K \cdot y \cdot \sqrt{\Sigma}$, where K is a parameter of how loose we want the constraints to be (typically, we use the value 1). This approach is easily applied to linear problems that utilize thermodynamic constraints, and use the standard Gibbs energies provided by Component Contribution. The vector of $\Delta_r G°$ for a given problem should be constrained to:

$$\Delta_r G^0 = \Delta_r G^0{}_{cc,\bar{X}} \cdot \sqrt{\Sigma}_{cc,\bar{X}} \quad (3)$$

Supplementary text S3. Presence and heterologous expression of carbon fixation enzymes in *E. coli.*

Here we review native presence and successful expression of all required enzymes for the carbon fixation pathways in *E. coli.*

For the Calvin cycle two required non-native enzymes have already been functionally expressed *in vivo* in *E. coli*. Several groups successfully expressed the enzymes ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and phosphoribulokinase (PrkA), mostly originating from *Synechococcus*

species in *E. coli* (Gong et al., 2015; Greene et al., 2007; Mueller-Cajar and Whitney, 2008; Mueller-Cajar et al., 2007; Parikh et al., 2006; Zhuang and Li, 2013). Probably a third enzyme of the Calvin cycle is not present in *E. coli* natively, i.e. sedoheptulose-1,7-bisphosphate aldolase. However, the native *E. coli* fructose-1,6-bisphosphatase may also have activity as an sedoheptulose-1,7-bisphosphate aldolase. For the native *E. coli* fructose-1,6-bisphosphatase aldolase it was shown that it also has sedoheptulose-1,7-bisphosphate aldolase activity (Nakahigashi et al., 2009). The enzyme with both fructose-1,6-bisphosphatate aldolase and sedoheptulose-1,7-bisphosphate aldolase activity from *Bacillus methanolicus* has been heterologously expressed in *E. coli* and activity was shown *in vitro*, but not yet *in vivo* (Stolzenberger et al., 2013).

The rTCA cycle is for a large part already present in *E. coli*, as it has the oxidative citric acid cycle natively, which is basically the reverse of the rTCA cycle. However, for two irreversible steps of the citric acid cycle, heterologously expressed enzymes are needed. These two enzymes are a 2-oxoglutarate:ferredoxin oxidoreductase synthase (OGOR) and ATP-citrate lyase (ACL). Recently, ACL was functionally expressed in *E. coli* by overexpressing AclBA from *Chlorobium tepidum* (Mainguet et al., 2013). Two OGOR enzymes from the thermophilic bacterium *Hydrogenobacter thermophilus* were expressed in *E. coli* and then purified and shown to be functional (Yamamoto et al., 2003). However, the activity was only demonstrated *in vitro* at 70 °C. So, *in vivo* functional expression of a mesophilic OGOR in *E. coli* has not been demonstrated so far. However, many mesophilic rTCA-harboring autotrophs exist, which can offer potential OGOR candidate genes for expression in *E. coli* (Berg, 2011).

Of the enzymes in the 3HP-4HB cycle, a large number is not present in *E. coli* natively. In the 3HP-4HB pathway malonyl-CoA is converted to R-methylmalonyl-CoA by four different enzymes. All these four enzymes have been functionally expressed in *E. coli* in an attempt to engineer the 3-hydroxypropionate bicycle, another natural carbon fixation pathway (Mattozzi et al., 2013). For the conversion of succinyl-CoA to 4-hydroxybutyrate another two enzymes need to be expressed heterologously. Both these enzymes were expressed successfully for 1,4-butanediol production in *E. coli* (Yim et al., 2011). The next step in the pathway is the ATP-dependent conversion of 4-hydroxybutyrate to hydroxybutyryl-CoA; the enzyme catalyzing this conversion, 4-hydroxybutyryl-

CoA synthetase, has not been expressed in *E. coli* so far. However, an alternative that has been used successfully in *E. coli* is 4-hydroxybutyryl-CoA transferase (Yim et al., 2011). This reaction, however, is coupled to the conversion of acetyl-CoA to acetate instead of the phosphorylation by ATP. The generated acetate from the transferase could be recycled again to acetyl-CoA by native *E. coli* acetyl-CoA synthetase, which results in the same net ATP costs. Conversion of 4-hydroxobutyryl-CoA to crotonyl-CoA by 4-hydroxybutyryl-CoA dehydratase has, to our knowledge, not been achieved yet *in vivo* in *E. coli*, but the functional mesophilic enzyme from *Clostridium aminobutyricum* was purified after expression in *E. coli* (Gerhardt et al., 2000). The last three enzymes required for the conversion of crotonyl-CoA into acetyl-CoA are all natively present in *E. coli*, but also have been heterologously expressed from *Clostridium acetobutylicum* for butanol production (Inui et al., 2008). In total, engineering the 3HP-4HB cycle would involve heterologous expression of eight enzymes in *E. coli*, the highest number for all analyzed pathways. Engineering of the complete 3HP-4HB cycle has to our knowledge not been tried yet in *E. coli*, however functional engineering of a small part of this cycle was already demonstrated for the thermophilic heterotroph *P. furiosus* (Keller et al., 2013). For the synthetic 3HP-4HB variant, the PyrS-PEPC-4HB cycle, only 4 enzymes need to be introduced into *E. coli* for the conversion of succinyl-CoA to crotonyl-CoA.

For the synthetic PyrS-PEPC-Glx bicycle eleven enzymes are needed, of which three are non-native in *E. coli*. These three enzymes are all part of the so-called reverse glyoxylate shunt: malatethiokinase, malyl-CoA lyase and ACL, the latter is also part of the rTCA cycle. These three enzymes have been heterologously expressed together to form a functional reverse glyoxylate shunt in *E. coli* (Mainguet et al., 2013). For another enzyme in the PyrS-PEPC-Glx bicycle: pyruvate synthase (PyrS), it was for a long time not clear if it is natively present in *E. coli* and what gene codes for this enzyme. However, PyrS activity has been detected already some time ago in soluble extracts of *E. coli* cultured under aerobic and anaerobic conditions (Blaschkowski et al., 1982) and later *ydbK* was identified as a hypothetical gene coding for PyrS (Reed et al., 2003). This enzyme was identified as a pyruvate:flavodoxin oxidoreductase (Nakayama et al., 2013). Overexpression of this gene in *E. coli*, gave indirect proof that ydbK may also use ferredoxin instead of flavodoxin for the reverse PyrS reaction (Akhtar and Jones, 2009; Eremina et al., 2010). So probably this *ydbK* gene can be overexpressed for a functional PyrS-PEPC-Glx bicycle in *E. coli*. For the PyrS-PyrC-Glx bicycle the same heterologous enzymes as for the PyrS-PEPC-Glx variant are required.

In addition, this pathway requires heterologous expression of a PyrC enzyme in *E. coli*. Heterologous expression of PyrC for succinate production in *E. coli* has been demonstrated (Gokarn et al., 1998; Wang et al., 2006). So, for both synthetic pathways all required enzymes have already been expressed in *E. coli* before.

Supplementary Excel Table S1. Reactions and related data used for FBA, MDF and kinetic analysis.
Available online: doi:10.1371/journal.pone.0157851.s017

Supplementary Excel Table S2. Reactions, constraints and results of FBA and FVA.
Available online: doi:10.1371/journal.pone.0157851.s018

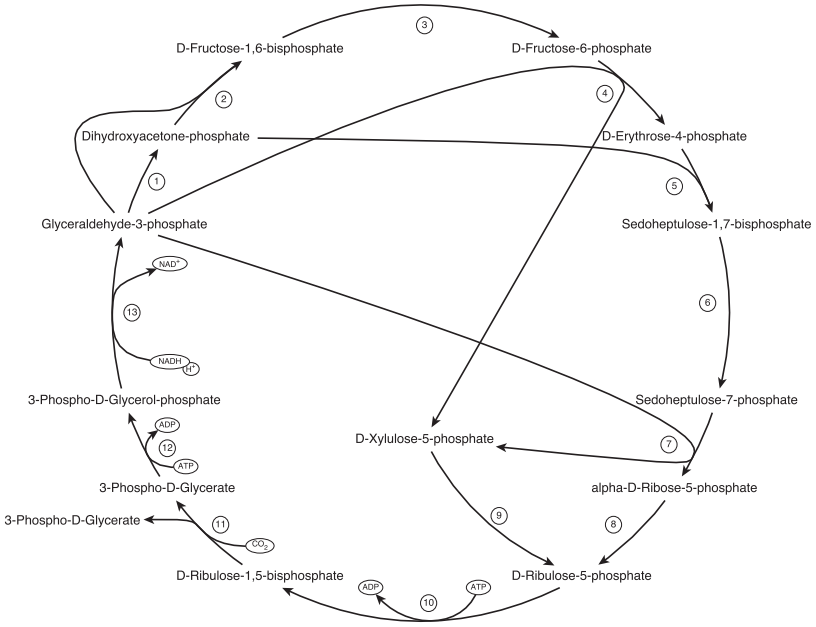


Figure S1. Calvin cycle. Reactions are: 1) Triose-phosphate isomerase, 2) Fructose-bisphosphate aldolase, 3) Fructose-bisphosphatase, 4) Transketolase, 5) Sedoheptulose-1,7-bisphosphate/fructose-1,7-bisphosphate aldolase, 6) Sedoheptulose-bisphosphatase, 7) Transketolase, 8) Ribose-5-phosphate isomerase, 9) Ribulose-phosphate 3-epimerase, 10) Phosphoribulokinase, 11) Ribulose-bisphosphate carboxylase, 12) Phosphoglycerate kinase and 13) Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating).

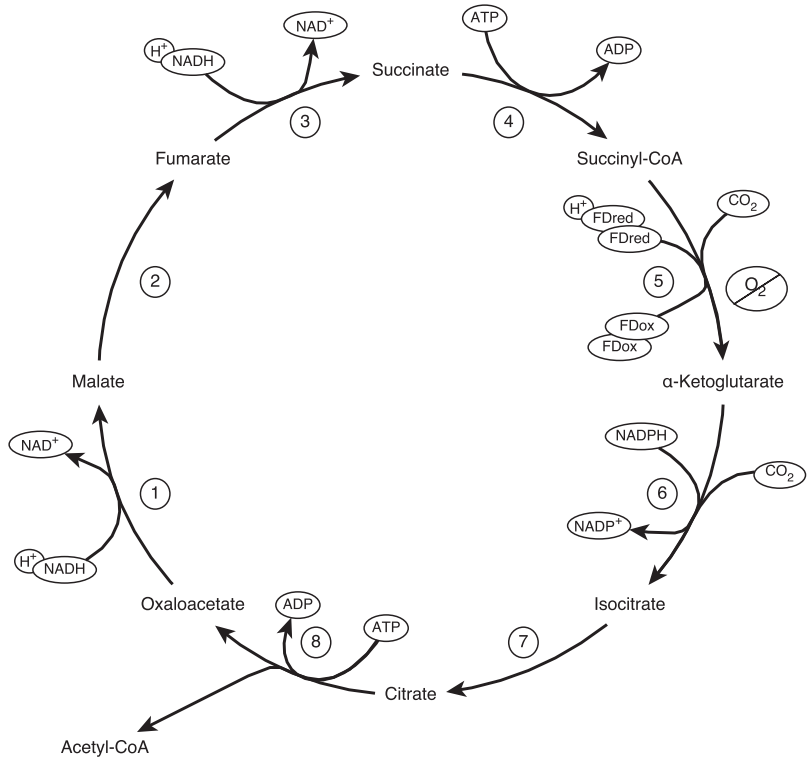


Figure S2. rTCA cycle. Reactions are: 1) Malate dehydrogenase, 2) Fumarate hydratase, 3) fumarate reductase, 4) Succinate-CoA ligase, 5) 2-oxoglutarate:ferredoxin-oxidoreductase synthase, 6) Isocitrate dehydrogenase, 7) Aconitate hydratase and 8) ATP citrate lyase.

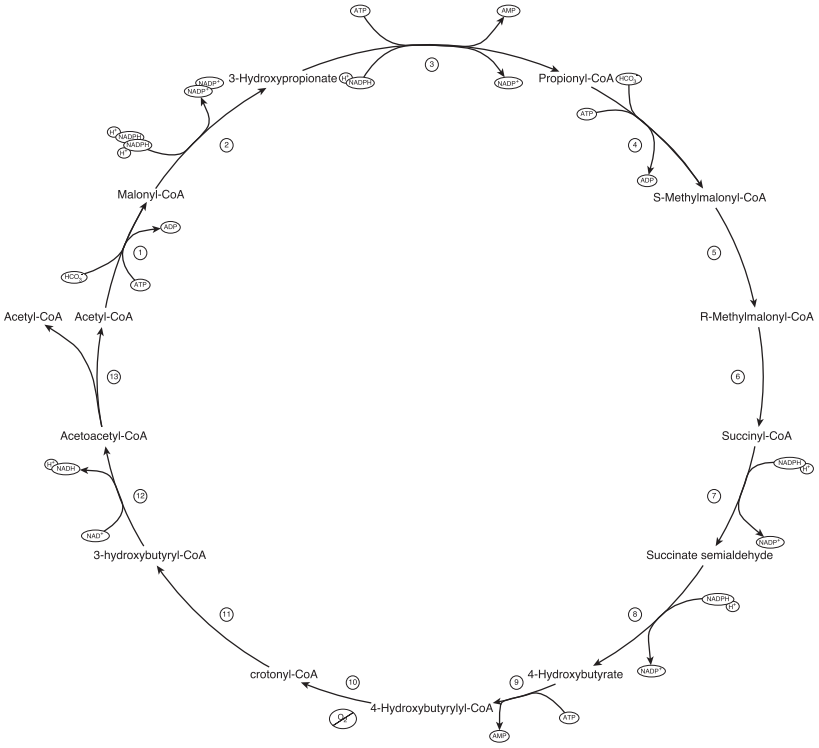


Figure S3. 3HP-4HB cycle. Reactions are: 1) Acetyl-CoA carboxylase, 2) Malonyl-CoA reductase (bifunctional enzyme), 3) Propionyl-CoA synthase (tri-functional enzyme), 4) Propionyl-CoA carboxylase, 5) Methylmalonyl-CoA epimerase, 6) Methylmalonyl-CoA mutase, 7) Succinyl-CoA reductase, 8) Succinate-semialdehyde reductase, 9) 4-hydroxybutyryl-CoA synthetase, 10) 4-hydroxybutyryl-CoA dehydratase, 11) 3-hydroxyacyl-CoA dehydratase (3-hydroxybutanoyl-CoA), 12) 3-hydroxyacyl-CoA dehydrogenase (acetoacetyl-CoA) and 13) Acetyl-CoA C-acetyltransferase.

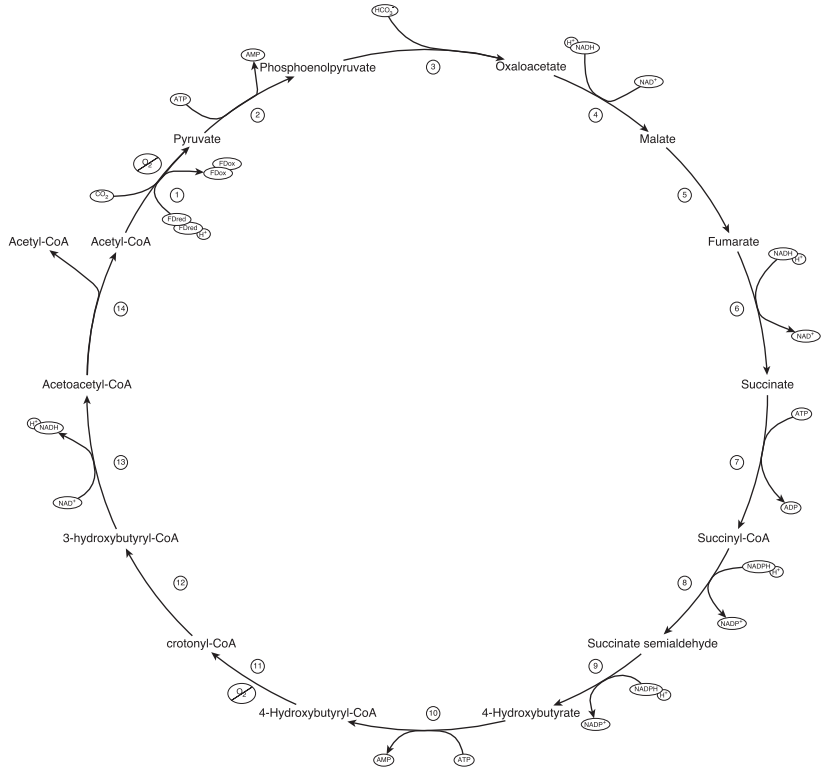


Figure S4. DC-4HB cycle. Reactions are: 1) Pyruvate synthase, 2) Pyruvate water dikinase, 3) Phosphoenolpyruvate carboxylase, 4) Malate dehydrogenase, 5) Fumarate hydratase, 6) fumarate reductase, 7) Succinate-CoA ligase, 8) Succinyl-CoA reductase, 9) Succinate-semialdehyde reductase, 10) 4-hydroxybutyryl-CoA synthetase, 11) 4-hydroxybutyryl-CoA dehydratase, 12) 3-hydroxyacyl-CoA dehydratase (3-hydroxybutanoyl-CoA), 13) 3-hydroxyacyl-CoA dehydrogenase (acetoacetyl-CoA) and 14) Acetyl-CoA C-acetyltransferase.

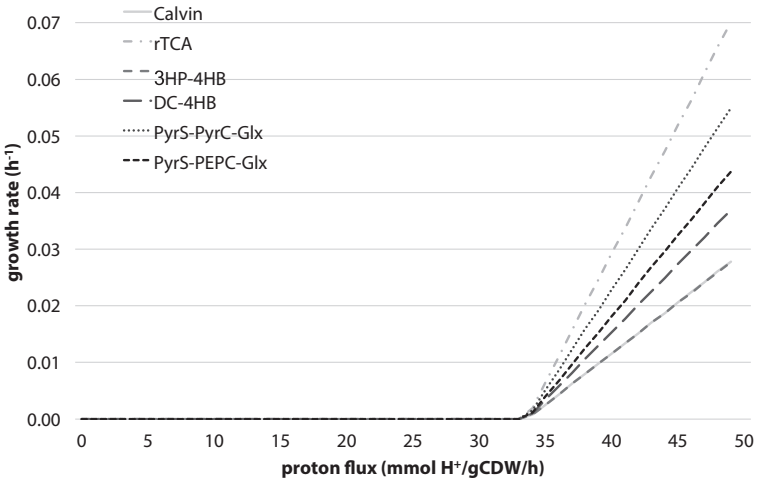


Figure S5. Growth rates for photo-electro-autotroph *E. coli* with high maintenance. Predicted by FBA for different carbon fixation pathways with a NGAM of 8.39 mmol ATP/gCDW/h. PPR proton flux was varied between 0 and 50 mmol H^+/gCDW/h.

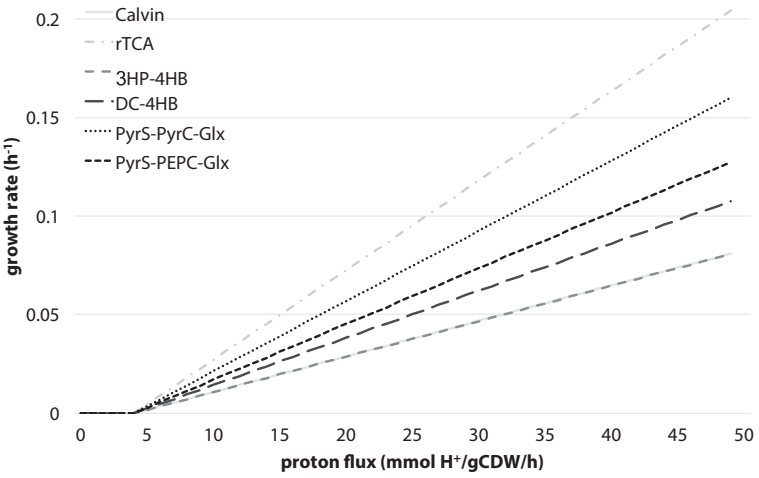


Figure S6. Growth rates for photo-electro-autotroph *E. coli* with low maintenance. Predicted by FBA for different carbon fixation pathways with a NGAM of 1.00 mmol ATP/gCDW/h. PPR proton flux was varied between 0 and 50 mmol H^+/gCDW/h.

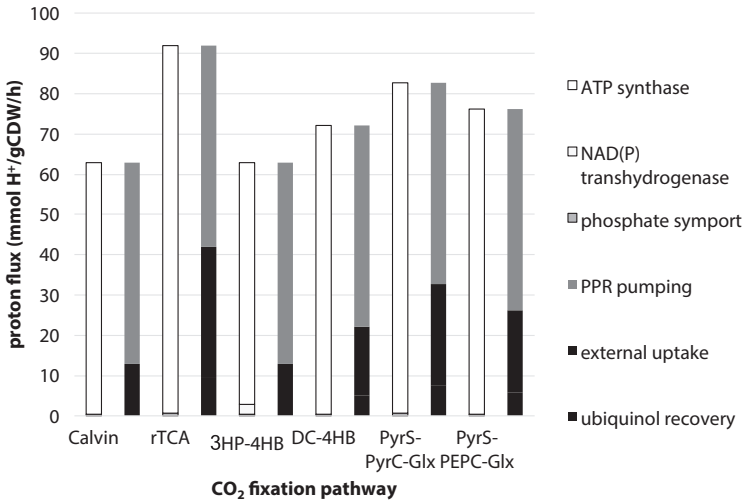


Figure S7. Proton fluxes related to proton import through ATP-synthase. Major proton fluxes in and out of the cell during at a proton-pumping flux of 50 mmol H^+/gCDW/h and a NGAM of 1.00 mmol ATP/gCDW/h.

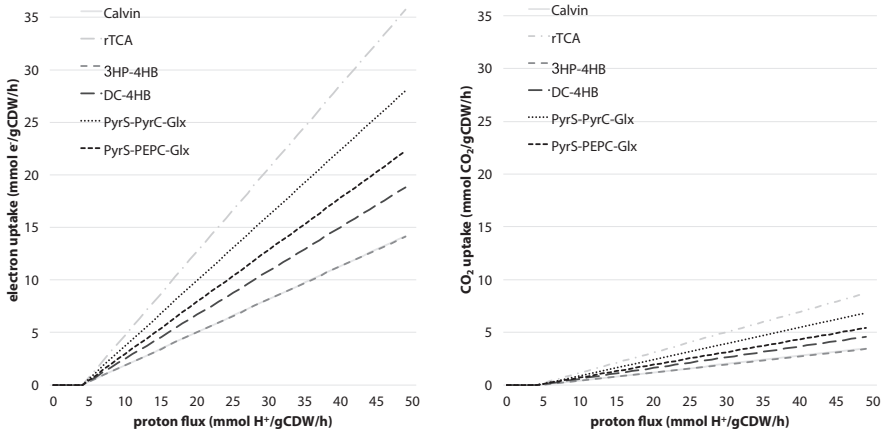


Figure S8. Electron and CO_2 uptake predicted by FBA for different carbon fixation pathways with a NGAM of 1.00 mmol ATP/gCDW/h. Proton pumping was varied between 0 and 50 mmol H^+/gCDW/h.

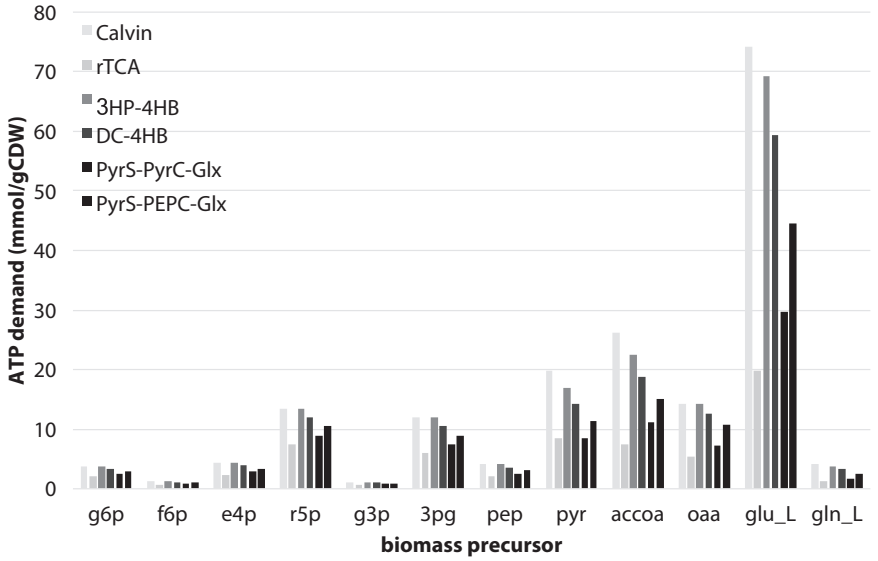


Figure S9. ATP requirements for the biomass precursors of the *E. coli* core model for different carbon fixation pathways. Biomass precursors are: D-glucose-6-phosphate (g6p), D-fructose-6-phosphate (f6p), D-erythrose-4-phosphate (e4p), *a*-D-ribose-5-phosphate (r5p), glyceraldehyde-3-phosphate (g3p), 3-phospho-D-glycerate (3pg), phosphoenolpyruvate (pep), pyruvate (pyr), acetyl-CoA (accoa), oxaloacetate (oaa), L-glutamate (glu_L) and L-glutamine (gln_L).

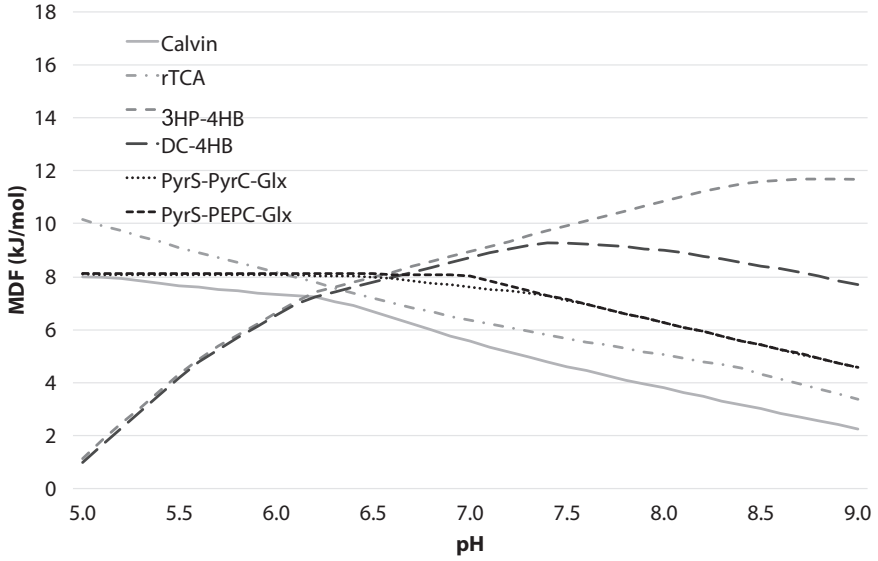


Figure S10. Max-Min Driving Force (MDF) for varying pH values.

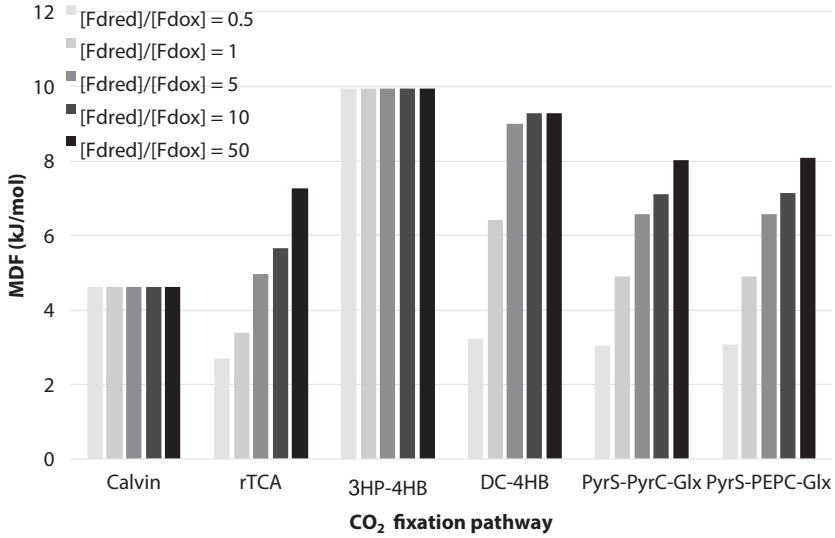


Figure S11. Max-Min Driving Force (MDF) for varying ratios of reduced ferredoxin and oxidized ferredoxin (FD_{red}/FD_{ox}). Bar labels indicate which reaction results in this MDF values, so which reaction forms the thermodynamic bottleneck. Bottleneck reactions are: glyceraldehyde-3-phosphate dehydrogenase (GAPD), isocitrate dehydrogenase (ICDHyr), 2-oxoglutarate:ferredoxin-oxidoreductase synthase (OGOR), succinyl-CoA synthetase (SUCOAS), pyruvate synthase (PYRS), (S)-malyl-CoA lyase (MCL), 3-hydroxyacyl-CoA dehydrogenase (HACD1), 3-hydroxyacyl-CoA dehydratase (ECOAH1).

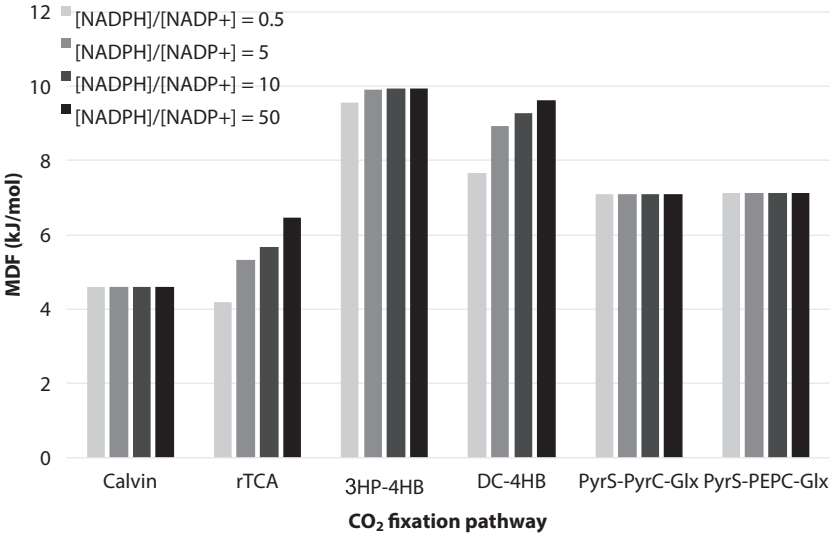


Figure S12. Max-Min Driving Force (MDF) for varying ratios of NADPH/$NADP^+$. Bar labels indicate which reaction results in this MDF values, so which reaction forms the thermodynamic bottleneck. Bottleneck reactions are: glyceraldehyde-3-phosphate dehydrogenase (GAPD), isocitrate dehydrogenase (ICDHyr), succinyl-CoA synthetase (SUCOAS), pyruvate synthase (PYRS), 3-hydroxyacyl-CoA dehydrogenase (HACD1), acetyl-CoA C-acetyltransferase (ACACT1r).

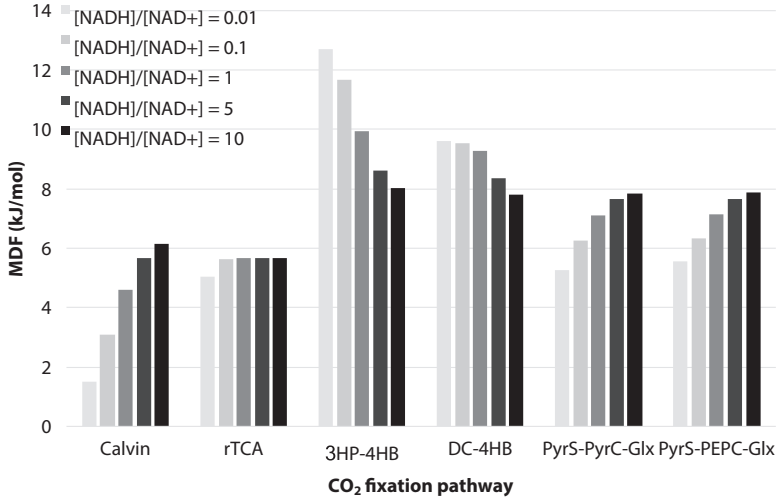


Figure S13. Max-Min Driving Force (MDF) for varying ratios of NADH/NAD$^+$. Bar labels indicate which reaction results in this MDF values, so which reaction forms the thermodynamic bottleneck. Bottleneck reactions are: glyceraldehyde-3-phosphate dehydrogenase (GAPD), ribose-5-phosphate isomerase (RPI), transketolase (TKT2), isocitrate dehydrogenase (ICDHyr), pyruvate synthase (PYRS), pyruvate carboxylase (PYRC), ATP citrate lyase (ACL), 3-hydroxyacyl-CoA dehydrogenase (HACD1), 3-hydroxyacyl-CoA dehydratase (ECOAH1).

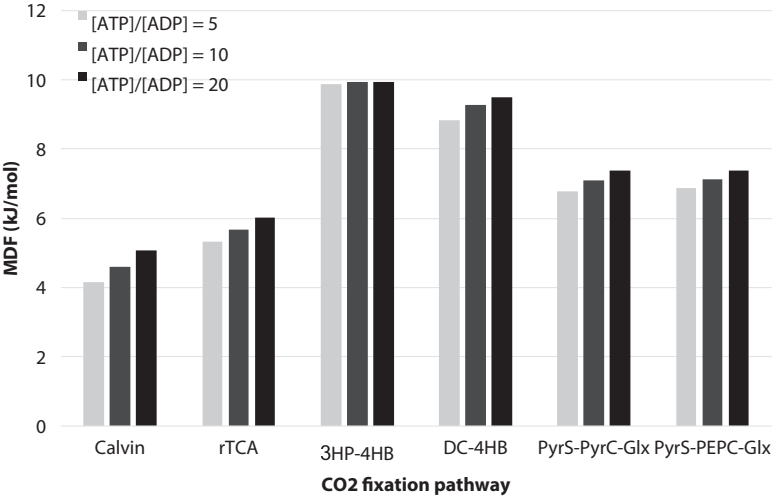


Figure S14. Max-Min Driving Force (MDF) for varying ratios of ATP/ADP. Bar labels indicate which reaction results in this MDF values, so which reaction forms the thermodynamic bottleneck. Bottleneck reactions are: ribose-5-phosphate isomerase (RPI), glyceraldehyde-3-phosphate dehydrogenase (GAPD), succinyl-CoA synthetase (SUCOAS), isocitrate dehydrogenase (ICDHyr), malate dehydrogenase (MDH), pyruvate synthase (PYRS), malyl-CoA synthetase (MTK), 3-hydroxyacyl-CoA dehydrogenase (HACD1), 4-hydroxybutyryl-CoA dehydratase (HB4COADH).

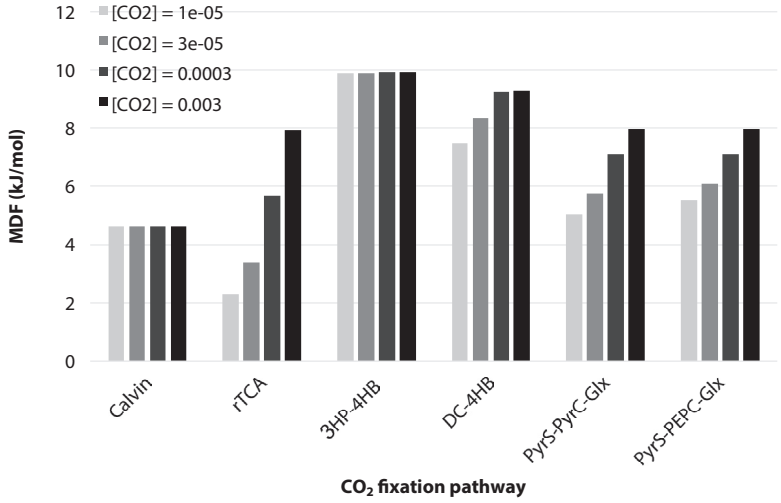


Figure S15. Max-Min Driving Force (MDF) for varying CO_2 concentrations (M). Bar labels indicate which reaction results in this MDF values, so which reaction forms the thermodynamic bottleneck. Bottleneck reactions are: glyceraldehyde-3-phosphate dehydrogenase (GAPD), ribulose-5-phosphate 3-epimerase (RPE), isocitrate dehydrogenase (ICDHyr), pyruvate synthase (PYRS), malate dehydrogenase (MDH), citrate hydrolyase (CITHL), 3-hydroxyacyl-CoA dehydrogenase (HACD1), 3-hydroxyacyl-CoA dehydratase (ECOAH1).

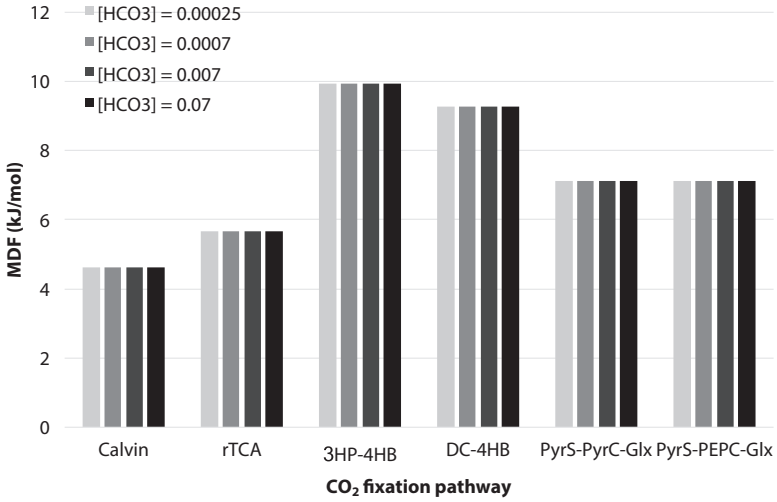


Figure S16. Max-Min Driving Force (MDF) for varying HCO_3^- concentrations (M). Bar labels indicate which reaction results in this MDF values, so which reaction forms the thermodynamic bottleneck. Bottleneck reactions are: glyceraldehyde-3-phosphate dehydrogenase (GAPD), isocitrate dehydrogenase (ICDHyr), malate dehydrogenase (MDH), (S)-malyl-CoA lyase (MCL), pyruvate synthase (PYRS), 3-hydroxyacyl-CoA dehydratase (ECOAH1), 3-hydroxyacyl-CoA dehydrogenase (HACD1).

Chapter 4

Tuning the functional production of rhodopsin photosystems with echinenone antennae in *Escherichia coli*

Nico J. Claassens[1]*, Max Finger Bou[1], Sem van Velden[1], Marta Vallvé Odena[1], Shreyans Chordia[1], Vitor A.P. Martins dos Santos[2,3], Willem M. De Vos[1,4], and John van der Oost[1]

*first author
[1]Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE, Wageningen, The Netherlands
[2]Laboratory of Systems and Synthetic Biology, Wageningen University, Stippeneng 4, 6708 WE, Wageningen, The Netherlands
[3]LifeGlimmer GmbH, Markelstraße 38, 12163 Berlin, Germany
[4]Research Programme Unit Immunobiology, Department of Bacteriology and Immunology, Helsinki University, Haartmaninkatu 3, 00014 Helsinki, Finland

Abstract

Transplanting and optimizing the production of functional photosystems in heterologous hosts is a major challenge. Proton-pumping rhodopsins are relatively simple, proton-motive-force-generating photosystems. Here, the heterologous production of the proton-pumping rhodopsins from *Gloeobacter violaceus* (GR) and from *Thermus thermophilus* JL18 (TR) was optimized in *Escherichia coli*. High constitutive production of GR and TR, up to ~$5*10^4$ rhodopsins/cell or 300 nmole rhodopsins/gCDW was achieved by fine-tuning transcription and translation. Besides the canonical retinal pigment, GR has the ability to bind a light-harvesting antennae pigment, echinenone. In addition, the here obtained absorption spectra from TR suggest that this thermophilic rhodopsin may be also able to bind echinenone. After optimization of the heterologous biosynthesis pathways for echinenone ~500 nmole/gCDW of this pigment was produced in *E. coli,* sufficient to reconstitute all GR or TR. *E. coli* cells co-expressing GR or TR with echinenone biosynthesis obtained broadened absorption spectra of their rhodopsins, indicating the potential to harvest blue light. Ideally both retinal and echinenone pigments should be co-produced in *E. coli* to generate a fully functional photosystem. Expression optimization for the genes encoding the branched biosynthesis pathway, to co-produce both pigments was attempted by using a small degenerate, library of ribosome binding sites. However, this resulted in dominant production of retinal throughout a large number of library clones screened. This indicates a catalytic bottleneck for the enzyme catalyzing the final step of the echinenone biosynthesis (CrtO), as compared to the enzyme for retinal biosynthesis (Blh). Therefore, appropriate expression tuning or the kinetic properties of the enzymes at the branching point in the pathway have to be tackled further. In general, the here described synthetic biology approach towards improved functional production of rhodopsin photosystems in *E. coli* and their pigments may prove useful for heterologous production of complex systems, for example involving branched pathways and/or membrane proteins.

Keywords

Dual-pigment rhodopsins: Heterologous pigment biosynthesis; RBS engineering; *Escherichia coli*

Introduction

Photosystems are key biological systems for converting the abundantly available solar energy into biochemical energy carriers such as ATP and reduced NAD(P)H. Photosynthetic microorganisms harboring photosystems, such as cyanobacteria and microalgae, are promising hosts for the future production of value-added products directly from sunlight and CO_2. However, the practical application of those photosynthetic organisms is still hampered by their limited solar-to-product energy conversion efficiencies (Claassens et al., 2016) (**Chapter 8**). To increase the energy-conversion efficiency, basically by capturing a larger part of the solar light spectrum, ambitious redesigns of photosystems have been suggested, requiring 'plug-and-play' of photosystem components (Blankenship et al., 2011; Ort et al., 2015). It is tempting to apply such a synthetic biology approach for engineering photosystems towards more efficient photosynthesis (Claassens et al., 2016; Kerfeld, 2015; Tikh and Schmidt-Dannert, 2013). However this 'plug-and-play vision' seems far from reality, given the complexity of the photosystems, involving many components such as pigments, membrane-embedded proteins (reaction centers and electron carriers), and also the limited genetic toolboxes of photosynthetic organisms. To make steps in this direction of engineering complex photosystems, we set out to express the genes for the relatively simple dual-pigment proton-pumping rhodopsin photosystems from the cyanobacterium *Gloeobacter violaceus* (GR), as well as the proton-pumping rhodopsin from *Thermus thermophilus* (TR), in the heterotrophic model bacterium *Escherichia coli*.

Proton-pumping rhodopsins (PPRs) consist of a 7-transmembrane helical apoprotein, which can covalently bind to an all-trans retinal pigment molecule (hereafter retinal). Excitation of the retinal pigment by a photon causes a conformational change of the retinal and a cascade of proton donations in the protein, which eventually results in the transport of one proton across the cytoplasmic membrane, out of the cell. As such, PPRs transduce light energy to an electrochemical energy potential: proton-motive force (PMF), which can be used to drive ATP synthesis and active transport. PPRs have been heterologously expressed in *E. coli* and a few other heterotrophic hosts in several studies (reviewed in Claassens et al. 2013) (**Chapter 2**) and in the cyanobacterium *Synechocystis* PCC6803 as an additional photosystem to

harvest green light (Chen et al., 2016). However, in most cases only relatively small physiological benefits at the level of ATP synthesis (Martinez et al., 2007), hydrogen production (Kim et al., 2012b; Kuniyoshi et al., 2015), or growth have been reported (Chen et al., 2016; Wang et al., 2015).

In general, PPRs are not as efficient as the more complex the chlorophyll-pigmented photosystems, which can pump 2-3 protons per photon in addition to producing reduced electron carriers (Kirchman and Hanson, 2013). To increase the efficiency of PPRs, one could select PPRs with fast pumping kinetics and/or broad light absorption (Claassens et al., 2013). Some of the so-called xanthorhodopsins are interesting PPRs, as some members of this group can non-covalently bind a second pigment in addition to retinal, to harvest a broader part of the light spectrum. Two PPRs with an additional light-harvesting pigment have been described so far, one in the halophilic bacterium *Salinibacter ruber* (Balashov et al., 2005) and one in the cyanobacterium *Gloeobacter violaceus*. *Salinibacter* xanthorhodopsin (SX) and *Gloeobacter* rhodopsin (GR) naturally harbor respectively the keto-carotenoids salinixanthin and echinenone as additional pigment. These light-harvesting antennae allow for additional absorbance of blue light and a higher absorption cross-section. Furthermore, GR appears to be a faster proton-pumping rhodopsin than the commonly used proteorhodopsin, when analysed in starved *E. coli* whole-cell proton-pumping assays (Ganapathy et al., 2015). Also the recently discovered, thermophilic rhodopsin from *Thermus thermophilus* (TR) may, based on homology with the dual-pigment xanthorhodopsins, potentially bind a carotenoid as light-harvesting antennae (Tsukamoto et al., 2013a, 2016).

Most *in vivo* PPR expression studies described so far, supplemented retinal in the medium. However, this pigment is both expensive and instable. The additional light-harvesting antennae pigment of GR, echinenone, is very hydrophobic and barely dissolves in growth media. Probably for these practical reasons, echinenone was not reported to be supplemented in *in vivo* studies on *E. coli* cells harboring GR. Retinal and echinenone are both directly derived from β-carotene, a pigment that can effectively be produced in *E. coli* by a heterologous pathway. *In vivo* biosynthesis of both pigments in *E. coli* requires engineering of a branched extension of the pathway.

Here, we test recently developed expression optimization approaches to achieve high, constitutive production of the membrane-embedded GR and TR proteins, and balanced production of the two enzymes for co-biosynthesis of both pigments. We applied the BIOFAB library with standardized expression units, both promoters and translational coupling elements, to optimize GR and TR production (Mutalik et al., 2013b). To optimize the production of the enzymes for the branched pigment pathway, we re-engineered the synthetic biosynthesis operon using Ribosome Binding Site (RBS) engineering based on the novel RedLibs (Reduced Libraries) algorithm (Jeschek et al., 2016). This algorithm allows for generating small libraries of RBS variants to fine-tune expression of multiple genes in an operon. By these approaches, we demonstrate the optimization towards increased production of GR, TR and co-production of retinal and echinenone in *E. coli*. The production of the GR photosystem with biosynthesis of echinenone has the potential to broaden the absorption spectrum of GR.

Results and Discussion

Optimizing for high constitutive production of Gloeobacter rhodopsin

In this work we aimed to constitutively produce a high level of membrane-embedded GR photosystems into the *E. coli* cytoplasmic membrane. The core of the photosystem, the GR apo-protein, is a membrane protein. For this protein, as for most membrane proteins, probably properly fine-tuned expression will be needed to achieve high level production and to avoid oversaturation of the membrane translocation systems in *E. coli* (Schlegel et al., 2014). Hereto we tested the expression of *GR* from 7 different plasmids, as available in the BIOFAB library, for different constitutive levels of expression (Mutalik et al., 2013b). The BIOFAB expression plasmids allow for testing a range of expression strengths through different combinations of constitutive promoters and translational coupling elements, so-called bicistronic design (BCD) elements (**Figure 1a,b**). The BCD elements consist of two translationally coupled RBSs. The first RBS has a strong translation initiation activity and initiates the translation of a short leader peptide. Within the 3'-end of the coding sequence of this short peptide, a second RBS is encoded that can vary in sequence and related translation initation strength. The translation initiation from this second RBS initiates the

translation of the protein of interest, here GR, and this RBS is translationally coupled to the translation of the upstream short leader peptide. In other words, part of the ribosomes that translate the short peptide will hop on to the second RBS and from there initiate the translation of GR. This double RBS design was demonstrated to be a reliable systems for tuning expression with a low dependence on the gene sequence context for the gene of interest (Mutalik et al., 2013b). This is a major advantage compared to using the standard, single RBS designs, where the translation initiation is highly unpredictable because of the variable secondary structure of the transcript at the 5′ UTR:gene sequence junction (Kudla et al., 2009b; Salis et al., 2009).

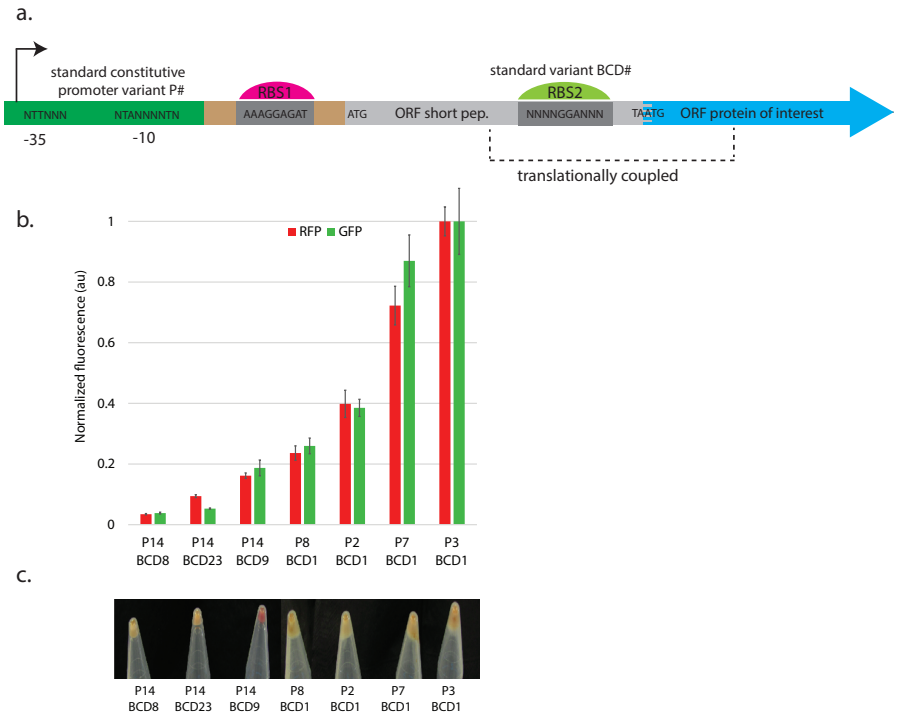


Figure 1. BIOFAB expression units and performance. a| Design of BIOFAB expression units including standard promoters and BCD elements, allowing for translation coupling. b| Relative expression strength of different BIOFAB promoter + BCD combinations assessed for both cytoplasmic GFP and RFP reporter production levels, performed by (Mutalik et al., 2013b). c| Visual production of GR from different promoter + BCD combinations as screened in *E. coli* MG1655 with retinal supplemented in the medium, the high-expressing combination was shown to give robustly high production in *E. coli* BL21(DE3) and *E. coli* DH10B (not shown).

Seven promoter and BCD combinations spanning a broad range of expression levels were selected to assess production of GR (**Figure 1b**). One of the intermediate strength combinations, pFAB-P14-BCD9, appeared to give a high constitutive production level of GR based on the high level of redness of the cell pellet (**Figure 1c**). Weaker and stronger promoter-BCD variants had a low or hardly discernible levels of red color in the cell pellet. The plasmid variant with the highest visual production, contained the medium-strength promoter *P14*, a constitutive variant of the P_{trc} promoter, and the medium-strength translational coupling element BCD9. High functional production from medium strength transcription and translation elements is in accordance with general findings that for high production of functional membrane proteins, intermediate expression levels are often optimal (Schlegel et al., 2014).

The GR level was quantified for *E. coli* BL21(DE3) grown at 37°C in LB to be 182±5 nmole GR/gCDW, which is estimated to be equivalent $3.3*10^4$ rhodopsins per cell (**Table 1**). Growth in minimal medium increased the GR level per cell up $5.5*10^4$, however the volumetric productivity dropped substantially because of the lower cell density reached. These found cellular levels are in the range of native and heterologous protein levels of $1\text{-}5*10^4$ rhodopsins per cell, reported by others (Béja et al., 2001; Johnson et al., 2010; Yoshizawa et al., 2012). Also a volumetric yield ~7 mg GR protein/L reached in LB at 37°C seems at the upper end of the range as reported for PPR production in *E. coli*. As a comparison, an extensive optimization for proteorhodopsin production in *E. coli* with different media, temperatures and inducer concentrations in a controlled reactor reached a maximum of 5 mg purified PPR/L (Gourdon et al., 2008). Often membrane protein yields increase when growing at lower temperatures, however, in LB medium at 30°C the GR yield for BCD 9 was slightly lower. Still higher yields may be feasible at 30°C by re-screening for an optimal promoter/BCD combination for this temperature. However, the approach applied allowed for high production at the physiological temperature of *E. coli*, 37 °C, which was preferred for operation of an engineered photosystem in this host.

The highest found maximum GR protein level per cell (in minimal medium at 37 °C) occupies ~1.6 % of the *E. coli* total proteome, equivalent to approximately ~16% of the cytoplasmic membrane proteome, so a substantial part of the membrane (**Table 1**). However, the obtained level may, based on *in silico*

predictions, still be too low to allow for the potential physiological benefits of PPRs (Claassens et al., 2013; Kirchman and Hanson, 2013; Volpers et al., 2016).

Table 1. Overview of maximum GR and TR production achieved. Expression in *E. coli* BL21(DE3) from constitutive promoter *P14* and BCDs as indicated.

Rhodopsin (BCD#)	*Growth conditions*	*Molecular weight (kDa)*	*Cellular production nmole /gCDW*	*Volumetric production mg /L*	*Cell yield (gCDW/L)*	*Estimated PPRs/cell**	*Estimated cytoplasmic membrane proteome occupancy ***
GR (BCD9)	LB 37°C	32.3	182 ± 5	7.4 ± 0.1	1.27 ± 0.05	$3.3*10^4$	9.8%
GR (BCD9)	LB 30°C	32.3	122 ± 3	5.5 ± 0.05	1.39 ± 0.04	$2.2*10^4$	6.6%
GR (BCD9)	M9 37°C	32.3	305 ± 7	1.0 ± 0.1	0.10 ± 0.01	$5.5*10^4$	16.4%
TR (BCD12)	LB 37°C	29.0	239 ± 6	6.9 ± 0.2	1.04 ± 0.04	$4.2*10^4$	11.1%

*Assuming a ~ $3*10^{-13}$ gCDW/cell for *E. coli* (Milo et al., 2009) **Assuming ~60% protein content in *E. coli* cell dry weight (Bremer and Dennis, 2009) and that 10% of the *E. coli* proteome resides in the cytoplasmic membrane (Liu et al., 2014)

P14-BCD libraries for high production* of *Thermus thermophilus rhodopsin

The applicability of tuning by BCD elements for high membrane protein production was also demonstrated for expression of the *TR* gene in *E. coli.* A slightly modified approach to optimize TR production is based on a pooled cloning for 22 different BCD elements (**Chapter 7**). This also yielded high constitutive expression, in a similar range as found for GR, for the combination of promoter *P14* with another medium-strength translational coupling element, BCD12 (**Table 1**). Production of PPRs with a constitutive promoter and tuning by BCD-elements generally seems to facilitate inducer-free, high expression in *E. coli* (**see also Chapter 7**).

Sufficient β-carotene production by non-induced BglBrick P_{RhaBAD} vector

For full operation of the GR photosystem, sufficient production of both pigments, retinal and echinenone, is required. Both pigments can be produced from β-carotene as a direct precursor (**Figure 2a**). In turn, β-carotene can be produced in *E. coli* by heterologous expression of the *crtEBIY* operon from *Pantoea ananatis* (Das et al., 2007; Misawa et al., 1990). The BioBrick *crtEBIY* operon (Bba_K274200) was converted into the BglBrick assembly format for expression from a BglBrick vector (Lee et al., 2011), with a replication origin and resistance marker compatible with pFAB-P14-BCD9-GR. This allowed for co-expression of the pigment biosynthesis pathway from one plasmid and *GR* from another plasmid.

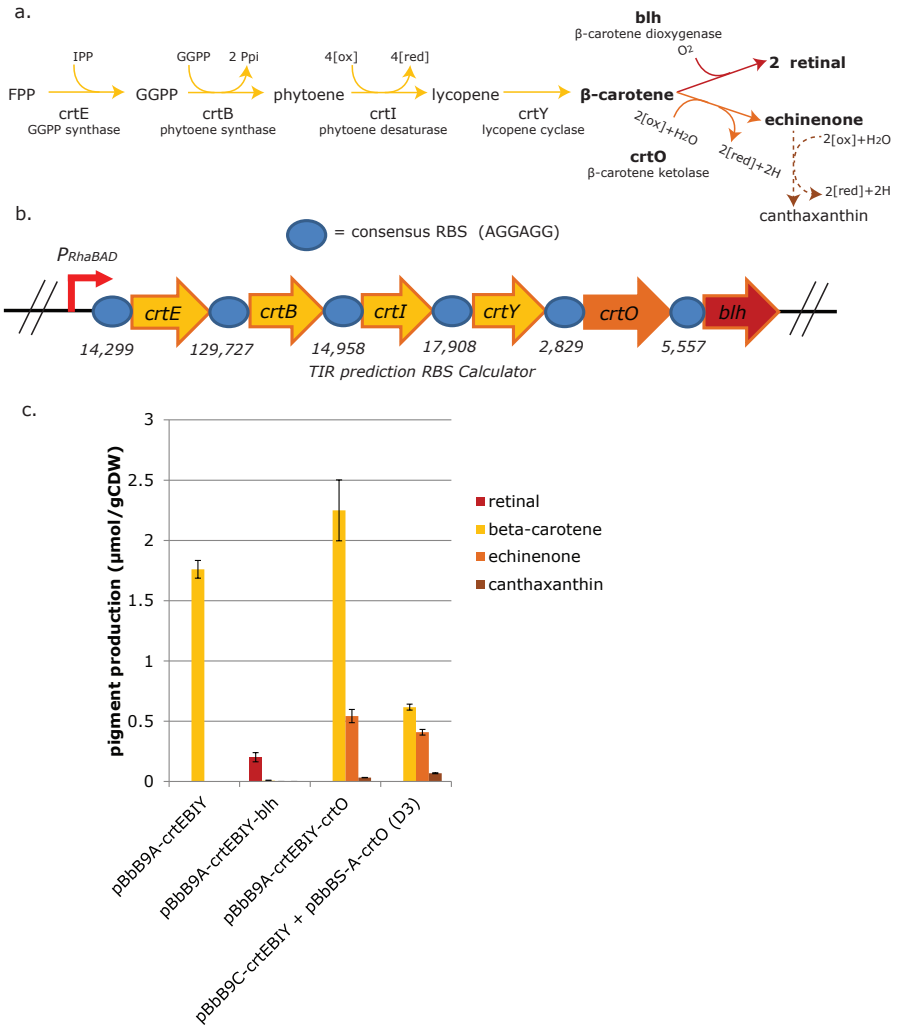


Figure 2. Pigment biosynthesis. a| Heterologous, branched pathway required for pigment biosynthesis of retinal and echinenone in *E. coli*. b| Synthetic full operon for pigment biosynthesis under control of P_{RhaBAD} promoter and standard consensus RBSs. c| Pigment production for *E. coli* BL21(DE3) with different pigment production plasmids, co-expressed with pFAB-P14-BCD9-GR, measured in biological duplicates. FPP: farnesyl disphosphate, IPP: isopentenyl disphosphate, GGPP: generanylgeranyl diphosphate, Ppi: diphosphate, [ox]: oxidized electron carrier, [red]: reduced electron carrier.

As fine-tunable control of the pigment biosynthesis pathway was anticipated to be be desirable, the well-titratable P_{RhaBAD} promoter was chosen. Therefore, the BglBrick vector system was expanded with the P_{RhaBAD} promoter, which has been demonstrated to allow for accurate control using the cheap inducer

L-rhamnose (Giacalone et al., 2006; Wegerer et al., 2008). This promoter was designated number 9 in the BglBrick promoter collection, resulting in pBbB9A and variants thereof. This vector indeed allowed for well-titratable expression, as was confirmed for mRFP (**Supplementary Figure S1**).

However, for pigment biosynthesis from pBbB9A-crtEBIY, no L-rhamnose induction was required. Leaky expression already gave 1.8 µmole β-carotene/gCDW, which probably provides a large enough precursor pool for the production of echinenone and retinal; for the obtained GR protein levels, ~0.2 µmole /gCDW, equimolar amounts of echinenone and retinal per cell are required. Successful non-induced expression for β-carotene production was demonstrated before for other vectors as well, while higher expression may give rise to toxicity issues (Jang et al., 2011).

Pigment biosynthesis of retinal and echinenone in non-optimized synthetic operons

For retinal biosynthesis, the codon-optimized *blh* gene, from the proteorhodopsin-harboring γ-proteobacterium SAR86, was added to the *crtEBIY* operon. *blh* encodes β-carotene 15,15-dioxygenase and was shown before to catalyze retinal biosynthesis in *E. coli* (Jang et al., 2011). The encoded Blh protein is predicted to have 7 transmembrane helices (TMHMM v. 2.0). Expression from pBbB9A-crtEBIY-blh without GR production resulted in non-detectable levels of retinal or β-carotene (data not shown). Probably all produced β-carotene is converted into retinal, which subsequently is degraded rapidly, especially in aerated conditions, as reported before (Jang et al., 2011). However, for co-expression of the retinal pathway with GR, red pigmentation of the cell pellet was observed, indicating retinal re-constituted GR. After bleaching covalently bound retinal from the GR by hydroxylamine, the detected levels of the disassociated pigment (retinaloxime) were in the same range as for GR apo-protein production, ~0.2 µg/gCDW (**Figure 2c**). This confirms that retinal is protected for degradation by covalent binding to GR, and that the non-induced expression of *crtEBIY-blh* already results in biosynthesis of sufficient retinal to reconstitute all GR proteins.

To allow for co-synthesis of the echinenone pigment, a codon-optimized *crtO* gene, encoding a β-carotene ketolase, from *Synechocystis* PCC6803 was

included in the vector pBbB9A-crtEBIY-crtO-blh. However, no echinenone production could be detected by HPLC, while retinal was detected and red pigmentation was observed (data not shown). This indicates highly efficient conversion of β-carotene to retinal, but not from β-carotene to echinenone. When *blh* was left out of the operon, some echinenone could be detected, although the major product was β-carotene (**Figure 2c**). Also minor amounts of the double ketolated product of β-carotene, canthaxanthin, were present. For *Synechocystis* CrtO produced in *E. coli*, some low ketolase activity, using already ketolated echinenone as substrate, resulting in canthaxanthin has been described before (Breitenbach et al., 2013). However, echinenone is dominantly produced compared to canthaxanthin. It has not been investigated yet if canthaxanthin may also bind GR and act as a light-harvesting antennae pigment.

The high levels of the β-carotene precursor indicated inefficient conversion of β-carotene into echinenone by CrtO. The latter conclusion might also explain why co-production of CrtO with the competing Blh enzyme only results in formation of retinal and not of echinenone. So, somehow the ketolase activity of CrtO is a bottleneck. This may be related to the kinetic performance of CrtO, also relative to the performance of Blh. The available *in vitro* kinetic parameters for CrtO (Kcat=1.30 min^{-1}, Km = 41.6 μM) (Breitenbach et al., 2013) and Blh (Kcat = 3.60 min^{-1}, Km = 37.0 μM) (Kim et al., 2009) suggest that the kinetic activities are in a similar range, with a seemingly slight kinetic advantage for Blh. However, those parameters were determined for different *in vitro* conditions, e.g. for Blh at 40°C and CrtO at 28°C, and they may in general not represent the actual *in vivo* conditions in *E. coli*.

The thermodynamic driving force for both reactions was also assessed using Gibbs Free-Energy changes calculated by EQuilibrator (Flamholz et al., 2012). Blh probably has a high driving force, given the ΔrG'° of -364.6 kJ/mole. The ΔrG'° of CrtO is highly dependent of which electron acceptor is used; for NAD^+ as electron acceptor ΔrG'° is +57.8 kJ/mole, which would be thermodynamically infeasible, however for ubiquinone as acceptor the ΔrG'° is -111.7, incidating a feasible reaction. However, for both enzymes it is important to note that those calculations are based on standard 1 mM concentrations in the aqueous phase for all substrates and products, while β-carotene substrate and both pigment

products have a very low solubility in water. So effective concentrations are much lower, and reactions may actually catalyzed in the lipid membrane environment. So even though the thermodynamic driving force for CrtO seems indeed smaller, this is hard to confirm, given the lack of knowledge about the physiological electron acceptor and concentrations in the reaction environment.

Apart from thermodynamic driving forces and kinetics, also differential expression levels of the *crtO* and *blh* genes could partly explain why only retinal is formed by co-expression of both from the same operon. *In silico,* RBS calculator tool indeed predicts a lower translation initiation rate (TIR) for the RBS of *crtO* than for the RBS of *blh* (**Figure 2b**), which may indicate insufficient CrtO protein levels compared to those of Blh. Therefore these RBS translatiation initation strengts were varied as discussed below.

Towards pigment co-production using RedLibs RBS engineering

To optimize the expression of both *crtO* and *blh* for simultaneous production of both echinenone and retinal from β-carotene, an RBS library was generated by the RedLibs algorithm (Jeschek et al., 2016). RedLibs employs RBS Calculator, using RBS sequences and their sequence context as input, to predict their TIRs. Based on these TIR predictions RedLibs generates sets with a limited number of RBS variants covering a uniform range of TIRs, which can be encoded by a single degenerate sequence. To aim for pigment co-production, degenerate RBS sequences for *blh* and *crtO* were selected, covering a TIR range of 0-100,000 with 12 different variants for both genes (**Figure 3a**). The total library size, 12^2 = 144, has been probed appropriately by screening 96 clones in a microtiter plate for pigment biosynthesis.

Initially a one-pot assembly for RBS libraries with both *blh* and *crtO* genes in a backbone with the *P14* promoter was attempted. However, no viable clones were obtained in *E. coli* DH10B, a *RecA*$^-$ strain, normally applicable for efficient transformation of plasmid libraries. However, gel electropheris analysis indicated that the Gibson assembly of all fragments was successful. This suggested that the constitutive expression levels of *crtO* and/or *blh* from the medium-strength *P14* promoter were detrimental for cell growth. After exchanging this *P14* promoter to one of the weaker constitutive promoters

from the BIOFAB library, *P13*, correct transformants could be obtained for *E. coli* DH10B. This successfully assembled library was retransformed to a strain harboring a helper plasmid for β-carotene precursor production and a third plasmid for GR expression.

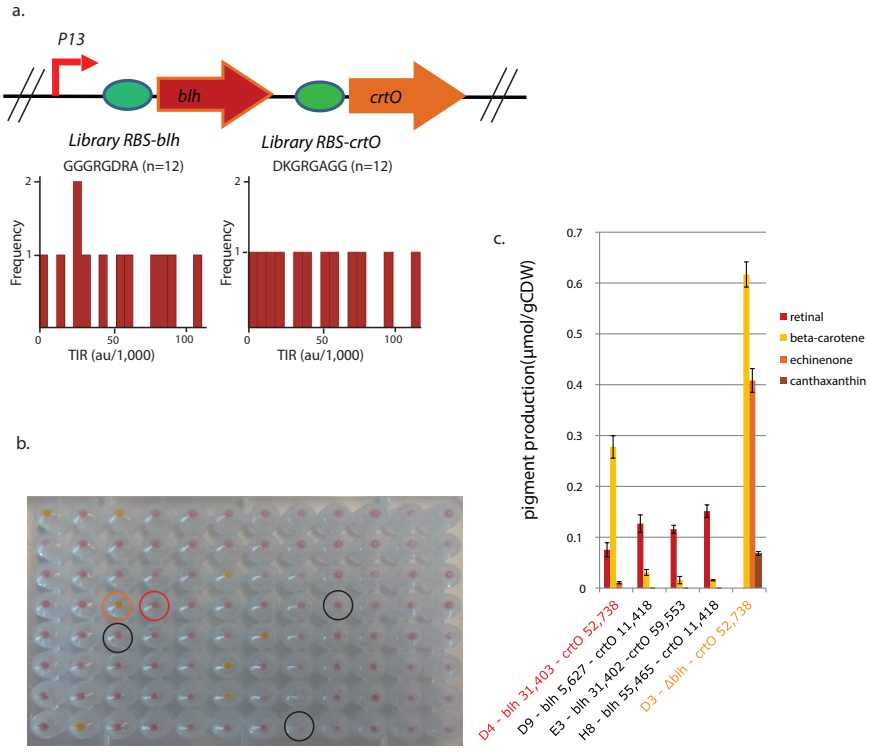


Figure 3. Varying branched pigment biosynthesis pathway enzyme expression by a RedLibs library. a| degenerate RedLibs RBS library for varying Blh and CrtO production, the operon is expressed from a weak BIOFAB P13 promoter, inserted in pBbS0A. b| Color-based screening of the *blh-crtO* RedLibs library in *E. coli* DH10B, harboring pFAB-P14-BCD9-GR and helper plasmid pBbB9C-crtEBIY, after 24 hours of cultivation, clones selected for HPLC are encircled. D3: lacking *blh* (orange circle); D9, E3, H8: randomly pick red clones (black circles); and D4: co-producing retinal, beta-carotene and some echinenone (red circle). c| Quantitative HPLC analysis of carotenoids and retinal production of co-producing clone D4, *blh* deletion clone D3 and three randomly selected red clones, all in *E. coli* DH10B, only D3 library variant was assessed in *E. coli* BL21(DE3).

After deep-well plate cultivation, 96 clones were screened by eye (**Figure 3b**). Then all those clones were analyzed by semi-quantitative HPLC. Some diversity was revealed, 8 clones results in very orange or yellow cell pellets; by semi-quantitative HPLC it was shown that the yellow clones mainly produced

β-carotene, while the orange clones also produced substantial amounts of echinenone in addition to β-carotene. Unexpectedly, none of those orange and yellow colonies produced retinal. Sequence analysis revealed that these clones were miss-assemblies lacking *blh*. The remaining majority of the clones (87) was red-pigmented, for 5 randomly selected red-pigmented clones it was demonstrated by sequencing that *blh* and *crtO* and RedLibs RBSs were correctly assembled.

Semi-quantitative HPLC revealed that only one out of 87 correctly assembled clones, D4, produced some detectable amounts of echinenone in addition to retinal and β-carotene (described in more detail below). All others only produced retinal, and from some clones minor amounts of β-carotene were detected. Based on the dominance of retinal-only producers for correctly assembled vectors, we conclude that despite varying *crtO* and *blh* expression, apparently the production level and/or enzyme kinetics of Blh still result in a flux from β-carotene almost exclusively into retinal.

By further quantitative HPLC analysis of some clones (**Figure 3c**), it was demonstrated that the only co-producing clone (D4) produced 77 nmole/gCDW of retinal, only 10 nmole/gCDW of echinenone, and a relatively high concentration of β-carotene (278 nmole/gCDW). Especially the level of echinenone was likely insufficient to reconstitute all GR (generally around ~200 nmole/gCDW). This co-producing clone had a lower predicted TIR for the *blh* RBS (31,403) than for the *crtO* RBS (52,738), indicating that higher expression of CrtO than for Blh may be required for co-production of both pigments. However, another randomly selected red-clone (E3), with an even higher predicted TIR for the *crtO* RBS (59,552) and the same TIR for the *blh* RBS as in clone D4 (31,403) resulted unexpectedly in retinal-only production. So based on a single clone co-producing echinenone and retinal and its TIR predictions, it is hard to generalize rules for further improved echinenone and retinal co-production. It also has to be mentioned that single TIR predictions cannot be considered highly accurate as the biophysical model used by RBS Calculator has some limitations (Reeve et al., 2014).

However, based on the dominance of retinal only-producing clones, a lower TIR range for *blh* and a higher range for *crtO* may be the way to further improve

co-production of the two pigments. Noteworthy, however, for all constructs in which echinenone production is observed, substantial amounts of residual β-carotene are present, while β-carotene is hardly detected when retinal is produced. Also in literature for production of CrtO in *E. coli* residual β-carotene was observed after 24 hours of cultivation (Breitenbach et al., 2013), while for the employed Blh variant after 24 hours no residual β-carotene was reported (Jang et al., 2011).

The results described above suggest that a difference for *in vivo* enzyme kinetics for Blh and CrtO may complicate efficient co-production. The use of different natural or engineered enzyme variants for Blh and CrtO with a more similar kinetic performance could facilitate the desired pigment co-production. A strategy could be to apply the original Blh and CrtO variants from *G. violaceus*, as this organism should natively be able to co-produce both retinal and echinenone. However, the CrtO from *G. violaceus* has been heterologously produced in *E. coli* before, and was observed to result in less echinenone production than the here employed *Synechocystis* CrtO (Scaife et al., 2009). Based on all CrtO variants included in that screen in *E. coli,* the here employed CrtO variant is by far the best for echinenone production (Scaife et al., 2009). The applied SAR86 Blh also was the best variant in a screen of several retinal-producing enzyme variants in *E. coli* (Jang et al., 2011). Hence, actually less efficient variants than the here employed Blh variant may provide a competitive advantage for echinenone production.

Given the challenge of realizing sufficient co-production of echinenone and retinal, further analysis of the GR photosystem is performed for strains harboring the pathway for *in vivo* echinenone biosynthesis only and retinal was further supplemented in the medium. Hereto we used one of the mis-assembled plasmids, derived from the non-retinal producing, *Δblh clone* D3, that gave relatively high echinenone production (TIR: 52,738), and low residual β-carotene production. This plasmid resulted in substantially less residual β-carotene, and similar echinenone production levels compared to the initially used echinenone production plasmid pBbB9A-crtEBIYO with consensus RBSs (TIR *crtO* RBS: 2,829) (**Figure 2c**).

In vivo echinenone biosynthesis broadens GR absorption spectrum

E. coli BL21(DE3), harboring pBbB9C-crtEBIY and pBbS-A-P13-RBS$_{TIR52,738}$-crtO (*Δblh from* clone D3), which gave sufficient stoichiometric echinenone biosynthesis for GR, with low residual β-carotene (**Figure 2c**), was supplemented with retinal during growth. A membrane extract of this strain was spectroscopically assessed, which showed a difference spectrum with distinct peaks in the blue spectrum at ~460 nm and ~490 nm, in addition to the typical GR peak at ~540 nm (**Figure 4**). Peaks around these wavelengths were also observed before for *in vitro* reconstitution of GR with echinenone and indicate specific binding of echinenone to GR (Balashov et al., 2010). Although, the peaks of this difference spectrum are not as pronounced and high as observed in an *in vitro* assay with echinenone supplementation. This can be explained, as in the here performed bleaching assay echinenone will still be present and have its own specific absorption ~470 nm, after being released from GR after bleaching. In the difference absorption spectrum this resulted in a lower height of the specific peaks at 460 and 490 nm.

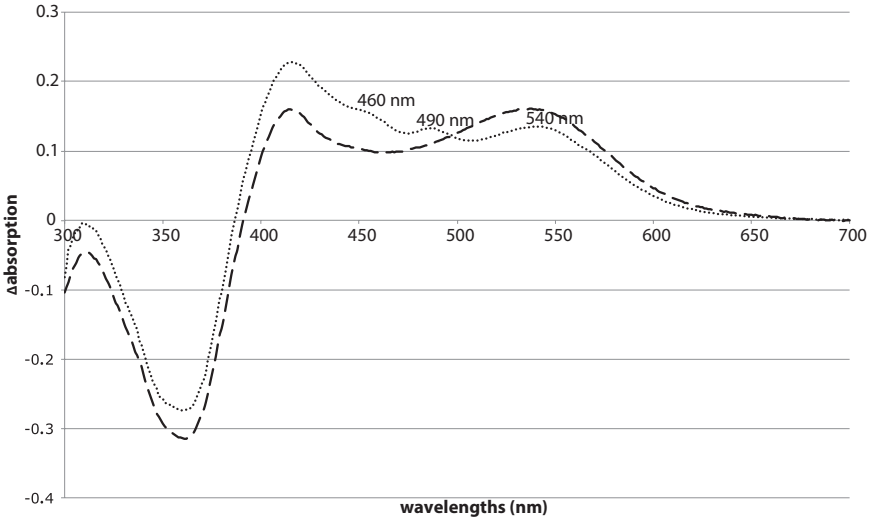


Figure 4. Spectral analysis of membrane extracts of GR-containing *E. coli* with and without echinenone biosynthesis. Difference absorbance spectra (before bleaching – after bleaching) for GR only (dashed) and GR co-expressed with *in vivo* echinenone biosynthesis (dotted). Membranes were respectively extracted from *E. coli* BL21(DE3) pFAB-P14-BCD9-GR and *E. coli* BL21(DE3) pFAB-P14-BCD9-GR + pBb9C-crtEBIY + pBbS-A-P13-RBS$_{TIR52,738}$-crtO. Absorption differences are obtained through bleaching with 0.1 M hydroxylamine, single curves are depicted for clarity, but biological duplicates gave similar curves.

An additional peak is observed in the difference spectra at ~415 nm, both for samples with and without echinenone biosynthesis. This peak was also observed for difference spectra of control samples without GR or retinal present (data not shown), indicating this peak is not related to GR, but probably to changes in the Soret peak for *E. coli* cytochromes.

Effect on ATP regeneration in blue light with and without echinenone antennae
To demonstrate the physiological impact of light on retinal-supplemented, GR-containing *E. coli* cells, with and without *in vivo* produced echinenone, ATP assays of two types of starved cells were performed. Cells were hereto grown in M9 minimal medium to have the highest possible levels of GR per cell (see above). After stationary phase cells were starved for 2-3 hours in PBS, they were exposed to green light for 30 minutes at 37˚C. Then it was observed that ATP levels for green light-exposed samples were reproducibly increased. The observed increase in green light, however, is not significantly different with and without echinenone biosynthesis (**Figure 5a**). This is most likely because the echinenone antennae do not substantially contribute to absorbance in the green part of the spectrum, but only in the blue part (Balashov et al., 2010). Surprisingly, when exposing cells to blue light, the ATP levels in all light-treated cells dropped significantly compared to cells kept in similar, but dark, conditions. The observed drop in ATP levels is probably caused by toxicity of high-energetic blue light, causing a drop in cell viability (**Figure 5b**). This complicates assessing the potential physiological effects of the light-harvesting echinenone antennae in blue light. However, for the GR plus echinenone samples, a significantly smaller drop in ATP is observed than for the GR only samples. This may indicate a partly counter-acting, ATP-producing effect in blue light, potentially driven by blue photon absorption by the echinenone antennae pigment. However, more extensive characterization is required to further confirm the effect of echinenone on light-harvesting of GR in *E. coli*. The negative impact of blue light may be mitigated by using lower light intensities or lowering the assay temperature, as some preliminary tests have indicated (data not shown). Alternatively, the physiological effect of blue light and the echinenone-antennae could also be assessed by proton-pumping assays or by cell survival experiments in different light conditions. However, attempts in this direction have so far failed to give reproducible, properly quantifiable results.

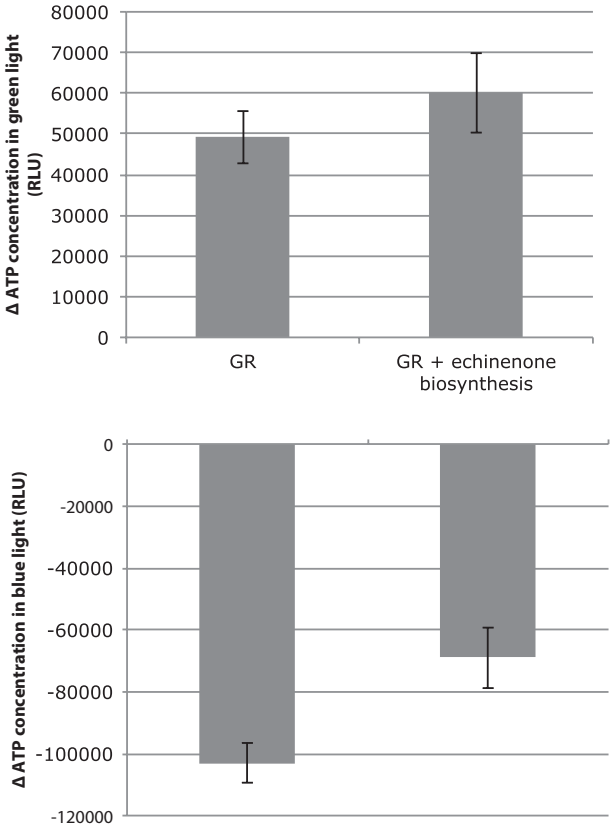


Figure 5. ATP production in GR-containing *E.coli* cells. ATP differences from light vs. dark exposed starved cells to 100 µmole photons m^{-2} s^{-1} of a| green LED light (530 nm peak) and b| blue LED light (447 nm peak). Starved cultures of *E. coli* BL21(DE3) pFAB-P14-BCD9-GR (GR) and *E. coli* BL21(DE3) pFAB-P14-BCD9-GR + pBb9C-crtEBIY + pBbS-A-P13-$RBS_{TIR52,738}$-crtO (GR + echinenone biosynthesis) which were grown in the presence of retinal were used. Data were obtained from at least two independent biological duplicates measured in technical triplicates.

In vitro and in vivo reconstitution of TR with echinenone

In the course of this study high production of *Thermus* rhodopsin (TR) has been achieved in *E. coli* . The TR primary structure is highly homologous to that of the GR and the *Salinibacter* xanthorhodopsin (SX), with a homology to both of about ~50%, and for both these xanthorhodopsins dual-pigment binding has been demonstrated in *in vitro* reconstitution experiments (Balashov et al., 2010; Imasheva et al., 2011) and for membrane preparations from *Salinibacter ruber* (Balashov et al., 2005, 2008). Based on the homology of TR with SX and GR, especially for some key residues related to keto-carotenoid binding, TR may be

able to bind a keto-carotenoid pigment as well (Tsukamoto et al., 2016). *Thermus thermophilus* is known to synthesize carotenoids, mostly thermozeaxanthins (Tian and Hua, 2010), which were suggested to act as a light-harvesting antennae of TR (Tsukamoto et al., 2016). However, it is not clear if *T. thermophilus* can synthesize specific keto-carotenoids similar to echinenone or salinixanthin; for GR the presence of a keto-group in keto-carotenoids was observed to be crucial for binding (Balashov et al., 2010). However, TR may also be capable of binding non-native keto-carotenoids as a light-harvesting antennae. A similar capability was observed for GR that can use salinixanthin as a functional light-harvesting antennae, even though this pigment is not natively produced by *Gloeobacter violaceus* (Imasheva et al., 2009).

The demonstrated high production of TR, and biosynthesis of echinenone, opens up possibilities to test the hypothesis that TR can bind echinenone and use it to harvest more light. To more clearly observe if TR can bind echinenone, first an *in vitro* reconstitution experiment was performed with his-tag purified, DDM-solubilized TR. Three minutes after reconstitution of purified TR with echinenone a relatively unstructured spectrum is observed. However, after 24 hours (not shown) and even more so after 72 hours, higher absorbance by a sharpened peak around 490 nm is observed (**Figure 6**). This indicates an interaction between TR and echinenone. For reconstitution of GR and SX with keto-carotenoids also sharpened peaks were formed, although for those double sharpened peaks were observed around 490 and 460 nm. For the reconstitution of TR with echinenone merely a weak shoulder at 460 nm seems to appear. In conclusion, this reconstitution experiment provides an indication that echinenone is converted into a different conformation upon associating with TR. However, to confirm the role of echinenone as an active light-harvesting antennae for TR, further biophysical studies are required, such as circular dichroism and fluorescence excitation spectroscopy.

Also the effect of *in vivo* echinenone biosynthesis with co-expression of *TR* was assessed. Similar as for GR with echinenone biosynthesis, *E. coli* BL21(DE3), harboring pBbB9C-crtEBIY and pBbS-A-P13-$RBS_{TIR52,738}$-crtO, was used with co-expression of pFAB-p14-BCD12-TR. The resulting difference absorbance spectrum also indicates a broadened absorption in the blue spectrum, but no clear peaks additional to the 525 nm peak are formed. This observation

corresponds to observations in the *in vitro* reconstitution experiment, where a relatively low peak is formed at 490 nm, which could in a difference spectrum, where echinenone is released by bleaching, become a shoulder to the retinal peak at 525 nm.

In general it is relevant that echinenone, which can be produced heterologously at least in *E. coli*, could be a light-harvesting antennae for several xanthorhodopsins. This opens up options for more fundamental studies and potential applications, such as for optogenetics, of broader light-harvesting, dual-pigment PPRs, also potentially in high temperature conditions and thermophilic hosts with thermostable TR.

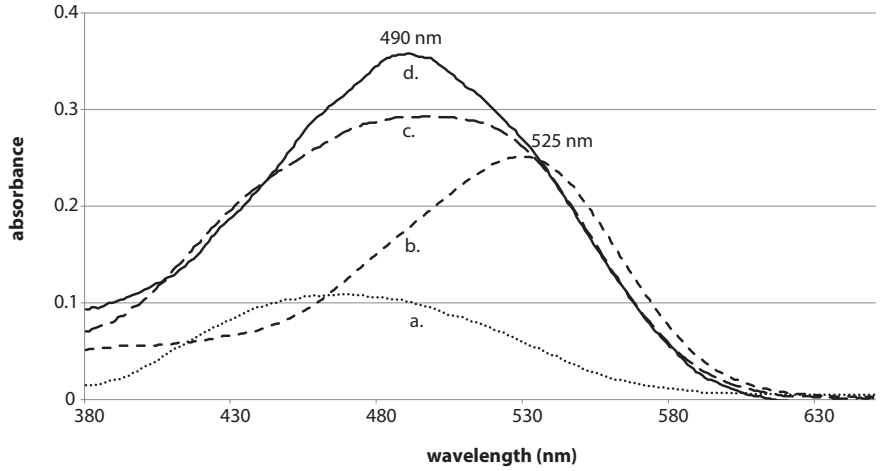


Figure 6. Spectral analysis of purified TR reconstituted *in vitro* with echinenone. Absorbance spectra for *in vitro* reconstitution of purified TR with echinenone: a| 4 µM echinenone (dotted); b| 4 µM TR (short dashes); c| 4 µM TR reconstituted with 4 µM echinenone after 3 minutes (long dashes) and d| after 72 hours (full line).

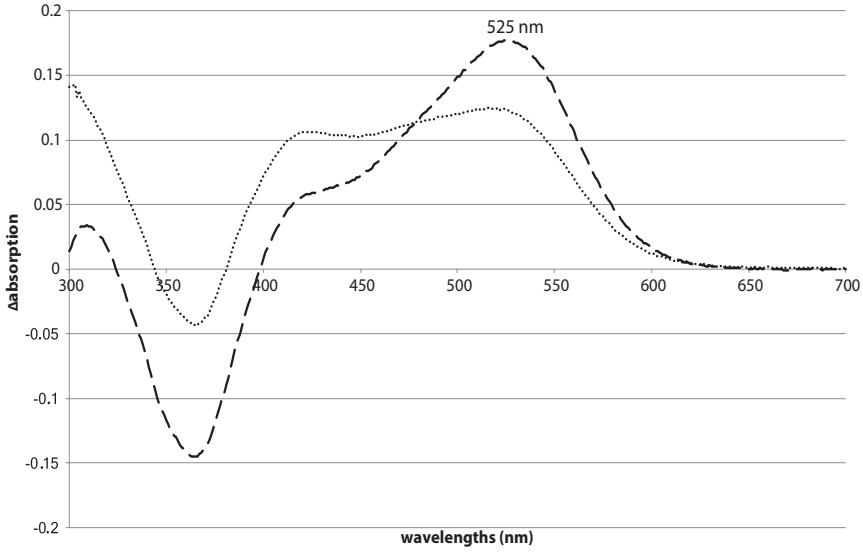


Figure 7. Spectral analysis of membrane extracts of TR-containing *E.coli* cells with and without echinenone biosynthesis. Difference absorbance spectra (before bleaching – after bleaching) for TR only (dashed) and TR co-expressed with *in vivo* echinenone biosynthesis (dotted). Membranes were respectively extracted from *E. coli* BL21(DE3) pFAB-P14-BCD12-TR and *E. coli* BL21(DE3) pFAB-P14-BCD12-TR + pBb9C-crtEBIY + pBbS-A-P13-$RBS_{TIR\text{-}52,738}$-crtO. Absorption differences are obtained through bleaching with 0.1M hydroxylamine, single curves are depicted for clarity, but biological duplicates gave similar curves.

Conclusions

In the present study we describe the optimization of the functional production of membrane-embedded GR and TR in *E. coli* through tuning of the translation initiation by varying the promoters and especially varying translational coupling elements. The obtained production levels appeared to be relatively high for PPR expression in heterotrophic hosts, and occupy a substantial part of up to approximately 16% of the *E. coli* membrane proteome. However, the membrane-embedded production of PPRs probably needs to be increased several-fold, based on earlier *in silico* analyses, to allow for sufficient ATP regeneration to efficiently stimulate growth in *E. coli* (Volpers et al., 2016). However, given the already high fraction of PPRs in the membrane proteome, further increasing of production levels in *E. coli* may require challenging strategies to increase the limited membrane surface (Arechaga et al., 2000; Eriksson et al., 2009) or to lower the expression of redundant membrane proteins.

Echinenone is the light-harvesting antennae of GR. This study describes its successful biosynthesis by GR-containing *E. coli* cells. Moreover, our *in vitro* and *in vivo* spectra suggest that it also may acts as a light-harvesting antennae for TR. Echinenone can broaden the absorption spectrum of GR and TR into the blue-spectrum, which could be promising to exploit for light-harvesting or optogenetic applications of PPRs. However, the physiological impact of blue-light absorption by GR has so far been hard to demonstrate, also because of the negative effect of blue light on *E. coli* viability and ATP regeneration. Co-production of echinenone and retinal by varying expression of *crtO* and *blh* through small, degenerate RBS libraries was achieved, but achieved very low amounts of echinenone in addition to higher retinal and β-carotene concentrations. Kinetically better matching enzyme variants, or further optimization with RBS libraries in a different range may be required for the expression of the pigment enzymes used in this study.

Summarizing, optimization of constitutive high-level, membrane-embedded production of rhodopsins together with echinenone biosynthesis has been demonstrated here. The here employed constitutive promoter and RBS tuning approaches may find wider applicability for engineering complex, membrane-bound protein systems and branched pathways, such as for transplanting diverse types of photosystems.

Materials and methods

Strains and culture conditions

E. coli strains (**Table 2**) were routinely cultured in Lysogeny Broth (LB) (Sambrook et al., 1989) , with appropriate antibiotics, kanamycin (50 µg/mL), carbenicillin (100 µg/mL) or chloramphenicol (35 µg/mL). M9 medium was prepared based on (Sambrook et al., 1989) supplemented with 0.4% glucose or 2% glycerol and appropriate antibiotics, kanamycin (25 µg/mL), carbenicillin (50 µg/mL) or chloramphenicol (17.5 µg/mL). All-trans retinal (Sigma) was added from a 20 mM ethanol stock to a final concentration of 20µM and re-added after 2-5 hours of cultivation to compensate for degradation.

E. coli cultures for quantification of rhodopsin expression or ATP assays were performed in 10 mL M9 medium 0.4% glucose in 50 mL Greiner tubes, to which 2% O/N pre-culture was added, they were cultivated for 22-24 hours at 37°C, at 180 rpm (unless specified otherwise).

Cultures for HPLC analysis were grown in 50 mL LB in 125 mL baffled erlenmeyer shake flasks, inoculated with 1% O/N pre-culture. They were cultured for 24 hours at 37°C at 180 rpm.

RedLibs library clones were pre-cultured (100 µL/well) and cultured (500 µL/well) in MASTERBLOCK® Deep 96-Well plates (Greiner), kept in a plastic bag with a humid atmosphere to prevent evaporation, for 24 hours at 37°C at 900 rpm.

Table 2. Strains used in this study.

Strain	*Description*	*Origin*
E. coli DH10B	*F⁻ λ⁻ Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14 Φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	NEB
E. coli DH5α	*F⁻ λ⁻ fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	NEB
E. coli MG1655	*F⁻ λ⁻ ilvG⁻ rfb-50 rph-1*	ATCC
E. coli BL21(DE3)	*F⁻ λ⁻ fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS*	NEB

Plasmid construction

Plasmids used for expression in this study are in **Table 3**, further plasmids used for constructing are in **Supplementary Table S1**, primers used are in **Supplementary Table S2**. PCR products were generated using either Phusion polymerase (Thermoscientific) or Q5 polymerase (NEB) according to manufacturer's protocols. Cloning for the BglBrick standard was performed using EcoRI, BlgII, BamHI and XhoI restriction enzymes (Thermoscientific or NEB) and ligated using T4 ligase (Thermoscientific or NEB). Golden Gate assembly for the BIOFAB and BCD assemblies was performed using BsaI (Thermoscientific or NEB) and subsequently ligated by T4 ligase (Thermoscientific or NEB). When appropriate, PCR samples were treated with DpnI (Thermoscientific or NEB) to remove template DNA. Gibson Assembly was performed using the NEBuilder® HiFi DNA Assembly (NEB) according to the manufacturer's protocol. Agarose gel DNA purifications and DNA purifications were respectively performed

using Zymoclean™ Gel DNA Recovery Kit and DNA Clean & Concentrator (Zymo Research). Assemblies were confirmed by restriction digests and/or by Sanger sequencing (GATC Biotech).

The genes encoding GR, Blh and CrtO were codon-optimized for *E. coli* and adapted to the BglBrick format, and synthesized by GeneArt (**Supplementary data 1**). *TR* was codon-harmonized using our algorithm (**Chapter 6**) based on (Angov et al., 2008) and synthesized and subloned into pGFPe (Drew et al., 2006) by GeneArt. pGFPe was a gift from Jan-Willem de Gier.

For construction of the BIOFAB vectors expressing GR, *GR-GFP* was PCR amplified from pBbE0A-GR-GFP (BG5971, BG5972) for introduction into seven BIOFAB plasmids with different expression unit strengths (pFAB3912, pFAB3999, pFAB3913, pFAB3833, pFAB3677, pFAB3737, pFAB3689), which were all a gift from Drew Endy (Addgene kit #1000000037) (Mutalik et al., 2013b). The BIOFAB backbones were amplified using primers (BG5973, BG5975, BG5976 and BG5981) for insertion of GR-GFP using Golden Gate assembly. As the GFP fusions at the C-terminus of GR were hampering proper reconstitution of GR with retinal, those were removed from the constructs using phosphorylated primers BG6162 and BG6163 and plasmid recircularization with T4 ligase.

For construction of the BIOFAB expression library for TR, *TR* was amplified from pGFPe-TR (BG7340, BG7337) and introduced into a BIOFAB-p14 backbone (pFAB3913 amplified from BG7339, 7341) by Gibson assembly. Subsequently the created plasmid pFAB-P14-TR was amplified (BG7505, BG7506), to allow for pooled introduction of 22 different RBS2 variants. These RBS2 variants were ligated into the pFAB-P14-TR PCR product, as a pooled library of annealed oligos (BG7291-BG7334). Oligos were phosphorylated using T7 polynucleotide kinase (NEB) according to the manufacturer's protocol. Then, pairs of oligos for each RBS variants were annealed, by heating to 95°C for 3 min and gradual cooling to RT in 30 min. The annealed oligo pairs were pooled to a total concentration of 10 ng/µL. One µl of this mixture was ligated with BsaI-digested, dephosphorylated, PCR product pFAB-P14-TR.

The initial pigment production operons were inserted in BglBrick plasmids (Lee et al., 2011). Several of the BglBrick vectors were a gift from Jay Keasling

and obtained through Addgene. The BglBrick collection was expanded with the P_{RhaBAD} promoter part. For introduction of the P_{RhaBAD} promoter into the BglBrick system, the promoter region was synthesized (GeneArt), including the upstream transcriptional terminator stemloop, and with a mutation to remove the forbidden BglBrick EcoRI restriction site from the promoter region (**Supplementary data 1**). The promoter of pBbB1A and pBbS1A (*Ptrc*), was replaced by the P_{RhaBAD} promoter, using AatII and EcoRI sites, generating pBbB9A-RFP and pBbS9A-RFP.

The Biobrick *crtEBIY* (BBa_K274200) was retrieved from the iGEM Registry of parts and modified into BglBrick format by removing the BamHI restriction by side-directed mutagenesis (BG470, BG4711). Subsequently the synthetic operon was introduced in BglBrick vector backbone (BG4714, BG4715), resulting in pBbE0A-crtEBIY. *blh* and *crtO,* including consensus RBSs, were also introduced into this standard backbone using BG4712/BG4713 and BG4760/BG4761, respectively. All complete synthetic operons were assembled using BglBrick assembly into pBb9A.

The helper plasmid pBbB9C-crtEBIY for the RedLibs libraries, was generated by replacing the Amp resistance cassette of pBbB9A-crtEBIY by a Cam resistance cassette from pBbE1C, using standard BglBrick sites AatII and SacI. The backbone for the RedLibs library was generated by moving the BIOFAB cassette, containing the *P13* promoter from pFAB3893 (*P13*-BCD1-RFP) into pBbS0A-RFP using BglBrick sites EcoRI and XhoI. The RedLibs libraries were generated by PCR amplification of *crtO* (BG7901, BG7538) and *blh* (BG7899, BG7900), with forward primers having degeneracy for the designed RBS libraries. These were assembled by Gibson assembly into the pFAB-P13 backbone, amplified by (BG7532, BG7898).

RedLibs library design

Initial 5'UTR regions of 25 nucleotides, preceding the coding sequences of *blh* and *crtO* were designed by RBS Calculator, set for the highest achievable TIR. These were used as inputs to the RedLibs algorithm as described elsewhere (Jeschek et al., 2016). In short, libraries were generated covering a range of TIRs of 0 till approximately 100,000. Subsequently, from the 10 most uniformly distributed RBS libraries in this range, with a degeneracy of 12, one library was

picked for *blh* and one for *crtO*. The related degenerate RBS sequence was used for degenerate primer design to employ in Gibson library assembly.

Table 3. Plasmids used in this study.

Plasmid name	*Antibiotic marker*	*Origin of replication*	*Important components*	*Reference*
pFAB-P14-BCD9-GR	Kan	*P15a*	*P14, BCD9, GR*	This work
pFAB-P12-BCD12-TR	Kan	*P15a*	*P14, BCD12, TR*	This work
pFAB-P#-BCD#-GR	Kan	*P15a*	*Varying promoters and BCDs, GR*	This work
pBbB9A-RFP	Amp	*BBR1*	P_{RhaBAD}, *mRFP*	This work
pBbS9A-RFP	Amp	*SC101*	P_{RhaBAD}, *mRFP*	This work
pBbB9A-crtEBIY	Amp	*BBR1*	P_{RhaBAD}, *crtEBIY*	This work
pBbB9A-crtEBIY-crtO	Amp	*BBR1*	P_{RhaBAD}, *crtEBIY*	This work
pBbB9A-crtEBIY-blh	Amp	*BBR1*	P_{RhaBAD}, *crtEBIY-blh*	This work
pBbB9A-crtEBIY-crtO-blh	Amp	*BBR1*	P_{RhaBAD}, *crtEBIY, blh, crtO*	This work
pBbB9C-crtEBIY	Cam	*pBBR1*	P_{RhaBAD}, *crtEBIY*	This work
RedLibs library: pBbS-A-P13-blh-crtO	Amp	*SC101*	*P13, RedLibs library for blh and crtO*	This work

Quantification and spectral measurements of GR and TR membrane extracts

For rhodopsin quantification 10 mL of culture was resuspended in 295 µL extraction buffer as in (Engqvist et al., 2015) and frozen for at least 1 hour. Cells were thawed and additional 295 µL extraction buffer, supplemented with 6 mg/mL lysozyme and 0.4 mg/mL DNAase was added. For cell lysis, this suspension was incubated at room temperature for 30 minutes. Rhodopsins were extracted from the crude cell extract by addition of 2.5% (w/v) dodecyl-maltoside (DDM, Sigma) at 180 rpm for 24-48 hours in dark. The extraction for GR was performed at room temperature, to increase the extraction efficiency of TR, this was performed at 65°C.

After 24-48 hours the extraction mixture was spun down to check for the color of the pellet, if a colorless pellet was obtained the supernatant fraction was used for spectroscopic quantification. 200 µL of supernatant was transferred to a black flat microtiter plate with a transparent bottom (Greiner) and the absorption spectrum was measured (Synergy MX BioTek). Next, 1.2 µL hydroxylamine/100 µL supernatant was added to bleach the retinal from the rhodopsin for 1 hour, with gentle shaking at room temperature, in dark. Then, the absorption spectrum was measured again. From this the difference

absorption spectrum could be generated, and from differential absorption at 540 nm (GR) or 525 nm (TR), the rhodopsin concentration was determined, assuming an extinction coefficient of 50,000 (M cm^{-1}) for both rhodopsins.

TR his-tag purification and in vitro reconstitution with echinenone

50 mL of *E. coli* BL21(DE3) pFAB-P14-BCD12-TR was grown in LB supplemented with retinal for 24 hours and membranes were solubilized using the extraction procedure described above. Membrane extracts in 2.5% DDM were loaded to a column with 250 µL Ni-NTA Agarose (Qiagen). The column was equilibrated with 2 mL of buffer 1, after sampling loading, the column was washed 3 times with 2 mL buffer 1. TR-his was eluted in three volumes of 250 µL buffer 2. Buffer 1 is composed of 20 mM Tris, pH 8, 500 mM NaCl, 0.02% DDM and buffer 2 is composed of 20 mM Tris, pH 8, 400 mM imidazole, 500 mM NaCl, 0.02% DDM.

The eluted TR fraction was quantified by spectroscopy and diluted to 4µM in buffer 2, for reconstitution equimolar amounts of echinenone were added. Spectra were obtained at room temperarature using UV-VIS spectroscopy (UV-2501PC, Shimadzu).

HPLC for carotenoid and retinal quantification

Carotenoids were extracted from cell pellets of 0.5 mL cultures in 100 µL acetone (semi-quantitative HPLC) or from 10 mL cultures in 300 µL acetone (quantitative HPLC). Extraction was performed for 30 minutes at 55°C.

Retinal was detected based on a protocol adapted from (Chen et al., 2016). Cell pellets from 10 mL cultures were resuspended in 1 mL in a 1:1 mix of methanol and 1M aqueous hydroxylamine. Suspension were incubated for bleaching at 30 °C for 15 minutes. Retinaloxime was subsequently extracted with 400 µL petroleum ether (boiling point 40-60 °C) for 5 minutes, which was repeated 3 times. Extraction fractions were pooled and evaporated under N_2. The extracted material was dissolved in 300 µL acetone .

Extraction fractions were analysed using an HPLC system (Accela, Thermo scientific) equipped with a C18-column (Grace™ GraceSmart™ RP18). The HPLC protocol was adapted from (Martinez et al., 2007); the mobile phase (0.5 mL/min) was starting with eluent A for 6 minutes, followed by a gradient to 100% eluent

B in 19 minutes and 14 minutes of eluent B. Eluent A contains methanol:water in 3:1 volume ratio, eluent B consists of methanol:dichloromethane in 4:1 volume ratio. Carotenoids were detected by UV-VIS detection at 460 nm and retinaloxime was detected at 386 nm.

Concentrations were calibrated with standard curves of acetone-dissolved β-carotene, canthaxanthin (Sigma-Aldrich) and echinenone was a kind donation of DSM. A standard curve for retinaloxime was obtained by bleaching all-trans retinal (Sigma) with hydroxylamine in a similar way as described for the cell extracts above.

Light-driven ATP production assay

Cells from 24 hour cultures were resuspended in PBS to an OD600 of ~0.5 for 2-4 hours of starvation at 37°C, 180 rpm, in dark conditions. Then, 1 mL aliquots of these starved cultures are added to transparent tubes, some aliquots are exposed to green (540 nm) or blue (447 nm) LED light (LED Light Source SL 3500, Photon Systems Instruments) at a light intensity of 100 µmol photons m^{-2} s^{-1}, measured by a light meter (LI-250A, LI-COR). Control cultures are tubes wrapped in aluminum foil to keep them in dark. All tubes were kept at 37°C in an air-temperature controlled incubator.

In order to measure ATP content of the samples, a luciferase assay is used (BacTiter-Glo™, Promega) as performed before (Martinez et al., 2007). Technical triplicates of 100 µL of dark and light exposed cell cultures are added to a white 96-well microtiter plate (NunclonTM Delta Surface, Thermo Scientific) and directly 100 µL of the BacTiter-Glo reagent is added to each well. The plate is left for exactly 10 minutes and measured for luminescence in a plate reader, with an integration time of 1s and gain 135 (Synergy MX BioTek).

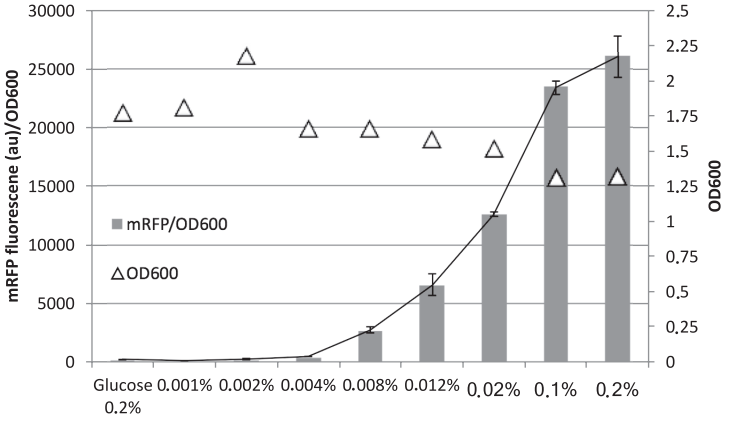
Author contributions

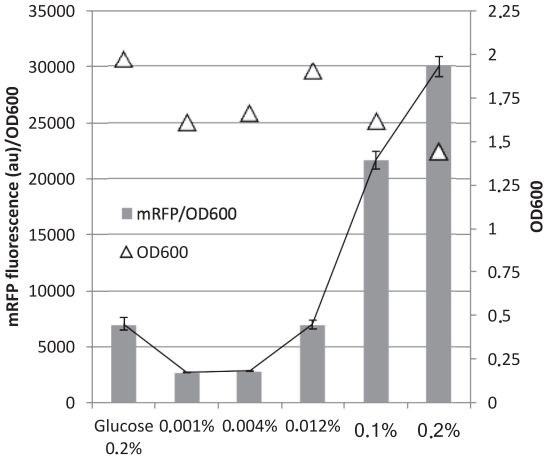
NJC, VAPMdS, WMdV and JvdO conceived the idea and designed this study. NJC, SvV, MVO, SC performed the experimental work and data analysis, where NJC coordinated and supervised the work of the latter three. NJC drafted the manuscript, which was critically revised by VAPMdS, WMdV and JvdO.

Acknowledgements

We would like to acknowledge Daniel Gerngross, Markus Jeschek and Sven Panke for their great help with generating the RedLibs libraries. Echinenone was a kind donation of Dr Luis Pasamontes from DSM-DNP. BlgBrick plasmids were a gift from Jay Keasling and BIOFAB plasmids were a gift from Drew Endy, all obtained through Addgene. pGFPe was a gift from Jan-Willem de Gier.

Supplementary information





Supplementary Figure S1. mRFP production and cell density from L-rhamnose inducible BglBrick vectors. a| pBbS9A-mRFP expression, a low-copy number plasmids (SC101 ori) in *E. coli* MG1655. b| pBbB9A-mRFP expression, a broad-host range medium copy number plasmid (pBBR1 ori). Expression was tested in *E. coli* MG1655 for a glucose and a range of L-rhamnose concentrations in M9 minimal medium with 2% glycerol. All data are from biological duplicates

Supplementary Table S1. Other plasmids construct in this study

Plasmid name	*Antibiotic marker*	*Origin of replication*	*Important components*	*Reference*
pBbE0A-RFP	Amp	*colE1*	*mRFP*	(Lee et al., 2011)
pBbE0A-GR-GFP	Amp	*colE1*	*GR-GFP-his fusion protein*	This work
pGFPe	Kan	*pBR322*	*pET28 derived for N-term GFP-his fusions*	(Drew et al., 2006)
pGFPe-TR	Kan	*pBR322*	P_{T7}, *TR-GFP-his fusion*	GeneArt/this work
pBbE0A-crtEBIY	Amp	*colE1*	*RBS + crtEBIY*	This work
pBbE0A-crtO	Amp	*colE1*	*RBS + crtO*	This work
pBbE0A-blh	Amp	*colE1*	*RBS + blh*	This work
pBbB1A-RFP	Amp	*BBR1*	*pTrc, RFP*	(Lee et al., 2011)
pBbS1A-RFP	Amp	*SC101*	*pTrc, RFP*	(Lee et al., 2011)
pBbS0A-RFP	Amp	*SC101*	*RFP*	(Lee et al., 2011)
pFAB3912	Kan	*P15a*	*P14, BCD8, RFP*	(Mutalik et al., 2013b)
pFAB3999	Kan	*P15a*	*P14, BCD23, RFP*	(Mutalik et al., 2013b)
pFAB3913	Kan	*P15a*	*P14, BCD9, RFP*	(Mutalik et al., 2013b)
pFAB3833	Kan	*P15a*	*P8, BCD1, RFP*	(Mutalik et al., 2013b)
pFAB3677	Kan	*P15a*	*P2, BCD1, RFP*	(Mutalik et al., 2013b)
pFAB3737	Kan	*P15a*	*P7, BCD1, RFP*	(Mutalik et al., 2013b)
pFAB3689	Kan	*P15a*	*P3, BCD1, RFP*	(Mutalik et al., 2013b)
pFAB3893	Kan	*P15a*	*P13, BCD1, RFP*	(Mutalik et al., 2013b)

Supplementary table S2. Oligos used in this study

Oligo nr.	*Sequence*	*Purpose*
BG4710	CAAGCTGCGGGGATTCCCGTCTTACTG	Del BamHI site crtEBIY fw
BG4711	CAGTAAGACGGGAATCCCCGCAGCTTG	Del BamHI site crtEBIY rv
BG4712	TATATAGATCTAGGAGGTACTAGATGGGTCTGATGCTGATTGATTG	Blh fw, BglII
BG4713	ATTTACTCGAGTTTGGATCCTTATTAGTTTTTGATTTTGATGCGGC	Blh rv, XhoI
BG4714	TATATAGATCTAGGAGGTACTAGATGACGGTCTG	crtEBIY fw, BglII
BG4715	ATTTACTCGAGTTTGGATCCTTATTAACGATGAGTCGTCATAATGG	crtEBIY rv, XhoI
BG4760	GCTCGAGATCTAGGAGGTACTAGATGATTACC	crtO fw, BlgII
BG4761	GGAGCCTCGAGTTTGGATCCTTACCAAAAACGACGCTGCTGTTTC	crtO rv, XhoI
BG5971	TTTTTTGGTCTCAAATGCTGATGACCGTTTTT	GR-GFP Fw amplification, BsaI
BG5972	TTTTTGGTCTCCTCAGTGGTGGTGGTGGTG	GR-GFP Rv amplification, BsaI
BG5973	TTTTTGGTCTCNCTGAGGATCGGTTGTCGAGTAAGG	pFAB Fw, BsaI
BG5975	TTTTTTGGTCTCCCATTAGAAACGGTCCGATG	pFAB3912 Rv, BsaI
BG5976	TTTTTTGGTCTCCCATTAGAAACTGTCCTTCGC	pFAB 3999 Rv, BsaI
BG5978	TTTTTTGGTCTCCCATTAGAAAGACTCCTCTGCA	pFAB 3913 Rv, BsaI
BG5981	TTTTTTGGTCTCCCATTAGAAAGTCTCCTGTGCA	All pFAB with BCD1 Rv, BsaI
BG6162	TGAGGATCGGTTGTCGAGTA	GFP deletion from GR-GFP fusion vectors Fw

Supplementary table S2. continued

Oligo nr.	*Sequence*	*Purpose*
BG6163	GCTAATCAGGCTACCACCGC	GFP deletion from GR-GFP fusion vectors Rv
BG7337	GATCCTTAGTGGTGGTGGTGGTGCTCTAACGC-CGCTTTAGCTTCC	TR-rv, Gibson
BG7340	CGTACTGAAACAAGAGACCTATAAGTAGCCAGGTCTCAAATGCG-GATGTTACCCGAA	TR-fw, Gibson
BG7341	CACCACCACCACCACT	pFAB-P14-rv, Gibson
BG7339	TGGCTACTTATAGGTCTCTTGTTTCAGTACGAAAATTGCTTTCATT	pFAB-p14-fw, Gibson
BG7505	TATATTGGTCTCAAATGCGGATGTTA	pFAB-P14-TR fw, BsaI
BG7506	CTTATAGGTCTCTTGTTTCAGTACG	pFAB-P14-TR rv, BsaI
BG7291	AACATCTTAATCATGCACAGGAGACTTTCT	BCD1-fw
BG7292	CATTAGAAAGTCTCCTGTGCATGATTAAGA	BCD1-rv
BG7293	AACATCTTAATCATGCTAAGGAGGTTTTCT	BCD2-fw
BG7294	CATTAGAAAACCTCCTTAGCATGATTAAGA	BCD2-rv
BG7295	AACATCTTAATCATGCAGGGGAGGGTTTCT	BCD5-fw
BG7296	CATTAGAAACCCTCCCCTGCATGATTAAGA	BCD5-rv
BG7297	AACATCTTAATCATGCGCCGGAGGTTTTCT	BCD6-fw
BG7298	CATTAGAAAACCTCCGGCGCATGATTAAGA	BCD6-rv
BG7299	AACATCTTAATCATGCTGGGGAGGGTTTCT	BCD7-fw
BG7300	CATTAGAAACCCTCCCCAGCATGATTAAGA	BCD7-rv
BG7301	AACATCTTAATCATGCATCGGACCGTTTCT	BCD8-fw
BG7302	CATTAGAAACGGTCCGATGCATGATTAAGA	BCD8-rv
BG7303	AACATCTTAATCATGCAGAGGAGTCTTTCT	BCD9-fw
BG7304	CATTAGAAAGACTCCTCTGCATGATTAAGA	BCD9-rv
BG7305	AACATCTTAATCATGCGGAGGATCGTTTCT	BCD10-fw
BG7306	CATTAGAAACGATCCTCCGCATGATTAAGA	BCD10-rv
BG7307	AACATCTTAATCATGCGGGGGAGTGTTTCT	BCD11-fw
BG7308	CATTAGAAACACTCCCCCGCATGATTAAGA	BCD11-rv
BG7309	AACATCTTAATCATGCTGCGGAGGGTTTCT	BCD12-fw
BG7310	CATTAGAAACCCTCCGCAGCATGATTAAGA	BCD12-rv
BG7311	AACATCTTAATCATGCAATGGAGGCTTTCT	BCD13-fw
BG7312	CATTAGAAAGCCTCCATTGCATGATTAAGA	BCD13-rv
BG7313	AACATCTTAATCATGCGGTGGAGGGTTTCT	BCD14-fw
BG7314	CATTAGAAACCCTCCACCGCATGATTAAGA	BCD14-rv
BG7315	AACATCTTAATCATGCGGGGGAGTCTTTCT	BCD15-fw
BG7316	CATTAGAAAGACTCCCCCGCATGATTAAGA	BCD15-rv
BG7317	AACATCTTAATCATGCTTAGGAGTCTTTCT	BCD16-fw
BG7318	CATTAGAAAGACTCCTAAGCATGATTAAGA	BCD16-rv
BG7319	AACATCTTAATCATGCGGAGGAGGGTTTCT	BCD17-fw
BG7320	CATTAGAAACCCTCCTCCGCATGATTAAGA	BCD17-rv
BG7321	AACATCTTAATCATGCGACGGAGCGTTTCT	BCD18-fw
BG7322	CATTAGAAACGCTCCGTCGCATGATTAAGA	BCD18-rv
BG7323	AACATCTTAATCATGCTATGGAGGTTTTCT	BCD19-fw
BG7324	CATTAGAAAACCTCCATAGCATGATTAAGA	BCD19-rv

Supplementary table S2. continued

Oligo nr.	*Sequence*	*Purpose*
BG7325	AACATCTTAATCATGCTGAGGAAAGTTTCT	BCD20-fw
BG7326	CATTAGAAACTTTCCTCAGCATGATTAAGA	BCD20-rv
BG7327	AACATCTTAATCATGCGAGGGATGGTTTCT	BCD21-fw
BG7328	CATTAGAAACCATCCCTCGCATGATTAAGA	BCD21-rv
BG7329	AACATCTTAATCATGCCTAGGAAGTTTTCT	BCD22-fw
BG7330	CATTAGAAAACTTCCTAGGCATGATTAAGA	BCD22-rv
BG7331	AACATCTTAATCATGCGAAGGACAGTTTCT	BCD23-fw
BG7332	CATTAGAAACTGTCCTTCGCATGATTAAGA	BCD23-rv
BG7333	AACATCTTAATCATGCGATGGACGGTTTCT	BCD24-fw
BG7334	CATTAGAAACCGTCCATCGCATGATTAAGA	BCD24-rv
BG7901	CAGAAGATTT DKGRGAGGACTAACATGATTACCACCGATGTTGT	crtO RedLibs Fw
BG7538	ACAACCGATCCTTACCAAAAACGACG	crtO RedLibs rv
BG7899	CATTTCAAAGGGRGDRAGTAAGTATGGGTCTGATGCTGATTG	Blh RedLibs fw
BG7900	GTCCTCYCMHAAATCTTCTGCTTAGTTTTTGATTTTGATGCGGC	Blh RedLibs rv
BG7532	TTTTGGTAAGGATCGGTTGTCGAGT	S0A-P13 vector Fw
BG7898	TTACTYHCYCCCTTTGAAATGTTTCCACACATTATACGAGCC	S0A-P13 vector rv
BG7901	CAGAAGATTTDKGRGAGGACTAACATGATTACCACCGATGTTGT	crtO RedLibs Fw

Fw: forward primer; Rv: reverse primer; P:promoter

Supplementary data 1

underlined = 5'UTR

codon-optimized GR +his

ATGCTGATGACCGTTTTTAGCAGCGCACCGGAACTGGCACTGCTGGGTAGCACCTTTGCACAG-
GTTGATCCGAGCAATCTGAGCGTTAGCGATAGCCTGACCTATGGTCAGTTTAATCTGGTGTATAACG-
CATTTAGCTTTGCCATTGCAGCAATGTTTGCAAGCGCACTGTTTTTTTTTCAGCGCACAGGCACTG-
GTTGGTCAGCGTTATCGTCTGGCCCTGCTGGTGAGCGCAATTGTTGTTAGCATTGCAGGCTATCAT-
TATTTCCGCATTTTCAATAGCTGGGATGCAGCATATGTTCTGGAAAATGGTGTTTATAGTCTGACCAGC-
GAGAAATTCAATGATGCCTATCGTTATGTTGATTGGCTGCTGACCGTTCCGCTGCTGCTGGTTGAAAC-
CGTTGCAGTTCTGACCCTGCCTGCAAAAGAAGCACGTCCTCTGCTGATCAAACTGACCGTTGCAAG-
CGTTCTGATGATTGCAACCGGCTATCCGGGTGAAATTAGTGATGATATTACCACCCGTATTATTTGGG-
GCACCGTTAGCACCATTCCGTTTGCATATATTCTGTATGTTCTGTGGGTTGAACTGAGCCGTAGCCTG-
GTTCGTCAGCCTGCCGCAGTGCAGACCCTGGTGCGTAATATGCGTTGGTTACTGCTGCTGAGCTG-
GGGTGTTTATCCGATTGCATATCTGCTGCCGATGCTGGGTGTGAGCGGCACCAGCGCAGCAGTTG-
GTGTTCAGGTTGGTTATACCATTGCAGATGTTCTGGCCAAACCTGTTTTTGGTCTGCTGGTTTTTG-
CAATTGCCCTGGTTAAAACCAAAGCAGATCAAGAAAGCAGCGAACCGCATGCAGCAATTGGTGCAG-
CAGCAAATAAAAGCGGTGGTAGCCTGATTAGCTGA

codon-harmonized TR + his

ATGCGGATGTTACCCGAACTGAGCTTTGGAGAATATTGGTTAGTCTTTAACATGCTGAGCCT-
GACCATTGCGGGCATGTTAGCGGCGTTTGTCTTTTTTCTGTTAGCTCGGAGCTATGTGGCGCCG-
CGTTATCATATTGCGCTGTATCTGAGCGCGCTGATTGTCTTCATTGCGGGCTATCATTATTTAAG-

GATTTTCGAAAGCTGGGTGGGCGCGTATCAGTTACAGGATGGCGTATATGTGCCCACTGGCAAAC-
CGTTTAACGATTTTTATCGTTATGCGGATTGGCTGCTGACCGTGCCGTTACTGCTGTTAGAACT-
GATTTTAGTCCTAGGTCTTACCGCTGCGCGTACCTGGAACCTAAGCATTAAACTTGTGGTGGCGT-
CAGTCTTAATGTTAGCGCTTGGCTATGTGGGAGAAGTGAACACTGAACCGGGACCGCGGACCTTA-
TGGGGCGCGTTAAGCAGCATACCGTTTTTTTATATTCTGTATGTGCTGTGGGTGGAATTAGGTCAGG-
CGATTCGCGAAGCTAAATTTGGTCCGCGGGTGTTAGAATTATTAGGTGCGACCCGTCTGGTCCTGT-
TAATGAGCTGGGGTTTTTATCCGATTGCGTATGCGTTAGGTACCTGGCTGCCGGGAGGCGCTGCG-
CAGGAAGTGGCGATTCAGATAGGTTATAGCCTTGCTGATTTAATTGCGAAACCGATTTATGGTTTATT-
AGTCTTTGCGATTGCGCGCGCGAAAAGCCTGGAAGAAGGTTTTGGTGTGGAAGCTAAAGCGGC-
GTTAGAGCACCACCACCACCACCACTAA

BglBrick RBS + codon-optimized blh

GAATTCATGAGATCTTTTAAGAAGGAGATATACAAATGGGTCTGATGCTGATTGATTGGT-
GTGCACTGGCACTGGTTGTTTTTATTGGTCTGCCGCATGGTGCACTGGATGCAGCAATT-
AGCTTTAGCATGATTAGCAGCGCAAAACGTATTGCACGTCTGGCAGGTATTCTGCT-
GATTTATCTGCTGCTGGCAACCGCATTTTTTCTGATTTGGTATCAGCTGCCTGCAT-
TTAGCCTGCTGATCTTTCTGCTGATTAGCATTATTCATTTTGGCATGGCCGATTTTAATG-
CAAGCCCGAGCAAACTGAAATGGCCTCATATCATTGCACATGGTGGTGTTGTTACCGT-
TTGGCTGCCGCTGATTCAGAAAAATGAAGTTACCAAACTGTTCAGCATTCTGACCAAT-
GGTCCGACCCCGATTCTGTGGGATATCCTGCTGATTTTTTTCCTGTGTTGGAGCATTGGT-
GTTTGCCTGCATACCTATGAAACCCTGCGTAGCAAACATTATAACATTGCCTTTGAACT-
GATCGGCCTGATTTTTCTGGCATGGTATGCACCGCCTCTGGTTACCTTTGCAACCTAT-
TTTTGTTTTATTCATAGCCGTCGCCACTTTAGCTTTGTTTGGAAACAGCTGCAACA-
CATGAGCAGCAAAAAAATGATGATTGGTAGCGCCATTATTCTGAGCTGTACCAGCT-
GGCTGATTGGTGGTGGTATTTACTTTTTCCTGAACAGCAAAATGATTGCAAGCGAAG-
CAGCACTGCAAACCGTGTTTATTGGCCTGGCAGCACTGACCGTTCCGCACATGATTCT-
GATCGATTTTATCTTTCGTCCGCATAGCAGCCGCATCAAAATCAAAAACTAAGGATC-
CAAACTCGAG

BglBrick RBS + codon-optimized crtO

GAATTCATGAGATCTTTTAAGAAGGAGATATACAAATGATTACCACCGATGTTGTTATTATTGGCG-
CAGGTCATAATGGTCTGGTTTGTGCAGCCTATCTGCTGCAACGTGGTCTGGGTGTTACCCTGCT-
GGAAAAACGTGAAGTTCCGGGTGGTGCAGCAACCACCGAAGCACTGATGCCGGAACTGAGTC-
CGCAGTTTCGTTTTAATCGTTGTGCAATCGATCACGAGTTTATCTTTCTGGGTCCTGTTCTGCAA-
GAACTGAATCTGGCACAGTATGGTCTGGAATACCTGTTTTGTGATCCGAGCGTTTTTTGTCCG-
GGTCTGGATGGTCAGGCATTTATGAGCTATCGTAGCCTGGAAAAAACCTGTGCACATATTG-
CAACCTATAGTCCGCGTGATGCAGAAAAATATCGTCAGTTTGTGAACTATTGGACCGATCTGCT-
GAATGCCGTTCAGCCTGCATTTAATGCACCGCCTCAGGCACTGCTGGACCTGGCACTGAATTATG-
GTTGGGAAAATCTGAAAAGCGTTCTGGCCATTGCAGGTAGCAAAACCAAAGCACTGGATTTTAT-
TCGTACCATGATTGGTAGCCCTGAAGATGTTCTGAATGAATGGTTTGATAGCGAACGTGTTAAAG-
CACCGCTGGCACGTCTGTGTAGCGAAATTGGTGCACCTCCGAGCCAGAAAGGTAGCAGCAGCG-
GTATGATGATGGTTGCAATGCGTCATCTGGAAGGTATTGCACGTCCGAAAGGTGGTACAGGTG-
CACTGACCGAAGCCCTGGTTAAACTGGTTCAGGCACAGGGTGGTAAAATCCTGACCGATCAGA-
CAGTTAAACGTGTTCTGGTTGAAAATAATCAGGCCATTGGTGTTGAAGTTGCAAATGGTGAA-
CAGTACCGTGCAAAAAAAGGTGTGATTAGCAATATTGATGCCCGTCGTCTGTTTCTGCAACTG-
GTTGAACCGGGTGCACTGGCAAAAGTTAATCAGAATCTGGGTGAACGTCTGGAACGTCGTAC-
CGTTAATAACAATGAAGCCATCCTGAAAATTGATTGTGCACTGAGCGGTCTGCCGCATTTTAC-
CGCAATGGCAGGTCCAGAGGACCTGACAGGCACCATTCTGATTGCAGATAGCGTTCGTCATGTT-
GAAGAAGCACACGCACTGATTGCACTGGGTCAGATTCCAGATGCAAATCCGAGCCTGTATCTG-
GATATTCCGACCGTTCTGGACCCGACAATGGCACCGCCTGGTCAGCATACCCTGTGGATTGAGT-
TTTTTGCTCCGTATCGTATTGCAGGTCTGGAAGGCACAGGTCTGATGGGTACAGGCTGGACCGAT-

GAACTGAAAGAAAAAGTTGCAGATCGCGTGATTGATAAACTGACCGATTATGCACCGAATCT-
GAAATCACTGATTATTGGTCGTCGTGTTGAAAGTCCGGCAGAACTGGCACAGCGTCTGGGTAGC-
TATAATGGTAATGTTTATCACCTGGATATGAGCCTGGATCAGATGATGTTTCTGCGTCCGCTGCCG-
GAAATTGCAAATTATCAGACCCCGATCAAAAATCTGTATCTGACAGGTGCAGGCACCCATCCGG-
GTGGCAGCATTAGCGGTATGCCTGGTCGTAATTGTGCACGTGTTTTTCTGAAACAGCAGCGTCGT-
TTTTGGTAA

BglBrick P_{RhaBAD} promoter

GACGTCTTGGAACCTCTTACGTGCCGATCAAGTCAAAAGCCTCCGGTCGGAGGCTTTT-
GACTTTCTGCTATGGAGGTCAGGTATGATTTAAATGGTCAGTATTGAGCGATATCTAGGAAAT-
TCGTCCACCACAATTCAGCAAATTGTGAACATCATCACGTTCATCTTTCCCTGGTTGCCAATGG-
CCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAATTGAGGCGCTTTTTAGACTGGTCGTAGG-
GAGACCACAACGGTTTCCCTCTAGTAATAATTTTGTTTAACGCGAATT

Chapter 5

Codon bias as a means to fine-tune gene expression

Tessa E.F. Quax[1,2*], Nico J. Claassens[1*], Dieter Söll[3], and John van der Oost[1]

*contributed equally
[1] Laboratory of Microbiology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, Netherlands
[2] Institut für Biologie II- Albert Ludwig Universität Freiburg , Schänzlestrasse 1, 79104 Freiburg, Germany
[3] Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut, United States of America

Chapter adapted from publication:
Molecular Cell (2015) 59:149-161 DOI: 10.1016/j.molcel.2015.05.035

Abstract

The redundancy of the genetic code implies that most amino acids are encoded by multiple synonymous codons. In all domains of life, a biased frequency of synonymous codons is observed at genome level, in functionally related genes (e.g. in operons), and within single genes. Other codon bias variants include biased codon pairs and codon co-occurrence. Although translation initiation is the key step in protein synthesis, it is generally accepted that codon bias contributes to the translation efficiency by tuning the elongation rate of the process. Moreover, codon bias plays an important role in controlling a multitude of cellular processes ranging from differential protein production to protein folding. Here we review currently known types of codon bias and how they may influence translation. We discuss how understanding the principles of codon bias and translation can contribute to improved protein production and developments in synthetic biology.

Keywords
Codon bias; Gene expression; Protein folding; Heterologous protein production

Introduction

The central dogma of molecular biology concerns the general principle of protein expression: DNA is transcribed to messenger RNA (mRNA), which is translated to protein. The key molecules of translation are the set of transfer RNAs (tRNAs), each providing a direct, specific link between a triplet of nucleotides and the corresponding amino acid. Ribosomes are the engines of translation that accommodate the tRNAs and mRNA (**Figure 1**). Deciphering the genetic code revealed that 61 codons (triplets) encode the standard 20 amino acids, whereas the remaining three are translation stop signals. The genetic code is nearly universal, meaning that almost all organisms use exactly the same codons for a specific amino acid. Because 18 out of 20 amino acids are encoded by multiple synonymous codons, the genetic code is called 'degenerate'.

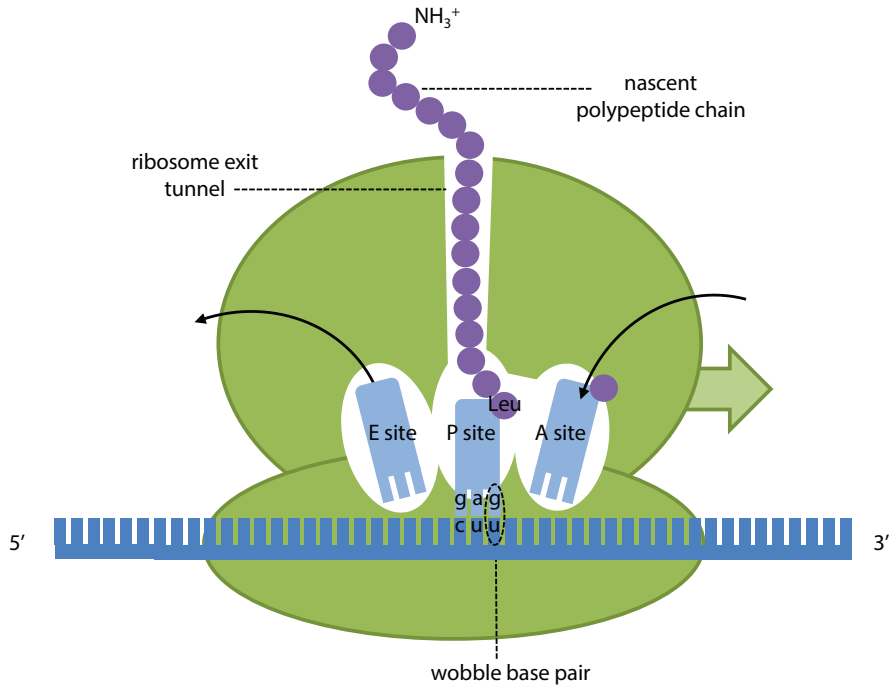


Figure 1. Translation in the ribosome and tRNA structure. Cartoon of the ribosome during translation of a wobbling codon-anticodon base-pair encoding a leucine amino acid. A-site: aminoacyl-tRNA site, P-site: peptidyl-tRNA site, E-site: exit site.

As synonymous mutations do not affect the identity of the encoded amino acid, they were originally thought to have no consequences for protein function or organismal fitness, and were therefore regarded as 'silent mutations'. However, comparative sequence analysis revealed a non-random distribution of synonymous codons in genes of different organisms. Each organism seems to prefer a different set of codons over others; this phenomenon is called codon bias (Sharp and Li, 1986).

Several important variations of codon bias have recently been discovered, such as the existence of a ramp of rare, slowly translated codons at the 5'-end of protein-coding sequences (Tuller et al., 2010), and the co-occurrence of certain codons (Cannarozzi et al., 2010). Apart from directly affecting general protein expression levels, it has been established that codon bias also influences protein folding (Pechmann and Frydman, 2013) and differential regulation of protein expression (Gingold et al., 2014). In addition to *in silico* analyses of codon bias, the development of ribosome density profiling has allowed experimental monitoring of the translation elongation rate at single-codon resolution (Ingolia, 2014). Partly based on ribosome profiling data, some studies conclude that codon bias plays an important role in translation efficiency. Alternatively, however, it is concluded that translation efficiency relies on other features of the coding sequence, such as mRNA secondary structure (Kudla et al., 2009a), and the presence of Shine-Dalgarno-like sequences (Li et al., 2012a).

Here we provide a comprehensive overview of distinct variations of codon bias. We discuss how codon bias can tune expression at multiple levels: genome, operon and gene level. Furthermore, we discuss how rules for codon bias may be further elucidated and applied to improve engineering projects, ranging from the biotechnological production of single proteins to more complex synthetic biology endeavors.

General codon bias variants

By the end of the 1970s, the development of DNA sequencing enabled comparisons of the rapidly growing number of gene sequences. Striking differences were observed between the preference of distinct organisms to

use certain synonymous codons over others (Grantham et al., 1980). It did not take long to discover that codon usage also differs between genes within one genome (Ikemura, 1985). Soon after, metrics for the frequency of optimal codons were proposed, such as the commonly used codon adaptation index (CAI) (Sharp and Li, 1987). The CAI for a certain organism is based on the codon usage frequency in a reference set of highly expressed genes, such as the ones encoding ribosomal proteins. The CAI for a specific gene can be determined by comparing its codon usage frequency to this reference set.

Wobbling and tRNA modifications

Analysis of the tRNA content of organisms in all domains of life showed that they never contain a full set of tRNAs with anticodons complementary to the 61 different codons; for example, 40 tRNAs with distinct anticodons are present in the bacterium *Escherichia coli*, 45 in the archaeon *Sulfolobus solfataricus* and 51 in the eukaryote *Homo sapiens* (**Table 1**). In some *Mycoplasma* species and related species the smallest sets are found, consisting of only 28 tRNAs with distinct anticodons (Grosjean et al., 2010). Translation of multiple synonymous codons by a single tRNA has been demonstrated to occur by wobble base-pairing: standard Watson-Crick base-pairing (A-U, G-C) is required at the first and second position of a codon, 'wobbling' (e.g. G-U) is allowed at the third position of a codon (corresponding to the 5' position of the anticodon, i.e. position 34 of a tRNA) (Crick, 1966; Soll et al., 1966). However, the affinity by which synonymous codons are recognized via wobble base-pairing is not similar. For instance, tRNAs with G in the 5' position of the anticodon have a higher binding affinity for C-ending codons than for U-ending codons (Crick, 1966; Soll et al., 1966).

Table 1. Codon frequencies and tRNA gene copy numbers for *E. coli (Eco), S. solfataricus (Sso)* and *H. sapiens (Has)*. Overview of codon frequency usage in all coding sequences and tRNA gene copy numbers for *E. coli* - K12, *S. solfataricus,* and *H. sapiens* (data obtained from: http://gtrnadb2009.ucsc.edu/), *suppressor tRNAs suppressing the stop codon, **tRNAs incorporating selenocysteine amino acid into selenoproteins, and suppressing the TGA stop codon.

		Second Base Codon																
		T				C				A				G				
		amino acid	*Eco* codon %	*Sso* codon % (tRNA #)	*Hsa* codon %	amino acid	*Eco* codon % (tRNA #)	*Sso* codon % (tRNA #)	*Hsa* codon % (tRNA #)	amino acid	*Eco* codon % (tRNA #)	*Sso* codon % (tRNA #)	*Hsa* codon % (tRNA #)	amino acid	*Eco* codon % (tRNA #)	*Sso* codon % (tRNA #)	*Hsa* codon % (tRNA #)	
First Base Codon		Phe	2.22	2.63	1.76	Ser	0.80	1.50	1.52	Tyr	1.60	3.10	1.22	Cys	0.50	0.40	1.06	T
			(0)	(0)	(0)		(0)	(0)	(11)		(0)	(0)	(1)		(0)	(0)	(0)	
			1.65	1.80	2.03		0.86	0.76	1.77		1.22	1.69	1.53		0.64	0.19	1.26	C
	T		(2)	(1)	(12)		(2)	(1)	(0)		(3)	(1)	(14)		(1)	(1)	(30)	
		Leu	1.39	4.11	0.77		0.71	1.55	1.22	Stop	0.21	0.18	0.10	Stop	0.09	0.10	0.16	A
			(1)	(1)	(7)		(1)	(1)	(5)		(0)	(0)	(2)*		(1)**	(0)	(3)**	
			1.36	1.59	1.29		0.89	0.48	0.44		0.03	0.07	0.08	Trp	1.53	1.05	1.32	G
			(1)	(1)	(7)		(1)	(1)	(4)		(0)	(0)	(1)*		(1)	(1)	(9)	
			1.10	1.52	1.32	Pro	0.70	1.23	1.75	His	1.29	0.82	1.09	Arg	2.09	0.17	0.45	T
			(0)	(0)	(12)		(0)	(0)	(10)		(0)	(0)	(0)		(4)	(0)	(7)	
			1.11	0.70	1.96		0.55	0.55	1.98		0.97	0.46	1.51		2.2	0.06	1.04	C
			(1)	(1)	(0)		(1)	(1)	(0)		(1)	(1)	(11)		(0)	(1)	(0)	
	C		0.39	1.91	0.72		0.84	1.60	1.69	Gln	1.54	1.56	1.23		0.35	0.13	0.62	A
			(1)	(1)	(3)		(1)	(1)	(7)		(2)	(1)	(11)		(0)	(1)	(6)	
			5.29	0.55	3.96		2.32	0.41	0.69		2.89	0.53	3.42		0.54	0.05	1.14	G
			(4)	(1)	(10)		(1)	(1)	(4)		(2)	(1)	(20)		(1)	(1)	(4)	

Third Base Codon

A	Ile	3.04	3.36	1.60	Thr	0.89	2.01	1.31	Asn	1.77	3.28	1.70	Ser	0.87	1.66	1.21	T
		(0)	(0)	(14)		(0)	(0)	(10)		(0)	(0)	(2)		(0)	(0)	(0)	
		2.52	1.11	2.08		2.34	0.68	1.89		2.16	1.67	1.91		1.60	0.73	1.95	C
		(3)	(1)	(3)		(2)	(1)	(0)		(4)	(1)	(32)		(1)	(1)	(8)	
		0.43	4.94	0.75		0.70	1.38	1.51	Lys	3.37	3.96	2.44	Arg	0.20	2.52	1.22	A
		(0)	(0)	(5)		(1)	(1)	(6)		(6)	(1)	(16)		(1)	(1)	(6)	
	Met Start	2.78	2.07	2.20		1.44	0.64	0.61		1.03	3.76	3.19		0.11	1.76	1.20	G
		(8)	(3)	(20)		(2)	(1)	(6)		(0)	(1)	(17)		(1)	(1)	(5)	
G	Val	1.83	2.76	1.10	Ala	1.53	2.24	1.84	Asp	3.22	3.42	2.18	Gly	2.48	2.19	1.08	T
		(0)	(0)	(11)		(0)	(0)	(29)		(0)	(0)	(0)		(0)	(0)	(0)	
		1.53	0.72	1.45		2.56	0.72	2.77		1.91	1.25	2.51		2.97	0.67	2.22	C
		(2)	(1)	(0)		(2)	(1)	(0)		(3)	(1)	(19)		(4)	(1)	(15)	
		1.09	2.81	0.71		2.02	1.92	1.58	Glu	3.96	3.84	2.90		0.79	2.58	1.65	A
		(5)	(1)	(5)		(3)	(1)	(9)		(4)	(1)	(13)		(1)	(1)	(9)	
		2.62	1.22	2.81		3.37	0.71	0.74		1.78	2.95	3.96		1.11	0.97	1.65	G
		(0)	(1)	(16)		(0)	(1)	(5)		(0)	(1)	(13)		(1)	(1)	(7)	

The influence of wobble base-pairing on decoding rates of codons by the ribosome is still unresolved and complex to analyze. The translation kinetics of different codon-anticodon pairs are complex as the following processes can play a role: (i) the diffusion kinetics of the matching tRNA, (ii) the relative codon-binding affinity of matching tRNAs over mismatching tRNAs (Gromadski et al., 2006), and (iii) the translocation kinetics of mRNA and tRNA through the ribosome, which is affected by anticodon-codon interactions (Khade and Joseph, 2011). Recently conflicting results were published that reported either slower (Stadler and Fire, 2011) or faster (Gardin et al., 2014) decoding of wobbling codons. This process deserves a more detailed analysis of both data sets and methods, and should also take into account the effect of tRNA modifications on wobble base-pairing.

Modified nucleotides present in tRNAs further extend the range of recognized synonymous codons by affecting the ability of these tRNAs to form wobble base-pairs (Agris et al., 2007). Some specific, key tRNA modifications are only present in some domains of life (Grosjean et al., 2010). Firstly, a key tRNA modification, present in eukaryotes and to some extent in bacteria, is the modification of adenine-34 to inosine-34, which allows non Watson-Crick pairing with adenine, cytosine and uridine. Secondly, exclusively in bacteria the key tRNA modifications of uridine-34 to hydroxy-uridine and derivatives are found, allowing for wobble pairing with adenine, guanosine and uridine. These key tRNA modifications explain many differences in the tRNAs sets that are present in archaea, bacteria and eukaryotes (**Table 1**) (Novoa et al., 2012).

Correlation of codon bias to tRNA pools

After the discovery of codon bias, a positive correlation was found between the frequency of codons and the concentration of tRNAs with complementary anticodons (**Figure 2a**) (Ikemura, 1985). This fact was established for several prokaryotes and unicellular eukaryotes (Kanaya et al., 1999). However, this correlation could initially not be identified for several, mostly multicellular eukaryotes.

To better analyze the relation between codon frequency bias and tRNA abundance in multicellular organisms, the tRNA adaptation index (tAI) was developed (dos Reis et al., 2004). This metric is based on the copy number of tRNA genes, assumed to be correlated to tRNA abundance in cells, and also

takes into account the efficiency of codon-anticodon binding, related to Crick's wobble rules (Crick, 1966). Based on computational analyses it was concluded that organisms with larger genomes have higher tRNA gene redundancy, which would decrease selection for specific codons (dos Reis et al., 2004). This explained why in multicellular organisms with larger genomes no positive correlation between codon usage and tRNA abundance could be identified in many studies. However, most studies at that time estimated tRNA abundance based on tRNA gene copy numbers.

Correlations based on tRNA copy numbers do not take into account that pools of distinct tRNAs and aminoacyl-tRNAs species are dynamic and can vary considerably in different conditions. For example, it was demonstrated by microarray analysis that tRNA expression abundance in humans varies widely among different tissues. This abundance could be statistically correlated to codon usage of highly expressed genes specific for those tissues (Dittmar et al., 2006). Furthermore, in bacteria it was found that the charging levels of different tRNAs recognizing synonymous codons vary drastically in response to amino acid starvation; i.e. while the pool of some synonymous tRNAs remains completely charged, the charged fraction of others can decline to zero (Dittmar et al., 2005; Elf et al., 2003).

So far, codon frequencies were mostly correlated to the total supply of tRNAs. However, when a more frequently used codon is recognized by a more abundant tRNA species, this codon will also compete for this tRNA with more codons. To take this into account, the normalized translation efficiency (nTE) metric was introduced, correcting for supply as well as demand rates of tRNAs (Pechmann and Frydman, 2013). The nTE considers codons to be more optimal if their relative tRNA abundance based on gene copy number (supply) exceeds their relative cognate codon usage (demand) based on codon frequencies in mRNA. Although the tAI and nTE already give a good indication of the availability of tRNAs for the translation of synonymous codons, this approximation could still be improved. The actual important value is the level of mature aminoacyl-tRNAs ready for amino acid delivery in the translation process. However, as the levels of tRNA expression and charging can undergo major fluctuations based on cellular conditions, it is not straightforward to take these values into account. In addition, it was demonstrated that codon frequency bias can be

better correlated to tRNA gene frequencies in all domains of life, if two major, domain-specific, tRNA modification types are taken into account (Novoa et al., 2012).

To summarize, highly expressed proteins are generally encoded by genes that contain a relatively high proportion of codons recognized by abundant, charged tRNAs with kinetically efficient codon-anticodon base-pairing. This explains to a large extent the observed codon frequency bias in many genes and genomes. In addition to codon frequency bias, two other general types of codon bias were identified in recent years: codon pair bias and codon co-occurrence bias; these will be discussed hereafter.

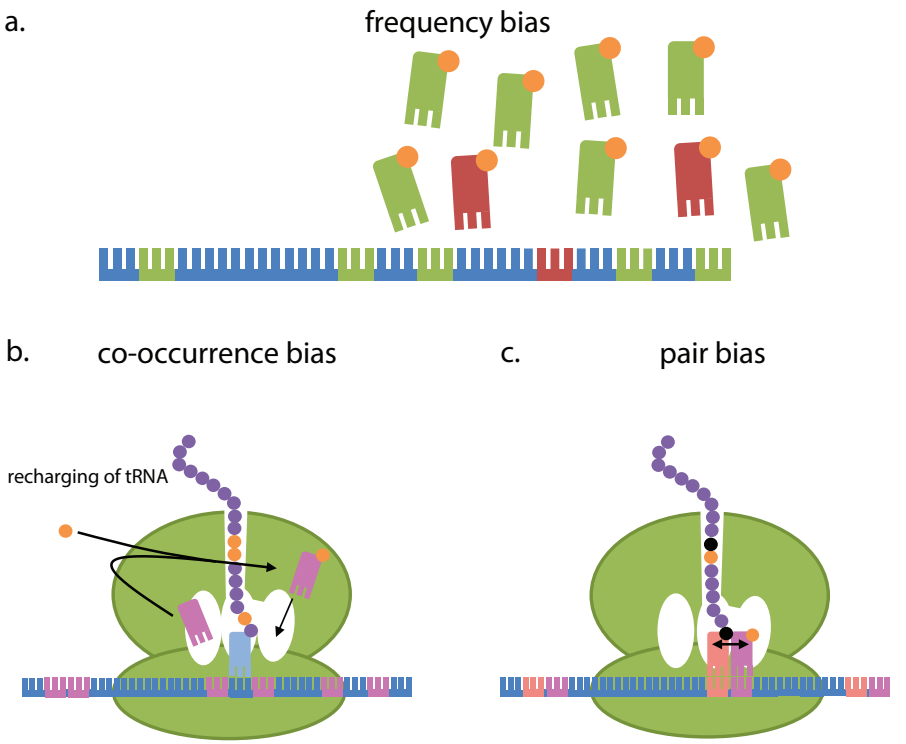


Figure 2. Different types of codon bias. a| Frequency bias will result in effective protein production when the frequency of used codons matches to the cellular tRNA population. b| Co-occurrence bias enhances protein expression, presumably due to tRNA recharging in the vicinity of the translating ribosome. c| Pair bias is probably selected because of more optimal interactions of tRNAs in the A-site and the P-site.

Synonymous codon co-occurrence bias

Recently it was shown that not only the overall frequency of synonymous codons is biased, but also the order in which they reside in a gene. While studying all coding sequences of *Saccharomyces cerevisiae,* a bias was revealed of clustered synonymous codons, called codon co-occurrence bias. Instead of a random distribution of synonymous codons on a coding sequence, there is a bias to cluster those synonymous codons that are recognized by the same tRNA, i.e. identical codons and isoaccepting codons (Cannarozzi et al., 2010) (**Figure 2b**). The effect of co-occurrence bias involves both frequent and rare codons and is most prominent in highly expressed genes that have to be rapidly induced, such as those involved in stress response (Cannarozzi et al., 2010).

It has been suggested that tRNAs remain in proximity to the translating ribosome after their exit from the E site, and that they are subsequently recharged by the corresponding aminoacyl-tRNA synthetases that somehow co-localize with the ribosome (Cannarozzi et al., 2010; Godinic-Mikulcic et al., 2014). At the next occurrence of the same or isoaccepting codon, the charged tRNA would be readily available for translation; this would have a positive effect on translation efficiency (Cannarozzi et al., 2010).

Co-occurence bias has been demonstrated in eukaryotes, bacteria and archaea (Cannarozzi et al., 2010; Shao et al., 2012; Zhang et al., 2013). However, co-occurrence of identical codons is strongly biased in all domains of life, while co-occurrence of non-identical isoaccepting codons is less prominent in prokaryotes than in eukaryotes (Shao et al., 2012; Zhang et al., 2013). The fact that co-occurrence of non-identical isoaccepting occurs more in eukaryotes most likely correlates with differences in affinity of codon-anticodon pairs between eukaryotes and prokaryotes (Shao et al., 2012). Domain-specific key modifications of tRNA result in differences in affinities of wobble base-pairing for certain synonymous codons. It has been hypothesized that only non-identical codon pairs that are recognized by a tRNA with similarly high affinity may result in co-occurrence bias (Shao et al., 2012). The described findings demonstrate that the use of identical and some isoaccepting codons in close proximity is generally advantageous for the translation process.

Non-synonymous codon pair bias

In addition to codon frequency and co-occurrence, also the context in which a codon resides is under selective constraint. Nucleotides neighboring a particular codon are distributed in a non-random manner (Buchan et al., 2006; Gutman and Hatfield, 1989). This phenomenon is called codon-pair bias (**Figure 2c**). For example, there are eight possible codon pairs to encode the adjacent amino acids alanine and glutamate. Based on codon frequencies, one would expect these amino acids to be encoded equally by GCC-GAA and GCA-GAG codon pairs. However, in humans the GCC-GAA pair is heavily under-represented compared to the expected frequency, even though it contains GCC, the most prevalent codon for alanine (Coleman et al., 2008). Some codon pairs are universally avoided or preferred; e.g. nnUAnn codon pairs are usually underrepresented while nnGCnn codon pairs are most preferred (Tats et al., 2008).

Although the exact mechanism by which codon pair bias might enhance translation efficiency is currently not well understood, it is assumed that tRNAs in the A and P site of the ribosome can interact and as such influence the efficiency of the translation process (**Figure 2c**) (Buchan et al., 2006). Several viral genomes also contain codon-pair bias which generally matches that of their host. Modification of this codon-pair usage in virulence-related genes of viruses and has been presented as an elegant strategy to produce vaccines with attenuated viruses (Coleman et al., 2008). However, it was recently suggested that this attenuation may be caused by an increased CpG and UpA dinucleotide bias rather than by a changed codon pair bias, since these dinucleotides are generally used at a low frequency in RNA and small DNA viruses infecting mammals and plants (Tulloch et al., 2014).

Translation efficiency correlated to codon bias?

The efficiency of translation, and the resulting protein production, is determined by both translation initiation and elongation rates. While the translation initiation rate controls how often a transcript is translated, the translation elongation rate controls the speed of this translation process.

Influence of the coding sequence on translation initiation

For initiating translation, ribosomes need to be sequestered on the mRNA and the start codon has to be recognized. This initiation process is facilitated by regulatory sequences. In prokaryotes, the binding strength between the Shine-Dalgarno sequence upstream of the coding sequence in the mRNA and the anti-Shine-Dalgarno sequence in the 16S rRNA regulate the efficiency of translation initiation. In eukaryotes, the Kozak sequence around the start codon is involved in interaction with the pre-initiation complex for translation (Tuller and Zur, 2014). Hence, the strength of mRNA folding around the regulatory initiation sequences and start codon can influence translation initiation efficiency; these 5′ mRNA secondary structures are also partially influenced by the 5′-end of the coding sequence. By analyzing libraries of synonymous variants of reporter genes in *E. coli* and *S. cerevisiae*, it was concluded that most of the observed variation in protein expression can be explained by differences in mRNA folding in the 5′-end of the mRNA (Bentele et al., 2013; Goodman et al., 2013; Kudla et al., 2009a). However, there is a lively debate on these results, as it has been argued that the effect of mRNA secondary structures is overestimated, because the aforementioned studies mainly rely on synthetic reporter gene variants with unusually strong mRNA secondary structures (Supek and Smuc, 2010; Tuller and Zur, 2014).

Translation elongation rate controlled by codon bias?

Generally it is assumed that translation efficiency is mostly determined by translation initiation. However, at least to some extent, the translation elongation rate also appears to contribute to translation efficiency. A recent report showed that at a high translation initiation rate, a high translation elongation rate is required for optimal expression. Low expression, on the other hand, can be caused independently either by low initiation rate or by low translation rates (Chu et al., 2014). Regarding translation elongation rates, it was assumed that more frequently used codons, recognized by abundant tRNAs, result in faster translation elongation and higher translation efficiency (Berg and Kurland, 1997). This assumption was confirmed by the observation that optimizing the overall codon sequence to more frequent codons does, at least in some cases, result in increased heterologous gene expression (Gustafsson et al., 2004).

Ribosome density profiling now allows for genome-wide analysis of ribosome distribution down to a single-codon resolution (**Figure 3**) (Ingolia, 2014; Ingolia et al., 2009). Ribosome density profiles are very well suitable for obtaining snapshots of genome-wide gene expression (translatome); hereby translation rate profiles within single genes can be obtained and potentially be correlated with codon usage (**Figure 3a**). Ribosome density profiles were also used to estimate intergenic translation efficiency differences (Quax et al., 2013). However, ribosome densities can only be used to compare translation efficiency of genes if the translation elongation rates for those genes are similar (**Figure 3b**) (Ingolia, 2014; Quax et al., 2013).

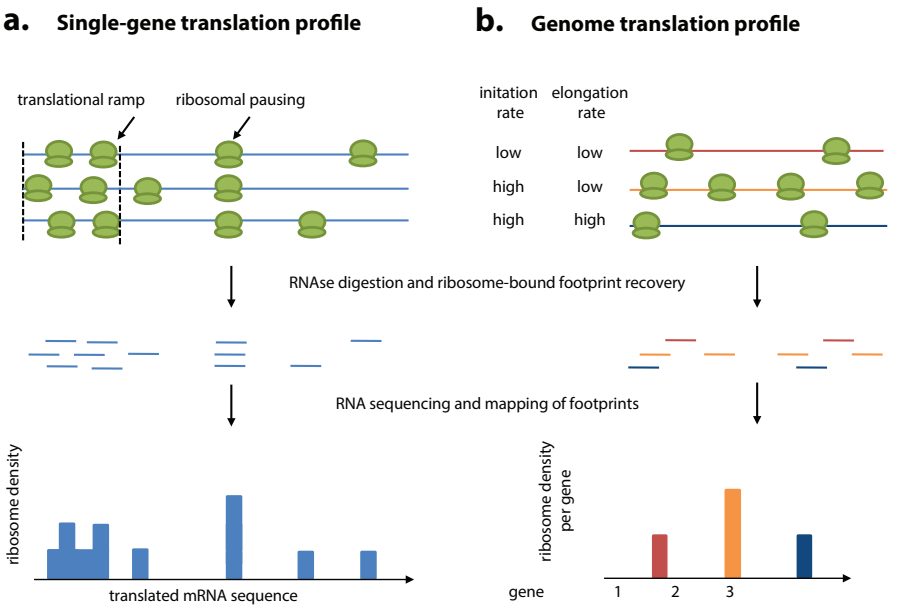


Figure 3. Ribosome density profiling to measure translation. a| Single-gene translation profiles based on ribosome density analysis are used to characterize intragenic fluctuations in translation elongation rates, e.g. corresponding to ramps of rare codons or pause sites at Shine-Dalgarno motifs. b| Genome translation profiles based on ribosome density analysis provide insight in intergenic differences in translation efficiency (translatome) by normalizing gene specific ribosome densities to their mRNA abundance. Importantly, intergenic differences in translation efficiency can be identified only if similar translation elongation rates can be assumed, e.g. based on a similar codon bias in the genes to compare).

Many studies on ribosome profiling data conclude that no correlations can be detected between high translation elongation rates and frequently used codons (Ingolia, 2014). On the other hand, correlations were found between

strong mRNA secondary structures and a local speed reduction of translating ribosomes in yeast (Pop et al., 2014; Shah et al., 2013). Furthermore, ribosomes were found to slow down at sequences encoding specific amino acid sequences, such as consecutive proline residues (Woolstenhulme et al., 2013) or positively-charged amino acids (Charneski and Hurst, 2013). In bacteria, ribosome profiling experiments identified ribosomal pausing at Shine-Dalgarno-like sequences (Li et al., 2012a). These studies reported no correlation between the codon usage frequency and ribosome speed. Some studies in eukaryotes reported that more frequent codons are translated at the same rate as rare codons (Pop et al., 2014; Qian et al., 2012). These studies suggest that codon bias is adapted to tRNA pools as a strategy to balance tRNA supply and demand, and hence to achieve optimal translation. In contrast, two recent studies concluded that rare codons with less abundant cognate tRNAs are decoded slower, thus resulting in decreased translation elongation rates (Dana and Tuller, 2014; Gardin et al., 2014). To find these conclusions these studies applied novel statistical methods for analyzing ribosome profiling data, either to remove bias caused by highly expressed genes (Gardin et al., 2014) or by extreme ribosome pause events that are unrelated to codon frequency (Dana and Tuller, 2014). The latter studies also optimized and compared different experimental protocols for ribosome profiling. Only for the mouse ribosome profiling data analyzed by the novel statistical method, still no correlation between codon decoding time and tRNA gene copy numbers (tAI) could be found (Dana and Tuller, 2014). However, this may be related to differential tRNA expression and/or charging levels in different tissues in multicellular organisms. By applying the novel statistical methods no correlation between codon decoding times and charged tRNAs could be seen (Dana and Tuller, 2014), yet aminoacyl-tRNA synthetase levels were used as an estimator for charged tRNAs, which is not a proven method to determine aminoacyl-tRNA pools.

In summary, the translation elongation rate is tuned by an interplay of multiple features of the coding sequences, including frequent and rare codon usage with their related tRNA pools, but also mRNA secondary structures, and Shine-Dalgarno-like sequences.

Intragenic codon landscapes and expression

In addition to codon frequency bias on genome level, local codon bias within genes is observed. Many genes have a locally biased distribution of rare and frequent codons, which results in a gene 'codon landscape'. As a consequence of this codon landscape, variable translation rates may occur along a gene's mRNA. Variable local translation rates can regulate the even distribution of translating ribosomes on the mRNA (Tuller et al., 2010), tune the protein co-translational folding process (Zhang et al., 2009), and facilitate protein translocation across membranes (Pechmann et al., 2014). Different types of intragenic codon landscapes that contribute to the modulation of elongation speed are discussed here.

Rare codon ramps to reduce ribosome jamming

Analysis of the distribution of frequent and rare codons within genes revealed a common pattern predominantly in highly expressed prokaryotic and eukaryotic genes: a ramp sequence immediately downstream of the start codon consisting of 30-50 relatively rare codons (Tuller et al., 2010) (**Figure 4a**). The translation rate of this codon ramp was predicted to be relatively slow, after which it would increase to reach a plateau during elongation of the remaining part of the gene. Experimental confirmation of the slower translation of the ramp came from inspection of ribosome density profiles of yeast, which revealed a higher ribosome density for the 50 5'-end codons of genes (Tuller et al., 2010). The only exception is the codon following the start codon; it is usually translated with high efficiency possibly to support fast release of the initiator $tRNA^{Met}$ (Tuller et al., 2010). The presence of the ramp is thought to increase the overall efficiency of protein synthesis. The relatively slow start of the elongation process appears to evenly space the ribosomes in order to reduce ribosome traffic jams during further elongation of highly expressed proteins with accordingly high ribosome densities on the mRNA (Tuller et al., 2010).

An alternative explanation is that the apparent ramp of rare codons is a side-effect of selection for reduced mRNA secondary structures at the 5'-end of coding sequences (Goodman et al., 2013; Shah et al., 2013). Some studies could not confirm slower translation elongation at the coding region's 5'-end; these investigators claimed this earlier observed phenomenon was either an

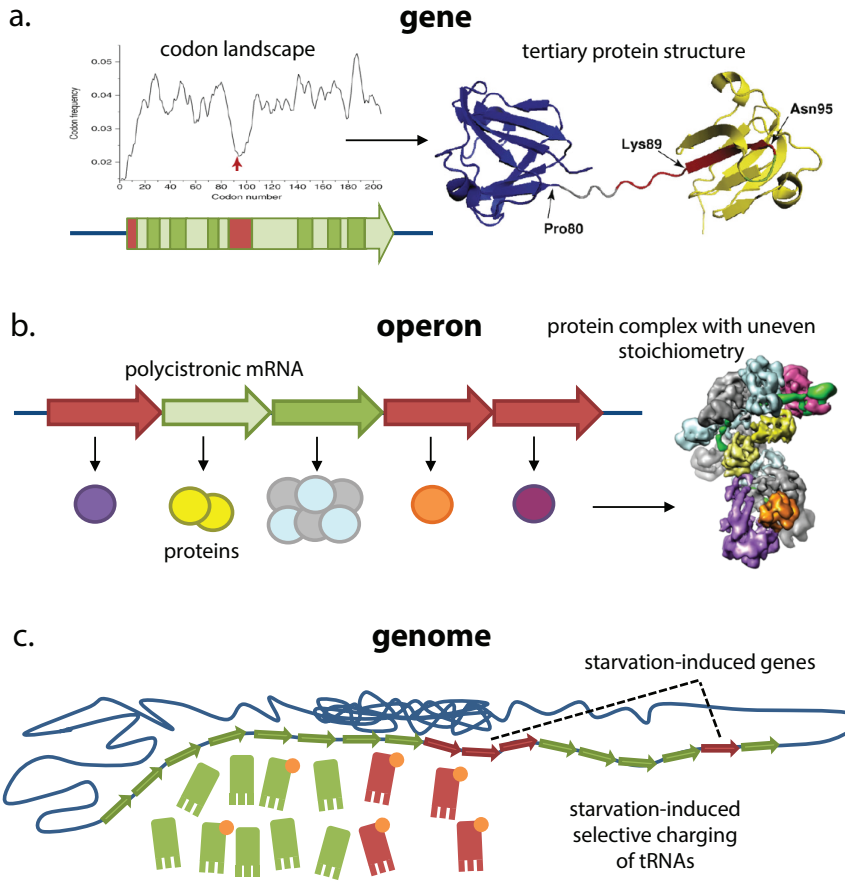


Figure 4. Codon landscapes and intergenic codon bias. a| The intragenic codon landscape can modulate the translation elongation rate, contributing to appropriate co-translational folding of the protein's secondary structure elements and domains. The red arrow indicates a stretch of rare codons encoding the red colored residues in the protein structure. Apart from a ramp sequence at 5'-end of the coding sequence, a valley in the codon landscape occurs between the two domains. The picture represents the bovine β-B2 crystallin, and is adapted from (Komar, 2009). b| Differential expression of genes residing in an operon is controlled by differential translation. This is mainly accomplished by enhanced initiation (as reflected by a higher ribosome density), and to some extend by adjusted elongation (as reflected by a more optimal codon bias). An example is shown of the operon encoding the CRISPR-associated Cascade complex from *E. coli* with uneven stoichiometry, crystal structure from (Wiedenheft et al., 2011). c| The expression of sets of functionally related genes can be co-regulated based on their codon bias. The available pools of aminoacyl-tRNAs can specifically change under certain conditions, which leads to improved translation of genes with adapted codon bias. As an example, genes encoding amino acid biosynthesis pathways are shown, which contain many rare codons translated by tRNAs that remain highly charged during amino acid starvation.

artifact of the ribosome profiling experiment (Ingolia et al., 2011) or an artifact of the data analysis because shorter genes have higher initiation rates (Shah et al., 2013). However, different experimental protocols and data normalization

for ribosome profiling did again confirm a slower translation elongation rate at the 5'-end of highly expressed genes (Tuller and Zur, 2014). Summarizing, multiple features of the 5' terminus of coding sequences play different roles in translation, which complicates efforts to separate the individual effects of each of these features on translation efficiency (Tuller and Zur, 2014).

Codon landscapes for protein translation across the membrane

Recently, rare codon clusters were identified in genes for membrane and secretory proteins; in yeast these clusters occur 35-40 codons downstream of the binding sites for the signal recognition particle (SRP) (Pechmann et al., 2014). The SRP recognizes secretory signal sequences or transmembrane protein segments as they emerge from the ribosomal exit tunnel. The 35-40 codon distance between these rare codon clusters and the SRP-binding sequence spans the length of the ribosomal exit tunnel. After SRP binding to the emerging peptide chain, the nascent chain ribosome complex is transferred to the membrane translocation machinery for co-translational transport of the protein across the membrane. The local translation slowdown by the rare codon clusters would provide additional time for the SRP to associate with the nascent chain ribosome complex and hence could facilitate the subsequent membrane translocation (Pechmann et al., 2014). Translation slowdown at these sites in yeast was confirmed by analyzing ribosome profiling data for transmembrane and secretory proteins (Pechmann et al., 2014).

Recently, ribosomes in *E. coli* were found to slow down during membrane protein targeting. This slowdown of the translation elongation was probably not caused by rare codons, but by Shine-Dalgarno-like sequences approximately 35 codons downstream of the transmembrane helices (Fluman et al., 2014). Apart from rare codons, slower translation elongation at the 5'-end of coding sequences of membranes proteins may also be caused by a bias for positively-charged amino acids found at the cytoplasmic N-terminus of membrane proteins (Charneski and Hurst, 2014).

Codon landscapes and co-translational protein folding

Clusters of rare codons within the codon landscape may modulate the translational rate to coordinate protein translation rate with co-translational protein folding (Purvis et al., 1987). A correlation between certain protein

structures and rare or frequent codons has been demonstrated experimentally for some proteins (Spencer et al., 2012; Zhang et al., 2009) and supported computationally for larger sets of proteins (Pechmann and Frydman, 2013; Saunders and Deane, 2010). Using the normalized translation efficiency (nTE) metric, it was concluded that frequent codons are depleted in regions that encode coils in protein structures. This may relate to the fact that coils contain loops that have keys roles in co-translational folding. For α-helices both enrichment of rare and frequent codon clusters is observed. At the start of α-helices a specific codon usage pattern is observed: rare codons before the helix and at positions 1 and 4, and frequent codons occur at positions 2 and 3; this may reflect the complex co-translational folding of α-helices, which already occurs within the ribosomal exit tunnel. β-sheets are encoded mainly by regions enriched in frequent codons (Pechmann and Frydman, 2013). The observed trends strongly suggest a correlation of codon bias with co-translational folding; however, complete understanding of the rules requires further analysis. In addition, it should be noted that certain structurally and/or functionally important regions (e.g. catalytic residues) may also be encoded by frequent codons not because of folding kinetics, but because these frequent codons may enhance translational accuracy of these important amino acids (Drummond and Wilke, 2008). It was experimentally demonstrated that substitution of rare codons by synonymous frequent codons can cause improper folding that results in either degradation or aggregation (inclusion bodies) (Spencer et al., 2012; Zhang et al., 2009), most likely because of a distorted balance of the protein folding process. In addition, synonymous mutations may also result in more subtle changes, probably by slightly affecting the folding process and resulting in minor changes in the protein structure or in post-translational modifications. Consequently, these 'silent' mutations may result in subtle but significant changes in protein functionality (Kimchi-Sarfaty et al., 2007; Zhou et al., 2013).

Intergenic codon bias and differential expression

Differences in codon bias do not only occur between regions within individual genes (intragenic, as discussed in the previous section), but also between sets of genes, either clustered in operons or scattered in a genome (**Figure 4b,c**).

The latter intergenic codon bias may facilitate differential expression of sets of genes. In this section some types of intergenic codon bias and related differential gene expression are discussed.

Starvation conditions and codon bias

The high number of rare codons in particular gene clusters that encode amino acid biosynthetic pathways appears counterintuitive. However, this observation makes perfect sense when the relative levels of aminoacyl-tRNA under starvation conditions are considered. These rare codons present in amino acid biosynthesis genes are recognized by tRNAs that remain relatively highly charged after starvation (Dittmar et al., 2005). Furthermore, during starvation, lowly charged tRNAs have a role in the 'transcriptional attenuation control' mechanism, found in bacterial amino acid biosynthesis operons (Elf and Ehrenberg, 2005). This control mechanism relies on the competition between ribosomes translating at the leader cistron of an mRNA and the RNA polymerase transcribing the cistrons downstream on the polycistronic mRNA. In the leader cistron of the mRNA, codons recognized by tRNAs with low charging levels during starvation are present, which during starvation results in ribosome stalling during translation of the leader. This stalling affects the secondary structure of a downstream sequence in the mRNA, and as such relieves transcription repression that results in further transcription of downstream cistrons encoding the biosynthetic pathway enzymes (Elf and Ehrenberg, 2005).

Cell cycle, differentiation and stress regulation by codon usage

Variations in tRNA expression during different states of a cell may enable differential expression of sets of genes with codons adapted to different tRNA pools. For example, concentrations of tRNAs and aminoacyl-tRNA synthetases have been demonstrated to oscillate during the human cell cycle (Frenkel-Morgenstern et al., 2012). As a consequence, gene sets that are expressed at different phases of the cell cycle have different codon usage (Frenkel-Morgenstern et al., 2012). This provides a codon bias strategy that supports cell cycle regulation. In humans and other vertebrates, tRNA concentrations differ in proliferating and differentiating cell types (Gingold et al., 2014). Genes specific for proliferation and differentiation processes have a corresponding codon bias; this implies that the two distinct translational programs that

operate during proliferation and differentiation are regulated by codon bias. The cyanobacterium *Synechoccus elongatus* uses codons to adjust its protein production to fluctuating environmental conditions. The genes encoding the circadian clock-associated oscillator proteins contain rare codons leading to low expression at low temperatures. This causes the desired suppression of circadian regulation at low temperatures (Xu et al., 2013).

Altered tRNA modification patterns present another strategy to adapt gene expression of large sets of genes to different conditions. RNA modifications can alter the codon-anticodon binding affinity of tRNA molecules. Translation of certain codons can be favored in this way, meaning enhanced expression of gene sets that contain elevated frequencies of these codons. Environmental factors have been reported to cause changes in RNA modifications. For example in yeast, stress induced by DNA damaging compounds (Begley et al., 2007) or oxidative stress (Chan et al., 2012) up-regulates specific tRNA modifying enzymes. Interestingly, in the genes required for coping with these stress factors the codons recognized by these modified tRNAs are over-represented (Begley et al., 2007; Chan et al., 2012). Thus, fluctuations of the (aminoacyl-) tRNA pools and tRNA modifications play a role in regulating expression of genes with adapted codon usage.

Differential expression within operons

As outlined in the previous sections, functionally related genes generally have a similar codon bias that allows for their co-regulation under specific conditions. However, significant differences in the degree of codon bias of related genes have been described as well. Even within prokaryotic operons, codon bias of individual genes may differ considerably. Within the ATP-synthase operon, genes encoding the highly abundant ATPase subunits were observed to be enriched in codons recognized by abundant tRNAs (Gouy and Gautier, 1982). In a recent comparative genomics analysis, the correlation between subunit stoichiometry and codon bias was demonstrated in many different operons in bacteria and archaea. These operons were selected as they encode protein complexes with established uneven subunit stoichiometry, and included highly expressed complexes (e.g. ribosome, ATPase) as well as poorly expressed complexes (e.g. CRISPR-associated Cascade complex) (Quax et al., 2013) (**Figure 4b**). Translation of several polycistronic messengers encoding protein

complexes with uneven stoichiometry were evaluated by ribosome density profiles (Li et al., 2012a); this analysis revealed that the cistrons coding for the abundant subunits had correlating high ribosome densities (Quax et al., 2013). This indicates that translation initiation of these cistrons allows for differential translation. In addition, a positive correlation was found between more highly expressed subunits and codon frequency bias and co-occurrence bias. This correlation strongly suggests a role of codon bias in tuning the elongation rate of highly expressed subunits. Differential translation is proposed as a universal control mode to tune differential production of operon encoded protein complexes with uneven stoichiometry (Quax et al., 2013). In addition to the aforementioned protein complexes, differential translation is important for other groups of (operon-encoded) related proteins of which differential production is required, including control systems and metabolic pathways (Li et al., 2014).

Selection pressure on codon usage

It is firmly established that codon usage is biased in the majority of living organisms. Two not mutually exclusive explanations on the evolution of this bias have been formulated: (i) non-randomness of mutation, and (ii) selection for codon bias (Hershberg and Petrov, 2008). Some nucleotides or codons may have higher mutation rates, resulting in lower frequencies of some codons and nucleotides. Some investigators have claimed that codon bias is mainly related to the non-random mutations caused by the global GC content of an organism, as this GC content seems determined for the complete genome, and not only for the coding part of the genome (Chen et al., 2004; Knight et al., 2001). However, as previously discussed, different types of codon usage bias were observed to be related to translation efficiencies; therefore this codon bias has to be under selective pressure during evolution and mutation rates alone cannot explain the various observations. Especially intragenic and intergenic codon bias cannot be explained by the mutational theory (Hershberg and Petrov, 2008).

Certain codons may be selected to achieve efficient and/or accurate translation, both may influence cellular fitness to a certain extent (Hershberg and Petrov, 2008). Higher translation accuracy may result from using codons that are

recognized by abundant tRNAs. Improved translation accuracy will avoid wasting resources and energy caused by production of non-functional proteins. The 'accuracy theory' is supported by the aforementioned detection of more stringent selection for codon bias at crucial positions in proteins, potentially to ensure high fidelity translation of these, structurally and/or functionally important residues (Drummond and Wilke, 2008; Stoletzki and Eyre-Walker, 2007; Zhou et al., 2009b). In addition, stronger codon bias was found for longer genes, most likely because of relatively higher resource costs for mistranslation of larger proteins (Stoletzki and Eyre-Walker, 2007).

Selection pressure on translation efficiency can act locally, because codon usage landscapes can affect protein folding. In addition, there will be selection for local coding sequence features such as mRNA secondary structures, as these allow for tuning of gene expression. Selection pressure may also act on codon usage at a global scale, since more optimal codons may give rise to higher global translation rates, thus keeping more ribosomes available and thereby increasing cellular fitness (Andersson and Kurland, 1990; Berg and Kurland, 1997). In the struggle of life that occurs in most natural ecosystems, the availability of limited resources will imply that the translation efficiency will impose a strong selective pressure on an optimal codon bias. This notion is supported by an enhanced frequency of optimal codons in highly expressed genes in rapidly growing organisms (Ran and Higgs, 2012). In summary, the different types of codon bias that have evolved in all domains of life, reflect an optimized combination of frequent and rare codons, that will allow for appropriate translation of a certain gene, in a certain organism, under certain conditions, and eventually in a certain tissue or organelle.

Applying codon bias as a means to improve protein production

Codon bias has been studied intensively during the past three decades, and some of the gained insights have been widely applied in biotechnology as a strategy to optimize gene expression for improving protein production rates and yields. The two major strategies are adjusting the expressed set of tRNAs, or adjusting the codon usage of a gene of interest.

Expressing additional tRNA genes in the production host

Additional copies of tRNA genes are introduced to enhance tRNA levels, aiming for the heterologous expression of genes that contain many rare codons. At present, this strategy has been mainly used in bacterial production systems. Several commercial bacterial strains expressing extra tRNAs from a plasmid are available for this purpose, e.g. *E. coli* Rosetta (pRARE plasmid with genes of tRNAs that recognize the following codons: AGG, AGA, AUA, CUA, CCC, GGA) and *E. coli* BL21-CodonPlus (pRIL plasmid with genes of tRNAs that recognize the following codons: AGG, AGA, AUA, CUA, CCC) (Gustafsson et al., 2004) (**Table 1**). There are many examples demonstrating that this strategy was successful, or at least resulted in improved functional protein production. However, in many other cases this approach did not enhance protein yields, for instance when slow translation of stretches of rare codons in a gene is required for proper protein folding (Zhang et al., 2009). The main problem with this strategy is that the protein-specific codon landscape (frequent as well as rare codons) is not taken into account.

Designing codon optimized genes

The decreasing costs of *de novo* DNA synthesis have allowed for synthesizing codon optimized genes. The potential space for synonymous codon substitutions throughout a gene is extremely large; for a protein of 300 amino acids more than 10^{100} different coding sequence variants are possible (Welch et al., 2009a). Therefore, automated codon optimization algorithms have been developed to design coding sequences optimized for increased expression in certain hosts (Gould et al., 2014). Most DNA synthesis companies offer codon optimization services, primarily based on confidential algorithms. Many of these algorithms optimize codon usage by maximizing the gene's CAI to match that of the expression host, along with optimizing for some sequence features. Sequence features regularly taken into account are: GC content, and avoidance of repeats and motifs such as RNAse recognition sites, transcriptional terminator sites, Shine-Dalgarno-like sequences, and sequences that lead to strong mRNA secondary structures (**Table 2**) (Gustafsson et al., 2012). There are many reports of successful codon optimization of coding sequences, which can lead up to 1000-fold increased gene expression by optimizing the coding sequence (Gustafsson et al., 2012). For expression of a large set of human genes in *E. coli* such algorithms were reported to be more successful than expressing

additional tRNAs (Maertens et al., 2010). However, for many genes expression was not improved by such algorithms (Gustafsson et al., 2012; Maertens et al., 2010). Thus, the output of these algorithms is not a guarantee for success; e.g., synonymous mutations might interfere with protein folding by altering the codon landscape.

Alternative approaches for synonymous codon design were also reported. The 'codon harmonization algorithm' adapts the codons in a way that the original codon landscape of the gene in the original host is maintained in the expression hosts (Angov et al., 2008). Hence, this algorithm retains a larger proportion of rare codons in comparison to other optimization algorithms. A successful application of this algorithm was the improved heterologous expression of a few proteins in *E. coli* (Angov et al., 2008; Spencer et al., 2012). A more systematic experimental and statistical approach to optimize coding sequences for heterologous gene expression in *E. coli* identified several crucial amino acids and specific synonymous codons that are essential for high expression. These crucial codons generally were not the most frequently used codons in highly expressed genes in the host, but some crucial codons were related to cognate tRNAs that remain highly charged under starvation conditions (Welch et al., 2009b). Evaluating different codon optimization strategies to improve protein expression is not straightforward. Firstly, generally only single case studies of optimizing coding sequences for different proteins are reported, and are thus difficult to evaluate (Welch et al., 2009a). Secondly, the many coding sequence features that influence expression (**Table 2**), and their uncertain hierarchy in different genes, in different hosts, and under different conditions pose a major challenge.

Challenges ahead

Unraveling codon bias and other factors influencing expression

A future challenge in studying the relation between coding sequences and protein production, is to perform a thorough comparative analysis of all currently known, and yet to be discovered, features of coding sequences that influence the translation process. This may be achieved by further improvements and integration of experimental approaches and statistical analyses. Experimental

RNA-sequencing data that determine mRNA and tRNA abundance, ribosome density profiling and proteomics should be integrally analyzed. To date, most coding sequence features affecting gene expression have been derived from natural expression data or overexpression of randomly generated libraries of reporter proteins. A systematic approach designing synthetic gene variants will be more efficient than generating and testing random gene variants (Gustafsson et al., 2012). The variant design should systematically vary coding sequence features that potentially affect expression, and whenever possible minimize the co-variance between individual features. Such a systematic approach would allow to reveal the effects of several coding sequence features with a relatively limited number of variants to be tested (Gustafsson et al., 2012).

Many studies that have addressed codon bias by analyzing libraries of synthetic reporter variants, mainly resulted in the identification of coding sequence features that affect translation initiation (Goodman et al., 2013; Kudla et al., 2009a). To analyze the effect of codon bias on translation elongation, it would be better to use a recently developed series of synthetic translation initiation elements that result in constant translation initiation rates independent of the 5'-end of a coding sequence (Mutalik et al., 2013b). To complement studies of overexpressed synonymous reporter variants, studies of synonymous codon variants of naturally expressed genes may reveal additional rules for translation efficiency. Chromosomal genes can be replaced by synthetic, strategically designed synonymous variants. Recently, a set of highly expressed chromosomal genes in *E. coli* has been replaced by codon-shuffled variants (Lajoie et al., 2013). However, this study was limited to only three variants of these genes, which does not yet allow for composing solid rules on how codon bias affects gene expression. Further probing the hierarchy of features hidden in the coding sequence that contribute to the efficiency of protein synthesis, requires the analysis of many different genes from different domains of life, cell types, and conditions.

Table 2. Coding sequence features relevant for synthetic gene design. Some features are already considered in several gene design algorithms, we argue that some features (bold) deserve more specific attention for synthetic gene design and/or better elucidation of their relevance for protein expression. Most factors are discussed in this review, some local sequence features, that may be avoided or specifically included in gene design were reviewed before (Gould et al., 2014; Gustafsson et al., 2012).

global host and condition related codon usage
codon usage frequence highly expression genes (CAI)
tRNA gene copy numbers (tAI)
balance tRNA supply/mRNA codon demand (nTE)
tRNA expression levels
charged tRNA levels
tRNA modification levels
synonomous codon co-occurence bias
non-synonomous codon pair bias
local gene-level codon usage landscape
rare codon ramps at 5'-end of coding sequences
codon frequencies related to protein co-translational folding
codon frequencies related to protein translcation across a membrane
features not directly related to codon usage
GC content
mRNA folding at 5'-end of coding sequences
mRNA folding throughout coding sequence
Shine-Dalgarno like sequences (prokaryotes)
transcriptional terminators
restriction sites
RNAse E sites
hidden stop codons
sequence repeats
transcription factor recognition sequences
CpG and other dinucleotide bias (eukaryotes)
polyadenylation signals (eukaryotes)
cryptic splicing signals (eukaryotes)

Further improving synthetic gene design

Natural selection formed optimal codon landscapes for different types of genes and organisms, and now the main challenge is to understand the rules on how to recreate these landscapes for high level production in heterologous production systems. The codon optimization field is gradually moving away

from the common concept that synthetic genes should contain as much as possible frequent codons in order to achieve high protein production. Many features are generally accepted to be important sequence for synthetic gene design (**Table 2**). To date several of these features have hardly been used in synthetic gene design and deserve more attention in future attempts. For example, a better experimental analysis of aminoacyl-tRNA abundance under protein overproduction conditions should provide the basis for better codon optimizations ensuring a balanced aminoacyl-tRNA supply for production of the protein of interest. This also implies that more accurate metrics than CAI, for example the nTE, for predicting optimal codon bias should be considered.
Given our deeper understanding of how the codon landscape influences the efficiency of translation and protein folding, novel design rules for synthetic genes should be considered. Especially for high level expression of secreted or membrane proteins, along with adjustment of the codon bias to the heterologous expression host, landscape features such as clusters of rare codons or other coding sequence pause sites need to be incorporated in the gene design. Further refinements of the 'codon harmonization' approach (Angov et al., 2008) would be promising for this purpose.

Applying codon bias as a tool in synthetic biology

Improved rational gene design underlies the synthetic biology attempts to create synthetic gene circuits, biosynthetic pathways, or even new genomes. In the design of circuits or pathways, expression of functionally related genes is crucial. For designing synthetic operons to express these circuits or pathways, one can learn from recently elucidated roles of differential codon bias and other factors in differentially expressed cistrons in prokaryotic operons (Li et al., 2014; Quax et al., 2013).

Synthetic biology now enables engineering at genome level. Rare codons have been replaced by synonymous codons on a genome-wide scale, and removed rare codons could be reassigned to encode non-natural amino acids (Lajoie et al., 2013; Mukai et al., 2015; Rovner et al., 2015). Furthermore, the nucleotide alphabet of life has recently been expanded by introducing two synthetic nucleotides in *E. coli* (Malyshev et al., 2014). For a sweeping modification of the codon alphabet a better understanding of codon bias will be advantageous. In addition, recent advances in DNA assembly permitted the assembly and

transplantation of complete synthetic genomes (Gibson, 2014). This technology practically enables redesigning of genomes from scratch. However, a more thorough understanding of codon bias is needed for rational design of an optimally functional synthetic genome with sensible choices of codon usage, related tRNA genes, and tRNA modifying enzymes.

Acknowledgements

The authors are grateful to Henry Grosjean for stimulating discussions. This work was supported by the Fonds voor Wetenschappelijk Onderzoek (FWO-Belgium) by an FWO Pegasus Marie-Curie grant to TQ, by the IP/OP program Systems Biology of Wageningen University to NC, by the National Institute of General Medical Sciences (GM22854) to DS, and by the Netherlands Organization for Scientific Research (NWO grant) by an ALW-TOP (854.10.003) to JO.

Chapter 6

Improving heterologous membrane protein production in *Escherichia coli* by combining transcriptional tuning and codon usage algorithms

Nico J. Claassens[1]*, Melvin. F. Siliakus[1,2]*, Sebastiaan K. Spaans[1], Sjoerd Creutzburg[1], Bart Nijsse[3], Tessa E. F. Quax[4], and John van der Oost[1]

*contributed equally
[1]Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE, Wageningen, The Netherlands
[2]present address: NIOZ, Royal Netherlands Institute for Sea Research, Department of Marine Microbiology and Biogeochemistry, 1790 AB Den Burg, Texel, the Netherlands
[3]Laboratory of Systems and Synthetic Biology, Wageningen University, Stippeneng 4, 6708 WE, Wageningen, The Netherlands
[4]Institut für Biologie II, Albert Ludwig Universität Freiburg, Schänzlestrasse 1, 79104 Freiburg, Germany.

Abstract

High-level, recombinant production of membrane-integrated proteins in *Escherichia coli* is extremely relevant for many purposes, but has also been proven challenging. Here we study a combination of transcriptional fine-tuning in *E. coli* LEMO21(DE3) with different codon usage algorithms for heterologous production of membrane proteins. The overexpression of 6 different proteins is compared for the wild-type gene codon usage variant, a commercially codon-optimized variant, and a codon-harmonized variant. We could show that transcriptional fine-tuning plays a major role in improving the production of all tested proteins. Moreover, different codon usage variants significantly improved production of several of the tested proteins. Not a single algorithm could successfully improve the membrane-integrated production of all 6 tested proteins. However, the harmonization algorithm seems generally the most robust algorithm for high membrane-embedded protein production, resulting in the highest production levels for 5 out of 6 tested proteins. In summary, this study shows the combined benefit of transcriptional tuning, and the additional smaller, but often significant, effect of different codon usage variants, including the rarely applied codon-harmonization algorithm, on membrane protein production in *E. coli*.

Keywords

Codon harmonization; Codon optimization; Membrane protein production; Proton-pumping rhodopsin; *E. coli* LEMO(21)

Introduction

Throughout the three domains of life (eukarya, bacteria and archaea), 15-30% of all genes encode integral α-helical membrane proteins (Wallin and von Heijne, 1998). This diverse group of proteins is involved in a variety of crucial processes, such as energy transduction, transport and signaling. To characterize membrane proteins, for example by biochemical assays or protein structure crystallography, overproduction of membrane proteins in recombinant hosts, such as *Escherichia coli*, is a key method. Also for metabolic engineering and synthetic biology endeavors in *E. coli* and other relevant organisms, functional, recombinant expression of membrane proteins, including transporters, sensors and enzymes is of utmost importance. Additionally, 70% of all drugs target human membrane proteins, and heterologous expression of these proteins is a crucial step in drug discovery and development (Lundstrom, 2007).

The recombinant production of membrane proteins, however, is often challenging, due to the fact that only low amounts of protein are properly folded and translocated into the membrane. Overproduced membrane proteins often end up as insoluble aggregates in the cytoplasm, accumulated in so-called inclusion bodies (Schlegel et al., 2014). For *E. coli* it has been demonstrated that this phenomenon can be partly related to the jamming of the membrane translocation systems, such as the Sec-translocon (Schlegel et al., 2014). Furthermore, systems available in the expression host for folding and membrane insertion, i.e. translocation systems and chaperones, may not be suitable to properly target, insert and/or fold some heterologous membrane proteins. To address these issues, some tools have been developed to improve membrane protein expression. Most of them for the frequently used expression host *E. coli*. Several *E. coli* strains have successfully been optimized for membrane protein expression, including the 'Walker strains', *E. coli* C41(DE3) and C43(DE3) (Miroux and Walker, 1996) and *E. coli* LEMO21(DE3) (Wagner et al., 2008). These strains are all based on reducing the high transcription rates from the T7 RNA polymerase (T7RNAP), which is commonly used in *E. coli* (DE3) strains to drive recombinant gene expression. The improved membrane protein expression levels in these Walker and LEMO strains, rely respectively on reduced expression of T7RNAP (Schlegel et al., 2015) or on fine tuning of the expression level of T7RNAP (see **Supplementary Figure S1**) (Wagner et

al., 2008). The protein production improvements of these strains are related to tuning the transcription rates of the recombinant mRNA, which can help to prevent the overload of chaperones and membrane insertion machineries.

On a translational level, codon usage also plays a key role for functional recombinant protein production. The fact that different organisms use different synonymous codons, is important to take into account when overexpressing heterologous proteins (Quax et al., 2015). To overcome problems in the expression of mostly eukaryotic genes, other *E. coli* strains have been developed, such as the Rosetta strains, which overexpress tRNA species for codons that are rare in *E. coli* (Maertens et al., 2010). In recent years, synthetic gene sequences with adapted codon usages have become another important tool to attempt to improve recombinant expression (Gustafsson et al., 2012). Hereto, typically coding regions are optimized by commercial vendors through mainly selecting codons that occur frequently in the expression host. In addition, several other factors are taken into account by many proprietary commercial, multi-parameter codon optimization algorithms. These include avoiding strong mRNA secondary structures in the 5'UTR, a desired GC-content and avoiding of certain undesired motifs, such as repeats, Shine-Dalgarno like sequences and RNase sites (Gould et al., 2014; Gustafsson et al., 2012).

Recent experimental and bioinformatics analyses of codon usage within genes have revealed that 'rare' codons can have an important role in functional production of proteins (Buhr et al., 2016; Pechmann and Frydman, 2013; Quax et al., 2015; Saunders and Deane, 2010; Spencer et al., 2012; Zhang et al., 2009). Rare codons are hypothesized to slow down translation in order to accommodate proper folding of certain protein domains, such as α-helices and β-sheets (Pechmann and Frydman, 2013). Also for membrane proteins, it has been suggested that clusters of rare codons may provide translational pauses that facilitate co-translational folding of specific domains and membrane insertion (Nørholm et al., 2012). Rare clusters of codons in genes encoding membrane proteins, e.g. in *Saccharomyces cerevisiae*, have been correlated to translocation of membrane proteins (Pechmann et al., 2014). The best algorithm so far, which takes the importance of rare codons into account, is the so-called 'codon harmonization' algorithm (Angov et al., 2008, 2011). This algorithm ensures that the frequency of a codon in the expression host, selected for the

synthetic coding sequence, is equal to the frequency of the original codon in the wild-type gene sequence in the native host. This algorithm has been applied for heterologous expression of a few eukaryotic and bacterial cytoplasmic and membrane proteins in *E. coli* and *S. cerevisiae*. In several cases it was reported that the codon harmonized variant gave increased heterologous production compared to production from the wild-type sequence variants (Angov et al., 2008; Buhr et al., 2016; Keniya et al., 2014; Sarduy et al., 2012; Van Zyl et al., 2014). Apart from causing higher expression levels, some studies that compare harmonized with wild-type or optimized gene variants report higher specific activities after expressing proteins from harmonized genes, presumably due to better folding (Buhr et al., 2016; Spencer et al., 2012; Vuoristo et al., 2015).

So far, the general performance of this codon harmonization algorithm on membrane protein production in *E. coli* has not been studied well; no studies have compared the expression several membrane proteins from different native organisms. Furthermore, to the best of our knowledge no studies compared the codon harmonization algorithm with commonly used commercial codon optimization algorithms, which primarily aim for frequent codons in the host. Therefore, in the current study the expression and concomitant membrane integration of 6 membrane proteins is analyzed, including some difficult-to-express membrane proteins. To this end, codon-harmonized, codon-optimized, and wild-type coding variants of the genes were fused to GFP at their C-termini for easy-monitoring of membrane-integrated expression in *E. coli* (Drew et al., 2001). In addition, expression of all these variants was fine-tuned on a transcriptional level using the *E. coli* LEMO21(DE3) strain.

Results and discussion

Applying harmonization and optimization algorithms

Six different integral membrane proteins were selected from bacteria, archaea and eukarya, to compare their heterologous, membrane-embedded production in *E. coli* from wild-type, optimized and harmonized gene variants (**Table 1**). Four of the selected membrane proteins are light-harvesting proton-pumping rhodopsins (PPRs) originating from all domains of life. PPRs are membrane proteins that harbor 7 transmembrane domains and covalently bind a retinal

pigment. The retinal pigment here functions to absorb a photon, leading to a conformational change of the pigment, which eventually leads to a proton being extruded from the cell, resulting in a proton motive force. These PPRs were specifically targeted in this study as they can function as simple energy-harvesting photosystems in many organisms, and through heterologous expression they can serve wide applications, which include optogenetic sensors in neuroscience (Zhang et al., 2011) and optogenetic control of bacteria or light-driven ATP regeneration in microorganisms (Claassens et al., 2013). Some of the selected PPRs have already been expressed relatively successful in *E. coli*, such as bacterial *Gloeobacter violaceus* rhodopsin (GR) (Imasheva et al., 2009; Lee and Jung, 2011) and archaeal *Haloarcula marismortui* rhodopsin (HR) (Fu et al., 2010), while others are not (yet) expressed in *E. coli* to appreciable levels, such as bacteriorhodopsin (BR) (Bratanov et al., 2015) and leptosphaeria rhodopsin (LR) (Waschuk et al., 2005).

In addition to PPRs, we tested two different integral membrane enzymes from different domains of life. Nitric oxide reductase (NorB) from the bacterium *Moraxella catarrhalis*, for which it was shown previously that the GeneArt codon-optimization in *E. coli* resulted in a significantly reduced production level compared to expression of the wild-type gene in *E. coli* (Schlegel et al., 2012). Furthermore, we included an archaeal 2,3-di-O-geranyl-geranyl-glycerylphosphate synthase (DGGGPs), an enzyme catalyzing ether bond formation in archaeal lipid synthesis. Successful heterologous expression of this integral membrane enzyme has been a major challenge for the transfer of the archaeal lipid biosynthesis pathway to the bacterium *E. coli* (Jain et al., 2014).

For all 6 gene candidates, a codon harmonized sequence was generated and codon-optimized was obtained from the proprietary GeneArt algorithm. The codon harmonization was based on the algorithm developed before (Angov et al., 2008, 2011). Our in-house script generates harmonized sequences, using the codon usage tables for the native and expression host as inputs; the algorithm the selects the codons for the synthetic sequence to most closely match the native codon frequency usage. For all the tested genes the so-called 'codon frequency landscapes' were generated. As intended, the codon landscapes of the harmonized variants for *E. coli* are comparable to the landscapes of

the wild-type variants for the native host (**Figure 1, Supplementary Figure S2**). Apart from assessing the codon landscapes graphically, they can also be evaluated quantitatively based upon a proposed Codon Harmonization Index (CHI). A CHI value close to 0 indicates a well-harmonized gene, all harmonized variants in this study have a CHI <0.1. All codon optimized and wild-type variants have codon landscapes in *E. coli* that deviate further from the native codon landscape and consequently when their CHI is calculated this result in higher values (>0.15). Especially the wild-type variants of archaeal DGGGPs and eukaryotic LR yield high CHI scores (>0.279).

The landscapes of the codon-optimized variants for *E. coli* generally are a 'high plateau' , because they mainly contain frequent codons (**Figure 1, Supplementary Figure S2**). However, due to some additional rules of the GeneArt multi-parameter algorithm, some less frequent codons are occasionally also included. Nevertheless, rare codons are hardly appointed by this algorithm, in sharp contrast to the harmonization algorithm. The preference for frequent codons in codon-optimized variants is also reflected by the overall high CAI scores, which are all above 0.94. The unmodified wild-type gene variants from the original organisms, have expectedly lower CAI scores based on *E. coli* codon usage; especially the wild-type sequences of the eukaryotic and archaeal genes result in lower CAI values (<0.758) (**Table 1**). These relatively low CAI scores can be related to larger phylogenetic distance to *E. coli* of those organisms from different domains of life. The harmonization algorithm also increases CAI scores compared to the wild-type variants, but generally to a lower extent than the optimization algorithm. Lower CAI increases from the harmonization algorithm are expected, as this algorithm deliberately includes rare codons, reducing overall CAI, to mimic the codon landscape of the wild-type gene in the native host.

Table 1. Overview of all 6 tested membrane proteins, for which the expression of different gene variants was studied and their analyzed codon usage parameters.

Gene abbreviation	*Protein name*	*Native host*	*Domain of life*	*Number of transmembrane domains*	*CHI for E. coli*			*CAI for E. coli codon table*			*CAI for wild-type variant and native host codon table*	*References**
					Wild-type	*Harmonized*	*Optimized*	*Wild-type*	*Harmonized*	*Optimized*		
GR	gloeobacter rhodopsin	*Gloeobacter violaceus*	Bacteria	7	0.221	0.099	0.339	0.776	0.799	0.979	0.636	1
BR	bacterio-rhodopsin	*Halobacterium salinarum*	Archaea	7	0.234	0.057	0.213	0.75	0.862	0.99	0.781	2
HR	halo-rhodopsin	*Haloarcula marismortui*	Archaea	7	0.219	0.041	0.190	0.749	0.866	0.984	0.786	3
LR	leptosphaeria rhodopsin	*Leptosphaeria maculans*	Eukarya	7	0.279	0.066	0.183	0.754	0.872	0.983	0.845	4**
NorB	nitric oxide reductase	*Moraxella catarrhalis*	Bacteria	14	0.250	0.083	0.197	0.815	0.878	0.984	0.664	5
DGGGPs	2,3-di-O-geranyl-geranyl-glycerylphosphate synthase	*Methanococcus maripaludis* C5	Archaea	7	0.281	0.056	0.230	0.758	0.829	0.946	0.772	6

*References; 1 (Imasheva et al., 2009); 2 (Bratanov et al., 2015); 3 (Fu et al., 2010); 4 (Waschuk et al., 2005); 5 (Schlegel et al., 2012; De Vries et al., 2010); 6 (Caforio et al., 2015); **not expressed before in *E. coli*

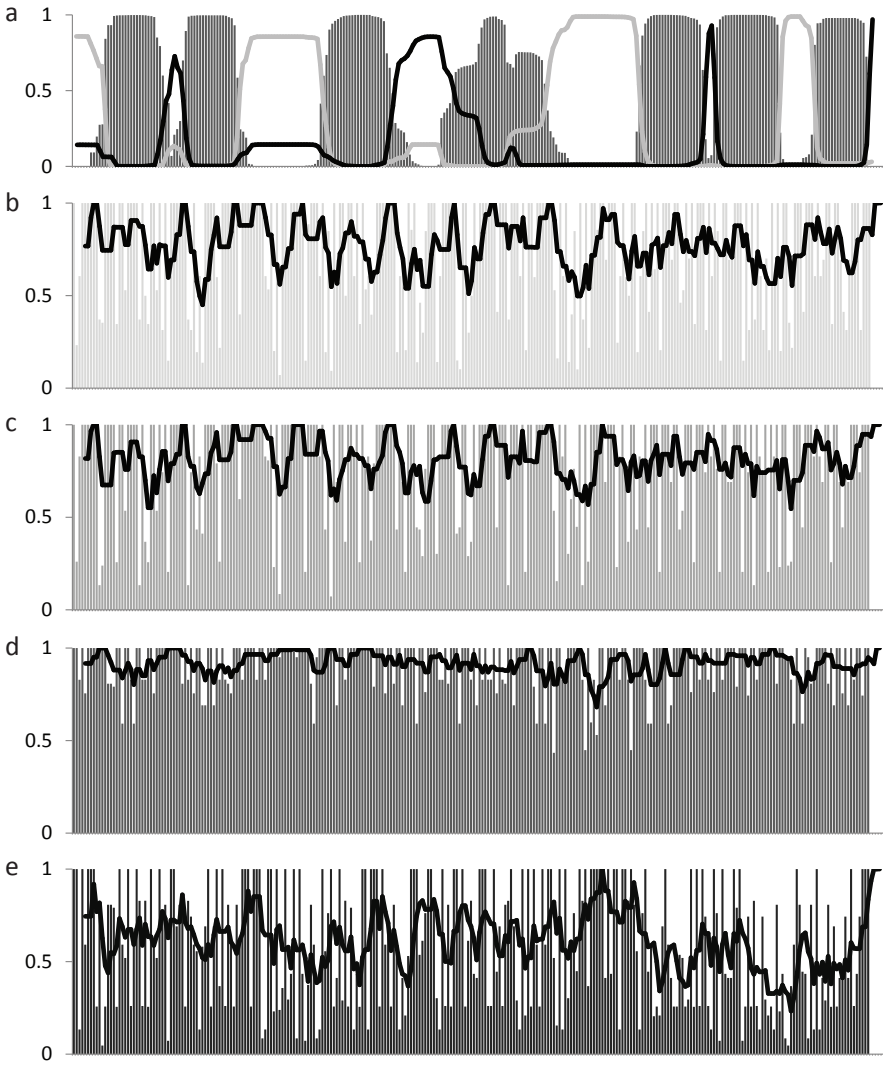


Figure 1. Transmembrane helix prediction and codon usage landscapes for the different variants for DGGGPs. a| Transmembrane helix prediction plot depicting the probability (Y-axis) of residues (X-axis) being in a transmembrane helix domain (dark grey bars), on the inside or cytosolic side of the membrane (light grey line) or outside of the membrane (black line) ((TMHMM v2.0). Codon usage landscapes are depicted in bars based on CAI scores (Y-axis) for individual amino-acids (X-axis) and a moving average over 5 codons (black line), for (b) the wild-type gene for native host codon usage (*M. maripaludis* C5); (c) the codon-harmonized gene variant for *E. coli* codon usage; (d) the codon-optimized gene variant for *E. coli* codon usage (e) the wild-type gene variant for *E. coli* codon usage.

Transcriptional tuning significantly improves membrane protein expression levels

To allow for easy quantification of membrane-integrated protein levels, GFP was used as a folding reporter and monitored by whole-cell fluorescence assays (Drew et al., 2006, 2001). The GFP protein was fused to the C-terminus of the membrane proteins. GFP will generally only be properly folded and generate a fluorescent signal when the fused membrane protein is integrated into the membrane (Drew et al., 2006). This method only works for membrane proteins with intracellular C-termini. However an alternative method is available for membrane proteins with an extracellular C-terminus, in this work only DGGGPs. Hereto one the pWarf vector was applied, which fuses an additional single transmembrane spanning domain in between the extracellular C-terminus of DGGGPs and the GFP-fusion (Hsieh et al., 2010). This allowed for intracellular localization and proper folding and fluorescence of the GFP report domain.

For all membrane proteins and their three gene variants, it was tested by in-gel fluorescence if the GFP signal originated from a single fusion protein (**Supplementary Fig S3**). Besides LR, for all the other proteins the signal comes mostly from a single fusion product that is of the correct size.

Expression by the *E. coli* LEMO(DE3) strain allowed us to optimize the level of each integrated-membrane protein of interest by transcriptional tuning by varying L-rhamnose in the common range (for mechanism see **Supplementary Figure S1**). For all tested protein candidates we observed that higher concentrations of L-rhamnose, i.e. down-tuning of transcription, resulted in significantly higher levels of fluorescence (**Figure 2**). For the wild-type codon variants of all studied membrane proteins optimization by transcriptional tuning led to 2-10 times higher expression. Also for the harmonized and optimized variants of those proteins, transcriptional tuning generally improved expression by similar orders of magnitude. However, it has to be noted that the optimal level of tuning, i.e. optimal concentration of L-rhamnose, frequently differs among different codon usage variants for the same protein. Previously, it was already demonstrated that the optimal level of transcription is specific for different proteins and expression conditions (Schlegel et al., 2012), and as we demonstrate here, this is also true for different codon usage variants.

For most of the membrane proteins and their codon variants, (down)-tuning of transcription appeared important to reach the highest level of properly folded, membrane-integrated protein. The role of tuning presumably lies in properly matching translation rates with the folding and translocation rates of heterologous protein into the *E. coli* membrane. This may facilitate proper integration into the membrane and avoid accumulation in inclusion bodies. One could expect that GeneArt codon-optimized gene variants, which consist mainly of frequently used codons, have the fastest translation rates and hence require more down-tuning. Consequently one could expect that the codon-optimized variants require generally higher L-rhamnose concentrations for optimal expression; however this relation cannot be clearly observed for most genes. This may have to do with the fact that also other factors determine the translation rates of different variants, such as the translation initiation rates and mRNA stability of different variants.

Interestingly, for membrane proteins that have been reported to be hard-to-express in *E. coli* such as DGGGPs and BR, it seems that tuning down to the lowest tested transcriptional level (2000 µM L-rhamnose) can substantially improve expression levels. Down-tuning of transcription increased membrane-integrated expression of the wild-type BR and wild-type DGGGP by 6-fold and 10-fold, respectively. However, to (further) increase the expression of these proteins and some others, it appears that the different codon usage variants can play an important role.

Different codon usage algorithms improve expression of some membrane proteins

In this study we compared the influence of three different codon usage variants on expression levels of membrane-integrated proteins. After optimization by transcriptional tuning, the optimal tuning conditions for each codon usage variants were compared. This gave rather mixed results on the success of different codon usage variants for different membrane proteins (**Figure 3**). For most PPRs there was no large or no significant difference in the maximum expression level between wild-type, harmonized or optimized variants. For the fungal PPR LR, the whole-cell fluorescence are not reliable because of the large band of loose GFP observed. However, in-gel fluorescence of LR for equal cell amounts indicates low levels of LR-GFP fusion production for both

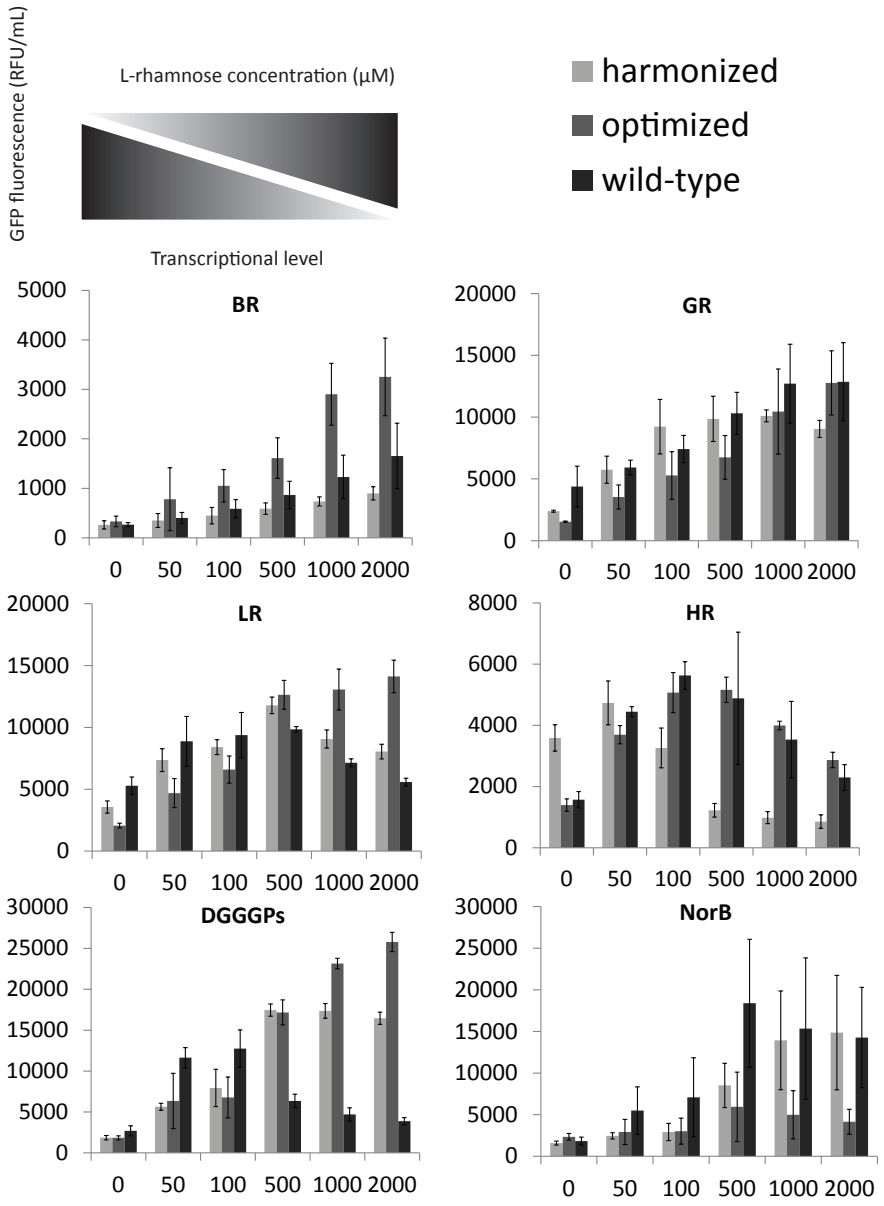


Figure 2. Membrane-integrated expression levels in *E. coli* LEMO21(DE3) for all codon usage variants of all genes based on whole-cell GFP-fluorescense at different transcriptional tuning by varying the L-rhamnose concentration (indicated in μM). All expression experiments were at least performed in biological triplicates. We note that data for LR do not well represent actual levels, especially for LR, given the detected bands of incomplete fluorescent products by in-gel fluorescence.

the harmonized and optimized variant, while this band is not detectable for the wild-type variant (**Supplementary Figure S3**). This indicates that for low-level production of this eukaryotic PPR both the optimized and harmonized variant are beneficial when compared to the wild-type variant. However, from the in-gel fluorescence it is not possible to determine if those two adapted codon variants give significantly different results. For BR a further improvement is achieved on top of the tuning for the codon-optimized variant, resulting in a further doubling of the expression level compared to using a wild-type variant and tuning expression. Surprisingly, for BR, the harmonized variant did not improve the expression compared to the wild-type variant, but instead decreased the expression slightly. So for hard-to-express BR, combining codon-optimization and transcriptional tuning seems the best strategy to improve the expression, however compared to other proteins its expression level is still generally low.

In addition to the 4 PPRs, the other two analyzed membrane proteins show that the harmonization algorithm is a fruitful strategy to increase membrane-integrated protein production. For both DGGGPs and NorB, the GeneArt optimized variants resulted in significantly reduced expression compared to the wild-type variants. As was already observed for some other membrane proteins (Schlegel et al., 2012), codon optimization may result in reduced membrane-integrated expression, possibly due to less efficient folding and/or translocation resulting from the dominant usage of frequent codons. This decrease in expression by optimization for these two proteins could be counteracted by the codon harmonization algorithm. Harmonization restored the expression of NorB to a similar level as the wild-type variant. More interestingly, codon harmonization of DGGGPs genes increased the expression by almost 50% compared to the wild-type variant.

As a general observation, we note that harmonization is beneficial for increasing membrane-embedded production compared to wild-type variants for some proteins, such as LR and DGGGPs, for which in this study the wild-type CHI score is also highest (>0.279). Hence, especially in cases in which the codon landscape of the wild-type gene in *E. coli* is far from the landscape in the native hosts, harmonization seems a promising strategy for improved membrane protein production.

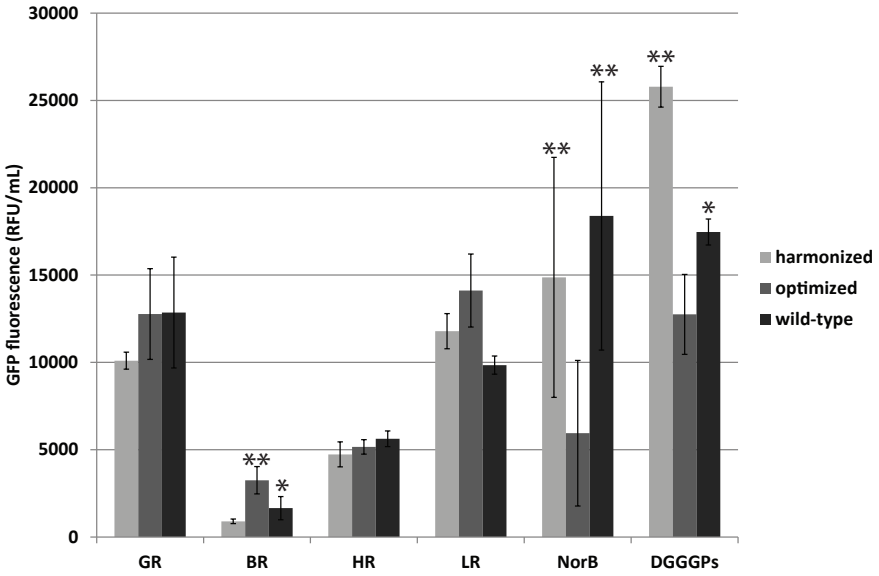


Figure 3: Comparison of the highest membrane-integrated production levels for different codon usage variants of all genes. All expression experiments were at least performed in independent triplicates. We note that data for LR do not well represent actual levels given the detected bands of incomplete fluorescent fusion proteins by in-gel fluorescence, so significant differences are not indicated for those. *indicates this variant expresses both significantly higher than the lowest expressing variant and significantly lower than the highest expressing variant for that protein(both $p<0.05$). ** indicates the expression levels of this or these variants are significantly higher from the lowest expressing variant for that protein ($p < 0.05$).

Conclusions

Here we demonstrate, using a set of different membrane proteins, that a combination of transcriptional tuning and different codon usage variants can be a successful approach to improve their heterologous expression.

The used GFP-folding reporter approach has been instrumental for this work and allowed for a convenient quantitative screen of membrane-integrated production for most of the proteins at different tuning-conditions. Though this study was limited to observing protein production in LB medium at 37°C, other conditions, such as commonly applied lower temperatures (20-30°C), could be assessed to determine optimal production conditions using the GFP-based screening as well.

For the codon usage algorithms, we expected that the relatively novel strategy of codon harmonization was a specifically promising strategy to improve the membrane-integrated expression. The underlying rationale was that stretches of more rare codons in the gene in the native host play an important role in proper folding and subsequent translocation of membrane proteins. These processes are often regarded as most crucial for the successful production of membrane-integrated proteins.

However, the somewhat mixed results of the codon optimization and harmonization algorithms for different proteins, again emphasize the complexity of optimizing codon usage for high-level protein production (Welch et al., 2009a) and specifically for membrane protein production (Nørholm et al., 2012). Codon usage and more general the mRNA sequence, have been shown to influence expression in many ways and new insights are still being elucidated (Boël et al., 2016; Quax et al., 2015). In general, studies to determine the influence of codon usage on both native and heterologous gene expression show a great complexity and interrelatedness of involved factors for different hosts, proteins and conditions. Important factors include frequent and rare codon usage, but also mRNA secondary structures, mRNA stability, concentrations of (charged) tRNA species, co-occurrence of specific codons and many more factors (Quax et al., 2015). Both the multi-parameter optimization algorithm and especially harmonization are based on simplified assumptions taking only some of these factors into account. Harmonization as employed here, only takes into account if specific codons in the wild-type gene are rare or frequent, relative to the overall codon usage in the native host. This codon frequency usage landscape is mimicked as well as possible by the harmonization algorithm, using the overall codon usage of the *E. coli* expression host. The harmonization approach could potentially be improved by also optimizing for some other important parameters, such as avoiding strong mRNA structures in the 5'UTR, as was shown to be useful for membrane proteins (Mirzadeh et al., 2015; Nørholm et al., 2013). To further improve the potential of the harmonization approach, systematic testing of algorithms that also take other parameters into account, could likely further improve functional protein expression (Quax et al., 2015)

Importantly, as shown in this study for several proteins, codon harmonization can lead to significant increases of membrane-embedded protein production compared to regularly chosen variants for heterologous protein production: the wild-type gene or a commercial codon-optimized variant. For 5 out of 6 tested membrane proteins, the harmonization algorithm gave either highest expression or the expression was similar to the highest observed expression variant.

In conclusion, our results indicate that, the wild-type gene for a membrane protein, can be used in initial attempts to optimize protein production in *E. coli* by transcriptional tuning, as this plays a major role in improved membrane protein production. However, if unsuccessful, especially if the CHI of the wild-type gene is in the high range (in this work > 0.279), codon harmonization is a promising algorithm for expression of synthetic gene variants.

Materials and methods

Generation of harmonized and optimized gene variants

Codon optimized sequences were designed using the GeneOptimizer algorithm of GeneArt for expression in *E. coli*, avoiding internal restriction sites required for cloning purposes. This algorithm is reported to be a multi-parameter sliding window algorithm, amongst other factors aiming for the usage of frequent codons for the expression host, a good GC-content and the avoidance of repeat sequences (Raab et al., 2010).

Codon harmonization was performed using the principle developed before (Angov et al., 2008, 2011). Native host and *E. coli* codon frequency tables were obtained from all annotated coding sequences for full genomes as deposited at NCBI. These frequency tables were converted to CAI (Codon Adaptation Index) scores (Sharp and Li, 1987). These CAIs scores were used to find the best matching synonymous codons for the harmonized gene variant (i.e. the synonymous codon with the CAI score in *E. coli* closest resembling the CAI score for that codon in the native host). For a limited number of cases, some internal restriction sites had to be removed for cloning purposes by choosing an alternative codon with the second closest CAI.

As a single metric to assess the extent of the harmonization we propose the Codon Harmonization Index (CHI):

$$CHI = \sum_{1}^{N} \mathrm{abs}(CAI_{harmonized} - CAI_{native})$$

In which N denotes the number of codons; $CAI_{harmonized}$ are the CAI scores of the codons in the harmonized variants based on the *E. coli* codon usage table (or possibly an alternative expression host); CAI_{native} are the CAI scores of codons in the wild-type gene variants based on the codon usage table of the native host. When CHI scores are close to 0, this means the codon landscape of a gene variant is close to that of the native landscape, indicating a well-harmonized codon landscape, the harmonization algorithm in fact tries to minimize the CHI score.

Strains and plasmids

All gene variants were synthesized by GeneArt (Thermo Fisher Scientific). Most synthetic genes were subcloned by GeneArt into the pET28+-based vector pGFPe (Drew et al., 2006), using XhoI and EcoRI sites. Only for the NorB gene variants XhoI and BamHI were used instead, pGFPe-NorB-wt and pGFPe- NorB -ga were a kind gift from Jan-Willem de Gier (Schlegel et al., 2012). DGGGPs is the only protein in this study with an extracellularly oriented C-terminus, therefore it was cloned into pWarf(+) (Addgene plasmid #34562) instead. This pWarf(+) vector introduces an additional transmembrane domain in between DGGGPs and GFP, allowing for intracellular localization of GFP, required for its maturation and fluorescence (Hsieh et al., 2010). Throughout the study, *E. coli* LEMO(DE3) (New England Biolabs) was generally used as an expression strain.

Culture conditions

Cultivation of *E. coli* strains for membrane protein expression was generally performed as described before (Hjelm et al., 2013). In short Lysogeny Broth (LB) (5 g/L yeast extract, 10 g/L NaCl and 10 g/L tryptone) was used throughout this study. Antibiotics were added for selection and maintenance of the pET expression vectors (kanamycin (50 µg/mL)) and pLEMO (chloramphenicol (34 µg/mL)).

Fresh transformants of *E. coli* were used to inoculate pre-cultures, as the use of re-streaked glycerol stocks may cause severe reduction of expression (Hjelm et al., 2013). Overnight pre-cultures (2 mL in 15 mL Greiner tubes, 37°C, 180 rpm) were used to inoculate 1:50 in 5 mL LB in 50 mL Greiner tubes with different L-rhamnose concentrations (0, 50, 100, 500, 1000 and 2000 µM). At an OD_{600} of 0.35-0.45, cells were induced with IPTG (isopropyl β-D-1-thiogalactopyranoside) at a concentration 0.4 mM. Cells were further incubated for 22 hours after induction (37°C, 180 rpm) and then harvested (13,000xg, 10 min, 4°C) for expression analysis. All data represented are derived from at least 3 independent cultivation experiments.

Whole-cell GFP fluorescence

Expression of membrane proteins was quantified using whole-cell GFP fluorescence as described before (Hjelm et al., 2013). In short, 1 mL of culture was resuspended in ice-cold 100 µL PBS and incubated at 4°C for at least 1 hour for further maturation of GFP. After this, suspensions were centrifuged (10 min, 13,000xg, 4°C) and resuspended in 100 µL PBS, which was transferred to a black 96-well microtiterplate with transparent bottoms (PerkinElmer). Fluorescence was directly measured using excitation at 485 nm and emission at 512 nm at a constant gain value (75) (BioTEK SynergyMX).

In-gel GFP fluorescence

To validate if the GFP signal originates from full-length fusions of the membrane protein with GFP, in-gel fluorescence was performed on the highest expressing samples found by transcriptional tuning, essentially as described before (Geertsma et al., 2008; Marino et al., 2015). Cell density was determined by measuring the absorbance at 600 nm (OD600) (WPA Biowave). Cultures were centrifuged for 5 minutes at 13,000xg and stored at -20°C. After thawing, pellets were resuspended to an estimated final concentration of 0.5 mg protein/100 µL in 50 mM kPi buffer (pH 7.5) (assuming 150 mg protein/L for OD_{600} of 1). This buffer was supplemented with 1 mM $MgSO_4$, 10% glycerol, 1 mM EDTA, 0.01 mg/mL DNaseI, 1 mg/mL lysozyme and protease inhibitor (Roche cOmplete™, EDTA free). Cells were lysed for one hour under mild shaking at room temperature and stored at -20°C for subsequent analysis. Twenty-five µL 4x Laemmli buffer (Bio-Rad) was added to 75 µL cell lysates, incubated for 5 minutes at 37°C only, as to prevent denaturation of GFP. Directly after resuspension, the samples were

shortly sonicated with three consecutive 0.1 ms pulses (Bandelin SONOPLUS HD 3100) to reduce sample viscosity for loading. Twenty-five µL of sample (~90 µg protein) was loaded and separated on a 12% Mini-PROTEAN® TGX™ protein gel (Bio-Rad). After electrophoresis, in-gel fluorescence was visualized on a Syngene G-box using a 525nm filter.

Author contributions

NJC, MFS, TEFQ and JvdO conceived the idea and designed this study. NJC performed the experimental work and data analysis with support from MFS. SKS, SC and BN were involved in setting up the codon harmonization algorithm. NJC drafted the manuscript with support from MFS. The manuscript was critically revised by SKS, TEQ and JvdO.

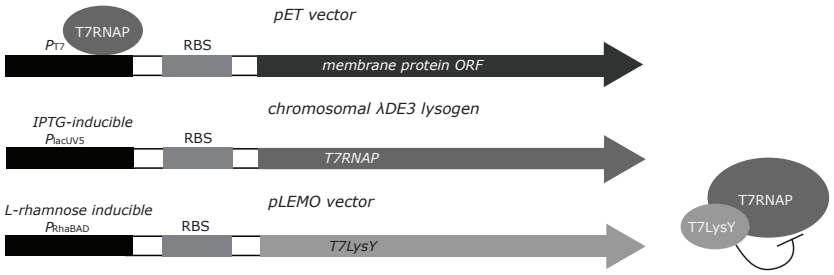
Acknowledgements

We would like to acknowledge Jan-Willem de Gier for good advice for this project and for sharing the pGFPe and the pGFPe-NorB-op and pGFPe-NorB-wt vectors. We thank Yuri Wolf and Eugene Koonin for sharing their thoughts and their great help with getting up-to-date codon usage tables for all organisms from the NCBI database. pWarf(+) was a gift from Jeff Abramson obtained through Addgene.

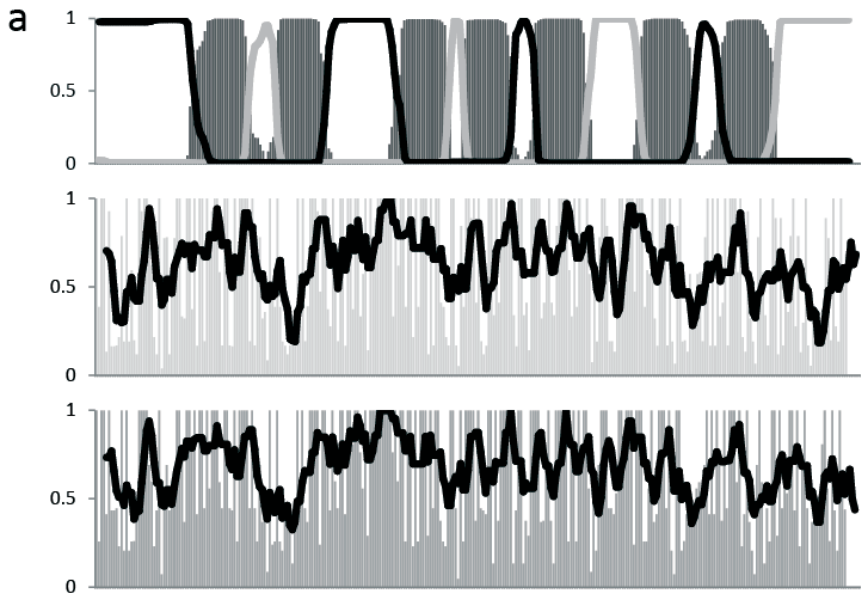
Supplementary information

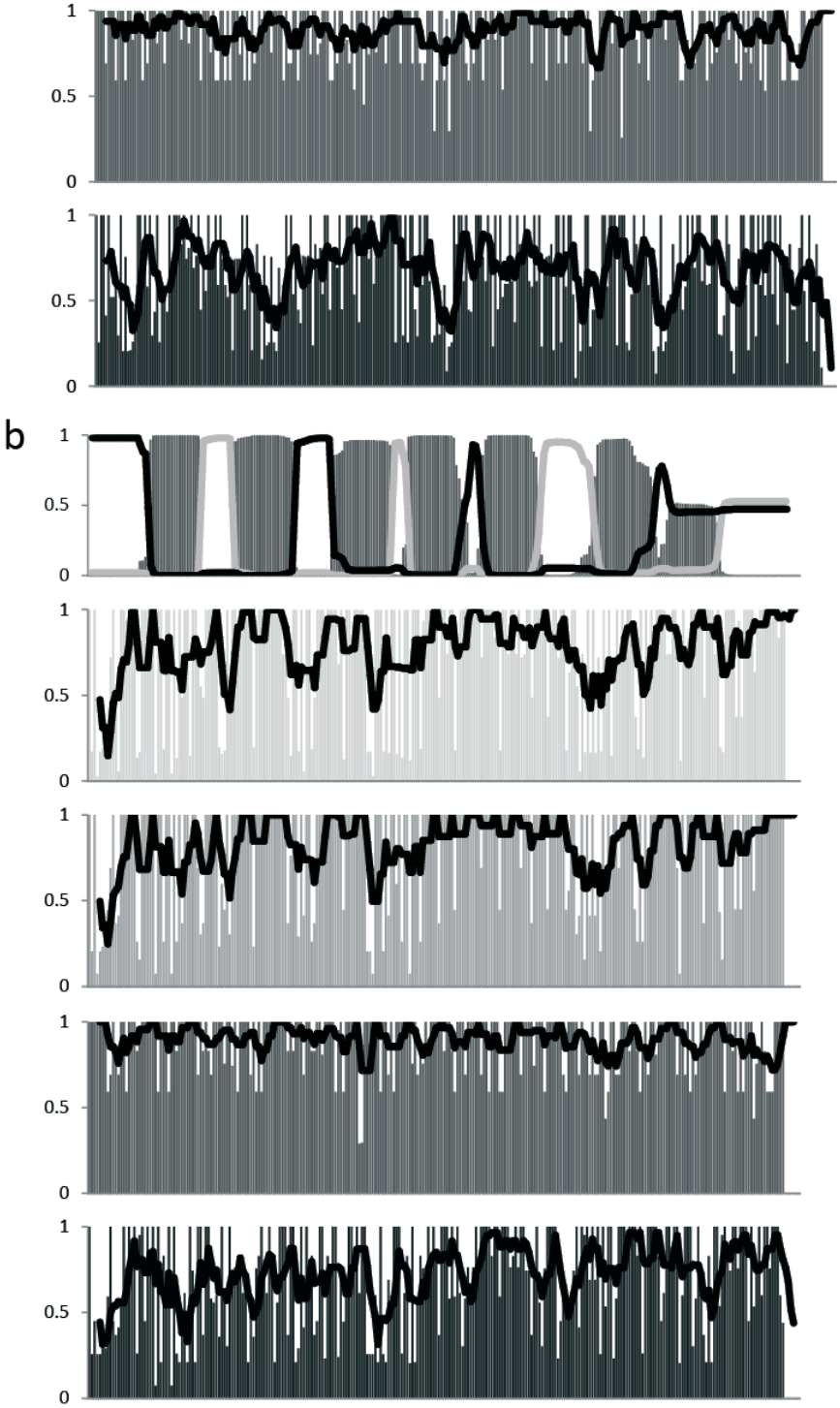
Supplementary data

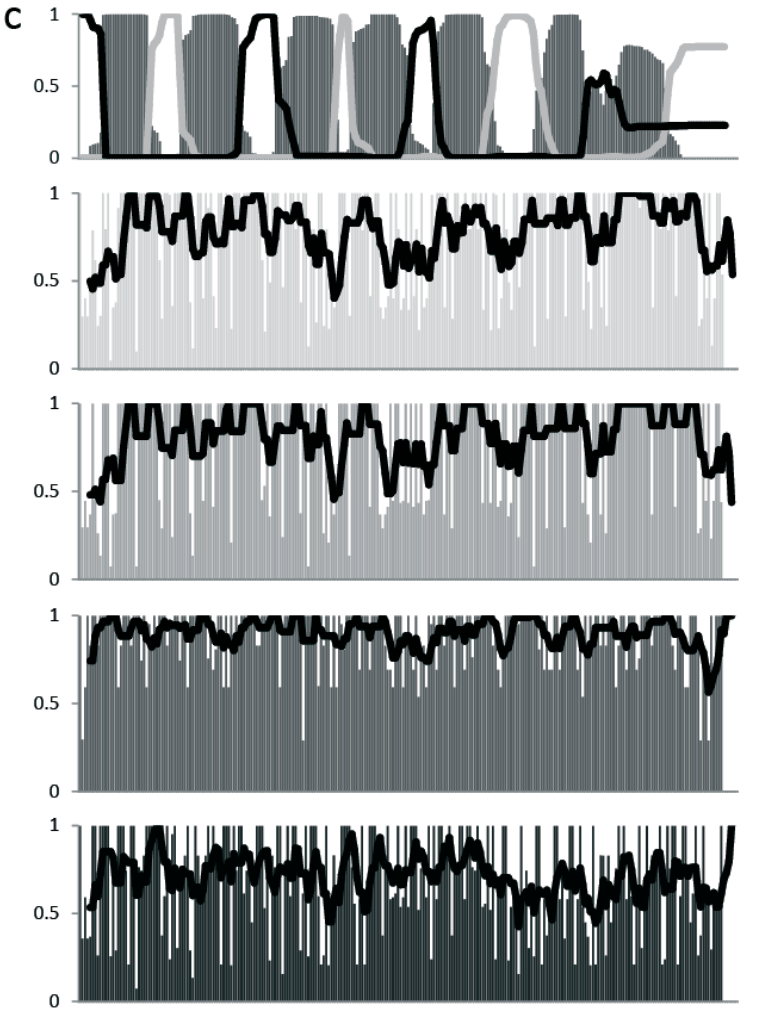
Gene sequences for codon-harmonized and codon-optimized variants in this work will be deposited to NCBI BankIt

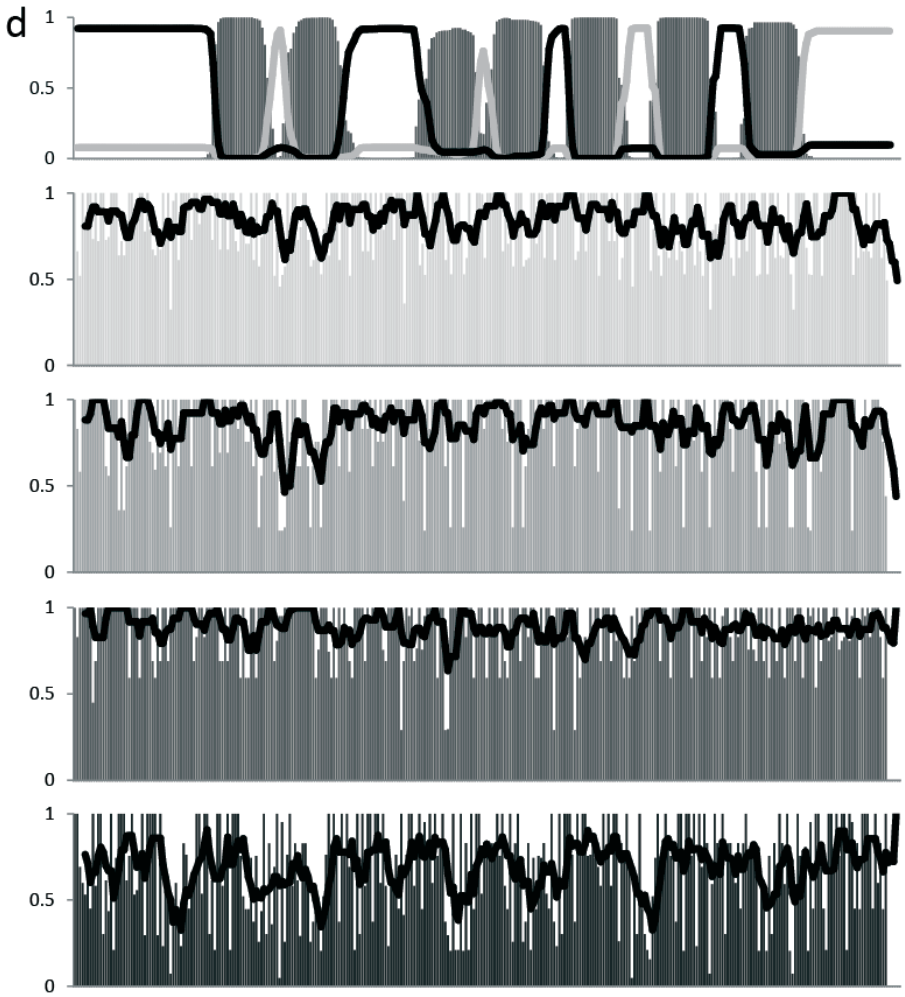


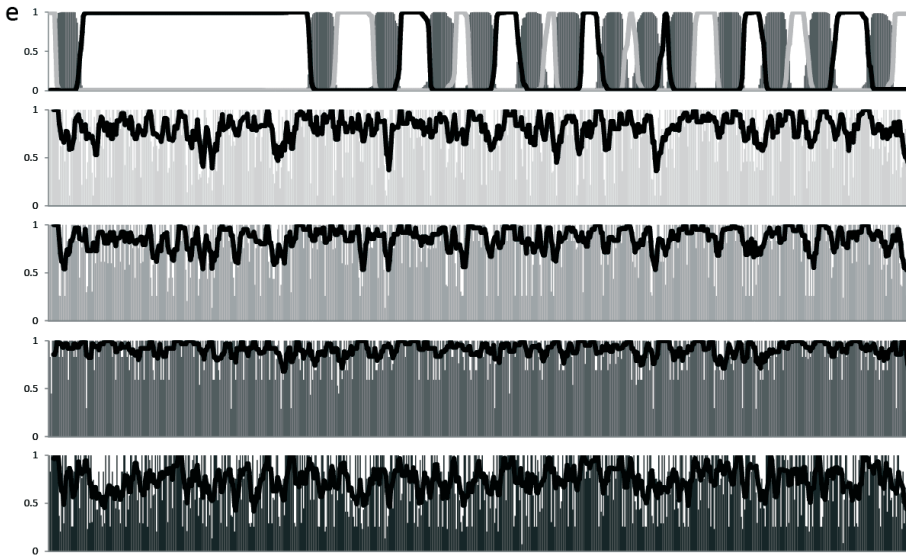
Supplementary figure S1. Schematic overview of the *E. coli LEMO21(DE3) system* (Schlegel et al., 2012; Wagner et al., 2008). The gene of interest is expressed from a pET vector from a T7 promoter. Transcription is driven by T7 RNA polymerase (T7RNAP), of which the gene is transcribed from an IPTG-inducible promoter, located on chromosomal locus (λDE3 lysogen). The mRNA transcript levels of the gene of interest are tuned by tuning the inhibition of T7RNA; this is accomplished through T7LysY, a T7 lysozyme, inhibiting T7RNAP activity. T7LysY is expressed from the pLEMO and can be tuned by L-rhamnose (P_{RhaBAD}).



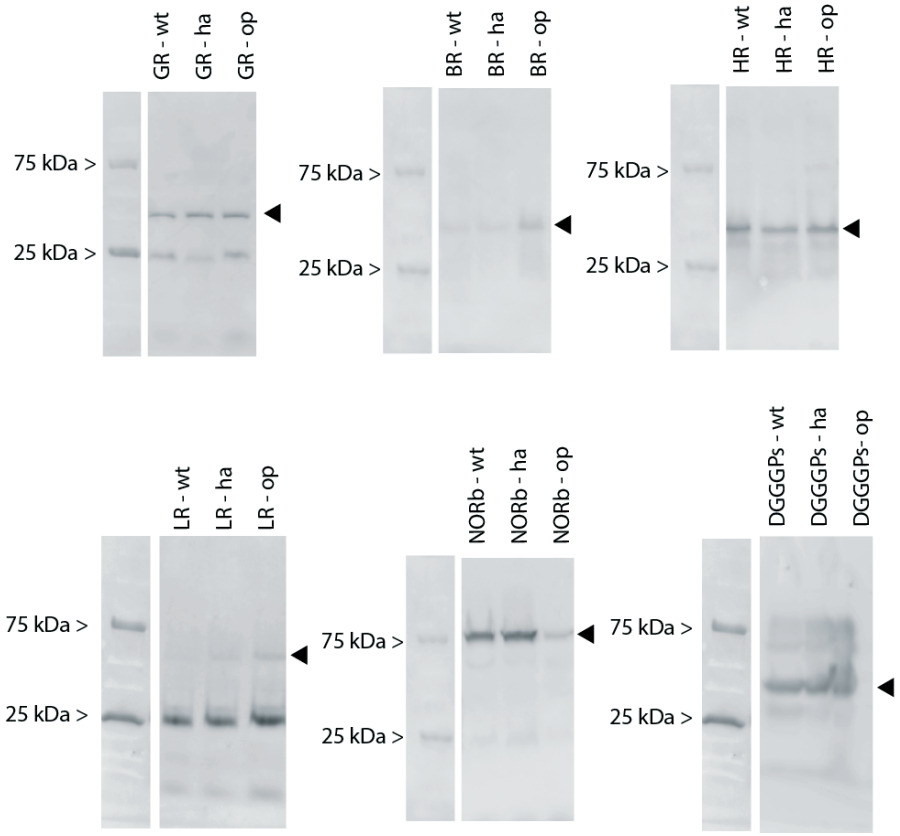

1
0.5
0
1
0.5
0
b
1
0.5
0
1
0.5
0
1
0.5
0
1
0.5
0
1
0.5
0


C
1
0.5
0
1
0.5
0
1
0.5
0
1
0.5
0
1
0.5
0


d
1
0.5
0
1
0.5
0
1
0.5
0
1
0.5
0
1
0.5
0



Supplementary figure S2. Transmembrane helix predictions and codon usage landscapes for the different variants of the membrane proteins in this study: a| GR; b| BR; c| HR; d| LR; and e| NorB. For the DGGGPs codon landscapes see Figure 1 in the main text. For each protein the upper graphs contain the transmembrane helix prediction plot, which predicts the probability (Y-axis) of residues (X-axis) being in a transmembrane helix domain (dark grey bars), on the inside or cytosolic side of the membrane (light grey line) or outside of the membrane (black line) ((TMHMM v2.0). In the next four graphs for each protein, codon usage landscapes are provided in bars based on CAI scores (Y-axis) for each residue (X-axis) and a moving average (black line) over 5 codons. The first graph (lightest grey bars) gives the codon landscape of the wild-type gene for the native host codon usage, secondly (light grey bars) the codon landscape of the codon-harmonized variant for *E. coli* codon usage; thirdly (dark grey bars) the codon landscape of the codon-optimized variant for *E. coli* codon usage, fourthly (darkest grey bars) the codon landscape of the wild-type gene variant for *E. coli* codon usage.



Supplementary figure S3. Membrane protein-GFP fusion integrity check by in-gel fluorescence. Integrity is analyzed for crude cell extracts of *E. coli* LEMO21(DE3), expressing the different gene variants at their optimal L-rhamnose concentrations. The Precision Plus Protein™ Dual Color marker was loaded on the gels, the fluorescent 25 and 75 kDa bands are indicated (open arrows). The right arrow (filled) indicates the bands most likely containing the protein of interest fused to GFP. It has to be noted that generally membrane protein-GFP fusion bands migrate lower than expected based on their molecular weight, as GFP is still in the folded state. For most variants there is one major fluorescent band representing the membrane protein – GFP fusion, for LR strong fluorescent bands are detected that are probably related to small or non-fused GFP product. Expected sizes full length sizes GR-GFP: 62.3 kDa; BR-GFP: 58.3 kDa; HR-GFP: 55.2 kDa; LR: 64.3 kDa; NorB-GFP: 115.2 kDa; DGGGPs-GFP: 61.0 kDa.

Chapter 7

High functional membrane protein production in *Escherichia coli* by tuning with translational coupling elements

Nico J. Claassens[1]*, Bart Scholten[1], Frederieke Muis[1], Jonas de Groot[1], Willem M. De Vos[1,2], and John van der Oost[1]

*first author
[1]Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE, Wageningen, The Netherlands
[2]Research Programme Unit Immunobiology, Department of Bacteriology and Immunology, Helsinki University, Haartmaninkatu 3, 00014 Helsinki, Finland

Abstract

The recombinant production of membrane proteins in *Escherichia coli* is a common approach towards their biochemical and structural analysis. Furthermore recombinant membrane proteins are often introduced in the model host *E. coli* as parts of synthetic metabolic modules. However, despite impressive progress over the last decades, high-level, functional membrane protein production in *E. coli* is still a major challenge. For improved membrane protein production, it has been demonstrated that precise tuning of the expression rate is crucial to avoid oversaturation of membrane translocation systems. Here we demonstrate improved, high-level production of several recombinant membrane proteins by fine-tuning on translational level, using translational coupling elements. Hereto we applied a library of recently developed Bicistronic Design (BCD) elements, these allow for translational coupling of a gene encoding a short leader peptide and the gene of interest that is under control of a variable ribosome binding site. An efficient cloning approach for assembling BCD elements with genes encoding membrane proteins is combined with microtiter plate screening, and allows for the efficient selection of high-expressing clones. For all four membrane proteins under conditions tested in this study, the BCD approach resulted in to 2 to 7-fold higher volumetric protein production levels, when compared to other state-of-the-art membrane protein production systems. The presented approach allows for inducer-free, constitutive, high-level production of membrane proteins *E. coli*, which is widely applicable for protein production as well as for synthetic biology endeavors.

Keywords
Membrane protein production; Bicistronic design elements; Translational tuning; *E. coli*

Introduction

About 15-30% of all genes encode integral membrane proteins, which have a wide plethora of important functions for the cell, such as transport, light-harvesting and signalling (Wallin and von Heijne, 1998). Compared to soluble, cytoplasmic proteins the production of high levels of recombinant functional membrane proteins in *Escherichia coli* has been demonstrated to be much more challenging (Schlegel et al., 2014). However, membrane protein production in *E. coli* still is one of the most convenient and commonly applied approaches to obtain sufficient protein for crystallography or for developing therapeutics targeting human membrane proteins (Snijder and Hakulinen, 2016). Also for engineering microbial cell factories or modules in synthetic biology the functional (over)production of membrane proteins in *E. coli* is very important. Relevant membrane proteins for cell factories or synthetic modules include transporters for substrates or products, and photosystems for energy-harvesting or light-sensing. The problematic production of functional membrane proteins in heterologous hosts is often related to the oversaturation of folding chaperones and translocation systems, such as the Sec-translocon in *E. coli* (Schlegel et al., 2014)

Generally membrane proteins are overexpressed in *E. coli* using pET vectors with the T7 promoter from which transcription is driven by the high activity of the T7 RNA polymerase (T7RNAP). However for many membrane proteins this system results in too high transcription and translation rates, resulting in low levels of functional membrane protein and accumulation of non-functional protein in inclusion bodies. Several *E. coli* strains have been evolved or developed to increase membrane protein production from the T7 promoter (Schlegel et al., 2014, 2016a). Notably this includes the *E. coli* Walker strains (C41, C43) (Miroux and Walker, 1996) and the *E. coli* LEMO21(DE3) strain (Schlegel et al., 2012; Wagner et al., 2008). The success of these strains relies on down-tuning of expression by lowering the levels of T7RNAP. Especially the LEMO21(DE3) strain has been developed as a versatile, successful system for the fine-tuning of membrane protein production (Hjelm et al., 2013; Lee et al., 2014; Schlegel et al., 2012). This strain has an indirect control of the T7RNAP levels through its inhibitor, T7-lysozyme (LysY), the expression of which is controlled by the L-rhamnose inducible promoter P_{rhaBAD}. This system works well, but typically requires two

plasmids (pLEMO and pET) and the addition of two different inducer compounds: L-rhamnose and the expensive IPTG (isopropyl β-D-1-thiogalactopyranoside). In addition, some alternative, well-titratable, inducible promoter systems are being develop for direct fine-tuning of expression of the gene of interest, such as direct tuning by P_{rhaBAD} (Giacalone et al., 2006; Hjelm, 2015; Kelly et al., 2016) or the m-toluate inducible XylS-*Pm* system (Balzer et al., 2013).

However, even with inducible promoter systems some membrane proteins have been reported difficult-to-express. This includes some membrane proteins for which it has been shown that there is a strong mRNA secondary structure in the region of the translational start site, i.e. the 5' untranslated region (5'UTR) and the adjacent part of the coding sequence (Mirzadeh et al., 2015; Nørholm et al., 2013). Such stem-loop structures can prevent or decrease the rate of ribosome association with the Ribosome Binding Site (RBS) (Kudla et al., 2009a; Reeve et al., 2014), which can in some cases result in detrimentally low translation initiation rates. To overcome these translation initiation issues, at least to some extent, well-expressed short peptides and their 5'UTR have been fused as tags to the N-terminus of some difficult-to-express membrane proteins (Kim et al., 2012a; Nørholm et al., 2013; Vazquez-Albacete et al., 2016). Although not extensively explored, libraries of N-terminal tags with different translation initiation strenghts may allow for fine-tuning of expression (Goltermann et al., 2011). However, as is the case with any tag used, the N-terminal fusion peptide may have undesired side-effects on the membrane protein of interest, as it may affect the protein structure and functionality (Nørholm et al., 2013).

As an alternative to N-terminal fusions, the Nørholm lab has shown that problems caused by undesired secondary structures of the 5'UTR:coding sequence junction may be solved by a random mutagenesis approach. They have substantially increased the production of several membrane proteins in *E. coli* by randomly mutating the 6 spacer nucleotides between the RBS and start codon and/or by silent mutations of the first few codons after the start codon (Mirzadeh et al., 2015; Nørholm et al., 2013). Although this approach has been successful for high membrane protein production, the size of the generated variant libraries requires high-throughput screening of clones, preferably by a Fluorescence-Activated Cell Sorting (FACS) set-up, to which many labs have no easy access.

Here we aim to develop a novel approach for membrane protein production by combining a non-inducible, constitutive promoter. We break with the common practice of using inducible promoter systems for membrane protein production. We expect that the typically included initial growth phase prior to induction and protein production is not required and not necessarily optimal for membrane protein production. Provided that the membrane protein of interest does not seriously affect the growth of *E. coli*, constitutive expression may facilitate continuous translocation into the cell membrane during the entire growth phase. A high rate protein production phase coinciding with decreased growth, typically happening after induction, may not provide the optimal capacity of membranes and membrane translocation machinery for membrane-embbed protein production. Supporting this hypothesis, it was recently demonstrated that compared to the common IPTG-induced expression, non-induced, constitutive expression from the T7 system resulted in higher production of several membrane proteins (Zhang et al., 2015). Efficient non-induced production of cytoplasmic proteins from the leakiness of the T7 system had been observed before as well (Grossman et al., 1998).

In an attempt to further fine-tune the expression of genes encoding membrane proteins, we here combine the expression from a constitutive promoter with a library of translational coupling elements, so-called Bicistronic Design elements (BCDs) (Mutalik et al., 2013b). These BCD elements contain two RBSs, the first RBS results in strong translation initiation. From this first RBS the translation of a short leader peptide is initiated. The second RBS is encoded within the last part of the coding sequence of this short peptide. This second RBS can be varied for different translation initiation strengths to tune the translation initiation rate of a protein of interest. The initiation at the second RBS probably occurs by so-called translational coupling, i.e. some of the ribosomes that finished the translation of the upstream short leader peptide continue translation of the protein of interest by hopping to the second RBS. It was demonstrated for a large set of cytoplasmic proteins that these BCD elements to a large extent abolish the potential negative influence of the 5'UTR:coding sequence junction on expression. This is probably because of the intrinsic helicase activity of the ribosome, which can eliminate potential secondary mRNA structures that could inhibit translation initiation (Mutalik et al., 2013b; Takyar et al., 2005). When compared to the typically used single RBS variants, the BCD expression system seems much more precise for tuning of gene expression in general.

Here we screen a BCD library with constitutive promoters, for membrane protein production of four different membrane proteins, including AraH which has been demonstrated difficult-to-express before due to mRNA secondary structures in the 5'-end of the transcript (Mirzadeh et al., 2015; Nørholm et al., 2013) . For all tested proteins we identify BCD variants that result in high-level production. We demonstrate this system allows for easy, inducer-free, high production of membrane proteins, both suitable for protein purification and for synthetic biology applications.

Results and Discussion

Screening for optimal Gloeobacter rhodopsin production

First we aimed to transcriptionally and translationally fine-tune the expression in *E. coli* of a codon-optimized gene encoding the membrane protein *Gloeobacter* rhodopsin (GR). This protein is a proton-pumping rhodopsin from the cyanobacterium *Gloeobacter violaceus* (Choi et al., 2014). Proton-pumping rhodopsins are membrane proteins that harbor 7 α-helical transmembrane domains and covalently bind a retinal pigment. The retinal pigment can absorb light, leading to a conformational change of the pigment, which eventually leads to a proton being extruded from the cell, resulting in a proton motive force (Ernst et al., 2014). Heterologous production of proton-pumping rhodopsins may be promising for engineering simple energy-harvesting photosystems in several production organisms (Claassens et al., 2013), and for optogenetic regulation (Zhang et al., 2011). Only when the GR protein is properly folded into the host's cytoplasmic membrane it can bind the retinal molecule, which results in a typical red color. The typical absorption and red color of properly folded rhodopsins allows for a quick visual screening for high production as well as for quantitative spectroscopic analysis.

Initially, production of GR was tested by using 7 different constitutive expression units from the BIOFAB Addgene collection. This collection contains different combinations of constitutive promoters elements and so-called BCD translational coupling elements with varying RBS strengths (**Figure 1a**). The 7 selected expression units cover a ~30-fold range, as was determined for cytoplasmic reporter proteins GFP and RFP (**Figure 1b**) (Mutalik et al., 2013b).

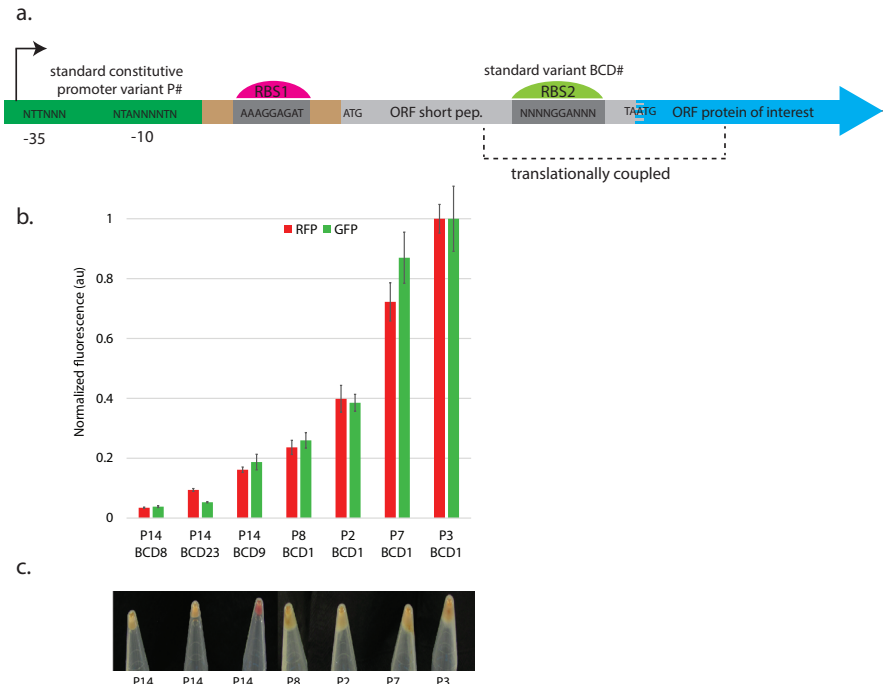


Figure 1. BIOFAB expression units and performance. a| Design of BIOFAB expression units including standard promoters and BCD element variants, allowing for translational coupling, ORF: open reading frame. b| Relative expression strength of different BIOFAB promoter + BCD variant combinations assessed for both GFP and RFP reporter production, performed by (Mutalik et al., 2013b). c| Production of GR from different promoter + BCD combinations in *E. coli* MG1655 that allowed for visual screening, the high-expressing combination (*P14*/BCD9) was shown to robustly result in high production in *E. coli* BL21(DE3) as well.

The initial screen in *E. coli* MG1655 resulted in the selection of only one expression unit that gave high visual production of GR (**Figure 1c**). This expression unit consisted of the medium strength promoter element *P14*, a constitutive promoter based on the *Ptrc* promoter, from which the operator site was removed. Furthermore, this expression unit contained the medium-strength translational coupling element BCD9. The two weaker and the four stronger expression units that were tested, resulted in substantially less or no detectable GR production as judged by both visual and spectroscopic inspection (data not shown). This indicated that the *P14*-BCD9 combination was a 'sweet spot' for high membrane-embedded production of GR, in the relatively wide expression range of expression units tested.

Fine-tuning of rhodopsin production by pooled cloning BCD elements

In an attempt to further optimize GR protein production, we fine-tuned its gene expression, while keeping the constitutive *P14* promoter as a constant element, with a library of BCD elements. The employed library consisted of 22 different BCD variants that cover a range of expression levels in combination with *P14* (**Figure 2a**). We designed an easy cloning approach for a pooled assembly of these BCD element variants to generate a vector library for *GR* expression. This approach is based on Golden Gate cloning (Engler et al., 2008), using type IIS restriction enzymes. A pool of 22 annealed BCD parts, containing RBS2, was introduced in the *P14*-GR expression vector. After library assembly, 87 clones were grown in medium supplemented with retinal in a deep-well plate (**Figure 2b**). In the cell pellets of 11 clones a red pigmentation could be detected by eye. Three of these high-producing clones were selected for further characterization. One of the selected clones contained BCD14, and two contained BCD15, which are both in the similar strength range of BCD9, which was found in the high-producing clone in the initial screen described above (Mutalik et al., 2013b). Spectral quantification of revealed that GR production of the three different BCDs was in the same range, although BCD14 seemed to result in slightly higher GR production than BCD15 and BCD9 (**Figure 3a**).

Using the same approach, we set out to optimize the production of a second rhodopsin, i.e. the thermophilic *Thermus* rhodopsin (TR) from *Thermus thermophilus* JL-18 (Tsukamoto et al., 2013a, 2016). Out of the 96 screened library clones, 11 red pigmented clones were observed. Three clearly red clones were selected for sequence analysis, revealing the presence of BCD12, BCD9 and BCD21. After culturing in a tube, the BCD12-clone gave the highest yield of TR, followed by BCD9. BCD21 showed substantially lower TR production in these tube culture conditions (**Figure 3b**). BCD12 and BCD9 are both medium translation initiation elements for GFP/RFP, while BCD21 was classified as a slightly weaker translation element (Mutalik et al., 2013b).

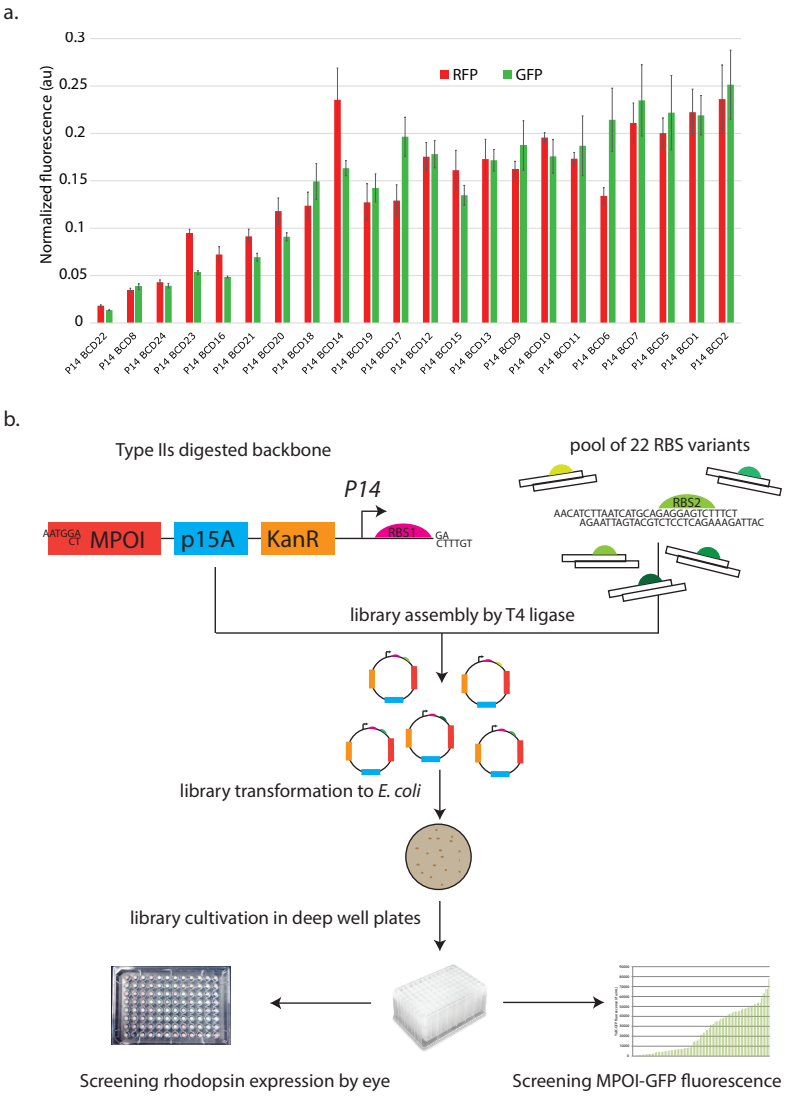


Figure 2. Tuning expression with a library of BCD elements. a| Relative expression strength of the *P14* promoter in combination with 22 different BCD elements, assessed for both GFP and RFP reporters (Mutalik et al., 2013b). b| DNA assembly and screening approach applied for generating the expression vector library with 22 BCD (RBS2) variants. *MPOI* = Membrane Protein Of Interest, *P15a* = medium copy number replication origin, *KanR* = kanamycin resistance gene.

Maximum volumetric PPR yields achieved for GR and TR were 5-7 mg protein/L. This is a high yield for PPRs, considering that an extensive optimization for proteorhodopsin production with different media, temperatures and inducer concentrations in *E. coli* resulted in a maximum yield of 5 mg purified PPR/L

(Gourdon et al., 2008). Though, it has to be noted our reported yields are not based on purified but on membrane solubilized PPR levels. As will be demonstrated below, the here achieved PPR yields are also substantially higher than found for optimized PPR production in *E. coli* LEMO21(DE3). So in summary, for the two tested rhodopsin membrane proteins, highly productive clones were obtained by easy screening ~90 clones from a pooled 22 BCD variant library.

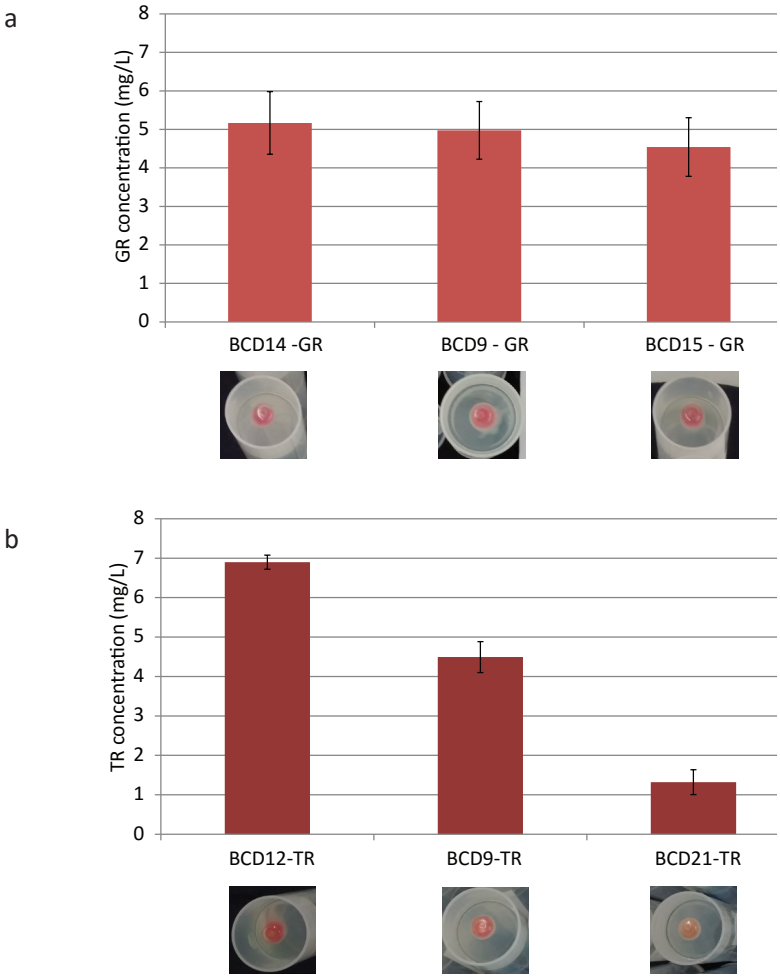


Figure 3. Rhodopsin production from high-producing BCD clones. Functional production of a| *Gloeobacter* rhodopsin (GR) and b| *Thermus* rhodopsin (TR) from selected high-expressing clones, grown in tubes and quantified by spectroscopy after solubilization in DDM. All culturing was performed at for 22 hours at 37˚C in LB. All data are at least from biological triplicates.

Fine-tuning the production of YidC-GFP

To test the translation tuning by BCD elements for other membrane proteins than rhodopsins, we could not rely on colorimetric screening and spectroscopic quantification. To quantify functional membrane protein production, C-terminal GFP-fusions are a well-established reporter systems (Drew et al., 2005, 2001). We optimized the overproduction of YidC fused to GFP by applying a BCD library. YidC is an *E. coli* integral membrane protein, functioning as a membrane-embedded chaperone. YidC fused to GFP has been routinely used as a model protein to test membrane protein production systems, and down-tuning of its transcription has been shown to be crucial for increasing functional YidC-GFP production in *E. coli* (Hjelm et al., 2013; Wagner et al., 2008; Zhang et al., 2015).

After BCD vector library assembly for YidC-GFP, 94 clones cultured in a deep-well plate were analyzed for GFP fluorescence, representing membrane-integrated YidC protein production. A large range of production levels was observed (**Figure 4a**). The top three of high-expressing clones from this plate were selected for further culturing in 10 mL medium in 50 ml tubes. Interestingly after 22 hours of growth in a tube, the BCD18 clone, which ranked third in the plate screen, produced ~3 times more YidC-GFP than the other two selected clones (BCD14, BCD5), which performed similarly (**Figure 4b**). So, it is noteworthy, that different relative performances BCD clones are observed in different culturing conditions, i.e. growth in deep-well plates (0.5 mL in 2 mL wells), versus growth in tubes (10 mL in 50 mL tube). The integrity of the YidC-GFP fusion protein and relative production levels were also confirmed by in-gel fluorescence, indicating that the whole-cell fluorescence data are reliable for quantification (**Figure 4c**).

Based on cytoplasmic reporter proteins, the selected BCDs for high YidC-GFP production are classified as intermediate, with BCD18 being somewhat weaker and BCD5 being somewhat stronger (5>4>18; **Figure 2a**). However, it also has to be mentioned that in the reliable and precise BCD expression system, still the strengths of BCD elements cannot be uniformly ranked for all genes. For example when BCD18 is compared to BCD14 for GFP production it gives 10% lower production levels. However, when comparing these BCDs for RFP, BCD18 is much weaker than BCD14, resulting in 50% lower production (**Figure 2a**) (Mutalik et al., 2013b). Even though the influence of the 5'UTR:GOI

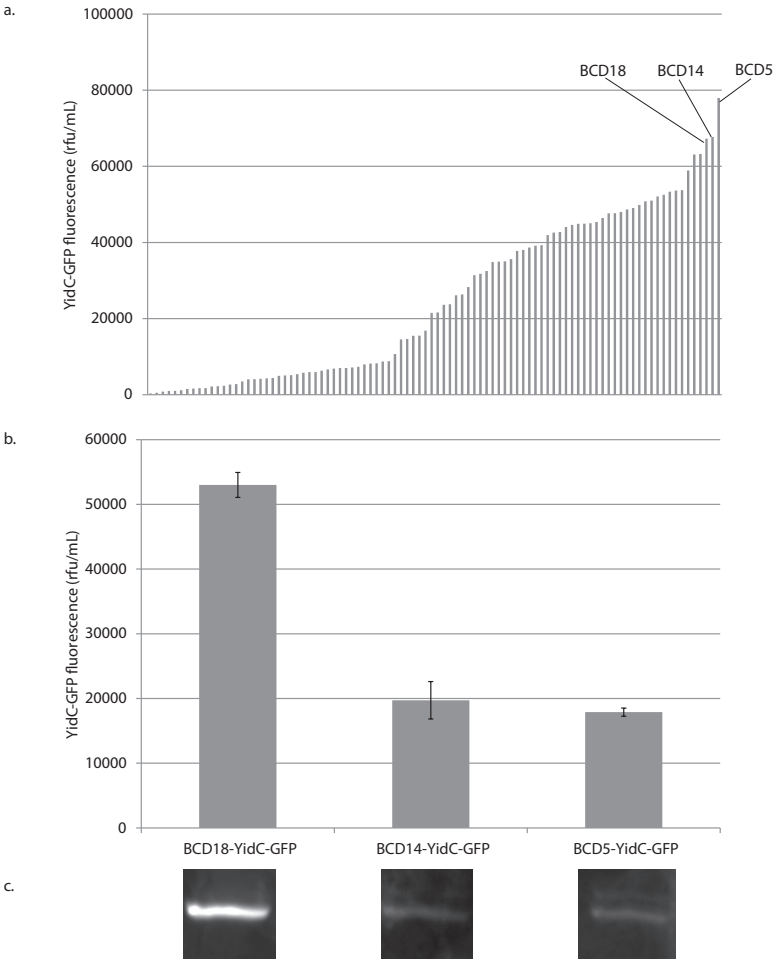


Figure 4. YidC-GFP production from BCD variants. a| Rank-ordered volumetric production of YidC-GFP for 94 BCD clones grown in deep-well plates based on whole-cell GFP fluorescence. b| Volumetric production of YidC-GFP from the top-3 highest expressing clones from a plate quantified after culturing in tubes, based on GFP fluorescence. c| In-gel fluorescence of YidC-GFP showed single bands, indicating integrity of the YidC-GFP fusions expressed from all three BCD variants, after cultivation in tubes, similar amounts of total cell protein were added to each lane based on cell density. Culturing was performed at for 22 hours at 37˚C in LB. The tube culture data are from biological duplicates.

combination on expression is very limited for the BCD system, still some small influences are present. This imperfection of the BCD systems has a relatively small impact on the preciseness of cytoplasmic protein production. However, at least observed here for YidC-GFP, similar strength BCD elements may result in a few-fold higher membrane protein production in some conditions.

In general, we demonstrated that for YidC-GFP constitutive expression, tuned with BCD variants can yield high membrane protein production.

Increasing the production of difficult-to-express AraH-GFP

For some native *E. coli* membrane proteins high recombinant overexpression from the typically used pET vectors was demonstrated before to be problematic (Daley, 2005). One of those membrane proteins is AraH, a component of the *E. coli* arabinose ABC transporter. For AraH-GFP it has been demonstrated that the expression from a pET vector in *E. coli* BL21(DE3) pLysS can be increased many-fold by optimizing the UTR:coding sequence junction. Optimal clones were found by screening random mutations in this junction region (Mirzadeh et al., 2015; Nørholm et al., 2013). High-expression clones were calculated to have relatively low-strengths mRNA secondary in the 5'-end of the mRNA. In this work the pET28a-AraH-GFP vector with an optimized junction resulting in the highest expression was used as a reference to compare the performance of BCD clones for AraH-GFP production.

After BCD vector library assembly for AraH-GFP, 95 clones were cultured in a deep-well plate at 30°C, as this lower temperature may increase production levels. Similar as for YidC-GFP, a large range of production levels was observed (**Figure 5a**). The two highest expressing clones both contained BCD9 and the third contained BCD14. So both BCD9 and BCD14 clones were characterized further in tube cultures. BCD9 clones resulted in the highest volumetric production levels, which were 5 times higher than volumetric production levels from pET28a-AraH-GFP with an optimized junction cultivated at 25°C (**Figure 5b**). For BCD14 the volumetric production in tubes was only half that of BCD9. The very high productivity of BCD14 is interesting when compared to the optimized junction pET28a-vector, also considering the productivity per cell density (**Figure 5b**). Though, it has to be noted that the BCD clones were cultivated longer (22 h at 30°C) than the strain harboring pET28a-opt.junc-AraH-GFP, which was cultivated according to the original published protocol (5 h at 25°C after induction). This difference may decrease the 5-fold improvement somewhat.

In summary, based on these results, the BCD system also has a large potential to overcome the expression problems of difficult-to-express proteins in pET vectors, and potentially can even further improve production levels and cell growth beyond those of strain harboring optimized junction pET vectors.

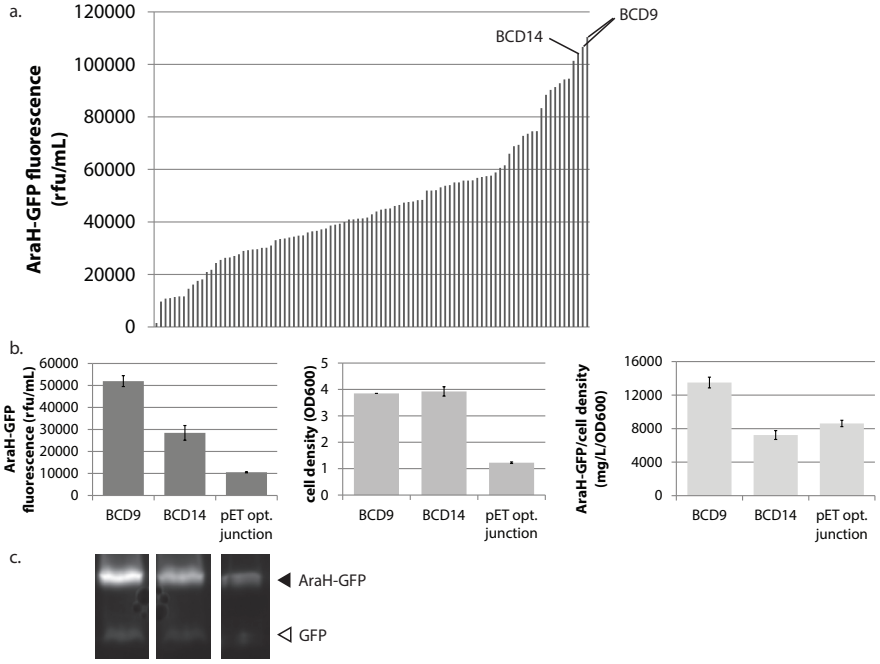


Figure 5. AraH-GFP production from BCD variants and pET28a vector with optimized junction. a| Rank-ordered volumetric production of AraH-GFP for 95 BCD clones grown in deep-well plates based on whole-cell GFP fluorescence. b| Production of AraH from the highest expressing *E. coli* BL21(DE3) BCD clones and from *E. coli* BL21(DE3) pLysS with the pET28a-AraH-GFP vector with an optimized junction (Mirzadeh et al., 2015), quantified after culturing in tubes (for 22 hours at 30°C in LB for BCD clones, for pET28a expression for 5 hours after 1.0 mM IPTG induction at 25°C in LB) c| In-gel fluorescence of AraH-GFP showed major bands for the fusion product and only minor bands indicating some loose GFP, similar amounts of total cell protein from tube cultures were added to each lane based on cell density. All tube culture data are at least from biological duplicates.

Translational versus transcriptional tuning

The protein production of the highest-producing BCD clones for GR, TR and YidC was compared to optimally tuned production for those proteins in *E. coli* LEMO21(DE3). Membrane protein production in *E. coli* LEMO21(DE3) is fine-tuned by transcriptional tuning, using different concentrations of L-rhamnose to indirectly control the level of T7RNAP (Wagner et al., 2008).

To quantify the optimal production by *E. coli* LEMO21(DE3), an initial screen was performed to find the optimal L-rhamnose concentrations for production of each of those proteins, expressed from pET28a vectors. For example for YidC-GFP production at 37°C in *E. coli* LEMO21(DE3) the optimal concentration of L-rhamnose was identified to be 2000 µM, corresponding to results from the De Gier lab for this protein (Schlegel et al., 2012).

After quantifying production at the optimal L-rhamnose concentration, it could be concluded that all highest-producing BCD clones gave higher membrane-integrated protein production than for the LEMO system (**Figure 6**). The best-performing BCD clones for both GR and TR gave about 2 – 3 fold more membrane-embedded protein, and for YidC-GFP the product yield of the best BCD clone is even 7 times higher than for *E. coli* LEMO21(DE3). This better performance may be partly related to the higher final cell densities reached by the *E. coli* BL21(DE3) having the BCD expression vectors, compared to the *E. coli* LEMO21(DE3) expression strains. However, also per cell density higher amounts of membrane-embedded proteins are produced (**Figure 6**). Here we also like to point out once more the different nature of the constitutive expression from the BCD vectors, versus the induced expression of the pET vectors in *E. coli* LEMO21(DE3). Induction of membrane protein production in mid-log growth phase, generally slows down growth and gives lower final biomass (Zhang et al., 2015). This may limit the further formation of cell membranes to translocate and accommodate recombinant membrane proteins.

In summary, the tuning by the BCD elements allows for constitutive high level membrane protein production, which,at least for the 3 proteins under the tested conditions, performs several-fold better than transcriptional tuning using the LEMO system.

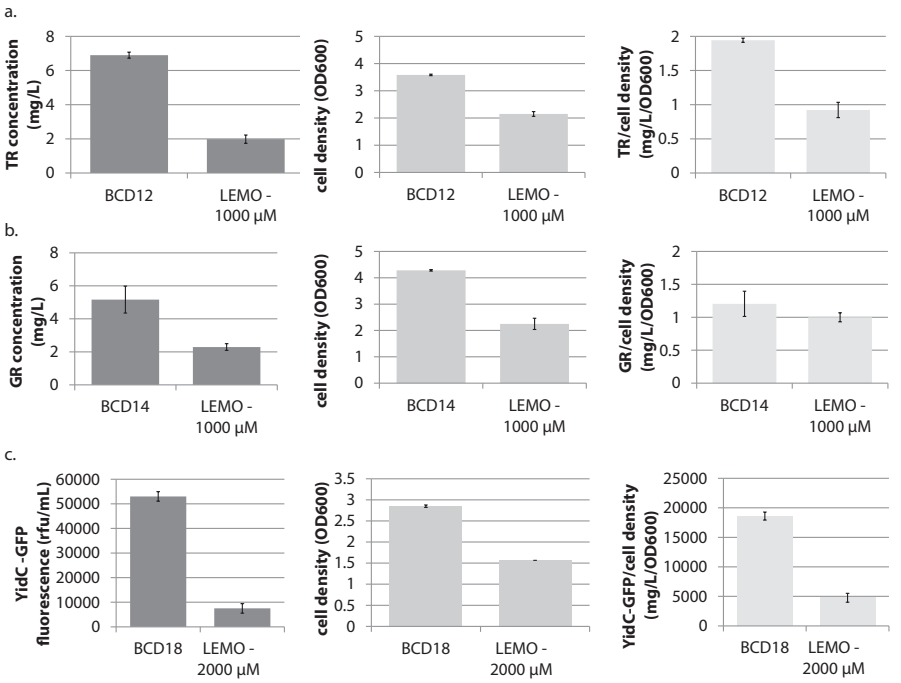


Figure 6. Comparing BCD expression versus LEMO expression. Production levels of a| GR, b| TR and c| YidC-GFP, final cell densities and production/cell density for the highest-producing BCD variants in *E. coli* BL21(DE3) and for the optimally tuned L-rhamnose concentration in *E. coli* LEMO21(DE3). Data obtained from at least three biological replicates grown in 10 mL in 50 mL tubes for 22 hours at 37˚C.

Conclusions

We tested protein production levels for four membrane proteins of from a constitutive, medium-strength promoter and a library of translational couplers, so-called BCDs. For all tested membrane proteins, which typically require fine-tuning for optimal expression, high constitutive production levels were found. For three tested proteins, higher production levels were achieved than those found in a parallel analysis of transcriptional fine-tuning using the well-developed LEMO system. For the fourth protein, AraH-GFP, which was demonstrated difficult to express in a standard pET vector, also high membrane-embedded production was demonstrated, surpassing the earlier demonstrated increased production from an optimized junction pET vector.

The BCD elements have been reported to largely eliminate effects of the 5'UTR:coding region junction on gene expression, which is a key issue for high production of proteins (Kudla et al., 2009a), including membrane proteins, as for example observed for AraH-GFP (Marino et al., 2015; Mirzadeh et al., 2015). Compared to other RBS engineering approaches (Bonde et al., 2016; Kosuri et al., 2013; Salis et al., 2009), these BCDs elements were reported to be most predictable for fine-tuning expression. However, despite their high precision for expressing several cytoplasmic proteins (Mutalik et al., 2013b), we observed substantial differences in the membrane-embedded production of YidC-GFP membrane in the present study for BCDs variants that were described to be very close in expression strength. This may be related to remaining BCD:coding sequence junction influences on expression, which are possibly not completely abolished by the translational coupling. The small 'imperfections' of the BCD system, i.e. differences in BCD strength ranking, observed when comparing the expression of various cytoplasmic proteins (Mutalik et al., 2013b), are probably further amplified for membrane protein production. This iis probably as for optimal membrane production there is a mostly a relatively narrow optimal window of expression.

It has to be noted that functional membrane protein production is typically optimal at other conditions than those tested here mostly, such as lower temperatures (e.g. 25 or 30 ˚C) and in large volume, well-controlled reactors. As was observed, different growth conditions, such as growth in wells versus growth in well-plates, requires different optimal tuning levels and different optimal RBS2 variants. So ideally screening is done in the envisioned final culture conditions. Especially for larger volume cultivation conditions or expression screening without an easy reporter, screening of the tested 22-member BCD library with sufficient coverage (e.g. 4 times coverage, 88 clones) becomes impractical. Therefore one may, alternatively to the here described pooled cloning approach, clone the 22 variants or even a smaller subset individually, which is easily feasible with the proposed standardized Golden Gate workflow. This will limit screening efforts to 22 or even less clones , and it will still allow for preferred complete mining of the desired library to identify the best tuning level for the desired conditions.

We envision that the here described BCD system for optimizing membrane protein production may find wide applicability for membrane proteins, and potentially also for secretory proteins or other proteins, which require fine-tuning (Hjelm, 2015). Based on the previously established BCD system, here we developed a convenient method for library assembly combined with efficient screening. This approach will be generally feasible for any molecular biology laboratory, and could also easily be automated. It can be used to achieve high-level membrane protein production without the need for induction and helper plasmids or specific *E. coli* strains. In general, we think these features make the presented approach widely applicable for high-level functional membrane protein production in *E. coli*, including for synthetic biology modules requiring functional membrane proteins. In addition similar, precise translational coupling systems may be also be developed and applied for high membrane protein production in other relevant bacterial hosts (Zobel et al., 2015).

Methods

Plasmid construction

Plasmids used for expression are in **Table 1,** additional plasmids used during construction are in **Supplementary Table S1**, oligos used are in **Supplementary Table S2**. PCR products were generated using either Phusion polymerase (Thermoscientific) or Q5 polymerase (NEB) according to manufacturer's protocols. Golden Gate assembly a was performed using type IIS restriction enzymes BsaI or BbsI (Thermoscientific and NEB) and subsequent assembly by T4 ligase (Thermoscientific and NEB). When appropriate PCR samples were treated with DpnI (NEB, Thermoscientific) to digest template DNA. Gibson Assembly was performed using the NEBuilder® HiFi DNA Assembly (NEB) according to the manufacturer's protocol. Agarose gel DNA purifications and DNA purifications were respectively performed using Zymoclean™ Gel DNA Recovery Kit and DNA Clean & Concentrator (Zymo Research). Assemblies were confirmed by restriction digests and/or by Sanger sequencing (GATC Biotech).

The gene encoding GR was codon-optimized for *E. coli* and synthesized by GeneArt for insertion into BglBrick vector pBbE0A, *TR* was codon-harmonized using our tool (**Chapter 6**) based on (Angov et al., 2008), and synthesized by

GeneArt (**Supplementary data 1**).

For construction of the BIOFAB vectors expressing *GR*, *GR-GFP* was PCR amplified from pBbE0A-GR-GFP (BG5971, BG5972) for introduction into seven BIOFAB plasmids with different expression unit strengths (pFAB3912, pFAB3999, pFAB3913, pFAB3833, pFAB3677, pFAB3737, pFAB3689), which were all a gift from Drew Endy (Addgene kit #1000000037) (Mutalik et al., 2013b). The BIOFAB backbones were amplified using primers (BG5973, BG5975, BG5976 and BG5981) for insertion of *GR-GFP* using Golden Gate assembly. As the GFP fusions at the C-terminus of GR were hampering proper retinal incorporation into GR, those were removed from the constructs using phosphorylated primers BG6162 and BG6163 and subsequent plasmid recircularization with T4 ligase. For construction of the BIOFAB-*P14* plasmid for TR, *TR* was amplified from pGFPe-TR (BG7340, BG7337) and introduced in a BIOFAB-*P14* backbone (pFAB3913 amplified from BG7339,7341) by Gibson assembly. For construction of the BIOFAB-*P14* plasmid for YidC-GFP, *yidc-gfp* was amplified from pET28a(+)-YidC-GFP (BG7342,BG7343) and assembled into a BIOFAB-*P14* backbone (amplified from pFAB3913 by BG7338,BG7339) via Gibson Assembly. For the construction of the BIOFAB-*P14* plasmid for AraH-GFP, *arah-gfp-his* was amplified from pET28a(+)-AraH-GFP-his (BG8448,BG7937), and together with one BCD oligo pair, introduced in the PCR amplified BIOFAB-*P14* backbone (BG7933, BG7934) by three part Golden Gate assembly.

For the construction of the BCD expression library for *GR, TR, yidc* and *arah* the above constructed plasmids were PCR amplified to introduce BsaI or BbsI type IIs restriction sites (BG7335,B7336 for *GR*, BG7505,BG7506 for *TR*, BG7784,B7785 for *yidc* and BG8448, BG7934 for *arah*). For *yidc* BbsI was used, as BsaI gave digestion issues. The BsaI/BbsI-digested PCR products were subsequently assembled with the pooled RBS2 variant library (**Figure 2b**). These RBS2 variants were constructed by a pooled library of phosphorylated, annealed oligos (BG7291-BG7334). Oligos were phosphorylated using T7-polynucleotide kinase (NEB) according to the manufacturer's protocol. Then, pairs of oligos for each RBS variants were annealed, by heating to 95°C for 3 min and gradual cooling to RT in 30 min. The annealed oligo pairs were pooled at a total concentration of 10 ng/µL. 0.5 µl of this mixture was ligated with ~100 ng BsaI/BbsI-digested, dephosphorylated PCR products. Those ligation mixes

were transformed to *E. coli* DH5α (for *TR* and *GR, arah-GFP*) or directly to *E. coli* BL21(DE3) for *yidc-gfp*.

For testing membrane protein production of GR and TR in *E. coli* LEMO21(DE3), pET28a(+)-expression vectors were generated. *GR* and *TR* were subcloned into the pET28a(+)-derived pGFPe and the C-terminal GFP fusion was removed by PCR amplification with phosphorylated primers (BG6696, BG6697 for *GR*, BG6696, BG8454 for *TR*) and recircularization by T4 ligase. pET28a(+)-YidC-GFP and pGFPe were a kind gift from Jan-Willem de Gier (Drew et al., 2006; Schlegel et al., 2012).

The pET28a(+)-AraH-GFP vector with an optimized sequence at the 5'UTR:CDS junction was reconstructed based on the optimal junction found before (Mirzadeh et al., 2015) by PCR amplification of pET28(+)-AraH-GFP by phosphorylated primers (BG8565,BG8566) and recirculation by T4 ligase. pET28a(+)-AraH-GFP was a kind donation of Daniel Daley (Daley, 2005).

Table 1. Plasmids used in this study.

Plasmid name	*Antibiotic marker*	*Origin of replication*	*Important components*	*Reference*
pFAB-P14-BCD9-GR	Kan	*P15a*	*P14, BCD9, GR*	This work
pFAB-P#-BCD#-GR	Kan	*P15a*	*Varying promoters and BCDs, GR*	This work
pFAB-P14-BCD#-GR-his	Kan	*P15a*	*P14, varying BCDs, GR-his*	This work
pFAB-P14-BCD#-TR-his	Kan	*P15a*	*P14, varying BCDs, TR-his*	This work
pFAB-P14-BCD#-YidC-GFP	Kan	*P15a*	*P14, varying BCDs, YidC-GFP*	This work
pFAB-P14-BCD#-AraH-GFP-his	Kan	*P15a*	*P14, varying BCDs, AraH-GFP-his*	This work
pET28a(+)-GR-his	Kan	*pBR322*	P_{T7}, *GR-his*	This work
pET28a(+)-TR-his	Kan	*pBR322*	P_{T7}, *TR-his*	This work
pET28a(+)-YidC-GFP	Kan	*pBR322*	P_{T7}, *YidC-GFP fusion*	(Schlegel et al., 2012)
pET28(+)-opt.junc.-AraH-GFP-his	Kan	*pBR322*	P_{T7}, *opt junction, AraH-GFP-his*	Reconstructed as published (Mirzadeh et al., 2015)

Strains and culture conditions

Table 2. Strains used in this study.

Strain	*Description*	*Origin*
E. coli DH5α	*F⁻ λ⁻ fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	NEB
E. coli MG1655	*F⁻ λ⁻ ilvG⁻ rfb-50 rph-1*	ATCC
E. coli BL21(DE3)	*F⁻ λ⁻ fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS*	NEB
E. coli LEMO21(DE3)	*F– λ– fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo(CamR)*	NEB
E. coli BL21(DE3) pLysS	*F⁻ λ⁻ fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/pLysS (CamR)*	Promega

E. coli strains (**Table 2**) were routinely cultured in Lysogeny Broth (LB) (Sambrook et al., 1989) with appropriate antibiotics, kanamycin (50 µg/mL) and chloramphenicol (35 µg/mL). All-trans retinal (Sigma) was added from a 20 mM ethanol stock to a final concentration of 20 µM and re-added after a 2-4 hours of cultivation to compensate for degradation.

Pre-cultures (100 µL/well) and expression cultures (500 µL/well) for library screening of *E. coli* DH5α or *E. coli* BL21(DE3) were performed in MASTERBLOCK® Deep 96-Well plates (Greiner). Plates were kept in a plastic bag with a humidified atmosphere to prevent evaporation, and grown for 22 hours at 37°C (or 30°C for AraH-GFP) at 900 rpm.

E. coli cultures for quantification of constitutive rhodopsin, YidC-GFP or AraH-GFP production, from BIOFAB-BCD vectors, were performed in 10 mL LB medium in 50 mL Greiner tubes. 1%-2% O/N pre-culture was added, they were cultivated for 22 hours at 250 rpm, at 37°C, or 30°C for AraH-GFP .

For *E. coli* LEMO21(DE3) cultures, fresh transformants were used to inoculate pre-cultures, as the use of re-streaked glycerol stocks was reported to lead to severely reduced recombinant protein production (Hjelm et al., 2013). Overnight pre-cultures were used to inoculate 1%-2% in 10 mL LB medium in 50 mL Greiner tubes with different L-rhamnose concentrations (0, 50, 100, 500, 1000 and 2000 µM). At an OD_{600} of 0.35-0.45, cells were induced with IPTG (isopropyl β-D-1-thiogalactopyranoside) at a concentration 0.4 mM, and for functional rhodopsin production all-trans retinal was added as well, and re-added 2-4 hours after induction. Cells were further cultivated for 22 hours after induction (37°C, 250 rpm) and then harvested for expression analysis.

For the expression of *arah-gfp* from *E. coli* BL21(DE3) pLysS pET28(+)-opt.junc.-AraH-GFP-his was cultured as published before (Mirzadeh et al., 2015). 2% O/N-pre-culture was inoculted in 10 mL LB in 50 mL tubes, and incubated at 37°C, 180 rpm till OD600 of 0.25-0.35 was reached. Then the cultures were induced with 1.0 mM IPTG and incubated for 5 hours at 25°C at 180 rpm .

Rhodopsin quantification

For rhodopsin quantification 10 mL of culture was resuspended in 295 µL extraction buffer, as in (Engqvist et al., 2015) and frozen for at least 1 hour. Cells were thawed and additional 295 µL extraction buffer, supplemented with 6 mg/mL lysozyme and 0.4 mg/mL DNase, was added. For cell lysis, this suspension was incubated at room temperature for 30 minutes. Rhodopsins were extracted from the crude cell extract by addition of 2.5% (w/v) dodecyl-maltoside (DDM, Sigma) and incubation at 180 rpm for 24-48 hours in dark. The extraction for GR was performed at room temperature, to increase the extraction efficiency of TR, the extraction hereof was performed at 65°C.

After extraction the mixture was spun down to check for the color of the pellet, if a colorless pellet was obtained the supernatant fraction was used for spectroscopic quantification. 200 µL of supernatant was transferred to a transparent flat bottom 96-well plate (Greiner) and the absorption spectrum (300-700 nm) was measured (Synergy MX BioTek). Next, 0.1 M hydroxylamine was added to bleach the retinal from the rhodopsin for 1 hour in dark at room temperature with gentle shaking. Then, the absorption spectrum was measured again. The difference absorption spectrum could be generated, and from differential absorption at 540 nm (GR) or 525 nm (TR), the rhodopsin concentration was determined, assuming an extinction coefficient of 50,000 ($M\ cm^{-1}$) for both rhodopsins.

Whole-cell GFP quantification

Production of YidC-GFP and AraH-GFP was quantified using whole-cell GFP fluorescence as described before (Hjelm et al., 2013). In short, 1 mL of culture (or 0.5 mL from deep-well cultivations) was resuspended in ice-cold 100 µL PBS (phosphate buffer saline) and incubated at 4°C for at least 1 hour for further maturation of GFP. After this, suspensions were centrifuged (10 min, 13,000xg, 4ºC) and resuspended in 100 µL PBS, which was transferred to a black 96-well

plate with transparent bottoms (Greiner). Fluorescence was directly measured using excitation at 485 nm and emission at 512 nm at a constant gain value (75) (BioTEK SynergyMX).

In-gel fluorescence

To validate if the GFP signal originated from full-length fusions of YidC-GFP and AraH-GFP, in-gel fluorescence was performed, essentially as described before (Geertsma et al., 2008; Marino et al., 2015). Cultures were centrifuged for 5 minutes at 13,000xg and stored at -20°C. After thawing, pellets were resuspended to an estimated final concentration of 0.5 mg protein/100 µL in 50 mM kPi buffer (pH 7.5) (assuming 150 mg protein/L for OD_{600} of 1). This buffer was supplemented with 1 mM $MgSO_4$, 10% glycerol, 1 mM EDTA, 0.1 mg/mL DNase and 10 mg/mL lysozyme. Cells were lysed for 1 hour under mild shaking at room temperature and stored at -20°C for later analysis. 4x Laemmli buffer (Biorad) was added to the cell lysate, incubated for 5 minutes at 37°C only, to prevent denaturation of folded GFP. Directly after the resuspension, the samples were shortly sonicated with three 0.1 ms pulses (Bandelin SONOPLUS HD 3100) to reduce sample viscosity for loading. Twenty-five µL of sample (~90 µg protein) was loaded and separated on a 12% Mini-PROTEAN® TGX™ protein gel (Biorad). After running in-gel fluorescence was visualized using a Syngene G-box using a 525nm filter.

Author contributions

NJC conceived the idea and designed this study with support from WMdV and JvdO. NJC, BS, FM and JdG performed the experimental work and data analysis, where NJC coordinated and supervised the work of the latter three. NJC drafted the manuscript, which was critically revised by WMdV and JvdO.

Acknowledgements

We would like to thank Jan-Willem de Gier for his advice and insights on membrane protein production and for sharing the pGFPe and pGFPe-YidC vectors. We thank Daniel Daley for sharing pGFPe-AraH-his. All BIOFAB plasmids were a kind gift of Drew Endy.

Supplementary information

Supplementary Table S1. Other plasmids generated for this study.

Plasmid name	*Antibiotic marker*	*Origin of replication*	*Important components*	*Reference*
pBbE0A-RFP	Amp	*colE1*	*mRFP*	(Lee et al., 2011) Addgene # 35372
pBbE0A-GR-GFP	Amp	*colE1*	*GR-GFP-his fusion protein*	This work
pGFPe	Kan	*pBR322*	*pET28a(+) derived for N-term GFP-his fusions*	(Drew et al., 2006)
pGFPe-TR	Kan	*pBR322*	*P_{T7}, TR-GFP-his fusion*	GeneArt/this work
pGFPe-GR	Kan	*pBR322*	*P_{T7}, GR-GFP-his fusion*	GeneArt/this work
pFAB-P14-YidC-GFP	Kan	*P15a*	*P14,RBS1, YidC-GFP*	This work
pFAB-P14-TR	Kan	*P15a*	*P14,RBS1, TR*	This work
pGFPe-AraH (pET28(+)-AraH-GFP-his)	Kan	*P15a*	*P_{T7}, AraH-GFP-his fusion*	(Daley, 2005)
pLEMO	Cam	*P15a*	*PrhaBAD, lysY*	(Wagner et al., 2008)

Supplementary Table S2. Oligos used in this study.

Oligo nr.	*Sequence*	*Purpose*
BG5971	TTTTTTGGTCTCAAATGCTGATGACCGTTTTT	GR-GFP Fw amplification, BsaI
BG5972	TTTTTGGTCTCCTCAGTGGTGGTGGTGGTG	GR-GFP Rv amplification, BsaI
BG5973	TTTTTGGTCTCNCTGAGGATCGGTTGTCGAGTAAGG	pFAB Fw, BsaI
BG5975	TTTTTTGGTCTCCCATTAGAAACGGTCCGATG	pFAB3912 Rv, BsaI
BG5976	TTTTTTGGTCTCCCATTAGAAACTGTCCTTCGC	pFAB 3999 Rv, BsaI
BG5978	TTTTTTGGTCTCCCATTAGAAAGACTCCTCTGCA	pFAB 3913 Rv, BsaI
BG5981	TTTTTTGGTCTCCCATTAGAAAGTCTCCTGTGCA	All pFAB with BCD1 Rv, BsaI
BG6162	TGAGGATCGGTTGTCGAGTA	GFP deletion from GR-GFP fusion vectors Fw
BG6163	GCTAATCAGGCTACCACCGC	GFP deletion from GR-GFP fusion vectors Rv
BG6696	CACCACCACCACCACTGAC	pET-delGFP fw
BG6697	GTGCTCGAGTGCGGCCGCGCTAATCAGGCTACCACCGC	pET-GR-delGFP-rv
BG7291	AACATCTTAATCATGCACAGGAGACTTTCT	BCD1-fw
BG7292	CATTAGAAAGTCTCCTGTGCATGATTAAGA	BCD1-rv
BG7293	AACATCTTAATCATGCTAAGGAGGTTTTCT	BCD2-fw
BG7294	CATTAGAAAACCTCCTTAGCATGATTAAGA	BCD2-rv
BG7295	AACATCTTAATCATGCAGGGGAGGGTTTCT	BCD5-fw
BG7296	CATTAGAAACCCTCCCCTGCATGATTAAGA	BCD5-rv
BG7297	AACATCTTAATCATGCGCCGGAGGTTTTCT	BCD6-fw
BG7298	CATTAGAAAACCTCCGGCGCATGATTAAGA	BCD6-rv

Supplementary Table S2. continued

Oligo nr.	*Sequence*	*Purpose*
BG7299	AACATCTTAATCATGCTGGGGAGGGTTTCT	BCD7-fw
BG7300	CATTAGAAACCCTCCCCAGCATGATTAAGA	BCD7-rv
BG7301	AACATCTTAATCATGCATCGGACCGTTTCT	BCD8-fw
BG7302	CATTAGAAACGGTCCGATGCATGATTAAGA	BCD8-rv
BG7303	AACATCTTAATCATGCAGAGGAGTCTTTCT	BCD9-fw
BG7304	CATTAGAAAGACTCCTCTGCATGATTAAGA	BCD9-rv
BG7305	AACATCTTAATCATGCGGAGGATCGTTTCT	BCD10-fw
BG7306	CATTAGAAACGATCCTCCGCATGATTAAGA	BCD10-rv
BG7307	AACATCTTAATCATGCGGGGGAGTGTTTCT	BCD11-fw
BG7308	CATTAGAAACACTCCCCCGCATGATTAAGA	BCD11-rv
BG7309	AACATCTTAATCATGCTGCGGAGGGTTTCT	BCD12-fw
BG7310	CATTAGAAACCCTCCGCAGCATGATTAAGA	BCD12-rv
BG7311	AACATCTTAATCATGCAATGGAGGCTTTCT	BCD13-fw
BG7312	CATTAGAAAGCCTCCATTGCATGATTAAGA	BCD13-rv
BG7313	AACATCTTAATCATGCGGTGGAGGGTTTCT	BCD14-fw
BG7314	CATTAGAAACCCTCCACCGCATGATTAAGA	BCD14-rv
BG7315	AACATCTTAATCATGCGGGGGAGTCTTTCT	BCD15-fw
BG7316	CATTAGAAAGACTCCCCCGCATGATTAAGA	BCD15-rv
BG7317	AACATCTTAATCATGCTTAGGAGTCTTTCT	BCD16-fw
BG7318	CATTAGAAAGACTCCTAAGCATGATTAAGA	BCD16-rv
BG7319	AACATCTTAATCATGCGGAGGAGGGTTTCT	BCD17-fw
BG7320	CATTAGAAACCCTCCTCCGCATGATTAAGA	BCD17-rv
BG7321	AACATCTTAATCATGCGACGGAGCGTTTCT	BCD18-fw
BG7322	CATTAGAAACGCTCCGTCGCATGATTAAGA	BCD18-rv
BG7323	AACATCTTAATCATGCTATGGAGGTTTTCT	BCD19-fw
BG7324	CATTAGAAAACCTCCATAGCATGATTAAGA	BCD19-rv
BG7325	AACATCTTAATCATGCTGAGGAAAGTTTCT	BCD20-fw
BG7326	CATTAGAAACTTTCCTCAGCATGATTAAGA	BCD20-rv
BG7327	AACATCTTAATCATGCGAGGGATGGTTTCT	BCD21-fw
BG7328	CATTAGAAACCATCCCTCGCATGATTAAGA	BCD21-rv
BG7329	AACATCTTAATCATGCCTAGGAAGTTTTCT	BCD22-fw
BG7330	CATTAGAAAACTTCCTAGGCATGATTAAGA	BCD22-rv
BG7331	AACATCTTAATCATGCGAAGGACAGTTTCT	BCD23-fw
BG7332	CATTAGAAACTGTCCTTCGCATGATTAAGA	BCD23-rv
BG7333	AACATCTTAATCATGCGATGGACGGTTTCT	BCD24-fw
BG7334	CATTAGAAACCGTCCATCGCATGATTAAGA	BCD24-rv
BG7335	ATTATAGGTCTCAAATGCTGATGACCGTTTTTAGCA	pFAB-P14-GR-fw
BG7336	TAATTAGGTCTCATGTTTCAGTACGAAAATTGCTTTCATT	pFAB-P14-GR-rv
BG7337	GATCCTTAGTGGTGGTGGTGGTGCTCTAACGC-CGCTTTAGCTTCC	TR-rv, Gibson
BG7338	GATGAGCTCTACAAATAAGGATCGGTTGTCGAGTAAG	pFAB-P14-fw2, Gibson
BG7339	TGGCTACTTATAGGTCTCTTGTTTCAGTAC-GAAAATTGCTTTCATT	pFAB-P14-rv, Gibson

Supplementary Table S2. continued

Oligo nr.	*Sequence*	*Purpose*
BG7340	CGTACTGAAACAAGAGACCTATAAGTAGCCAGGTCTCAAATGC-GGATGTTACCCGAA	TR-fw, Gibson
BG7341	CACCACCACCACCACT	pFAB-P14-fw, Gibson
BG7342	AAACAAGAGACCTATAAGTAGCCAGGTCTCAAATGGATTCG-CAACGCAATC	YidC-GFP fw, Gibson
BG7343	CGATCCTTATTTGTAGAGCTCATCCATGCC	YidC-GFP rv, Gibson
BG7937	TTTATTGGTCTCTATCCGACAGTGCGTTTCGCTTTTTG	AraH-GFP rv, BsaI
BG7505	TATATTGGTCTCAAATGCGGATGTTA	pFAB-P14-TR fw, BsaI
BG7506	CTTATAGGTCTCTTGTTTCAGTACG	pFAB-P14-TR rv, BsaI
BG7784	TATATTGAAGACTTAATGGATTCGCAACGCA	pFAB-P14-YidC-GFP fw, BbSI
BG7785	TTATTAGAAGACTATGTTTCAGTACGAAAATTGCTTTC	pFAB-P14-YidC-GFP rv, BbSI
BG8454	GTGCTCGAGCGCCGCTTTAGCTTCCACAC	pET-TR-delGFP-rv
BG8565	ATTACTATGATGTCGTCTGTTTCTACATCGGGGTCTGGC	Construction of optimized junction from pET28a-AraH-GFP Fw
BG8566	TCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGG-GAATTGTTATCC	Construction of optimized junction from pET28a-AraH-GFP rv
BG7933	TATTATGGTCTCAGGATCCGAAAACCTGTACTTCCAG	pFAB-P14-fw, BsaI
BG7934	TAATTAGGTCTCATGTTTCAGTACGAAAATTGCTTTCATTGT	pFAB-P14-fw, BsaI
BG7937	TTTATTGGTCTCTATCCGACAGTGCGTTTCGCTTTTTG	AraH-GFP rv, Golden gate
BG8448	TATTATGGTCTCTAATGATGTCTTCTGTTTCTACATCGGG	AraH-GFP fw, BsaI

Supplementary data 1

codon-optimized GR +his

ATGCTGATGACCGTTTTTAGCAGCGCACCGGAACTGGCACTGCTGGGTAGCACCTTTGCACAG-
GTTGATCCGAGCAATCTGAGCGTTAGCGATAGCCTGACCTATGGTCAGTTTAATCTGGTGTATAAC-
GCATTTAGCTTTGCCATTGCAGCAATGTTTGCAAGCGCACTGTTTTTTTTCAGCGCACAGGCACT-
GGTTGGTCAGCGTTATCGTCTGGCCCTGCTGGTGAGCGCAATTGTTGTTAGCATTGCAGGCTAT-
CATTATTTCCGCATTTTCAATAGCTGGGATGCAGCATATGTTCTGGAAAATGGTGTTTATAGTCT-
GACCAGCGAGAAATTCAATGATGCCTATCGTTATGTTGATTGGCTGCTGACCGTTCCGCTGCTGCT-
GGTTGAAACCGTTGCAGTTCTGACCCTGCCTGCAAAAGAAGCACGTCCTCTGCTGATCAAACT-
GACCGTTGCAAGCGTTCTGATGATTGCAACCGGCTATCCGGGTGAAATTAGTGATGATATTAC-
CACCCGTATTATTTGGGGCACCGTTAGCACCATTCCGTTTGCATATATTCTGTATGTTCTGTGGGTT-
GAACTGAGCCGTAGCCTGGTTCGTCAGCCTGCCGCAGTGCAGACCCTGGTGCGTAATATGCGTTG-
GTTACTGCTGCTGAGCTGGGGTGTTTATCCGATTGCATATCTGCTGCCGATGCTGGGTGTGAGC-
GGCACCAGCGCAGCAGTTGGTGTTCAGGTTGGTTATACCATTGCAGATGTTCTGGCCAAACCT-
GTTTTTGGTCTGCTGGTTTTTGCAATTGCCCTGGTTAAAACCAAAGCAGATCAAGAAAGCAGC-
GAACCGCATGCAGCAATTGGTGCAGCAGCAAATAAAAGCGGTGGTAGCCTGATTAGCTGA

codon-harmonized TR + his

ATGCGGATGTTACCCGAACTGAGCTTTGGAGAATATTGGTTAGTCTTTAACATGCTGAGCCT-
GACCATTGCGGGCATGTTAGCGGCGTTTGTCTTTTTTCTGTTAGCTCGGAGCTATGTGGCGCCG-
CGTTATCATATTGCGCTGTATCTGAGCGCGCTGATTGTCTTCATTGCGGGCTATCATTATTTAAG-
GATTTTCGAAAGCTGGGTGGGCGCGTATCAGTTACAGGATGGCGTATATGTGCCCACTGGCAAAC-
CGTTTAACGATTTTTATCGTTATGCGGATTGGCTGCTGACCGTGCCGTTACTGCTGTTAGAACT-
GATTTTAGTCCTAGGTCTTACCGCTGCGCGTACCTGGAACCTAAGCATTAAACTTGTGGTGGCGT-
CAGTCTTAATGTTAGCGCTTGGCTATGTGGGAGAAGTGAACACTGAACCGGGACCGCGGACCTTA-
TGGGGCGCGTTAAGCAGCATACCGTTTTTTTATATTCTGTATGTGCTGTGGGTGGAATTAGGTCAG-
GCGATTCGCGAAGCTAAATTTGGTCCGCGGGTGTTAGAATTATTAGGTGCGACCCGTCTGGTCCT-
GTTAATGAGCTGGGGTTTTTATCCGATTGCGTATGCGTTAGGTACCTGGCTGCCGGGAGGCGCT-
GCGCAGGAAGTGGCGATTCAGATAGGTTATAGCCTTGCTGATTTAATTGCGAAACCGATTTATG-
GTTTATTAGTCTTTGCGATTGCGCGCGCGAAAAGCCTGGAAGAAGGTTTTGGTGTGGAAGCTA-
AAGCGGCGTTAGAGCACCACCACCACCACCACTAA

Chapter 8

Harnessing the power of microbial autotrophy

Nico J. Claassens[1]*, Diana Z. Sousa[1], Vitor A.P. Martins dos Santos[2,3], Willem M. de Vos[1,4], and John van der Oost[1]

*first author
[1]Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE, Wageningen, The Netherlands
[2]Laboratory of Systems and Synthetic Biology, Wageningen University, Stippeneng 4, 6708 WE, Wageningen, The Netherlands
[3]LifeGlimmer GmbH, Markelstr. 38, Berlin, Germany
[4]Research Programme Unit Immunobiology, Department of Bacteriology and Immunology, Helsinki University, Haartmaninkatu 3, 00014, Helsinki, Finland

Chapter adapted from publication:
Nature Reviews Microbiology (2016) 14:692-706
DOI: 10.1038/nrmicro.2016.130

Abstract

Autotrophic microorganisms convert CO_2 into biomass by deriving energy from light or inorganic electron donors. These CO_2-fixing microorganisms have a large, but so far limitedly realized, potential for the sustainable production of chemicals and biofuels. Productivities have been improved in autotrophic hosts by introducing production pathways and modifying autotrophic systems through genetic engineering. In addition, approaches are emerging in which CO_2 fixation pathways and energy-harvesting systems are transplanted into heterotrophic model microorganisms. Alternative, promising concepts are hybrid production systems of autotrophs and heterotrophs, and bio-inorganic hybrids of autotrophic microorganisms with electrocatalysts or light-harvesting semiconductor materials. In this Review, we discuss recent advances and bottlenecks for engineering microbial autotrophy and elaborate on novel strategies, which will pave the way towards improved microbial autotrophic production platforms.

Keywords
Microbial autotrophy; CO_2 fixation; Photosystems; Synthetic Biology; Microbial chemical and fuel production

Introduction

Autotrophic microorganisms have the ability to fix carbon dioxide (CO_2) and convert CO_2 into biomass and potentially into products of interest in an industrial setting. To reduce CO_2, autotrophs require an energy source, either light (photoautotrophy) or an external inorganic electron donor (chemolithoautotrophy) (**Figure 1a**). Autotrophic production potentially offers a more sustainable, carbon-neutral alternative to chemical production processes that are based on fossil resources, and is potentially more energy efficient than the heterotrophic conversion of organic substrates into similar products (**Figure 1b**) (Conrado et al., 2013; Hawkins et al., 2013). However, so far, the biotechnological application of autotrophic microorganisms is limited. Autotrophic microorganisms evolved their pathways and systems to produce biomass and thrive in natural conditions. Efficient production of chemicals by autotrophs in industrial conditions requires major adaptations of these microorganisms.

To improve industrial performance, native systems of autotrophic microorganisms, such as CO_2 fixation pathways and photosystems, have been modified or replaced by using genetic engineering approaches (Beckmann et al., 2009; Shih et al., 2014; Straub et al., 2014). In addition, pathways to generate products of interest have been introduced (Angermayr et al., 2015; Gimpel et al., 2015). However, the available genetic engineering tools for autotrophic hosts are still limited compared with the advanced toolboxes that exist for model heterotrophs. Therefore partial, and recently even complete CO_2 fixation pathways and other systems related to autotrophy have been transplanted into heterotrophs (Antonovsky et al., 2016; Bonacci et al., 2012; Guadalupe-Medina et al., 2013; Martinez et al., 2007; Mattozzi et al., 2013). These transplantations provide useful insights into the possibilities and challenges of engineering autotrophic systems, which may also be applied to improve natural autotrophs in the future. However, as several systems are required for complete autotrophy, converting a microbial heterotroph into an autotroph still remains a major challenge. An alternative promise of introducing CO_2 fixation pathways or other autotrophic systems into heterotrophs is that their product yields can be increased, when they are grown as mixotrophs on both CO_2 and organic substrates (Gong et al., 2015; Guadalupe-Medina et al., 2013).

This Review discusses recent advances in the genetic and metabolic engineering of microorganisms to pave the way for efficient autotrophic production platforms. We discuss hosts with a potential to become autotrophic or mixotrophic microbial cell factories and describe important system related to autotrophy in more detail. We then assess the potential of diverse natural and synthetic CO_2 fixation pathways as well as photosystems and systems for taking up external electron donors. Finally, this Review highlights how synthetic biology principles, such as model-driven design and modularization, enable better quantitative design and experimental implementation of engineered autotrophic microbial production platforms.

Potential solar-to-product efficiencies

So far, most progress in genetic and metabolic engineering for the production of chemicals and biofuels has been made for heterotrophic microorganisms. Heterotrophs can convert plant-derived biomass, such as carbohydrates or lignocellulosic materials, into products. Production from plant biomass is a theoretically carbon-neutral alternative to fossil-resource-based processes. However, the overall solar-to-product energy conversion efficiency from plant biomass is generally very low (**Figure 1b**). In a typical fermentation process of sugarcane to ethanol, the final product contains only ~0.2% of the available solar energy (Conrado et al., 2013). This low efficiency is mainly due to the ineffective plant photosynthesis (Zhu et al., 2010), but also to energy losses in the subsequent processing of biomass and microbial fermentation (Conrado et al., 2013; Hawkins et al., 2013). In addition, plant biomass-based production generally competes with food and feed production for arable land and other limited resources, while food production still needs to grow to meet the demands of a growing world population (Bruinsma J., 2009). Furthermore, from an economic perspective, plant biomass substrates typically add up to more than half of the total costs of biobased production (Fast and Papoutsakis, 2012). Autotrophic microbial cell factories are potentially more efficient and sustainable than plant biomass-based chemical production. For example, photosynthetic microorganisms, such as cyanobacteria and microalgae, have a potentially higher photosynthetic efficiency than plants. While for plants the typical solar-to-biomass efficiency is only 1 - 2% (Zhu et al., 2010), for microalgae

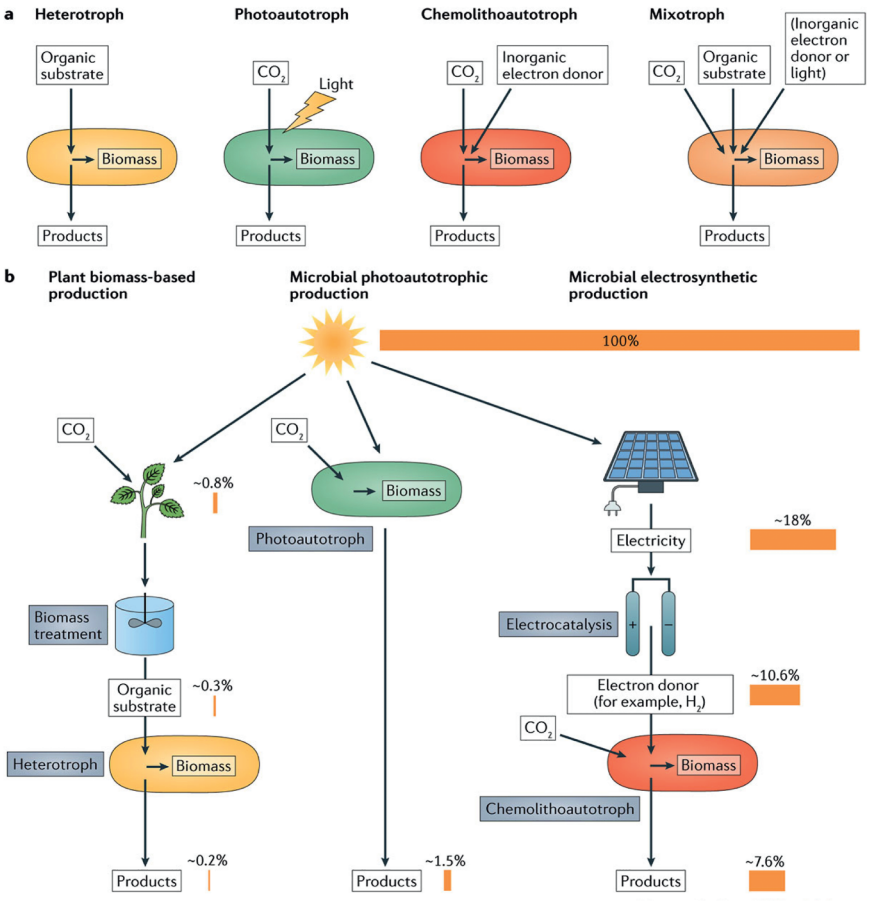


Figure 1. Production platforms based on different types of microbial trophic metabolism. a| Different types of microbial trophic metabolisms can be applied for microbial production. Heterotrophic microorganisms convert organic carbon substrates into biomass and products. Autotrophic microorganisms fix and convert carbon dioxide (CO_2) into biomass and products. To reduce CO_2, autotrophs require an energy source, either light (photoautotrophy) or an external inorganic electron donor (chemolithoautotrophy). Mixotrophs use both CO_2 and an organic carbon source to grow. All these different types of trophic metabolism can be exploited for the production of products of interest in an industrial setting. b| Diverse platforms can be envisioned for biotechnological chemical production from sunlight and CO_2. Heterotrophs can ferment plant-derived biomass, such as carbohydrates or lignocellulosic materials, into products. The overall solar-to-product energy conversion efficiency from such heterotrophic fermentations is generally low, the efficiency of typical sugarcane-to-ethanol fermentation in Brazil is about 0.2% (Conrado et al., 2013; Leal et al., 2012).

Photosynthetic microorganisms have potentially higher solar-to-product conversion efficiencies, for example~1.5% for oil production by microalgae (Wijffels and Barbosa, 2010). The efficiency of photovoltaic solar panels is much higher compared to biological photosynthesis, and these solar panels can be combined with electrocatalytic production of for example hydrogen, which can be subsequently used by chemolithoautotrophic cell factories. This so-called microbial electrosynthetic production has a demonstrated solar-to-product efficiency of ~7.6% for polyhydroxybutyrate production by *Cupriavidus necator* (Liu et al., 2016).

energy conversion efficiencies of 3% have been reported (Wijffels and Barbosa, 2010). Thus, there seems to be room for improving the photosynthetic efficiency of plants and photosynthetic microorganisms, in particular through synthetic biology approaches (Blankenship et al., 2011; Ort et al., 2015).

Compared to biological photosynthesis, the efficiency of photovoltaic solar panels is very high; currently available solar panels have solar-to-electricity efficiencies of ~18%, and new innovations may enable efficiencies above 40% (Blankenship et al., 2011; Conrado et al., 2013). This notion has led to the development of microbial electrosynthesis (Conrado et al., 2013; Rabaey and Rozendal, 2010), in which electrical energy generated by solar panels or other renewable energy technologies is used to efficiently generate electron donors for the production of chemicals and biofuels by chemolithoautotrophs (**Figure 1b**). An example is the energy-efficient generation of hydrogen (H_2) by electrocatalysis of water, and subsequent feeding of the generated H_2 as electron donor to chemolithoautotrophs. Using CO_2 as carbon source, these microorganisms can generate a range of products from inorganic electron donors such as H_2. Microbial electrosynthesis based on H_2 was recently demonstrated to enable very efficient solar-to-product conversion (Liu et al., 2016) (**Figure 1b**).

Apart from H_2 , some other non-carbon electron donors, such as cathodic electrons (Rabaey and Rozendal, 2010), ferrous iron (Fe^{2+}) or ammonium (NH_4), can be generated by electrical energy, and directly be metabolized by some chemolithoautotrophs (Nybo et al., 2015). In addition, some electrocatalysis systems can efficiently generate reduced one-carbon (C_1) compounds from CO_2 and electricity (Lu and Jiao, 2016), which can be metabolized by some microorganisms, including some autotrophs. Although this Review focusses on microbial conversions of CO_2 into products, similar promises and challenges are faced for microbial conversion of reduced C_1 compounds into products (**Box 1**).

Box 1. C_1 assimilation

Many microorganisms can use reduced one-carbon (C_1) compounds both as carbon and electron sources; C_1 compounds include carbon monoxide (CO), formic acid (CHOOH), methanol (CH_3OH) and methane (CH_4). C_1 compounds are relatively cheap substrates compared to plant-derived carbohydrates. For example CO mixed with H_2 and CO_2 (syngas) can be obtained by gasification of diverse waste streams(Latif et al., 2014) or from off-gases from the steel industry(Molitor et al., 2016). Methane and methanol can be obtained either by processing waste streams or directly from natural gas(Whitaker et al., 2015). In addition, reduced C_1 compounds can be produced by electrochemical processes. Especially formate and CO can be produced with high (> 80%) electricity-to-C_1 compound energy efficiency by electrocatalysis in the presence of CO_2 (Lu and Jiao, 2016). This makes these C_1 compounds ideal substrates for the so-called microbial electrosynthesis. In particular formate is an attractive substrate as it is more soluble in microbial growth media than gaseous compounds, such as H_2, CO_2, CO and CH_4 (Yishai et al., 2016).

Carbon monoxide, possibly mixed with H_2 and CO_2 in syngas, can be used as a substrate by various acetogens (Diender et al., 2015) (see also main text). Methanol, methane and formate can be substrates for various methylotrophic microorganisms (Chistoserdova et al., 2009). Analogous to engineering autotrophic organisms and their pathways, methylotrophic microorganisms could be further improved for high growth and productivity under industrial conditions. These organisms and their potential for biotechnological applications are extensively reviewed elsewhere (Dürre and Eikmanns, 2015; Kalyuzhnaya et al., 2015; Whitaker et al., 2015).

Alternatively, the transplantation of simple methylotrophic pathways into non-methylotrophs is a promising option (reviewed in (Bar-Even et al., 2013; Whitaker et al., 2015)). A successful proof-of-principle has been provided by the model-driven experimental implementation of a methylotrophic pathway in *E. coli*, which enabled efficient assimilation of methanol into biomass (Müller et al., 2015). A theoretically more efficient synthetic methanol assimilation pathway has been designed; this methanol condensation cycle was functional in an *in vitro* system with purified enzymes (Bogorad et al., 2014), but has not yet been functionally introduced into a microbial host.

Extensive analyses suggested several synthetic formate assimilation pathways based on natural enzymes that are promising to be implemented into *E. coli* (Bar-Even, 2016; Bar-Even et al., 2013). A synthetic pathway for aerobic formate assimilation, based on the non-natural, computationally-designed formolase enzyme, has been introduced into *E. coli*; however, the obtained overall formate assimilation activity for this pathway is still very low (Siegel et al., 2015). As many of the approaches, pathways and hosts applied for metabolic engineering of C_1 assimilation overlap with metabolic engineering of CO_2 fixation, developments in both fields are expected to reinforce each other.

Host selection

The selection of one or more host chassis is crucial for developing efficient autotrophic production platforms. A host chassis of choice should ideally be well-characterized in terms of physiology, biochemistry and genetics. Furthermore, metabolic engineering requires a high transformation efficiency and a powerful genetic toolbox. Hosts that best meet these criteria include the gram-negative bacterium *Escherichia coli* and the yeast *Saccharomyces cerevisiae*. However, as these organisms are strict heterotrophs, extremely challenging engineering strategies are required to convert them into complete autotrophs. Still, transplanting autotrophic (sub)systems into heterotrophs is promising as it may enable for higher 'mixotrophic' product yields by additional CO_2 assimilation and/or energy-harvesting. Mixotrophy has also been proposed to increase and stabilize yields and productivities of autotrophic hosts from CO_2, by supplementation with organic substrates (Fast and Papoutsakis, 2012; Fast et al., 2015; Wan et al., 2015). Recently, mixotrophy in an engineered cyanobacterium was shown to enable stable 2,3-butanediol production in a diurnal light-dark cycle, using both light and sugar as energy sources (McEwen et al., 2016).

Natural autotrophs

The wide variety of natural autotrophs provides a large pool of potential hosts to be converted into autotrophic cell factories. However, here we discuss only autotrophic hosts that have been shown to be transformable and for which some genetic tools are available.

Natural photoautotrophic microorganisms are either oxygenic or anoxygenic. Oxygenic photoautotrophs have water-splitting, oxygen-generating photosystems that can generate reducing power and a proton gradient for regeneration of ATP. For some oxygenic photoautotrophic, cyanobacterial strains of *Synechocystis* and *Synechococcus* sets of genetic tools have been developed (Berla et al., 2013; Ramey et al., 2015). In recent years, this has enable impressive progress in the metabolic engineering of both cyanobacterial species (Angermayr et al., 2015), for example the production of commodity chemicals such as isobutyraldehyde (Atsumi and Liao, 2009), L-lactic acid (Angermayr et al., 2014) and 2,3-butanediol (Oliver et al., 2013). So far cyanobacteria are the best-developed autotrophic cell factories, still their productivities lag far behind heterotrophic cell factories. The toolboxes for these cyanobacteria can also be applied to modify, complement or replace their autotrophic systems, possibly making them more efficient autotrophic cell factories (Gao et al., 2016). Other potential oxygenic photoautotrophic cell factories are eukaryotic microalgae, for example for the production of lipids and alkanes. However, modifications of most microalgae are so far severely hampered by poor transformation efficiencies and insufficient genetic tools, though some progress is being made for *Chlamydomonas reindhardtii* and a few other species (Gimpel et al., 2015; Wijffels et al., 2013).

Anoxygenic photolithoautotrophic microorganisms have photosystems that cannot split water; instead they require inorganic electron donors (such as H_2, hydrogen sulfide or sulfur) to generate reducing power. Genetic tools are available for a few anoxygenic bacteria, such as *Rhodobacter sphaeroides* (Nybo et al., 2015), which can be applied for metabolic engineering (Jaschke et al., 2011; Tikh et al., 2014b). In addition, *R. sphaeroides* has metabolic flexibility that enables aerobic, anaerobic, heterotrophic and autotrophic growth.

There is a major potential to improve the low light-harvesting efficiency of phototrophic microorganisms (Blankenship et al., 2011), which on the one hand is very challenging but on the other hand is very important as it is required to make large-scale industrial application of photosynthetic organisms economically feasible (Wijffels and Barbosa, 2010; Wijffels et al., 2013). Furthermore, there is large potential to improve the efficiency of CO_2 fixation pathways, especially for all oxygenic photoautotrophs and many

anoxygenic photoautotrophs as well as chemoautotrophs that harbor the ubiquitous Calvin cycle (also known as Calvin-Benson-Bassham or reductive pentose phosphate cycle) (**Figure 2**). This cycle is ATP-inefficient and the kinetic rates of several enzymes in this pathway are low (Bar-Even et al., 2010) (see below).

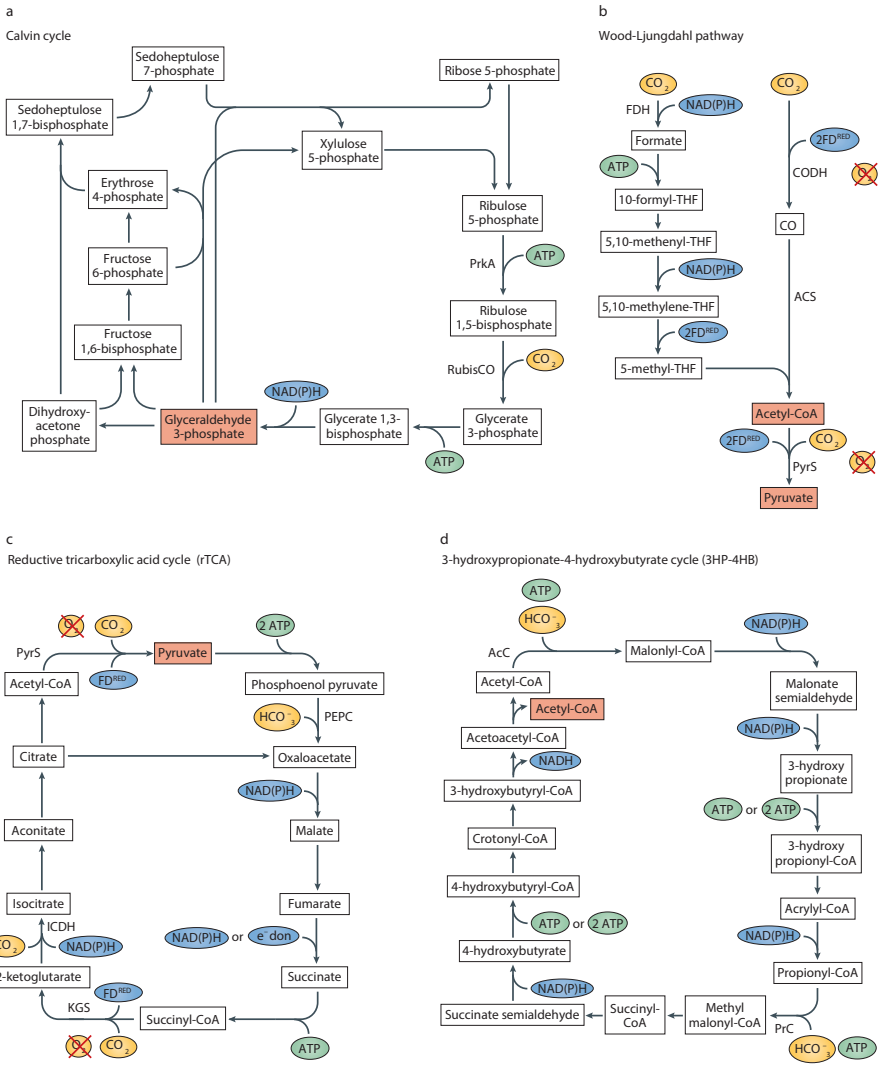
Alternative to light, inorganic electron donors are an energy source for a broad range of chemolithoautotrophic microorganisms. A promising chemolithoautotrophic candidate is the genetically accessible 'Knall-gas' bacterium *Cupriavidus necator* (formerly *Ralstonia eutropha*), which can oxidize H_2 or formate under aerobic conditions, while using the Calvin cycle for CO_2 fixation (Nybo et al., 2015). *C. necator* has already been successfully integrated with biocompatible electrochemical systems that generate H_2 (Liu et al., 2016; Torella et al., 2015) or formate (Li et al., 2012b) (**Figure 3a**), leading to high solar-to-product efficiencies for various products. *C. necator* can naturally accumulate poly-hydroxybutyrate (PHB), a storage polymer and bioplastic precursor, in amounts up to 70% of its biomass weight (Nybo et al., 2015). In addition, genetic engineering of heterologous production pathways into *C. necator* has led to the efficient production of compounds like branched-chain alcohols (Lu et al., 2012) and alkanes (Bi et al., 2013).

Another interesting group of chemolithoautotrophs includes bacteria that can exchange electrons directly with an electrode (Sydow et al., 2014). For example, *Acidithiobacillus ferrooxidans* is an aerobic bacterium, which harbors the Calvin cycle and can take up electrons from electrochemically produced Fe^{2+} or directly from a cathode in a bioelectrochemical system (Sydow et al., 2014). Genetic tools for *A. ferrooxidans* are emerging, which recently enabled the introduction of production pathways for isobutyric acid and heptadecane (Kernan et al., 2016).

A promising, special group of chemolithoautotrophs is formed by the acetogens, which can grow anaerobically on H_2 and CO_2, but also on many other compounds, such as C_1 compounds (Schiel-Bengelsdorf and Dürre, 2012) (**Box 1**). All acetogens harbor the Wood-Ljungdahl pathway for CO_2 fixation, which is the only CO_2 fixation route that can conserve ATP, but it operates near the border of thermodynamic feasibility (Bar-Even et al., 2010; Schuchmann

and Müller, 2014) (**Figure 2b**). To generate sufficient ATP, acetogens have a high production rate of acetate, which is coupled to ATP regeneration through acetate kinase. In addition to acetate, some acetogens, such as *Clostridium ljungdahlii* and *Clostridium autoethanogenum*, can naturally form other products, such as ethanol or 2,3-butanediol (Schiel-Bengelsdorf and Dürre, 2012).

Even though transformation efficiencies and genetic tools are generally limited for acetogens, recent progress has been made in metabolic engineering of especially *C. ljungdahlii*, *C. autoethanogenum* and *Acetobacterium woodii* to enable the production of , for example, butanol, butyrate and acetone (Dürre and Eikmanns, 2015; Hoffmeister et al., 2016; Köpke et al., 2010; Ueki et al., 2014). Furthermore, autotrophic acetate production by *A. woodii* was improved by overexpression of some key enzymes of the Wood-Ljungdahl CO_2 fixation pathway (Straub et al., 2014). Advances in the metabolic engineering of acetogens already led to the first large-scale industrial applications of *C. authoethanogenum* for the production of ethanol from syngas (mixture of H_2, CO_2 and CO) (Dürre and Eikmanns, 2015). However, the low ATP regeneration capacity of strictly anaerobic acetogens limits the feasible product yields of more ATP-demanding products (Fast and Papoutsakis, 2012). This may be circumvented by introducing more ATP efficient pathways or additional ATP regeneration system. Alternatively, as recently demonstrated, acetogens can be integrated with heterotrophs that can aerobically convert acetate into more ATP-demanding products, as will be discussed below (Hu et al., 2016; Liu et al., 2015).



Heterotrophs

Model heterotrophs, like *E. coli* and *S. cerevisiae*, lack functional CO_2 fixation pathways, photosystems or functional external electron uptake systems. Some of these systems have been (partly) successfully transplanted in those heterotrophic hosts, enabling mixotrophy. Some other heterotrophs natively contain external electron uptake systems required for chemolithoautotrophy. For example, some heterotrophic bacteria can naturally take up or donate

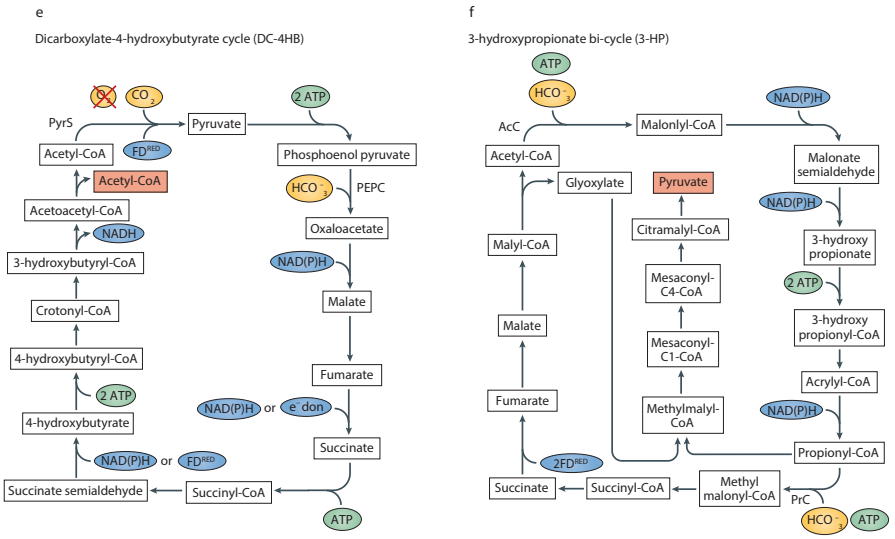


Figure 2. Natural CO_2 fixation pathways. All six known natural CO_2 fixation pathways are shown (Berg, 2011). The red-crossed O_2 indicate the presence of oxygen-sensitive ferredoxin-dependent oxidoreductase enzymes. For clarity only reduced electron donors are depicted in the pathways. Key products directly formed by the pathways are in an orange box. Key enzymes, mostly carboxylases, are indicated in bold.
AcC, acetyl-CoA carboxylase; ACS, acetyl-CoA synthase; CODH, carbon monoxide dehydrogenase/acetyl-CoA synthase; FDH, formate dehydrogenase; KGS, 2-ketoglutarate synthase; PEPC, phosphoenol pyruvate carboxylase; PrC, propionyl-CoA carboxylase; PrkA, phosphoribulokinase; PyrS, pyruvate synthase; RubisCO, ribulose-1,5-bisphosphate carboxylase/oxygenase.
ATP, adenosine triphosphate; ADP, adenosine diphosphate; CoA, coenzym A; $2FD^{OX}$ oxidized ferredoxin pair; $2FD^{RED}$, reduced ferredoxin pair; NAD(P)H, Nicotinamide adenine dinucleotide (phosphate); THF, tetrahydrofuran; UQ^{RED}, reduced ubiquinol.

electrons or other electron donors. Both *Shewanella oneidensis* and *Geobacter sulfurreducens* are heterotrophic, electroactive bacteria for which genetic tools are available (Sydow et al., 2014). As they do not harbor a functional CO_2 fixation pathway, autotrophic growth requires the introduction and sufficient expression of all enzymes for such pathways. Another interesting heterotrophic host is the anaerobic, hyperthermophilic archaeon *Pyrococcus,* which already harbors hydrogenases for H_2 uptake (Atomi et al., 2011). Attempts are being undertaken to introduce a CO_2 fixation pathways in this thermophile (Keller et al., 2013), which is especially attractive for industrial chemical production at higher temperatures, for example to facilitate product recovery of volatile chemicals (Hawkins et al., 2011).

Integrating autotrophs and heterotrophs

Autotrophic production platforms may also integrate autotrophic and heterotrophic hosts to combine their beneficial characteristics. Two promising proofs-of-concept have recently been reported for the combination of non-engineered, autotrophic acetogens with engineered, heterotrophic production hosts. Firstly, the acetogen *Sporomusa ovata* was integrated with several engineered *E. coli* strains (Liu et al., 2015). Wild type *S. ovata* fixes CO_2 while taking up extracellular electrons from light-capturing, semiconductor nanowires in anaerobic conditions (**Figure 3b**). *S. ovata* then excretes acetate, which can be used by engineered *E. coli* strains in aerobic conditions to generate valuable products, dependent on the introduced pathway, for example, for *n*-butanol or PHB formation (Liu et al., 2015) (**Figure 3c**). A similar approach is the two-reactor integration of the H_2- and CO_2-consuming, thermophilic acetogen *Moorella thermoacetica* in one anaerobic reactor; the acetate product was subsequently fed to a strain of the yeast *Yarrowia lipolytica* that resides in a second aerobic bioreactor and was engineered for increased lipid production (Hu et al., 2016). These integrated systems have a great potential to efficiently convert CO_2 into valuable products, obviating the need to introduce all pathways, which are possibly incompatible (aerobic and anaerobic), into a single cell. High energy conversion efficiencies were reached in these studies; in general, acetogens have high electron donor-to-acetate conversion efficiencies of ~90% (Hu et al., 2016). Subsequent yields of aerobic heterotrophic production from acetate are dependent on the energy requirements of the production pathway of interest and on the microbial conversion efficiency; high acetate-to-product yields have been reported, for example, up to 52% for PHB-production by *E. coli* (Liu et al., 2015).

For photosynthetic microalgae and cyanobacteria, co-cultures with the diazotrophic, heterotrophic *Azotobacter vinelandii* have been reported, enabling growth and production by the co-culture based on the simultaneous fixation of CO_2 and nitrogen(Ortiz-Marquez et al., 2012; Smith and Francis, 2016). To create synthetic co-cultures, the cyanobacterium *Synechococcus elongatus* PCC 7942 has been engineered to overproduce and excrete sucrose, which can be converted to products of interest by a co-cultured heterotroph (Ducat et al., 2012; Smith and Francis, 2016).

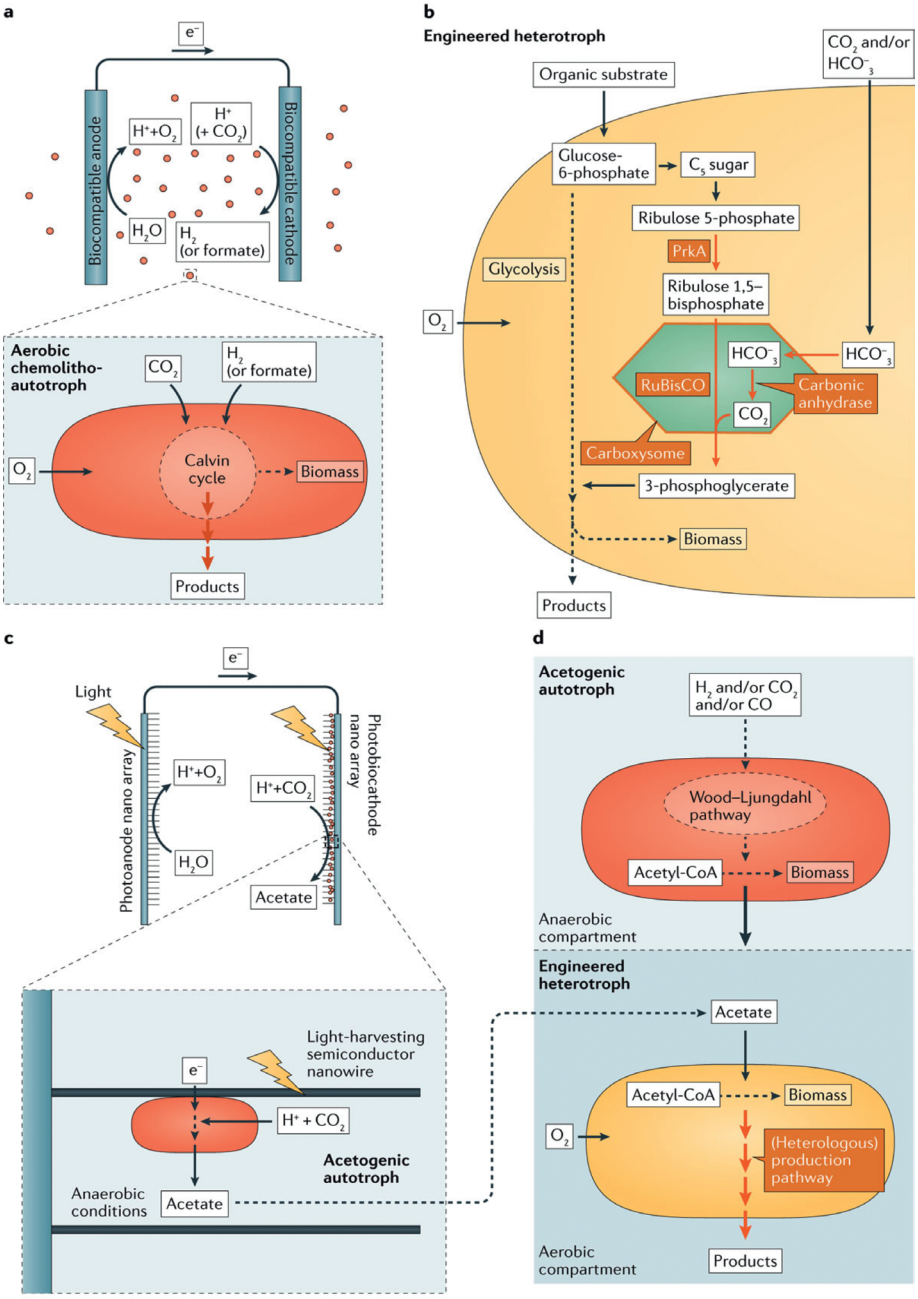


Figure 3. Key examples of engineering towards autotrophic microbial production platforms. a| Bio-inorganic hybrid systems of the aerobic chemolithoautotroph *Cupriavidus necator* cell factory integrated with electrodes producing H_2 or formate are very promising autotrophic production platforms, especially if the electrodes are designed to be biocompatible, for example by reducing the release of

chemical species toxic to *C. necator* (indicated by red circles), such as reactive oxygen species (Li et al., 2012b; Liu et al., 2016; Torella et al., 2015). b| The Calvin cycle enzymes ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) and phosphoribulokinase (PrkA) have been successfully introduced in the model heterotrophs *Saccharomyces cerevisiae* (Guadalupe-Medina et al., 2013) and *Escherichia coli* (Antonovsky et al., 2016; Gong et al., 2015) leading to mixotrophy. Additionally, carboxysome microcompartments (indicated by the hexagon) and carbonic anhydrase have been successfully transplanted in *E. coli* (Bonacci et al., 2012; Gong et al., 2015), leading to higher CO_2 concentrations in the proximity of RubisCO to enhance CO_2 fixation (Bonacci et al., 2012). Pathways depicted in orange are transplanted into model heterotrophs by genetic and metabolic engineering approaches.
c| Bio-inorganic hybrid systems of chemolithoautotrophic acetogens, such as *Sporumusa ovata*, with light-harvesting semiconductor systems, form promising light-driven autotrophic production platforms of acetate, as demonstrated recently (Liu et al., 2015; Sakimoto et al., 2016). The acetate produced can be subsequently converted by other (heterotrophic) organisms into various products, as was demonstrated (Liu et al., 2015) and as is depicted in panel d.
d| Chemolithoautotrophic, non-engineered acetogens (such as *S. ovata* (Liu et al., 2015) or *Moorella thermoacetica* (Hu et al., 2016) producing acetate efficiently from H_2 and CO_2 or CO, were coupled to heterotrophic strains engineered for product formation (such as. *Escherichia coli* (Liu et al., 2015) or yeast *Yarrowia lypolytica* (Hu et al., 2016)), leading to efficient autotrophic production platforms. Strictly anaerobic acetogens function in anaerobic reactors or compartments and generate acetate; heterotrophic production hosts can grow and convert this acetate in another aerobic reactor or compartment to form products of interest.

Autotrophic systems

A core system for each autotroph is the CO_2 fixation pathway (**Figures 2,4,5,**). Furthermore, an autotrophic microorganism possesses energy-harvesting photosystems and/or external electron donor uptake systems, which could be modified or transplanted to improve the performance of autotrophic cell factories.

Carbon fixation

Six natural autotrophic CO_2 fixation pathways and several variants thereof are currently known (Berg, 2011; Berg et al., 2010; Fuchs, 2011; Konneke et al., 2014) (**Figure 2**). Whereas the number of known natural CO_2 fixation pathways will probably increase even further, the available variants already offer an important reservoir for the metabolic engineering of CO_2 fixation pathways. In addition, an extensive *in silico* exploration based on known natural enzymes yielded many more attractive options for synthetic CO_2 fixation pathways (Bar-Even et al., 2010) **(Figure 4**). All these CO_2 fixation pathways differ in thermodynamic feasibility under different conditions, required electron donors (for example NADPH, NADH and/or reduced ferredoxin) and/or the type of carbon substrate (for example dissolved CO_2 and/or HCO_3^-) (Ducat and

Silver, 2012). The key features that determine the efficiency of a CO_2 fixation pathway are the pathway kinetics and ATP requirements, which are both to a large extent influenced by the carboxylase involved (**Box 2**). For example, the inefficient Calvin cycle has slow pathway kinetics and high ATP consumption, which is caused by the poor performance of its carboxylase: ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO). Apart from being a slow and energy-demanding catalyst, RubisCO has an undesired side-reaction with oxygen, which results in photorespiration.

Several synthetic pathways, based on alternative carboxylases, seem very appealing alternatives to introduce in autotrophic cell factories. For example, the proposed group of malonyl-CoA-oxaloacetate-glyoxylate (MOG) pathways, which do not include any oxygen-sensitive enzymes, are predicted to be kinetically superior compared to all natural aerobic cycles, at similar or potentially lower ATP costs (Bar-Even et al., 2010) (**Table 1, Figure 4b**). Although these synthetic MOG pathways have not yet been established as functional CO_2 fixation pathways, some sub-pathways of these synthetic cycles have recently been successfully introduced into *E. coli* (Mainguet et al., 2013). Another very promising, short, synthetic CO_2 fixation pathway includes the very ATP-efficient PyrS-PyrC-glyoxylate cycle, which is predicted to have fast pathway kinetics (Bar-Even et al., 2010) (**Table 1, Figure 4a**). However, this pathway only functions under anaerobic conditions because of its oxygen-sensitive carboxylase pyruvate synthase.

Box 2 Carboxylases

Carboxylases form a very diverse group of enzymes that catalyze the incorporation of CO_2 or HCO_3^- into organic molecules. As such, they are key enzymes in autotrophic CO_2 fixation pathways. Some carboxylases combine the carboxylation reaction with a reduction reaction and do not require ATP to drive the carboxylation, while other carboxylases have a non-reductive carboxylation mechanism in which their conversion is directly or indirectly coupled to ATP-hydrolysis. As a result, the types of carboxylases involved in a CO_2 fixation pathway has a major impact on the overall ATP-consumption of the pathway (Erb, 2011). In addition, carboxylases are often kinetic bottlenecks of CO_2 fixation pathways, and their specific activity substantially affects the overall kinetics of CO_2 fixation pathways (Bar-Even et al., 2010). Another crucial factor is the oxygen-sensitivity of some carboxylases. Especially the ferredoxin-dependent carboxylases are oxygen-sensitive, therefore most of these ATP-efficient carboxylases are unsuitable for aerobic conditions, however some oxygen-tolerant variants are known for example for pyruvate synthase (PyrS) (Erb, 2011).

A broad diversity of carboxylases can be found in nature, several of them are involved in autotrophic CO_2 fixation pathways, such as the most abundant enzyme on earth: ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) (Erb, 2011). RubisCO is the sole carboxylase in the ubiquitous Calvin cycle, however it is a slow catalyst and has a non-reductive mechanism, requiring coupling to 'energy expensive' ATP-hydrolysis (Erb, 2011). Furthermore, in aerobic conditions the efficiency of carbon fixation by RubisCO is further decreased due to the side-reaction of the enzyme with oxygen, which results in photorespiration that leads to a futile loss of fixed CO_2 and additional consumption of ATP and reducing power.

As alternative to RubisCO, other natural CO_2 fixation routes than the Calvin cycle use alternative carboxylases, such as the oxygen-tolerant and faster catalyst phosphoenol pyruvate carboxylase (PEPC) and acetyl-CoA carboxylase (ACC) (Bar-Even et al., 2010), making some natural pathways kinetically faster and/or more ATP-efficient alternatives to the Calvin cycle. These alternative autotrophic carboxylases and also carboxylases from non-autotrophic pathways may be considered for synthetic CO_2 fixation pathways (Bar-Even et al., 2010; Erb, 2011). Examples included the fastest carboxylating enzyme known at high CO_2 concentrations(Bar-Even et al., 2010): crotonyl-CoA carboxylase (CCR), which is found in the assimilatory ethylmalonyl-CoA pathway(Erb et al., 2007, 2009) or the fast pyruvate carboxylase (PyrC), which is an anaplerotic carboxylase found in many organisms. Several synthetic pathways, using for example PyrC, PEPC, ACC or the oxygen-sensitive PyrS (**Figure 4**), seem to be generally more attractive regarding their overall pathway kinetics and ATP-consumption than the Calvin cycle (Bar-Even et al., 2010).

Other attractive carboxylase variants may also be identified by functional screening of metagenomic libraries. A proof-of-principle screen was performed to identify undiscovered RubisCO variants by expression of metagenomic fragments in a cyanobacterium that harbors a native RubisCO knockout (Varaljay et al., 2015). An alternative approach for finding better carboxylases is protein engineering, as recently successfully applied by random mutagenesis of cyanobacterial RubisCO leading to improved kinetics and CO_2 specificity (Cai et al., 2014; Durão et al., 2015). Similarly, it has been shown by rational protein engineering of CCR that its substrate specificity for crotonyl-CoA can be extended to larger enoyl-CoA thioesters as well (Peter et al., 2015). In summary, the broad natural and synthetic repertoire of carboxylases can be explored for constructing more efficient synthetic CO_2 fixation pathways.

Table 1. ATP and electron donors required for all known natural and some synthetic CO_2 fixation pathways to produce common metabolites, that function as key precursors for product and biomass biosynthesis pathways#. $2FD^{RED}$, reduced ferredoxin pair; NAD(P)H, reduced Nicotinamide adenine dinucleotide (phosphate).

		ATP per precursor			Electron donors/ pyruvate#		
		acetyl-CoA	pyruvate	glycerate-3-phosphate	NAD(P)H	$2FD^{RED}$	Oxygen senstivie
Natural known CO_2 fixation pathways							
Calvin cycle*		7	7	9	5	0	no
Wood-Ljungdahl pathway**		<1	<1	4	2	3	yes
reductive tricarboxylic acid cycle		2	2	5	3	2	yes
3HP-4HB cycle***	Thaum-archaeota	4	5	8	5	0	no
	Cren-archaeota	6	9-10	11-12			
dicarboxylate-4-hydroxybutyrate cycle		5	5	8	3	2	yes
3-hydroxypropionate bicycle (3-HP)		7	7	10	5	0	no
Examples of promising synthetic CO_2 fixation cycles*							
PyrS-PyrC-glyoxylate cycle		3	3	6	3	2	yes
C4-PyrC-alanine MOG cycle		4	6	8	5	-	no

#Electron donor requirements per pyruvate are the types of donors most commonly used in those pathways and are the net requirements, excluding extra NAD(P)H requirements, which are compensated by the production of other electron donors elsewhere in the 3-HP and 3HP-4HB pathways(Bar-Even et al., 2010, 2012a; Berg et al., 2010). *For the Calvin cycle operating in aerobic conditions, the loss of fixed carbon through RubisCO oxygenation and photorespiration will result additional ATP and NAD(P)H on top of reported values (Bar-Even et al., 2010). **The Wood-Ljungdahl pathway consumes a bit less than 1 ATP, as some energy is conserved in this pathway via transmembrane ion gradient generation coupled to regeneration of some ATP (Schuchmann and Müller, 2014). ***For the 3HP-4HB cycle in the Crenarchaeal phylum, ATP requirements per pyruvate and glycerate-3-phosphate depend on the assimilation pathway taken, in the Thaumarchaeal phylum a more ATP-efficient alternative exists. **** Both synthetic cycles produce glyoxylate for which it is assumed that assimilation towards glyceraldehyde-3-phosphate occurs via the bacterial glycerate pathway (Bar-Even et al., 2010) and towards acetyl-CoA via the reverse glyoxylate cycle (Mainguet et al., 2013; Volpers et al., 2016). Pyruvate is produced either via the reverse glyoxylate cycle and oxygen-sensitive PyrS for the PyrS-PyrC cycle (Volpers et al., 2016) or through an oxygen-tolerant combination of the TCA cycle and the glyoxylate shunt for the C4-glyoxylate-PyrC-alanine MOG cycle (Bar-Even et al., 2012a).

So far, the most successful transplantations of CO_2 fixation pathways to heterotrophs concern the introduction of two key enzymes of the Calvin cycle, phosphoribulokinase (PrkA) and RubisCO, which are missing in heterotrophs; other Calvin cycle enzymes are present in many heterotrophs as they are also part of the oxidative pentose pathway (**Figure 3d**).

In the past decade several groups demonstrated the functional introduction of these two enzymes into *E. coli* (Gong et al., 2015; Mueller-Cajar and Whitney, 2008; Parikh et al., 2006) and *S. cerevisiae* (Guadalupe-Medina et al., 2013) which resulted in a functional linear pathway for carboxylation of C5-sugars by RubisCO. This linear pathway was shown to be beneficial for the yield of the typical glucose-to-ethanol fermentation of *S. cerevisiae*, during which an excess of NADH and CO_2 is produced. In the RubisCO-engineered yeast strain, excess CO_2 was reduced with excess NADH to mixotrophically yield 10% extra ethanol (Guadalupe-Medina et al., 2013). Recently, a complete, functional Calvin cycle has been engineered in *E. coli* by first introducing RubisCO, PrkA and carbonic anhydrase, followed by laboratory evolution (Antonovsky et al., 2016). Although so far the selected *E. coli* strains still require pyruvate to regenerate ATP and NADH, they are the first heterotrophs that can completely synthesize sugars from CO_2. In studies on cyanobacteria, which already harbor a functional Calvin Cycle, productivity and growth was improved by introducing heterologous variants and higher expression levels of RubisCO (Atsumi and Liao, 2009; Iwaki et al., 2006) and some other key Calvin cycle enzymes (Liang and Lindblad, 2016).

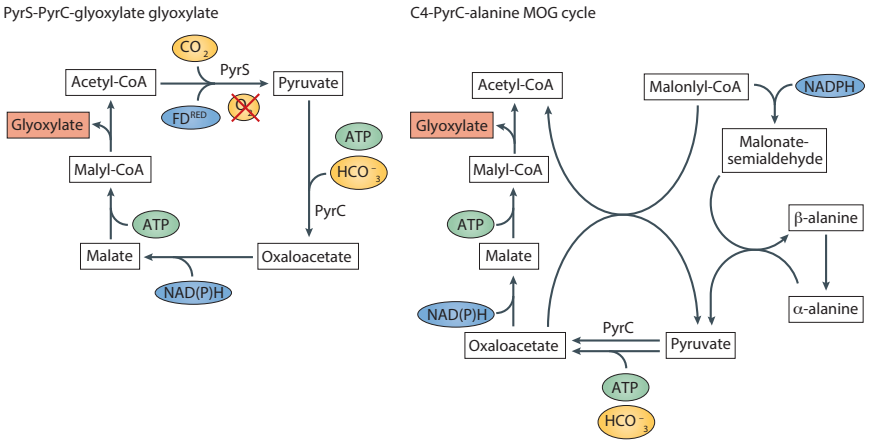


Figure 4. Two examples of promising synthetic CO_2 fixation cycles. Based on all natural enzymes many alternative synthetic CO_2 fixation pathways have been suggested, of which some are very promising candidates for the engineering of improved CO_2 fixation pathways(Bar-Even et al., 2010). These characteristics include low ATP-requirements, fast carboxylation kinetics and a low number of enzymes involved. The red-crossed O_2 indicate the presence of oxygen-sensitive ferredoxin-dependent oxidoreductase enzymes. For clarity only reduced electron donors are depicted in the pathways.
Key products directly formed by the pathways are in an orange box. Key enzymes, mostly carboxylases, are indicated in bold.

PyrC, pyruvate carboxylase; PyrS, pyruvate synthase; ATP, adenosine triphosphate; ADP, adenosine diphosphate; CoA, coenzym A; $2FD^{OX}$, oxidized ferredoxin pair; $2FD^{RED}$, reduced ferredoxin pair; NAD(P)H, Nicotinamide adenine dinucleotide (phosphate); THF, tetrahydrofuran; UQ^{RED}, reduced ubiquinol.

Part of an alternative natural CO_2 fixation pathway, the 3-hydroxypropionate/4-hydroxybutyrate (3HP-4HB) cycle (**Figure 2d**), was successfully transplanted into the thermophilic host *P. furiosus* to produce 3-hydroxypropionate from CO_2, H_2 and some organic substrate (Keller et al., 2013). Another ambitious attempt was made to the introduce the complete natural 3-hydroxypropionate (3-HP) bicycle into *E. coli* (**Figure 2f**). To achieve this, 10 heterologous enzymes had to be introduced in *E. coli*, and even though all introduced enzymes were expressed functionally, the full CO_2 fixation bicycle did not function in *E. coli*. This indicated that the heterologous expression requires further optimization and/or adjusted regulation of the native host metabolism through laboratory evolution is needed.

So far, all heterotrophs engineered with CO_2 fixation pathways still grow as mixotrophs, requiring an organic carbon source for generation of the starting substrates of linear CO_2 fixation pathways and for regeneration of ATP and/or electron donors. Converting them into true autotrophs would require the functional transplantation of complete CO_2 fixation cycles and the transplantation of and integration with energy-harvesting systems.

In autotrophs harboring the natural Calvin cycle, some energy-efficient heterologous pathways have been introduced to bypass less energy efficient CO_2-wasting photorespiration pathways. In the model plant *Arabidopsis thaliana* more efficient photorespiration pathway was introduced, consisting of three enzymes of the *E. coli* glycolate pathway, resulting in substantially increased plant growth and biomass formation (Kebeish et al., 2007). For the cyanobacterium *Synechococcus elongates* an even more energy-efficient, CO_2-fixing photorespiration pathway was designed, based on 6 enzymes of the natural 3-HP bicycle. Even though the enzymes were functionally expressed, no increase in growth was observed; potentially other enzyme variants with a lower temperature optimum are required to replace the thermophilic counterparts that were expressed in the mesophilic cyanobacterium(Shih et al., 2014).

In addition to the introduction of alternative pathways, carboxylation kinetics can be improved by increasing intracellular CO_2 or HCO_3^- concentrations. Natural CO_2 concentration mechanisms include carboxysomes, which are bacterial microcompartments that contain RubisCO and that have increased internal CO_2 concentrations. In addition, carbonic anhydrase is an enzyme that catalyzes intracellular conversion of HCO_3^- to CO_2, the required carbon species for RubisCO and other carboxylases. Both carboxysomes (Bonacci et al., 2012) and carbonic anhydrase (Antonovsky et al., 2016; Gong et al., 2015) were successfully transferred into *E. coli*, resulting in increased RubisCO-dependent CO_2 fixation (**Figure 3d**). Cyanobacterial carboxysomes are complex heteromeric protein complexes, to facilitate further engineering of carboxysome complexes, some of the proteins have been fused resulting in a streamlined, functional, carboxysome in a cyanobacterium (Gonzalez-Esquer et al., 2015). In addition, synthetic scaffolds are promising systems to bring carboxylases and carbon concentrating enzymes in close proximity (Siu et al., 2015). Alternatively, overexpressing a native HCO_3^- transporter has resulted in increased intracellular CO_2 levels in the cyanobacterium *Synechocystis* sp. PCC6803, which almost doubled biomass production when grown at atmospheric CO_2 concentrations (Kamennaya et al., 2015). Heterologous expression of a HCO_3^- transporter in the aforementioned RubisCO-engineered *E. coli* did not increase the CO_2 fixation flux, although this was only tested at bicarbonate concentrations far above ambient levels (Gong et al., 2015).

Photosystems

Various photosystems have evolved in natural phototrophic microorganisms to harvest light energy (Hohmann-Marriott and Blankenship, 2012). The simplest photosystems are the retinal pigment-based proton-pumping rhodopsins, which can generate a proton gradient from light that can be used to regenerate ATP (Claassens et al., 2013). Proton-pumping rhodopsin photosystems have been heterologously expressed in several hosts, occasionally including expression of the retinal pigment biosynthesis pathways (Claassens et al., 2013). These photosystems have been introduced into the heterotrophic hosts *E. coli* (Martinez et al., 2007; Walter et al., 2007; Wang et al., 2015) and *S. oneidensis* (Johnson et al., 2010), and more recently as an additional photosystem in the cyanobacterium *Synechocystis* (Chen et al., 2016). This resulted in increased cell survival and electricity production for *S. oneidensis* (Johnson et al., 2010),

and for *E. coli* in slightly increased ATP levels (Martinez et al., 2007), increased H_2 production (Kim et al., 2012b) and even slightly increased growth (Wang et al., 2015) in the presence of light. Also a *Synechocystis* strain with a proton-pumping rhodopsin photosystems had a slight increase in growth compared to a control strain with a non-proton-pumping mutant rhodopsin. However, the expression of the rhodopsins resulted in decreased growth for both engineered strains compared to wild-type *Synechocystis*, probably because rhodopsin overexpression results in an energetic burden and/or membrane space occupation (Chen et al., 2016). Thus, the physiological benefits of these proton-pumping rhodopsins achieved so far seem rather limited, and their ATP regeneration capacity per cell is low (Chen et al., 2016; Claassens et al., 2013; Kirchman and Hanson, 2013).

The more complex, chlorophyll-based photosynthetic reaction centers are more efficient in terms of energy transduction. Excitation of the chlorophyll pigments of these photosystems results in chlorophyll electrons being excited. These high-energy electrons can be used both to generate a proton gradient and to generate reducing power. Compared to rhodopsins, these reaction center photosystems can pump more than one proton per absorbed photon, and they absorb more light as they generally contain light-harvesting antennae consisting of, for example, chlorophylls and carotenoids (Tikh and Schmidt-Dannert, 2013). The chlorophyll-based reaction center photosystems consists of multiple subunits and their pigment biosynthesis pathways are more complex than the retinal biosynthesis pathway. Hence, it is much more complicated to introduce these systems in heterologous hosts, compared to the simpler rhodopsin photosystems. Nonetheless, carotenoid biosynthesis has been achieved in multiple heterologous hosts, such as *E. coli* and *S. cerevisiae* (Ye and Bhatia, 2012), and the first steps towards bacteriochlorophyll biosynthesis were demonstrated in *E. coli* (Tikh et al., 2014a). However, complete transfer of reaction center photosystems to heterotrophs has not been achieved yet. A general challenge for heterologous expression of photosystems is the involvement of membrane proteins, as membrane proteins are generally harder to express heterologously (Schlegel et al., 2014) and membrane surfaces heterotrophic hosts are limited, as they lack internal membranes, which are present in photoautotrophs (Liu, 2016).

In a much simpler manner, the native reaction center photosystems in photoautotrophs can be engineered by truncation to increase their light-harvesting efficiency. The truncation of the light-harvesting antennae systems results in less absorption by the individual cells and hence better distribution of high-intensity light over the cells in dense, industrial cultures (Work et al., 2012). This strategy was very successful for the microalgae *C. reinhardtii* (Beckmann et al., 2009) and cyanobacterium *Synechocystis* sp. PCC6803 (Kirst et al., 2014), in which a genetic mutation for truncated light-harvesting antennae under high-intensity light resulted in approximately 50% increased biomass production.

Another promising but challenging strategy would be to extend the wavelength of the light that can be harvested by photosystems. The chlorophyll-pigmented photosystems in natural oxygenic phototrophs only absorb light with a wavelength up to 700 nm, thereby not exploiting about 50% of available solar irradiation. This absorbance wavelength maximum can potentially be extended by using alternative photosystems or alternative light-harvesting pigments, for example from the photosystems of anoxygenic phototrophs, which have bacteriochlorophylls that absorb up to 1075 nm (Blankenship et al., 2011; Chen and Blankenship, 2011; Ort et al., 2015). Recently, chlorophyll *a* biosynthesis was demonstrated in the anoxygenic phototroph *R. sphaeroides*, as a basis for future attempts to broaden light-absorption by combining chlorophyll and native bacteriochlorophyll pigments (Hitchcock et al., 2016).

Alternative to natural photosystems, a recently direct coupling of efficient, light-harvesting semiconductor photosystems with non-photosynthetic bacteria was achieved. For the acetogens *S. ovata* and *M. thermoacetica* hybrid systems were described with silicon and titanium dioxide nanowire arrays (**Figure 3b**), and cadmium sulfide nanoparticles, respectively. It was shown that these hybrid systems of semiconductor-photosystem and bacteria produce acetate from light and CO_2 (Liu et al., 2015; Sakimoto et al., 2016). These studies reported solar-to-product efficiencies in the same range as described for algae or cyanobacteria. It is anticipated that these bio-inorganic hybrid systems can be improved well beyond natural photosynthetic efficiencies.

External electron donor uptake systems

Electron donors such as H_2, Fe^{2+} and cathodic electrons can be used by microorganisms to generate reducing power and to indirectly produce ATP

via respiration chains. H_2 is a common electron donor in the microbial world, therefore various hydrogenases are available for the uptake of H_2 (Vignais and Billoud, 2007). Hydrogenases can be divided in cytoplasmic hydrogenases, which can directly regenerate NAD(P)H, and membrane-bound hydrogenases, which generate electron donors for respiration chains to generate a proton gradient. Many autotrophic and heterotrophic microorganisms harbor hydrogenases, some of which catalyze efficient H_2 uptake. *E. coli* does not have this ability, but a substantial effort has been made to establish heterologous expression of hydrogenases in *E. coli*, so far mainly aiming at H_2-generating hydrogenases for the production of H_2 as a biofuel (Rousset and Liebgott, 2014). However, the functional heterologous expression of hydrogenases is challenging, as hydrogenases are generally oxygen-sensitive, and major steps are required to enable successful transplantation of these systems.

Most H_2-uptake hydrogenases reduce NAD^+ or $NADP^+$, and their co-factor specificity for either of those could be altered by protein engineering. However, instead of NAD(P)H, some CO_2 fixation pathways require ferredoxin as an electron donor. As the ferredoxin redox pair has a very low potential, the direct reduction of ferredoxin by H_2 or formate is thermodynamically not feasible in most conditions, therefore the introduction of mechanisms that can overcome this thermodynamic barrier have to be considered (Buckel and Thauer, 2013). A natural mechanism to solve this is electron bifurcation by cytoplasmic enzymes that couple the thermodynamically unfavorable reduction of ferredoxin to the favorable reduction of, for example, $NAD(P)^+$ (Buckel and Thauer, 2013; Wang et al., 2013). One of these bifurcating hydrogenases has already been functionally expressed in *E. coli* (Wang et al., 2010).

An alternative source of reducing power is direct electron uptake from a cathode, which is potentially very efficient but requires large cathodic surface areas (Lovley and Nevin, 2013; Rabaey and Rozendal, 2010). For example, several microorganisms have been described to perform extracellular electron transfer via extracellular cytochromes or microbial nanowires. However, these mechanisms are mostly used for donating electrons to external acceptors, rather than accepting electrons (Malvankar and Lovley, 2014). Furthermore, these mechanisms are still relatively poorly understood and mainly studied in two model organisms: *S. oneidensis* and *G. sulfurreducens*. Nonetheless, initial

attempts to heterologously express parts of the extracellular electron transport chain of *S. oneidensis* in *E. coli* have been performed, resulting in electron donation activity of *E. coli* to an electrode (Goldbeck et al., 2013; Jensen et al., 2010; TerAvest et al., 2014). Given the limited knowledge on direct electron donation mechanisms, future research should focus on identifying genes and proteins involved in these mechanisms and their maturation factors, to enable their improvement and transplantation into heterologous hosts. Alternatively, the use of natural hosts that harbor these electron transfer systems can be further explored(Rosenbaum and Henrich, 2014).

Key engineering strategies

So far only sub-systems related to autotrophy have been improved or transplanted. To realize complete, efficient autotrophic cell factories, novel strategies should be applied based on the principles of synthetic biology (Way et al., 2014). During the phase of designing improved autotrophic cell factories, computational tools are instrumental. For the experimental implementation of full autotrophic cell factories, modular approaches are becoming more important.

Design and quantitative analyses

For the engineering of autotrophic cell factories appropriate design settings must first be defined (Way et al., 2014). Based on factors such as the product of interest, genetic tools required and preferred conditions, such as process temperature, the most appropriate host microorganism can be selected. Major factors related to autotrophy that should be considered include the energy source, that is, light and/or electron donors. Another important condition for autotrophy to consider the absence or presence of oxygen. Anaerobic conditions may be advantageous for industrial-scale processes (Weusthuis et al., 2011) and are required for some oxygen-sensitive autotrophic systems. However, anaerobic conditions are incompatible with some production pathways or autotrophic systems that produce or require oxygen.

For subsequent quantitative design and comparison of autotrophic and production pathways computational analysis tools have been shown instrumental. The main types of quantitative analyses can be divided in stoichiometric, thermodynamic and kinetic analyses.

The theoretical biomass and/or product yields for autotrophic pathways can be calculated based on stoichiometric analyses. Such analyses require reaction equations for the CO_2 fixation pathways, the biomass formation pathways and the production pathways. Simplified analyses can be performed with lumped reactions for biomass production (Fast and Papoutsakis, 2012). More elaborate analyses, including the detailed metabolism of a host, can be performed by using core or genome-scale metabolic models, and methods such as flux balance analysis (Bar-Even et al., 2010, 2013; Boyle and Morgan, 2011; Orth et al., 2010; Pandit and Mahadevan, 2011). Stoichiometric methods have been applied to perform comparative analyses of autotrophic growth and production based on different natural and synthetic CO_2 fixation pathways and energy uptake systems (Boyle and Morgan, 2011; Fast and Papoutsakis, 2012; Pandit and Mahadevan, 2011; Volpers et al., 2016). Theoretical yields calculated from stoichiometric analysis can also be used to estimate energetic conversion efficiencies from a source (light, electron donors) into a product or biomass, which is an important parameter for evaluating the efficiency of autotrophic production platforms (Boyle and Morgan, 2011; Fast and Papoutsakis, 2012).

These stoichiometric analyses have demonstrated that the ATP-demand for precursors for biomass and products vary widely for diverse CO_2 fixation pathways (**Figure 5**). Hence, theoretical product and biomass yields heavily depend on the selected CO_2 fixation pathway, as well as the ATP-requirement of the production pathway and for host maintenance (Bar-Even et al., 2010; Boyle and Morgan, 2011; Fast and Papoutsakis, 2012; Volpers et al., 2016). The net reducing power required for a certain precursor metabolite for biosynthesis does not vary for different CO_2 fixation cycles, as it only depends on the reduction state of the metabolite itself. However, the types and ratios of required electron donors do vary (for example, NADPH, NADH or reduced ferredoxin) for various CO_2 fixation pathways and they have to be supplied in proper ratios by energy generating systems (Bar-Even et al., 2012a) (**Figure 5**). Furthermore the required electron donor flux may depend on the ATP-requirements, as ATP may be regenerated through respiration of electron donors.

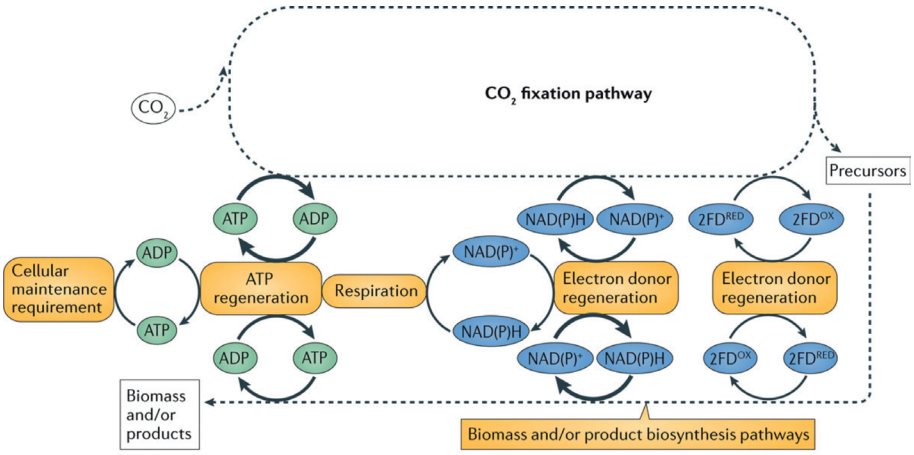


Figure 5. Balancing supply and demand for CO_2 fixation pathways. Supply and demand fluxes of ATP and electron donors have to be properly balanced between CO_2 fixation pathways and other pathways in autotrophic cells. In an autotroph, the CO_2 fixation pathway is the key consumer of ATP and reduced electron donors. The required ATP and reduced electron donors should be regenerated by appropriate systems. In addition, there has to be sufficient supply of ATP and reduced electrons for pathways towards biomass and products and ATP for cellular maintenance.

The highest yields identified by stoichiometric analysis generally are found for CO_2 fixation pathways with low ATP-consumption. However, some minimum input of ATP is required to overcome thermodynamically infeasible conversions (Bar-Even et al., 2010, 2012a). To assess thermodynamic feasibility, the Gibbs free energy of reactions can be calculated, for example using the EQuilibrator tool (Flamholz et al., 2012). However, this calculation requires physiological concentrations of substrates, products and co-factors that are involved in the reactions, and these metabolic concentrations are often hard to measure or predict. Alternatively, a range of metabolite concentrations or ratios can be used to estimate thermodynamic favorability of reactions. In addition, thermodynamic pathway optimization algorithms, such as the Max-min driving force method, can be applied to identify thermodynamic bottleneck reactions in a pathway (Noor et al., 2014). Thermodynamic analyses were used to identify thermodynamically favorable synthetic CO_2 fixation pathways (Bar-Even et al., 2010). Thereby it was also shown that very ATP-efficient natural pathways are not feasible under certain conditions, as for example the Wood-Ljungdahl pathway, which is not thermodynamically feasible under too low (ambient) CO_2 concentrations (Bar-Even et al., 2010, 2012a, 2012b).

Apart from thermodynamics, the practical feasibility of autotrophic production and growth is determined by the kinetic rates of enzymes involved, which in the end determines autotrophic growth rates and productivities. Exact quantitative assessment of these kinetics is complicated given the lack of *in vivo* kinetic parameters. However, for some enzymes *in vitro* measured kinetic parameters are available that can be applied to approximate kinetic favorability of enzymes and pathways (Davidi et al., 2016). Hereto a simplified framework was proposed to estimate and compare the upper limit pathway kinetics for many CO_2 fixation cycles (Bar-Even et al., 2010). This method revealed that some synthetic pathways have superior kinetics compared to most natural cycles, such as the widespread Calvin cycle (Bar-Even et al., 2010).

Unfortunately, the kinetic activities of energy-harvesting systems, such as hydrogenases and photosystems, are less well-characterized by *in vitro* studies, while these may be the kinetic bottlenecks for autotropic systems in many cases. In contrary to the strict limits of thermodynamics and stoichiometry, kinetics and related pathway fluxes can often be improved. Once kinetic bottleneck enzymes are identified by computational methods and/or by metabolomics experiments, these enzymes can be tackled either by improving their expression levels or by (metagenomics) mining or protein engineering for kinetically superior variants (**Box 2**).

Modular experimental implementation

Dependent on the selected microbial host, a set of metabolic engineering and synthetic biology tools is available to modify and introduce systems for autotrophy and production pathways. However, irrespective of the host, transplanting or modifying extensive pathways or multiple systems in one step is a major hurdle, and has so far limited the extensive engineering of autotrophy. To overcome this hurdle, systems and pathways needed for autotrophy and products of interest can be subdivided in modules or sub-pathways. Then these modules can be designed and built using metabolic engineering and synthetic biology tools (Boock et al., 2015; Boyle and Silver, 2012; Smanski et al., 2016). This includes the refactoring of the required native genes in synthetic DNA clusters (Smanski et al., 2016), as was shown for photosystem II and the carboxysome (Gimpel et al., 2016; Gonzalez-Esquer et al., 2015). Subsequently one can test modules for functionality, and based on this further optimize

them in iterative design-build-test cycles and further improve the module and interaction with the host metabolism by laboratory evolution approaches (Portnoy et al., 2011) (**Figure 6**).

A modular approach was taken by the most ambitious transplantation of a CO_2 fixation pathway published so far: the introduction of the natural 3-HP bicycle into *E. coli*. This bicycle was subdivided in 4 sub-pathways, and activity assays showed at least some activity for each of the sub-pathways (Mattozzi et al., 2013). However, the combination of all sub-pathways did not yet result in a functional CO_2 fixation bicycle (Mattozzi et al., 2013). Therefore, the functionality of the sub-pathways should first be further optimized in subsequent cycles, for example by including larger enzyme libraries, varying expression levels, and very importantly, adaptation of the host metabolism by laboratory evolution.
The aforementioned successful generation of a functional Calvin cycle in *E. coli* was based on an integrated approach involving rational design, modular engineering and selection of functional variants by laboratory evolution (Antonovsky et al., 2016). In an engineered *E. coli* strain, which harbored all Calvin cycle enzymes, some knockouts of enzymes from central metabolism were made to couple the growth rate to functionality of the Calvin cycle. Initially this strain still required xylose for growth, however after ~150 generations in a xylose-limited chemostat, *E. coli* strains evolved that did not require xylose anymore as they had acquired a functional Calvin cycle. In these evolved strains, several genomic mutations related to host metabolism were identified, which seemed to be crucial for Calvin cycle functionality. In general, autotrophic modules and pathways can often be coupled to growth, providing a large potential to apply laboratory evolution.

Outlook

Very promising breakthroughs for autotrophic production platforms have recently been realized, in particular for hybrid systems involving inorganic photovoltaic systems, electrocatalysts and chemolithoautotrophic microorganisms. These bio-inorganic hybrid systems achieved solar-to-product efficiencies that are up to ten times higher than conversions achieved based on natural photosynthesis (Liu et al., 2016), making them an attractive

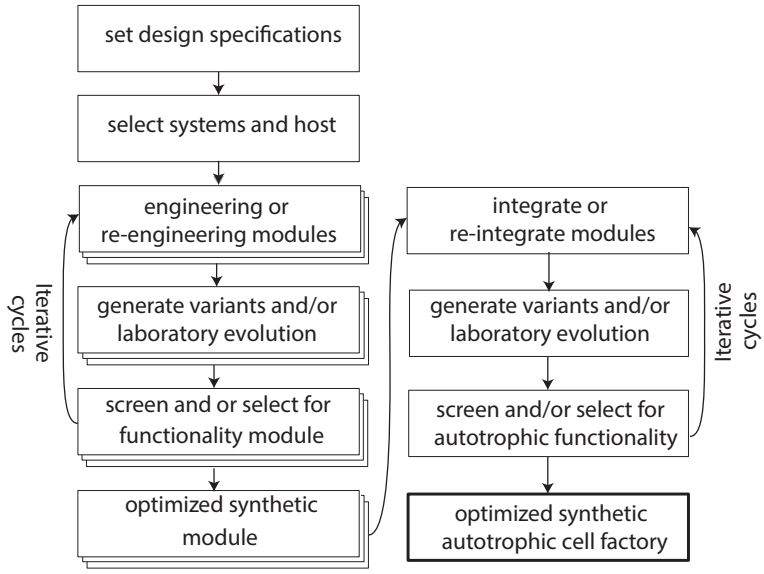


Figure 6. Proposed workflow for engineering autotrophic cell factories. To realize more efficient autotrophic cell factories, a workflow based on engineering principles from synthetic biology is considered instrumental. To tackle the complexity of engineering autotrophy, systems related to autotrophy can be transplanted or improved as modules. These modules should be designed and improved in an iterative manner, including laboratory evolution of the host metabolism through selection for improved module functionality. Eventually functional modules and improved host metabolism can be combined to achieve functional, efficient autotrophic cell factories.

alternative for platforms based on (microbial) photoautotrophy. However, to make (chemolitho)autotrophic production platforms economically feasible, further improvements of yield and productivity are required (Khan et al., 2014). To boost autotrophic strain performance, computational studies have provided promising blueprints for alternative (synthetic) pathways for CO_2 fixation, potentially based on non-natural enzymes, which could be linked to efficient production pathways. So far, these alternative CO_2 fixation pathways have been implemented with limited success in the heterotrophic model organisms *E. coli* and *S. cerevisiae*. The mere introduction of a set of missing genes to establish a complete synthetic CO_2 fixation pathway seemed to be an overly simplistic strategy. Such an approach overlooks complicating factors, such as enzyme kinetics, toxic intermediates, allosteric regulation and, very importantly, the appropriate integration in the native metabolism. To tackle all these issues, extensive engineering has to be performed, including testing of several (engineered) enzyme variants followed by laboratory evolution.

Furthermore, to create autotrophic strains that function optimally, it is crucial to engineer and evolve them for optimal performance in the envisioned production environment, for example in a bio-inorganic electrocatalytic hybrid system or a photobioreactor. This emphasizes the need for multidisciplinary engineering projects, involving integration of fields such as microbial physiology, biochemistry, metabolic engineering and protein engineering in combination with bioreactor engineering, electrochemistry and (bio) nanotechnology, as showcased by some recent, successful projects (Hu et al., 2016; Liu et al., 2015, 2016). Autotrophic production platforms should also be designed and implemented with the products of interest in mind. Autotrophic production seems especially promising for future, sustainable production of bulk chemicals and, if production costs can be sufficiently decreased, for biofuels (Liao et al., 2016). To realize future industrial-scale autotrophic production, major adaptations in the manufacturing infrastructure will be required; resource supplies and operational units will be completely different from those in current biomass-based fermentation industries.

Moreover, the lessons learned and to be learned from attempts of improving autotrophic microorganisms may contribute to yet another major scientific challenge: the engineering of plant photosynthesis for improved crop productivity (Ort et al., 2015). In conclusion, engineering towards more efficient autotrophic microbial platforms will be one of the grand promises and challenges in the coming years.

Acknowledgements

NJC was supported by the IP/OP program Systems Biology of Wageningen University (KB-17-003.02-024). We acknowledge support from the Netherlands Organization of Scientific Research (NWO) through grants in the TOP (714.015.001; JO), Gravitation (SIAM 024.002.002; WMV) and Spinoza (WMV) programs, as well as from the EU through ERC Grants (323009 and 250172, DS and WMV), the FP7 project SPLASH (311956: VDS) and H2020 project EmpowerPutida (635536: VDS). The authors declare no competing interest as defined by Nature Publishing Group.

Chapter 9

Thesis summary

The application of microbial and plant photosynthesis for biobased production on the one hand has a huge potential but on the other hand photosynthesis has serious limitations regarding its efficiency. Hence, studying both fundamental features of photosynthetic processes and engineering of photosystems is of paramount interest, exploring the engineering of photosystems is the overarching aim of this thesis. As described in **Chapter 1,** natural photosystems may be modified or transplanted to allow for more efficient conversion of solar light energy into biochemical energy. Hereto ambitious proposals to engineer photosystems have been made, and to realize those endeavors the disciplines of synthetic and systems biology are required. To explore how to apply and improve those disciplines hereto, the work described in this thesis has focused on the transplantation of simple photosystems (proton-pumping rhodopsins; PPRs) into the cell membrane of the heterotrophic model bacterium *Escherichia coli*. Both *in silico* analyses, including metabolic and thermodynamic modeling (Chapter 3) and a series of experimental studies (Chapters 4,6 and 7) on transplanting PPR photosystems and other membrane proteins have been executed, which identified several challenges, limitations and most importantly opportunities.

Chapter 2 provides more details on previously reported examples of heterologous expression of PPRs in several hosts, and on the physiological impact of these transplanted photosystems. Based on this evaluation, some suggestions are made to improve and further exploit the transplantation of these photosystems. Relatively low proton-pumping capacities, and also narrow light absorption spectra are identified as limitations for their implementation. Options are discussed to improve the physiological potential of PPRs, these include increasing the abundance of functional PPRs in the host membrane, engineering or screening for increased proton-pumping rates, and obtaining systems with broader absorption spectra, such as dual-pigment PPR variants.

In **Chapter 3** a systematic, integrated *in silico* analysis is made of anaerobic, photo-electro-autotrophic synthetic metabolism in *E. coli*, consisting of (i) a PPR photosystem for ATP regeneration, (ii) an electron uptake pathway, and (iii) a natural or synthetic carbon fixation pathway. Constraint-based metabolic modelling of *E. coli* central metabolism is used, in combination with kinetic and thermodynamic pathway analyses. The photo-electro-autotrophic

designs are predicted to have a limited potential for anaerobic, autotrophic growth of *E. coli*, given the relatively low ATP regenerating capacity of the PPR photosystems, and the relatively high ATP consumption due to maintenance. If the ATP regeneration by PPRs can be improved and/or the ATP consumption for maintenance can be lowered, the metabolic modeling indicates that efficient carbon fixation can for example be attractively realized via the natural reductive tricarboxylic acid (rTCA) cycle or via the synthetic pyruvate synthase-pyruvate carboxylase-glyoxylate bicycle. Both oxygen-sensitive carbon fixation cycles are very ATP efficient, while having estimated fast kinetics. The latter also results in relatively low estimated protein costs for these pathways. Furthermore, the mentioned synthetic bicycle is highly thermodynamically favorable under various conditions analyzed. In general these analyses illustrate the potential of *in silico* analyses to identify potential bottlenecks and solutions in complex designs for autotrophic and photosynthetic metabolism, as a basis for subsequent experimental implementation.

To tackle a main bottleneck of PPR systems: their functional production level, the heterologous production in *E. coli* of the proton-pumping rhodopsins from *Gloeobacter violaceus* (GR) and from *Thermus thermophilus* JL18 (TR) is quantified and experimentally optimized in **Chapter 4**. High constitutive production of both rhodopsin proteins is achieved by fine-tuning transcription and translation. Maximum levels of approximately 50,000 PPRs/cell or 7 mg PPR/L culture are achieved, which seems close to the maximum achievable for *E. coli*. Besides the canonical retinal pigment, the GR system has the ability to bind a light-harvesting antennae pigment, echinenone. After optimization of the heterologous pigment biosynthesis pathways for either retinal or echinenone production, appropriate quantities of retinal or echinenone for PPR reconstitution were detected in *E. coli*. Association of echinenone with GR broadens its absorption spectrum in *E. coli,* broadening the potential for light-harvesting also to blue light. Interestingly, we could observe in absorption spectra for TR that this PPR seems to be able to associate with echinenone as well, indicating a potential function as light-harvesting antennae. For GR it was attempted to detect increased ATP regeneration or other physiological advantages in *E. coli* in blue light, but they were not properly identified yet. Ideally both retinal and echinenone pigments could be co-produced in *E. coli* to generate a fully functional dual-pigment photosystem. Optimization of the

branched pathway for simultaneous biosynthesis of both retinal and echinenone has been attempted by using a smart library of variable Ribosome Binding Sites (RBSs) with varying strengths (RedLibs). This attempt has revealed a probable catalytic bottleneck for the enzyme (CrtO) catalyzing the final conversion of beta-carotene into echinenone, as compared to the more efficient conversion by Blh of beta-carotene into retinal. Therefore, production or kinetics of the enzymes at the branching point in the pathway have to be optimized further for balanced co-production of the two pigments. In general, the here described approaches towards improved functional production of rhodopsin photosystems in *E. coli* and their pigments may prove more widely applicable for heterologous production of more complex photosystems and other systems.

In **Chapter 5** an up-to-date overview is provided on how codon usage can influence functional protein production. In all domains of life, a biased frequency of synonymous codons is observed at different levels: genome, regulon, operon, and within genes. The fact that all known organisms have an incomplete set of tRNAs, indicates that codon usage could act as a general mechanism that allows for fine-tuning the translation speed. Although translation initiation is the key control step in protein production, it is broadly accepted that codon bias, especially in regions further downstream of the start codon, can contribute to the translation efficiency by tuning the translation elongation rate. Moreover, codon bias plays an important role in controlling a multitude of cellular processes, ranging from protein folding (intragenic bias) to differential protein production (intergenic bias). In naturally evolved systems, an optimal set of codons has been selected to allow for appropriate modulation of translation elongation, that, at least for a sub-set of proteins, is important for correct co-translational folding, and as such for functional protein production. Modulation of the translation speed depends on a combination of factors, including the secondary structure of the transcript (more or less RNA hairpins), the codon usage landscape (frequent and more rare codons) and for bacteria also RBS-like sequences at which ribosomes can pause. Despite these insights, it is intriguing that, more than half a century after the deciphering of the genetic code, we are still not well able to design good synthetic sequences for the heterologous expression of any gene of interest. The main problem is the complex combination of interdependent factors related to codon usage that can influence translation initiation and elongation. All in all, design of synthetic genes for heterologous

expression is still in its infancy, and despite the availability of some codon usage algorithms, it is often as well a matter of trial and error.

In **Chapter 6** the effect of different codon usage algorithms (optimization and harmonization) has been experimentally tested for heterologous production of membrane proteins. Apart from the codon usage algorithms also the combined effect of transcriptional fine-tuning in *E. coli* LEMO21(DE3) was assessed. The overproduction of six different membrane-embedded proteins, including 4 PPR variants (from bacteria, archaea and eukaryotes), was tested. For all these membrane proteins the recombinant production was compared for (i) the wild-type codon usage variant, (ii) a codon-optimized variant from a commercial algorithm, and (iii) a codon-harmonized variant. Transcriptional fine-tuning plays a major role in improving the production of all tested proteins. For production of tested PPR variants, the different codon usage algorithms hardly influenced production, while transcriptional tuning had a large impact on production levels. Interestingly, for the other two tested non-PPR membrane proteins, some codon usage variants significantly improved production. For both these proteins the codon-optimization algorithm reduced functional production below that of the wild-type codon variant, while the harmonization algorithm gave significantly higher production, equal or even higher than for the wild-type variant.

In **Chapter 7** it is demonstrated that a translational-tuning system can be used to successfully optimize the expression of several membrane proteins, based on initial findings presented in Chapter 4. The employed, recently developed Bicistronic Design (BCD) system, is based on translational coupling of a gene encoding a short leader peptide and the gene of interest that is under control of a variable ribosome binding site. A standardized library of 22 RBSs allows for precise, gene context-independent, fine-tuning of expression of this second gene, here encoding a membrane protein. The presented method is based on an easy library assembly approach, followed by microtiter plate screening, and allows for the efficient selection of high-producing clones. For all four membrane proteins tested in this study the BCD approach resulted in 3 to 7-fold higher protein levels than obtained by two other recently developed methods for optimizing membrane protein production. The presented approach allows for inducer-free, constitutive, high-level production of membrane proteins in

E. coli, which can be widely applicable for both membrane protein purification studies as well as for synthetic biology projects involving membrane proteins.

In **Chapter 8** a broad review and perspectives are provided on the potential of microbial autotrophs for the production of value-added compounds from CO_2. Both photoautotrophic and chemolithoautotrophic production platforms are discussed, and recent progress in improving their efficiency and production potential is highlighted. Transplantation efforts for photosystems, but also for CO_2 fixation pathways and electron uptake systems are discussed. An overview is provided on novel *in silico* and experimental approaches to engineer components related to autotrophy in heterotrophic and autotrophic model hosts, including approaches applied in this thesis. Furthermore, promising examples are discussed of hybrid production systems of autotrophs and heterotrophs, and bio-inorganic hybrids of autotrophic microorganisms with electrocatalysts or light-harvesting semiconductor materials. Future avenues are discussed for realizing more efficient autotrophic production platforms, including an integrated synthetic biology workflow that relies both on rational engineering and laboratory evolution approaches.

Chapter 10

General discussion

Production levels and limits of PPRs

The engineering of PPR photosystems and, more specifically, of the dual-pigment *Gloeobacter* rhodopsin (GR) photosystem into *E. coli*, is a key part of this thesis. For successful functioning of transplanted photosystems, high cellular levels of the membrane-embedded photosystem proteins are required.

In this work the functional production of GR, and also *Thermus* rhodopsin (TR) in *E. coli* was successfully increased by using a constitutive promoter and tuning with translational coupling elements (**Chapters 4,7**). The maximum levels of the membrane-embedded PPR levels were estimated to be in the order of ~50,000 PPRs/cell (**Chapter 4**). These PPR levels are relatively high, but still within the range of 1-5*10^4 PPRs per cell, as observed by others for native and heterologous PPR expression in several (photo)heterotrophic hosts (**Table 1 in Chapter 2**). The maximum volumetric productivity of ~7 mg PPR/L reached in this thesis, seems at the high end of the feasible range for PPR production in heterotrophs such as *E. coli* (**Chapter 4**). It has to be noticed that volumetric productivities can probably be further increased by optimizing fermentation conditions in controlled reactors, aiming for higher final cell densities. For high-producing, constitutively expressing PPR clones, after 24 hours an OD600 of ~4 was reached, which was higher than for optimized PPR expression in *E. coli* LEMO21(DE3), which reached a final OD600 of ~2. A published extensive optimization for proteorhodopsin production in *E. coli* with different media, temperatures and inducer concentrations in a well-controlled fermenter set-up resulted in a maximum of 5 mg purified PPR/L, but final cell density were not reported (Gourdon et al., 2008). However, these results indicate that the yield of 7 mg PPR/L obtained in this thesis by solely optimizing translational tuning, is already very high.

An open question is whether the cellular levels of PPRs can be further increased in *E. coli*. Given that all published studies report levels of 10^5 PPRs/cell or lower, it seems that a practical limit may be in place. A plausible explanation could be the limitedly available cytoplasmic membrane space to accommodate PPRs in (photo)heterotrophic organisms. As an illustration, the mass fraction of proteins in the cytoplasmic membrane in *E. coli* has been experimentally determined to consist of approximately 10% of the total proteome (Liu et al.,

2014). Based on our measurements, at maximum ~1.6% of the *E. coli* proteome consisted of PPRs (**Chapter 4**). This indicates that already a relatively high proportion of ~16% of the cytoplasmic membrane proteome consists of PPRs. Further increasing recombinant membrane protein levels may be hard, given the limited capacity of the protein translocation system and the membrane surface area. Furthermore, space competition with other membrane proteins and lipids may start to become a serious issue.

A possible solution to substantially improve the PPR content per cell could be a larger membrane surface area. To accommodate more PPRs or other membrane proteins, increasing the membrane surface by forming membrane invaginations or membrane vesicles in *E. coli* is an intriguing option. In fact, such phenomena have been observed in *E. coli* by the overexpression of the genes for some membrane proteins (Arechaga et al., 2000; Eriksson et al., 2009). However, membrane invaginations or vesicles have only been observed in case of membrane proteins with major cytoplasmic domains, or for membrane-attached, monotopic proteins (Eriksson, 2009). As PPRs do not have such features, membrane invaginations could be induced by co-expressing the PPR-encoding genes with those for other membrane proteins that have those features. However these 'helper membrane proteins' would probably also need to be produced in high levels, for which also membrane space and cellular resources are needed. Alternatively, a promising, but challenging approach would be to engineer or evolve the membrane synthesis and cell shape machinery of *E. coli* to achieve internal membrane formation.

Also removing redundant membrane proteins to generate some space could be an interesting avenue. To explore this, we performed some work in an *E. coli* strain in which three membrane-embedded cytochrome oxidase complexes were knocked out (Portnoy et al., 2010). Those complexes occupy a substantial part of the cytoplasmic membrane proteome (Liu et al., 2014), and are redundant for fermentative growth. However, similar GR production levels were found in this strain compared to that in anaerobically grown wild-type *E. coli* MG1655 (data not shown). Because of the very slow growth of this strain and problems with removing a chromosomal resistance marker, we did not further explore PPR engineering in this cytochrome knock-out strain.

However, the strategy of removing redundant membrane proteins deserves further attention. A recent quantification of the *E. coli* proteome revealed that, in addition to proteins related to oxidative phosphorylation, transporters form the largest part of the cytoplasmic membrane proteome (Liu et al., 2014). For well-defined culturing conditions, many of those constitutively present transporters may be redundant. Genetic engineering or evolving *E. coli* towards a minimal membrane proteome would be promising methods to explore for increasing the production of PPRs and other membrane proteins.

Compared to (photo)heterotrophic hosts, photosynthetic organisms generally have larger internal membrane surfaces available, such as thylakoids in cyanobacteria and extensive internal membranes in photosynthetic purple bacteria. In cyanobacteria this allows for higher levels of the larger reaction center photosystems than level found for PPRs, i.e. ~$2*10^5$ reactions centers per cell are found in *Synechococcus elongatus* (Mackenzie et al., 2004). Also the part of the total proteome allocated to photosystem proteins in cyanobacteria is larger than observed for PPRs, i.e. ~18% for *Synechocystis* PCC6803 (Liebermeister et al., 2014; Wegener et al., 2010).

Recently, proteorhodopsin has been produced in two model photosynthetic organisms with larger membrane surfaces: *Rhodobacter sphaeroides* and *Synechocystis* PCC6803. For the cyanobacterium *Synechocystis,* about 10^5 PPR proteins could be produced per cell (Chen et al., 2016), twice as much as the production levels in *E. coli* in this thesis. This production level in *Synechocystis* was achieved by screening the expression from three promoters with different strengths, of which a medium-strength promoter gave the highest produced level (Chen et al., 2016). However, there may be a potential to further increase the PPR production in *Synechocystis*, for example by further fine-tuning by other promoter or RBS variants. For *R. sphaeroides* it was demonstrated that additional membrane space could be generated for higher PPR production; hereto the expression of native anoxygenic photosystems and light-harvesting antennae was knocked out. These knockouts resulted in an increase of purified proteorhodopsin production from 1 to 3 mg purified PPR/L of culture (Tikh et al., 2014b). However, this is in the similar range as the production reported for *E. coli* in **Chapter 4**: ~7 mg PPR/L. It has to be noted, that the reported

proteorhodopsin production in *R. sphaeroides* was not optimized in that work (Tikh et al., 2014b), probably leaving room for improvement in this host with a 'large membrane surface'.

An alternative strategy could be to improve the packing of PPRs in the membrane in oligomers, as also was observed for efficient packing of bacteriorhodopsin in haloarchaea, potentially yielded ~10^5 PPRs/cell (**Chapter 1**). However, GR and TR produced in *E. coli* were observed to be already in trimeric state (Tsukamoto et al., 2013b, 2014), probably limiting the potential to improve the efficient packing of those.

Apart from increasing the abundance per cell, the performance of PPR systems could also be improved by targeting their kinetic properties. Among the known natural variants, GR seems to be a rhodopsin with relatively fast kinetics, and based on starved whole-cell assays of *E. coli* seems to have a pumping rate about 30% higher than for the commonly applied proteorhodopsin (PR) SAR86 variant (Ganapathy et al., 2015). Also based on its photocycle characteristics, GR seems to be a fast rhodopsin, which is also significantly faster than TR (Tsukamoto et al., 2013a). However, the measured activity of 4-5 protons GR^{-1} s^{-1} in whole-cell assays is still several-fold lower than the proton-pumping activity of reaction center photosystems (Kirchman and Hanson, 2013) and than that of the respiration chain in *E. coli* (**Chapter 2**).

The relatively small energy-transducing potential of PPRs (photophosphorylation) versus respiration (oxidative phosphorylation) in *E. coli* has been measured experimentally (Martinez et al., 2007) and can also be simulated *in silico*. Based on the apparent upper limit of 10^5 PPRs per cell and a rate of 5 protons PPR^{-1} s^{-1}, only a marginal (2%) increase of growth rate is predicted by modelling for optimal heterotrophic growth of *E. coli* on glucose (**Chapter 2**). Furthermore, it was predicted *in silico* that the capacity of PPR is likely to be insufficient to support ATP synthesis for anaerobic electro-photo-autotrophic growth in *E. coli* (**Chapter 3**). This limited potential for growth of PPRs was also observed experimentally by us and others, in several attempts to obtain light-stimulated growth of *E. coli*.

The potential for ATP regeneration and growth by PPRs

PPRs have the potential to boost the proton motive force (PMF) of a cell and potentially increase ATP regeneration and growth. Given the large challenges faced by others and by us to detect physiological impacts of PPRs in natural and engineered microorganisms, we did not pursue extensive further work in this direction, but still a reflection on some of our and others results is provided here.

Light intensity is obviously an important factor for the functionality of PPRs. Earlier observations using flagellar rotation as a read out for PPR activity in azide-starved *E. coli* showed that light saturation occured only at very high green light intensities of ~3000 µmole photons m^{-2} s^{-1} (Walter et al., 2007). However, as we observed in some of our experiments with higher light intensities, these intensities soon become toxic to growth or survival of *E. coli*. The light-toxicity is probably because of the formation of light-induced reactive oxygen species, for which the dark-adapted, gut bacterium *E. coli* is not well equipped (Khaengraeng and Reed, 2005).

In **Chapter 4**, green and blue light, generated by light emitting diodes (LEDs), of 100 µmole photons m^{-2} s^{-1} was applied to test ATP regeneration in starved *E. coli* cells at 37 °C. After 30 minutes of light exposure, ATP levels were measured by a luciferase-based assay and compared to the levels from duplicate samples kept in darkness. Interestingly, in green light ATP levels were increased, while for blue light-exposed samples ATP levels were reduced compared to samples kept in the dark. This suggests a negative impact of higher energy blue photons at this intensity on *E. coli* cell viability and indicates that it may be hard to detect physiological PPR effects in *E. coli* at higher light intensities, maybe excluding short experiments such as the aforementioned flagellar rotation experiment (Walter et al., 2007). Lower light intensities may be preferred for physiological tests in *E. coli*, especially with blue light, but a lower light intensity may compromise PPR pumping activity. In conclusion, these experiments indicate the need to carefully study the relation between light intensity and PPR activity, which at present has been investigated only limitedly (Walter et al., 2007).

Promising conditions for finding PPR-stimulated growth of *E. coli* in light are likely to include ATP-limited growth conditions, such as anaerobic, fermentative growth. Based on *in silico* predictions, realistic PPR fluxes (< 10 mmol h^{-1}/gCDW/h) could have a potential positive effect on the fermentative growth rate on some substrates (**Figure 1, Supplementary Figure S1**). Substantially PPR-increased growth rates were predicted for specific substrates, such as pyruvate, D-lactate, malate and glycerol, on which growth is related to low substrate level phosphorylation. This was experimentally tested with an initially used expression systems for GR, using an IPTG-inducible *Ptrc* promoter and a range of IPTG concentrations for tuning of GR production. Fermentative growth from several of these substrates was tested in microtiter plates. These experiments were executed in 100 µmole photons $m^{-1}s^{-1}$ green light. In addition, as cells grown in dark may not be the best controls, given potential light toxicity effects on growth, control cells were grown in light without retinal supplementation. Even though for some substrates the tested cells with retinal gave higher OD600 values at some time points, it was noticed that supplementation of retinal to wild-type *E. coli* MG1655 also gave increased OD600 reads. Although retinal is degraded quickly, it still seems to influence OD600 values and for low growth this can be a significant effect. In conclusion, we could not find increased growth for *E. coli* expressing GR in fermentation conditions, and these experiments were not further continued.

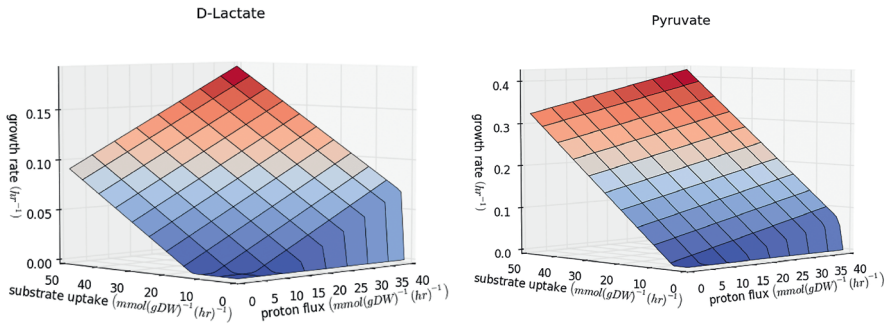


Figure 1. *In silico* predicted growth rates of *E. coli* harboring PPRs, for D-lactate and pyruvate substrates with varying ranges of PPR-pumping and substrate uptake. Grow rates were predicted by Flux Balance Analysis using standard model parameters for the *E. coli* genome-scale model *iJO1366* (see also **Chapter 3**). Color gradients from blue to red indicate increasing growth rates as depicted on the y-axis. Predictions for other substrates are in Supplementary Figure S1.

However, a recent study has provided an indication that fermentative growth of *E. coli* may be slightly stimulated by PPRs when exposed to light (Wang et al., 2015). The authors showed slightly increased growth in light for a fermenting, heme-biosynthesis knock-out *E. coli* strain, expressing proteorhodopsin. Under very high light, ~900 µmole photons m^{-2} s^{-1}, they observed a ~10% increased growth on minimal medium with D-lactate. For growth in minimal medium with 2% glucose, however, no growth difference was observed, still a small shift in fermentation products from glucose was found, i.e. less acetate and more lactate was formed in light (Wang et al., 2015). The authors of this study conclude that the relatively small physiological impacts may be increased by improving the functional production of proteorhodopsins in their strain; by using photoactive localization microscopy (PALM) they estimated a level of only ~300 PPRs/cell. However, this number of 300 may seems a major underestimation, given the observed physiological effects (including ATP regeneration) and the reported clear red pigmentation of the *E. coli* cell pellet. The commonly used, reliable spectroscopic method for PPR quantification, as also used in this thesis (**Chapters 4,7**), would be preferred to validate this number. In conclusion, there are some indications that light may slightly increase fermentative growth of *E. coli* with a PPR, at least when grown on lactate.

Surprisingly, a much more drastic growth impact of PPRs on *E. coli* growth was reported recently for aerobic growth on 0.3% glucose in low light (Na et al., 2015). For *E. coli* expressing GR, a more than two-fold higher biomass yield after 10 hours in light versus dark conditions was reported. Even though the light intensity in this work was not exactly reported ("two 13W light bulbs at 20 cm"), we repeated this experiment with the same *E. coli* W3110 strain, originally published GR expression vector and growth medium. In this experiment we found an increase in final OD600 for light growth compared to dark growth, but only 12% higher and not the previously reported two-fold higher biomass (**Figure 2**). However, it has to be noted that the exponential growth rates in light were not clearly higher as reported before (Na et al., 2015), also compared to a control without retinal. We only observed that the final biomass yield was slightly higher in light for retinal complemented cells. Even though the earlier reported two-fold effect on biomass yield could not be reproduced, the small effect that we observed is interesting. Moreover, rather surprisingly this small increase in biomass yield was observed for the energy-rich substrate glucose,

although at a relatively low substrate concentration, and in low light. In fact, also for *Pseudomonas putida* with heterologous proteorhodopsin expression, growth on low glucose concentrations (0.2%) in 120 µmole photons m^{-2} s^{-1} was reported in a PhD thesis to result in ~10 fold increased cell number yields, for growth in light versus dark (Buck, 2012). Though as the reported two-fold increase for *E. coli* with GR could not be repeated by use, these large effects of PPRs in *P. putida* preferably need further validation as well.

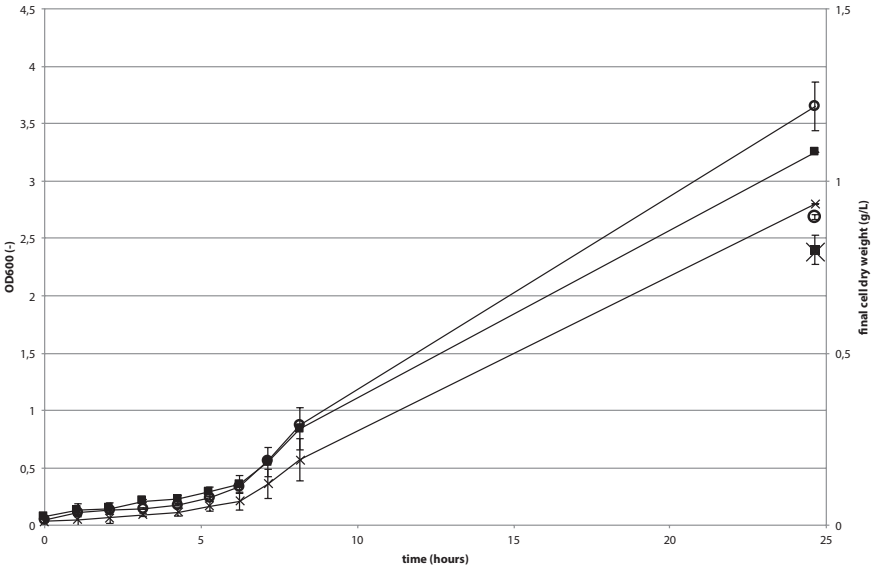


Figure 2. Growth of *E. coli* W3110 pKJ606-GR expressing GR in light conditions (open circles), dark conditions (black squares) and in light conditions without retinal (crosses). Separate points at t = 24 indicate cell dry weights (secondary axis) at the final time point (note error bar dry weight sample without retinal is too small to be vissible behind black square). Growth rates in the exponential phase are similar, the control without retinal has a longer lag phase. Final biomass yields for the light-grown culture with retinal is ~12% higher, confirmed by both OD600 optical density and dry weight measurments, unlike a ~100% higher biomass previously reported for this strain after 10 hours (Na et al., 2015). The even lower final OD600 for the control grown in light without retinal is not confirmed by the dry weight measurement, and may indicate a small effect of retinal on OD600 measurements. The strain was pre-grown and gornw grown under the conditions as published previously (Na et al., 2015), in short in 50 mL minimal M9 medium in 125 ml erlenmeyers, with 0.3% glucose, supplemented with thiamine, 0.1 mM IPTG and, when indicated, 20µM retinal in a shaking erlenmeyer in a 37 °C waterbath. Light cultures were exposed to 15 µmole photons m^{-2} s^{-1} green LED light from the top. The experiment was executed in biological duplicates. The GR overproduction plasmid pKJ606-GR was a kind donation by prof. Kwang-Hwan Jung (Kawanabe et al., 2009).

Apart from the unconfirmed two-fold increased biomass yield in light, it was reported in the original study that GR overproduction itself substantially decreased the growth rate and final biomass yield of *E. coli* when compared to an empty control vector (Na et al., 2015). In another recent study on *Synechocystis* PCC6803 haboring proteorhodopsin, it was also observed that the negative effect of proteorhodopsin production burden on growth is larger than the small increase in growth rate observed from light (Chen et al., 2016). Also our *E. coli* strains harboring PPRs grow at less than half of the wild-type growth rate and reach about half the final cell density compared to wild-type strains.

Overall, it is challenging to obtain light-increased growth with PPR-engineered microorganisms. This is in agreement with observations for natural microorganisms that natively harboring PPR systems, for which often no increased growth can be observed in light, or only in specific energy-limited conditions (DeLong and Beja, 2010; Gómez-Consarnau et al., 2007, 2015; Riedel et al., 2013; Steindler et al., 2011).

Summarizing, light-boosted growth of *E. coli* harboring PPRs is a major challenge. First of all, it seems hard to achieve sufficient PPR activity per cell. Secondly, it is a challenge to identify appropriate experimental conditions with good controls, to unequivocally demonstrate a stimulatory effect. Then the question also comes up how relevant achieving light-boosted growth for *E. coli* actually is. A case for which such phenotype would be very relevant is laboratory evolution experiments. Such evolution experiments are an intriguing option to obtain better functioning PPR variants, by selecting for improved light-stimulated growth rates. For such an approach, however, a proper, reproducible growth rate advantage is needed, under well-defined conditions (light intensity, substrate and aeration).

Alternative or complementary to light-growth based, evolutionary and selection approaches, more 'direct' assays of PPR activity can be used to screen their functionality. These assays, all with their own limitations, include ATP assays on starved cells as performed in this thesis, or alternatively proton-pumping assays on starved cells. Proton-pumping assays are relatively labor-intensive and generally hard to quantify. However, a nice step forward is the

recently developed multiplex 96-well plate proton-pumping assays, allowing for simultaneously screening many PPR variants in *E. coli* (Pushkarev and Béjà, 2016).

ATP assays are less labor intensive. However, their reproducibility and potential for quantitative assessments is alos limited, as ATP concentrations are influenced by other ATP-producing and consuming processes in cell. We observed optimizing for the proper starvation state (temperature and time of starvation) is important to detect reproducible effects of light on PPR-harboring cells. Furthermore reproducible ATP assays require quick sample processing after light exposure to detect all ATP formed through PPR activity.

Tuning protein production by RBS engineering

Tuning of gene expression has been a key challenge for the work described in this thesis. Commonly, systems for the fine-tuning of gene expression are based on transcriptional tuning using inducible promoter systems. However, even for *E. coli*, only a limited set of well-titratable inducible promoter systems is available (Lee et al., 2011). Furthermore, undesired cross-talk between inducers, and catabolite repression of inducible promoters, further complicates robust fine-tuning of multiple genes by multiple inducible promoters (Lee et al., 2011). Fortunately, apart from inducible promoters, several other systems for tuning expression are developed for *E. coli* (see **Table 1 in Chapter 1**). Inducible promoters may also not be strictly required for engineering photosystems or many other modules in synthetic biology. Modules adding novel cell functions to cells, such as photosystems, are preferably functioning all the time during growth of the cell and therefore should be continuously produced. Consequently, the production of PPR photosystems and other synthetic modules does not need to be started at a certain point by induction of an inducible promoter, instead many module genes can be continuously expressed. Hence, constitutive or non-induced promoters are considered preferable for heterologous production of PPR photosystems in this thesis. In addition, some other recent attempts to optimize production of synthetic modules in *E. coli* were based on constitutive promoters (Goldbeck et al., 2013; Schmidl et al., 2014).

Apart from using libraries of constitutive promoters, expression tuning in the here described projects was further refined at the translational level. Screening a library of translational coupling elements, so-called Bicistronic Designs (BCDs) (Mutalik et al., 2013b), was generally very successful for tuning for high membrane protein production (**Chapters 4, 7**). For three tested membrane proteins, BCD-tuned production levels were substantially higher than achieved by transcriptional fine-tuning with the commonly used *E. coli* LEMO21(DE3) strain (**Figure 3**). This success may be partly related to the higher final cell yields achieved for constitutive, BCD-tuned expression versus induced, transcriptionally-tuned expression in *E. coli* LEMO21(DE3). In addition, under tested conditions (LB, 37 °C) the membrane-embedded protein levels were even higher per cell. In general, the constitutive, BCD-tuned expression systems allow for easy, inducer-free and high level membrane protein production in multiple *E. coli* strains.

Application of a similar BCD library for fine-tuning of the branched pigment production pathway was also considered, by making constructs consisting of the assembly vector with a *P14* promoter, and two BCD elements for controlling *crtO* and *blh* separately in a single operon. For unknown reasons, however, no viable clones were obtained after successful Golden Gate assembly of these plasmids. Through personal communication with some other groups, we understood they also experienced difficulties in getting clones for operons with two BCDs elements. A possible explanation could be the repeated use of the strong RBS1 in the BCD elements, possibly sequestering too many ribosomes and thereby affecting the overall cell viability. Alternatively, the overproduction of small peptides has been suggested to be a major burden (Mutalik et al., 2013b). In addition, recombination issues may play a role if homologous BCD elements are repeated multiple times (Mutalik et al., 2013b). Another option is that the relatively strong *P14* promoter has caused the problem. This may agree with the fact that we obtained viable clones for the synthetic *blh-crtO* operons with RedLibs RBS libraries, but only after replacing *P14* by the weaker *P13* promoter. Hence, a potentially successful strategy for assembling a plasmid with multiple BCDs could be using weaker promoter variants.

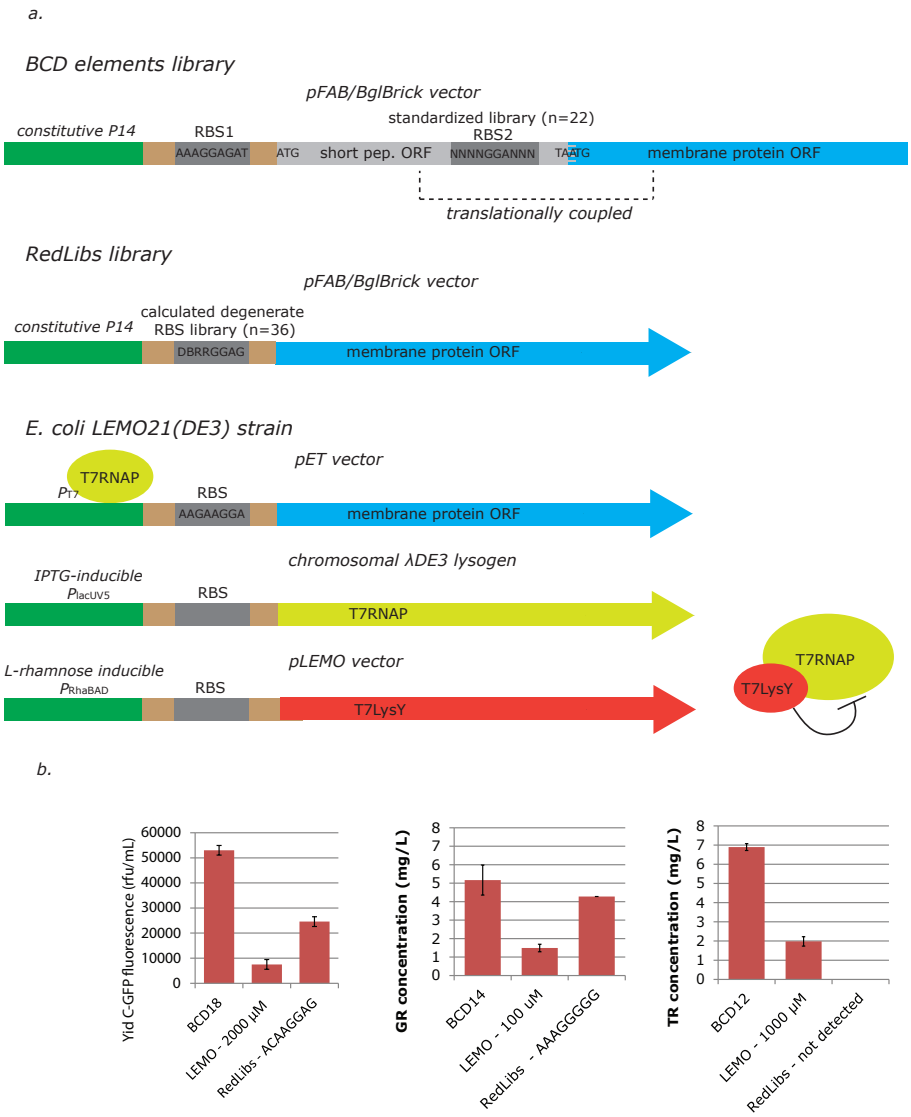


Figure 3. Comparing different tuning system for membrane protein production a| Three different systems applied in this thesis for tuning to optimize membrane protein production. b| Highest production levels obtained for three different membrane proteins using three the different tuning systems. Optimal L-rhamnose inducer levels for *E. coli* LEMO21(DE3) are indicated. The optimal RedLibs RBSs for GR and YidC-GFP have a predicted Translation Initiation Rate (TIR) of 28,213 and 220,102, respectively. For a wide RedLibs library for TR no visual detectable expression was achieved. All expression was measured in biological triplicates or in duplicates for RedLibs. T7LysY: T7 RNA polymerase lysozyme mutant Y; T7RNAP: TT RNA polymerase.

Given the potential issues with multiple BCDs, the RedLibs algorithm was chosen for varying expression for the pigment operon (**Chapter 4**). This algorithm generates small libraries of RBS sequences that cover a wide range of predicted translation initiation rates (TIRs), and which can be introduced by a single degenerate sequence (Jeschek et al., 2016). The functionality of this approach was recently demonstrated in *E. coli* for optimizing the production of deoxyviolacein and violacein by a branched pathway (Jeschek et al., 2016). For the engineered branched pathway in this thesis, towards retinal and echinenone production, the screened library yielded only one clone co-producing both pigments. This clone mainly produced retinal and β-carotene and only minor amounts of echinenone. No clear-cut conclusions could be drawn from the analyzed clones and the predicted translation initiation strengths for selected RBSs. This limited success may be related to potentially different *in vivo* kinetics of the CrtO and Blh enzymes in the *E. coli* cellular environment. Possibly, a second round with redesigned libraries (with a stronger TIR range for the *crtO* RBS and/or a weaker TIR range for the *blh* RBS) could shift the balance towards more echinenone production. Alternatively, varying CrtE production to vary β-carotene precursor availability may relieve the bottleneck for echinenone production. In addition, alternative natural or engineered variants of one or both enzymes for a better matching kinetic performance may be an option. In fact, a second round was also required to optimize violacein production, but the first library already provided a more diverse range of deoxyviolacein and violacein production phenotypes (Jeschek et al., 2016), compared to the in this thesis observed production phenotypes for echinenone and retinal.

Noteworthy, the basis of the RedLibs algorithm are translation initiation rate predictions from the RBS Calculator. This calculator has been shown to result in acceptable correlations between predicted translation initiation rates and actual protein production in *E. coli* and a few other bacteria (Farasat et al., 2014; Salis, 2011; Salis et al., 2009). However, compared to expression controlled by BCDs, RBS Calculator predicted RBSs have a substantially lower accuracy (Bonde et al., 2016). This may also explain the unsuccessful TR production by a 36-member RedLibs library, while a BCD library did result in successful TR production. Also for GR and YidC-GFP, 36-member RedLibs libraries were tested, but no clones were obtained that performed as good as the best variants of the BCD libraries (**Figure 3**). In summary, given the potential limitations of multiple

BCD units for simultaneously tuning multiple genes, using the RedLibs system may still be an interesting and viable approach. However, for RedLibs, more rounds of optimization seem to be needed to achieve optimal phenotypes, making it more labor intensive. For fine-tuning expression of single membrane proteins BCD-tuning appeared to be a successful method (**Chapters 4,7**).

A promising alternative to both BCDs and RedLibs for tuning expression in *E. coli* is the recently published EMOPEC library (Empirical Model and Oligos for Protein Expression Changes) (Bonde et al., 2016). This is a large 4096-member RBS library, for which production of GFP has been quantified in *E. coli*. From this large library, small subsets of ~10 RBS variants can be selected for precisely tuning the levels of proteins of interest. These small subsets are selected to cover a wide range of expression-strengths based on GFP data. Furthermore, only variants are selected that form no strong, predicted 5'-end mRNA secondary structures for the gene sequence of interest. It was shown that selected libraries gave precise linearly increasing protein levels for several expressed genes in *E. coli*, which were tuned almost as precise as for the BCD elements (Bonde et al., 2016). Especially for multi-gene systems this system may be more reliable and require even smaller test libraries than for RedLibs, while not having the repeated sequence issues or potential small peptide synthesis burden of BCDs. However, the applicability of this EMOPEC system for simultaneously varying the expression of multiple genes has not been demonstrated yet.

As observed for BCDs and EMOPEC, RBS engineering seems to become more reliable if the influence of 5'-end mRNA secondary structures on translation initiation can be reduced. Translational coupling is in principle a promising method to reduce this influence of secondary structures, as demonstrated for example for BCDs (Mutalik et al., 2013b). To further harness this benefit of translational coupling, multiple genes in an operon may be directly translationally coupled without introducing BCD elements in between, avoiding the need of small peptide synthesis. However, as one probably wants to avoid missense mutations in the end of the upstream coding region, the possibilities for incorporating RBS sequence variants into this coding sequence are limited. For *in silico* design and to explore options for direct translational coupling variants for synthetic operons, a recently developed predictive model for translational coupling may be usefull (Tian and Salis, 2015).

Approaches for high membrane protein production

High production of membrane protein still is far from routine, adding a challenge to the engineering of PPRs. This is also true for transplanting other membrane-embedded photosystems and for many other endeavors in molecular and synthetic biology involving membrane proteins. In the projects described in this thesis the influence was assessed of different codon usage variants as well as of transcriptional and translational tuning on membrane protein production (**Chapters 4,6,7**).

Synonymous codon usage can influence functional protein production in many ways, which is only partly understood (**Chapter 5**). However, gaining insight in this matter is relevant for improving recombinant production of proteins in general, and of membrane proteins in particular. Most membrane proteins have multiple hydrophobic α-helical transmembrane domains alternating with hydrophilic domains. Given this specific architecture, appropriate folding is a key issue for the production of functional membrane-embedded. Protein folding and translocation into the membrane often occurs co-translationally, and evidence has been provided that slowing down of translation at some points can be beneficial for correct folding of the nascent polypeptide chain. Genome-scale bioinformatics studies have revealed correlations between 'slower codons' or 'slower translation regions' with co-translational folding events, and with protein domain boundaries at certain loci, both in bacteria and eukaryotes (Fluman et al., 2014; Pechmann and Frydman, 2013; Pechmann et al., 2014; Saunders and Deane, 2010). In addition, synonymous codon substitutions have been reported to affect folding and activity of several proteins (Buhr et al., 2016; Spencer et al., 2012; Zhang et al., 2009). These regions of locally decreased translation elongation are probably, at least partly, caused by specific 'slower' synonymous codons, but also other factors, such as mRNA secondary structures and RBS-like sequences, may locally slow down the translation process. The limited understanding of factors related to synonymous codon usage and their interplay, complicates the design of synthetic gene sequences to ensure high production of properly folded proteins. However, it was hypothesized in this thesis that the relatively new codon harmonization algorithm (Angov et al., 2008, 2011) could capture some of these factors, and facilitate better folding and lead to higher levels of functional membrane

protein production. Therefore, functional membrane-embedded protein expression of codon-harmonized variants was compared to typically applied codon optimization, which optimizes for frequent codons, and with wild-type codon sequence variants (**Chapter 6**). The benefit of codon harmonization was indeed confirmed for some of the membrane proteins included in our study. However, codon harmonization did not clearly enhance the production of the PPRs included in this study.

To further improve the design of synthetic genes for high heterologous production of membrane proteins and other proteins, more insights are needed. High-throughput studies with multiple codon usage variants for the expression of various genes may provide relevant details, as wells as better characterization of charged tRNA-pools may provide better insights. Also detailed studies focusing on translation and folding kinetics of proteins are relevant. In a recent *in vitro* kinetic study on the translation of a human transmembrane protein, synonymous codons were demonstrated to be crucial for proper co-translational folding (Kim et al., 2015). These insights are leading to better models for translation (Nissley et al., 2016), which may facilitate future synthetic gene designs for efficiently produced and properly folded membrane proteins.

Compared to the challenging design of synthetic genes, tuning transcription and translation initiation for optimized membrane protein production is relatively straight forward. It is demonstrated in this thesis (**Chapter 6**) and elsewhere (Hjelm et al., 2013; Wagner et al., 2008) that tuning at the transcriptional level may be a good approach for improved membrane protein production. The benefit of transcriptional tuning for enhancing the production of several membrane proteins was generally larger than the contribution of different codon usage variants, as observed in **Chapter 6**. In this chapter tuning was performed by the *E. coli* LEMO21(DE3) system, based on transcriptional tuning. However, in **Chapter 7** an alternative, even more successful method is demonstrated for tuning, based on translational tuning, using BCD elements. For the four tested membrane proteins this method was successful for improving membrane protein production, also compared to transcriptional tuning in *E. coli* LEMO21(DE3) (**Figure 3**). The BCD tuning method should be

preferably further assessed for more membrane proteins, under different expression conditions, such as lower temperatures and potentially with various codon-harmonized, codon-optimized and other gene variants.

The presented BCD method may become highly valuable for increased membrane protein yields, especially for high-purification yields for diverse molecular biology studies. To optimize the simultaneous production of multiple membrane proteins, however, BCDs may not be the ideal method, as discussed in the section above, and alternative translational tuning methods should be considered. Desirably, also precise translational tuning tools will be developed and become available for membrane protein production in other important bacterial and eukaryotic hosts.

Apart from codon usage and transcriptional and translational tuning for improving membrane protein production, the expression host itself may be modified or evolved. Several examples exist of improved membrane protein production in engineered expression hosts, for example by overexpressing chaperones involved in folding and translocation (de Marco, 2013; Schlegel et al., 2014). As discussed before, also hosts may be engineered for increased, internal membranes surfaces or reduced membrane proteomes. Especially for *E. coli,* several strains have been engineered and evolved for elevated production of membrane proteins (Gialama et al., 2016; Schlegel et al., 2016b). The evolved strains often work well for the production of the target membrane protein for which they were evolved, but often their general application for high-level production of other membrane proteins is limited (Schlegel et al., 2016a). Furthermore many of the engineered *E. coli* strains, such as *E. coli* LEMO21(DE3), require helper plasmids, making them less suitable as flexible hosts for introducing other plasmids in synthetic biology projects.

In summary, tuning expression is definitely one of the easier approaches, and as confirmed in this thesis, a highly promising approach for improving membrane protein production.

Future prospects for PPR photosystems

The physiological impact of PPRs as light-harvesting photosystems for energy generation in microorganisms is generally limited. However, the insights gained in this thesis for PPR photosystem transplantation, may also be applied in the future for the transplantation of alternative, more efficient photosystems. In addition, PPRs may still be implemented in certain promising applications.

A relatively successful demonstration of PPR systems in *E. coli* is light-driven, increased hydrogen production. Heterologous production of hydrogenases together with proteorhodopsins, resulted in ~30% increased hydrogen production in cultures with retinal, exposed to medium intensity white light (70 – 150 µmole m^{-2}-s^{-1}), both aerobically (Kim et al., 2012b) and anaerobically (Kuniyoshi et al., 2015). This effect is probably related to a higher availability of periplasmic protons for the heterologous membrane-bound hydrogenases. This is an interesting avenue to explore further, also with improved and broader-spectrum absorbing PPRs, such as GR. However, given the low economic value of hydrogen and given the requirement of both organic substrate and light, it is questionable if such a system will become economically feasible.

Hydrogen is so far the only product for which yields have been improved in light by exploiting PPRs. The question whether other ATP-limited production processes can be boosted by PPR systems remains open. Therefore, it has to be noted that the anaerobic production of fermentation products is often coupled to ATP-regeneration (Weusthuis et al., 2011), and hence that extra ATP regeneration from a PPR system could in fact decrease production of those fermentation products. However, in some other production pathways ATP, or the available PMF, could be a bottleneck. The production from such pathways, could be interesting to improve by PPRs. However, general issues regarding the light-toxicity effects and the protein burden of PPR production need to be tackled, and limit the potential benefits of PPRs for boosting production.

Furthermore, the economic feasibility of PPR-based biotechnological production has to be properly assessed. PPR-based phototrophy requires both the input of light, requiring photobioreactor systems, and the addition of organic substrates or electron donors. Also the general energetic conversion

efficiency of PPRs has to be considered. PPRs mainly use wavelengths around ~540 nm, and can only use the energy of one photon (~2.3 eV) for pumping one proton against a typical electrochemical gradient of ~0.2 eV, resulting a in maximum energy conversion of ~10%. This conversion does not include some other typical losses associated with photosystems (**Figure 1 in Chapter 1**). So for PPR-boosted PMF or ATP-generation, relatively much light energy has to be invested. Reaction center photosystems are much more efficient on a photon-to-proton basis, they can export ~2-3 protons by absorbing 1 photon, and simultaneously have the ability to produce reduced electron carriers (Kirchman and Hanson, 2013).

Still, PPRs may have applications for biotechnology as an addition to reaction center photosystems in photosynthetic organisms. In such a combination, PPRs may play a role to harvest some of the unused wavelengths. However, the largest potential is in the wavelength range above 700 nm, for which only very recently proton-pumping activity with a retinal analog was demonstrated (Ganapathy et al., 2017). However, even with the availability of such >700 nm PPR systems, their energetic conversion efficiency will still remain still relatively low. The question remains than if the costs in cellular resources and membrane space are compensated by additional light-harvesting by the PPR.

A more promising future for PPRs may be in the rapidly developing field of optogenetics, which is currently mainly focusing on the precise, light-dependent control of eukaryotic neural cells (Deisseroth, 2015). The extended wavelength absorption for blue light by the echinenone light-harvesting antennae in GR, and potentially TR, may be an interesting feature for this field. Echinenone is probably not natively present in most eukaryotic cells and given the extremely low water solubility of this pigment, *in vivo* biosynthesis is probably crucial. This will at least require the introduction of a functional ketolase in the cell systems, as performed in this thesis for *E. coli*. Furthermore, also in the field of optogenetics the limited flux and production levels were observed to be a major challenge (Deisseroth, 2015). In this thesis, we show the benefit of translational tuning based on RBSs for improved production (**Chapters 4,7**). In eukaryotic hosts translation is not mediated through RBSs and hence promoter engineering has been the preferred method to tune expression. However, alternative eukaryotic mechanisms to tune translation

initiation, such as the Kozak motif (Kozak, 2005), and mRNA stability could be used for tuning in eukaryotes.

Future prospects for engineering photosynthesis

For the engineering of more efficient photosynthetic organisms, a promising strategy may be the blending of oxygenic photosystems with anoxygenic photosystems (**Chapter 1**) (Blankenship et al., 2011; Ort et al., 2015). However, transplanting reaction center photosystems, including multiple transmembrane photosystem proteins and complex pigment biosynthesis pathways, will be a major challenge. Some of the applied *in silico* analyses and explored experimental tools in this thesis, including fine-tuning through RBS modulation, can be instrumental hereto. However, given the large number of components to be optimized for transplanting reaction center photosystems, this will be even more challenging than the here demonstrated transplantation of the dual-pigment GR system. One solution to engineer large systems, as basically also performed in this thesis, is to split parts of the system, such as pigments synthesis in modules, and first engineer those modules separately.

Once engineered photosystems have the potential to provide a sufficient energetic benefit to the cell, laboratory evolution becomes a powerful approach. When growth of a cell can be directly coupled to the functionality of these photosystems, better functioning variants can be easily selected and further evolved in laboratory evolution approaches (Portnoy et al., 2011). Given the available genetic engineering tools, *E. coli* may be an interesting host to transplant and optimize photosystems. Regarding factors such as available membrane space, the presence of complementary autotrophic systems, and final cell factory applications, photosynthetic hosts may be preferred instead as a chassis. However, the number of tools for model photosynthetic organisms is still relatively limited. Specifically, tools for fine-tuning, such as precise RBS engineering approaches, will have to be further developed for photosynthetic bacterial hosts, such as cyanobacteria (Ramey et al., 2015). Microalgae and plants, with their limited genetic accessibility and tools, seem currently not yet suitable for ambitious, complex attempts to transplant and modify non-native photosystems.

It is also good to point out that anoxygenic phototrophs, such as the model purple bacterium *Rhodobacter sphaeroides*, should be considered as a chassis for engineering photosynthesis. Anoxygenic phototrophs have more efficient reaction center systems for proton-pumping than PPRs, which are capable of using wavelengths >700 nm. Electrochemically produced hydrogen, or possibly formate, could be interesting electron donors for those anoxygenic phototrophs in biotechnology. This could result in efficient anaerobic photo-electro-autotrophs, as proposed in **Chapter 3**. For these potential anaerobic photosynthetic cell factories, also the engineering of alternative CO_2 fixation pathways should be considered, as some oxygen-sensitive, but highly ATP-efficient alternatives for CO_2 fixation seem promising in these conditions (**Chapter 3**).

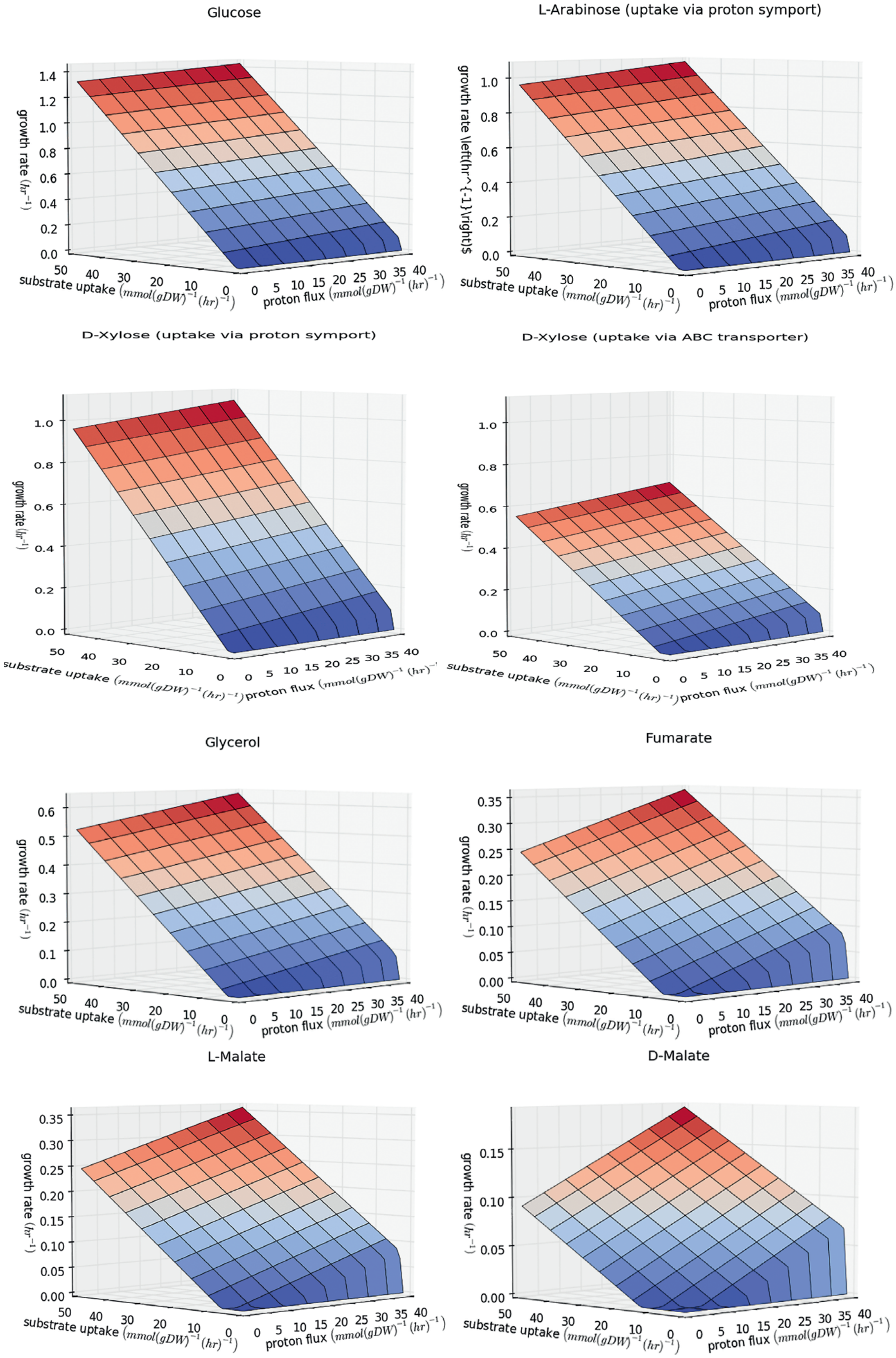
In general, engineering of CO_2 fixation pathways, the 'dark reactions' of photosynthesis, seems promising and deserves extensive exploration (**Chapter 8**). Recently, a groundbreaking example of adaptive laboratory evolution has been reported for engineering a functional CO_2 fixation cycle in *E. coli* (Antonovsky et al., 2016). Together with the recent *in vitro* demonstration of the first synthetic CO_2 fixation cycle (Schwander et al., 2016), these milestones provide further confidence for introducing more efficient synthetic CO_2 fixation pathways in heterotrophs and photoautotrophs.

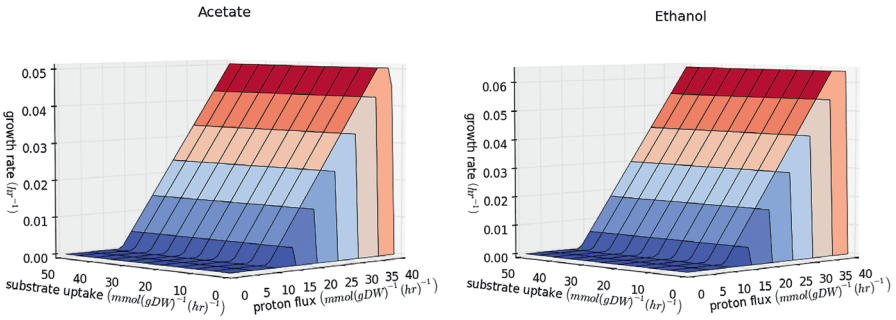
As a general critical point for both natural and improved biological photosystems, it should be mentioned that their currently demonstrated and predicted energy conversion efficiencies are outcompeted by far by state-of-the-art manmade photosystems: inorganic photovoltaic (PV) solar cells (Blankenship et al., 2011). Hence, a promising approach is to integrate these PV-cells with efficient electrocatalysis, producing electron donors, such as hydrogen or formate, which could subsequently be used to feed chemoautotrophic cell factories (**Chapter 8**). A very encouraging and highly efficient proof-of-principle for such a platform, based on the chemoautotroph *Cupriavidus necator*, was recently demonstrated (Liu et al., 2016). This work claims a light-to-biomass energy efficiency of 10%, largely exceeding the efficiencies of all demonstrated platforms based on biological photosynthesis. This development should be an important reference to compare photosynthetic production platforms, not only

regarding their energy efficiencies, but also with respect to their economic and large-scale feasibility.

Still, the field of improving photosynthesis may have an exciting future, also given the important challenge of improving photosynthesis for agricultural crops (Long et al., 2015). Many promising ideas for engineering both the 'light reactions' and 'dark reactions' of photosynthesis deserve a chance to be further explored. Novel and rational ideals will be helpful for improving photosynthesis, but realizing them will mostly require a lot of experimental effort. Therefore, we should definitely not forget the support these efforts can get from the power of laboratory evolution for realizing more efficient photosynthesis.

Supplementary information





Supplementary Figure S1: *in silico* predicted growth rates of *E. coli* for different substrates with varying ranges of PPR-pumping and substrate uptake. Grow rates were predicted by Flux Balance Analysis using standard model parameters for *E. coli* genome-scale model *iJO1366* (see also **Chapter 3**).

References

Agris, P.F., Vendeix, F.A.P., and Graham, W.D. (2007). tRNA's Wobble Decoding of the Genome: 40 Years of Modification. J. Mol. Biol. *366*, 1–13.

Akhtar, M.K., and Jones, P.R. (2009). Construction of a synthetic YdbK-dependent pyruvate:H2 pathway in *Escherichia coli* BL21(DE3). Metab. Eng. *11*, 139–147.

Andersson, S.G., and Kurland, C.G. (1990). Codon preferences in free-living microorganisms. Microbiol Rev *54*, 198–210.

Angermayr, S.A., van der Woude, A.D., Correddu, D., Vreugdenhil, A., Verrone, V., and Hellingwerf, K.J. (2014). Exploring metabolic engineering design principles for the photosynthetic production of lactic acid by *Synechocystis* sp. PCC6803. Biotechnol. Biofuels *7*, 99.

Angermayr, S.A., Gorchs Rovira, A., and Hellingwerf, K.J. (2015). Metabolic engineering of cyanobacteria for the synthesis of commodity products. Trends Biotechnol. *33*, 352–361.

Angov, E., Hillier, C.J., Kincaid, R.L., and Lyon, J.A. (2008). Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. PLoS One *3*, e2189.

Angov, E., Legler, P.M., and Mease, R.M. (2011). Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding. Methods Mol Biol *705*, 1–13.

Antonovsky, N., Gleizer, S., Noor, E., Zohar, Y., Herz, E., Barenholz, U., Zelcbuch, L., Amram, S., Wides, A., Tepper, N., et al. (2016). Sugar Synthesis from CO_2 in Escherichia coli. Cell *166*, 1–11.

Arechaga, I., Miroux, B., Karrasch, S., Huijbregts, R., de Kruijff, B., Runswick, M.J., and Walker, J.E. (2000). Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale over-production of the b subunit of F_1F_0 ATP synthase. FEBS Lett. *482*, 215–219.

Aro, E.-M. (2016). From first generation biofuels to advanced solar biofuels. Ambio *45*, 24–31.

Atamna-Ismaeel, N., Sabehi, G., Sharon, I., Witzel, K.-P., Labrenz, M., Jürgens, K., Barkay, T., Stomp, M., Huisman, J., and Beja, O. (2008). Widespread distribution of proteorhodopsins in freshwater and brackish ecosystems. ISME J. *2*, 656–662.

Atamna-Ismaeel, N., Finkel, O.M., Glaser, F., Sharon, I., Schneider, R., Post, A.F., Spudich, J.L., von Mering, C., Vorholt, J.A., Iluz, D., et al. (2012). Microbial rhodopsins on leaf surfaces of terrestrial plants. Environ. Microbiol. *14*, 140–146.

Atomi, H., Sato, T., and Kanai, T. (2011). Application of hyperthermophiles and their enzymes. Curr. Opin. Biotechnol. *22*, 618–626.

Atsumi, S., and Liao, J.C. (2009). Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. Nat Biotechnol *27*, 1177–1180.

Azai, C., Harada, J., and Oh-oka, H. (2013). Gene expression system in green sulfur bacteria by conjugative plasmid transfer. PLoS One *8*, 1–13.

Balashov, S.P., Imasheva, E.S., Boichenko, V.A., Antón, J., Wang, J.M., and Lanyi, J.K. (2005). Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. Science *309*, 2061–2064.

Balashov, S.P., Imasheva, E.S., Wang, J.M., and Lanyi, J.K. (2008). Excitation Energy-Transfer and the Relative Orientation of Retinal and Carotenoid in Xanthorhodopsin. Biophys. J. *95*, 2402–2414.

Balashov, S.P., Imasheva, E.S., Choi, A.R., Jung, K.-H., Liaaen-Jensen, S., and Lanyi, J.K. (2010). Reconstitution of gloeobacter rhodopsin with echinenone: role of the 4-keto group. Biochemistry *49*, 9792–9799.

Balzer, S., Kucharova, V., Megerle, J., Lale, R., Brautaset, T., and Valla, S. (2013). A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in *Escherichia coli*. Microb. Cell Fact. *12*, 26.

Bar-Even, A. (2016). Formate assimilation: The metabolic architecture of natural and synthetic pathways. Biochemistry *55*, 3851–3863.

Bar-Even, A., Noor, E., Lewis, N.E.N.N.E., and Milo, R. (2010). Design and analysis of synthetic carbon fixation pathways. Proc. Natl. Acad. Sci. U. S. A. *107*, 8889–8894.

Bar-Even, A., Noor, E., Milo, R., Flamholz, A., and Milo, R. (2012a). A survey of carbon fixation pathways through a quantitative lens. J. Exp. Bot. *63*, 2325–2342.

Bar-Even, A., Flamholz, A., Noor, E., and Milo, R. (2012b). Thermodynamic constraints shape the structure of carbon fixation pathways. Biochim. Biophys. Acta *1817*, 1646–1659.

Bar-Even, A., Noor, E., and Milo, R. (2012c). A survey of carbon fixation pathways through a quantitative lens. J. Exp. Bot. *63*, 2325–2342.

Bar-Even, A., Noor, E., Flamholz, A., and Milo, R. (2013). Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes. Biochim. Biophys. Acta *1827*, 1039–1047.

Beckmann, J., Lehr, F., Finazzi, G., Hankamer, B., Posten, C., Wobbe, L., and Kruse, O. (2009). Improvement of light to biomass conversion by de-regulation of light-harvesting protein translation in *Chlamydomonas reinhardtii*. J. Biotechnol. *142*, 70–77.

Begley, U., Dyavaiah, M., Patil, A., Rooney, J.P., DiRenzo, D., Young, C.M., Conklin, D.S., Zitomer, R.S., and Begley, T.J. (2007). Trm9-catalyzed tRNA modifications link translation to the DNA damage response. Mol Cell *28*, 860–870.

Béja, O., Spudich, E.N., Spudich, J.L., Leclerc, M., and DeLong, E.F. (2001). Proteorhodopsin phototrophy in the ocean. Nature *411*, 786–789.

Béjà, O., Aravind, L., Koonin, E. V, Suzuki, M.T., Hadd, A., Nguyen, L.P., Jovanovich, S.B., Gates, C.M., Feldman, R.A., Spudich, J.L., et al. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science *289*, 1902–1906.

Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van, S.J., Rabinowitz, J.D., Van Dien, S.J., Rabinowitz, J.D., Van, S.J., Rabinowitz, J.D., et al. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. Nat. Chem. Biol. *5*, 593–599.

Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., and Bluthgen, N. (2013). Efficient translation initiation dictates codon usage at gene start. Mol Syst Biol *9*, 675.

Berg, I.A. (2011). Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. Appl. Environ. Microbiol. *77*, 1925–1936.

Berg, O.G., and Kurland, C.G. (1997). Growth rate-optimised tRNA abundance and codon usage. J. Mol. Biol. *270*, 544–550.

Berg, I.A., Kockelkorn, D., Ramos-Vera, W.H., Say, R.F., Zarzycki, J., Hügler, M., Alber, B.E., and Fuchs, G. (2010). Autotrophic carbon fixation in archaea. Nat. Rev. Microbiol. *8*, 447–460.

Berla, B.M., Saha, R., Immethun, C.M., Maranas, C.D., Moon, T.S., and Pakrasi, H.B. (2013). Synthetic biology of cyanobacteria: unique challenges and opportunities. Front. Microbiol. *4*.

Bernsel, A., and Daley, D.O. (2009). Exploring the inner membrane proteome of *Escherichia coli*: which proteins are eluding detection and why? Trends Microbiol. *17*, 444–449.

Bi, C., Su, P., Müller, J., Yeh, Y.-C., Chhabra, S.R., Beller, H.R., Singer, S.W., and Hillson, N.J. (2013). Development of a broad-host synthetic biology toolbox for *Ralstonia eutropha* and its application to engineering hydrocarbon biofuel production. Microb. Cell Fact. *12*, 107.

Bickel-Sandkötter, S., Gärtner, W., and Dane, M. (1996). Conversion of energy in halobacteria: ATP synthesis and phototaxis. Archaeal Microbiol. *166*, 1–11.

Birge, R.R. (1990). Nature of the primary photochemical events in rhodopsin and bacteriorhodopsin. Biochim. Biophys. Acta, Bioenerg. *1016*, 293–327.

Blankenship, R.E., Tiede, D.M., Barber, J., Brudvig, G.W., Fleming, G., Ghirardi, M., Gunner, M.R., Junge, W., Kramer, D.M., Melis, A., et al. (2011). Comparing photosynthetic and photovoltaic efficiencies and recognizing the potential for improvement. Science *332*, 805–809.

Blaschkowski, H.P., Knappe, J., Ludwig-Festl, M., and Neuer, G. (1982). Routes of flavodoxin and ferredoxin reduction in *Escherichia coli*. Eur. J. Biochem. *123*, 563–569.

Blazeck, J., and Alper, H.S. (2013). Promoter engineering: recent advances in controlling transcription at the most fundamental level. Biotechnol. J. *8*, 46–58.

Boël, G., Letso, R., Neely, H., Price, W.N., Wong, K., Su, M., Luff, J.D., Valecha, M., Everett, J.K., Acton, T.B., et al. (2016). Codon influence on protein expression in E. coli correlates with mRNA levels. Nature *529*, 358–363.

Bogorad, I.W., Chen, C., Theisen, M.K., Wu, T., Schlenz, A.R., and Lam, A.T. (2014). Building carbon – carbon bonds using a biocatalytic methanol condensation cycle. Proc. Natl. Acad. Sci. U. S. A. *111*, 1592815933.

Boichenko, V. a, Wang, J.M., Antón, J., Lanyi, J.K., and Balashov, S.P. (2006). Functions of carotenoids in xanthorhodopsin and archaerhodopsin, from action spectra of photoinhibition of cell respiration. Biochim. Biophys. Acta, Bioenerg. *1757*, 1649–1656.

Bonacci, W., Teng, P.K., Afonso, B., Niederholtmeyer, H., Grob, P., Silver, P.A., and Savage, D.F. (2012). Modularity of a carbon-fixing protein organelle. Proc. Natl. Acad. Sci. U. S. A. *109*, 478–483.

Bonde, M.T., Pedersen, M., Klausen, M.S., Jensen, S.I., Wulff, T., Harrison, S., Nielsen, A.T., Herrgård, M.J., and Sommer, M.O.A. (2016). Predictable tuning of protein expression in bacteria. Nat. Methods.

Boock, J.T., Gupta, A., and Prather, K.L. (2015). Screening and modular design for metabolic pathway optimization. Curr. Opin. Biotechnol. *36*, 189–198.

Boyle, N.R., and Morgan, J.A. (2011). Computation of metabolic fluxes and efficiencies for biological carbon dioxide fixation. Metab. Eng. *13*, 150–158.

Boyle, P.M., and Silver, P.A. (2012). Parts plus pipes: synthetic biology approaches to metabolic engineering. Metab. Eng. *14*, 223–232.

Bratanov, D., Balandin, T., Round, E., Shevchenko, V., Gushchin, I., Polovinkin, V., Borshchevskiy, V., and Gordeliy, V. (2015). An Approach to Heterologous Expression of Membrane Proteins. The Case of Bacteriorhodopsin. PLoS One *10*, e0128390.

Breitenbach, J., Gerjets, T., and Sandmann, G. (2013). Catalytic properties and reaction mechanism of the CrtO carotenoid ketolase from the cyanobacterium *Synechocystis* sp. PCC 6803. Arch. Biochem. Biophys. *529*, 86–91.

Bremer, H., and Dennis, P.P. (2009). Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate. Escherichia Coli Salmonella Cell. Mol. Biol. *2*, 1527–1542.

Brown, L.S. (2013). Eubacterial rhodopsins – unique photosensors and diverse ion pumps. Biochim. Biophys. Acta, Bioenerg. *1837*, 553–561.

Brown, L.S., and Jung, K.-H. (2006). Bacteriorhodopsin-like proteins of eubacteria and fungi: the extent of conservation of the haloarchaeal proton-pumping mechanism. Photochem. Photobiol. Sci. *5*, 538–546.

Bruinsma J. (2009). The resource outlook to 2050: By how much do land, water and crop yields need to increase by 2050?

Bryant, D.A., and Frigaard, N.-U. (2006). Prokaryotic photosynthesis and phototrophy illuminated. Trends Microbiol. *14*, 488–496.

Buchan, J.R., Aucott, L.S., and Stansfield, I. (2006). tRNA properties help shape codon pair preferences in open reading frames. Nucleic Acids Res *34*, 1015–1027.

Buck, J.D. (2012). Physiological effects of heterologous expression of proteorhodopsin photosystems. Massachusetts Institute of Technology.

Buckel, W., and Thauer, R.K. (2013). Energy conservation via electron bifurcating ferredoxin reduction and proton/Na^+ translocating ferredoxin oxidation. Biochim. Biophys. Acta - Bioenerg. *1827*, 94–113.

Buhr, F., Jha, S., Thommen, M., Mittelstaet, J., Kutz, F., Schwalbe, H., Rodnina, M.V. V., and Komar, A.A.A. (2016). Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. Mol. Cell *61*, 341–351.

Caforio, A., Jain, S., Fodran, P., Siliakus, M., Minnaard, A., van der Oost, J., and Driessen, A.J.M. (2015). Formation of the ether lipids archaetidylglycerol and archaetidylethanolamine in *Escherichia coli*. Biochem. J. *470*, 343–355.

Cai, Z., Liu, G., Zhang, J., and Li, Y. (2014). Development of an activity-directed selection system enabled significant improvement of the carboxylation efficiency of Rubisco. Protein Cell *5*, 552–562.

Cameron, D.E., Bashor, C.J., and Collins, J.J. (2014). A brief history of synthetic biology. Nat. Rev. Microbiol. *12*, 381–390.

Cannarozzi, G., Schraudolph, N.N., Faty, M., von Rohr, P., Friberg, M.T., Roth, A.C., Gonnet, P., Gonnet, G., and Barral, Y. (2010). A role for codon order in translation dynamics. Cell *141*, 355–367.

Casini, A., Storch, M., Baldwin, G.S., and Ellis, T. (2015). Bricks and blueprints: methods and standards for DNA assembly. Nat. Rev. Mol. Cell Biol. *16*, 568–576.

Ceroni, F., Algar, R., Stan, G.-B., and Ellis, T. (2015). Quantifying cellular capacity identifies gene expression designs with reduced burden. Nat. Methods *12*, 415–418.

Chan, C.T., Pang, Y.L., Deng, W., Babu, I.R., Dyavaiah, M., Begley, T.J., and Dedon, P.C. (2012). Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. Nat Commun *3*, 937.

Charneski, C.A., and Hurst, L.D. (2014). Positive charge loading at protein termini is due to membrane protein topology, not a translational ramp. Mol. Biol. Evol. *31*, 70–84.

Charneski, C. a., and Hurst, L.D. (2013). Positively charged residues are the major determinants of ribosomal velocity. PLoS Biol. *11*, e1001508.

Chen, M., and Blankenship, R.E. (2011). Expanding the solar spectrum used by photosynthesis. Trends Plant Sci. *16*, 427–431.

Chen, Q., van der Steen, J.B., Dekker, H.L., Ganapathy, S., de Grip, W.J., and Hellingwerf, K.J. (2016). Expression of holo-proteorhodopsin in *Synechocystis* sp. PCC 6803. Metab. Eng. 1–12.

Chen, S.L., Lee, W., Hottes, A.K., Shapiro, L., and McAdams, H.H. (2004). Codon usage between genomes is constrained by genome-wide mutational processes. Proc Natl Acad Sci U S A *101*, 3480–3485.

Cheong, D.E., Ko, K.C., Han, Y., Jeon, H.G., Sung, B.H., Kim, G.J., Choi, J.H., and Song, J.J. (2015). Enhancing functional expression of heterologous proteins through random substitution of genetic codes in the 5' coding region. Biotechnol. Bioeng. *112*, 822–826.

Chistoserdova, L., Kalyuzhnaya, M.G., and Lidstrom, M.E. (2009). The Expanding World of Methylotrophic Metabolism. Annu. Rev. Microbiol. *63*, 477–499.

Choi, A.R., Shi, L., Brown, L.S., and Jung, K.-H. (2014). Cyanobacterial Light-Driven Proton Pump, Gloeobacter Rhodopsin: Complementarity between Rhodopsin-Based Energy Production and Photosynthesis. PLoS One *9*, e110643.

Chow, B.Y., Han, X., Dobry, A.S., Qian, X., Chuong, A.S., Li, M., Henninger, M. a, Belfort, G.M., Lin, Y., Monahan, P.E., et al. (2010). High-performance genetically targetable optical neural silencing by light-driven proton pumps. Nature *463*, 98–102.

Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M.F., and von der Haar, T. (2014). Translation elongation can control translation initiation on eukaryotic mRNAs. EMBO J *33*, 21–34.

Claassens, N.J., Volpers, M., Martins dos Santos, V.A.P., van der Oost, J., and de Vos, W.M. (2013). Potential of proton-pumping rhodopsins: engineering photosystems into microorganisms. Trends Biotechnol. *31*, 633–642.

Claassens, N.J., Sousa, D.Z., Santos, V.A.P.M. dos, de Vos, W.M., and van der Oost, J. (2016). Harnessing the power of microbial autotrophy. Nat. Rev. Microbiol. *14*, 692–706.

Coleman, J.R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E., and Mueller, S. (2008). Virus attenuation by genome-scale changes in codon pair bias. Science *320*, 1784–1787.

Conrado, R.J., Haynes, C.A., Haendler, B.E., and Toone, E.J. (2013). Electrofuels: A New Paradigm for Renewable Fuels. In Advanced Biofuels and Bioproducts, pp. 1037–1064.

Corcelli, A., Lattanzio, V.M.T., Mascolo, G., Papadia, P., and Fanizzi, F. (2002). Lipid-protein stoichiometries in a crystalline biological membrane: NMR quantitative analysis of the lipid extract of the purple membrane. J. Lipid Res. *43*, 132–140.

Crick, F.H. (1966). Codon-anticodon pairing: the wobble hypothesis. J. Mol. Biol. *19*, 548–555.

Croce, R., and van Amerongen, H. (2014). Natural strategies for photosynthetic light harvesting. Nat Chem Biol *10*, 492–501.

Daley, D.O. (2005). Global Topology Analysis of the Escherichia coli Inner Membrane Proteome. Science *308*, 1321–1323.

Dana, A., and Tuller, T. (2014). The effect of tRNA levels on decoding times of mRNA codons. Nucleic Acids Res. *42*, 9171–9181.

Das, A., Yoon, S.H., Lee, S.H., Kim, J.Y., Oh, D.K., and Kim, S.W. (2007). An update on microbial carotenoid production: Application of recent metabolic engineering tools. Appl. Microbiol. Biotechnol. *77*, 505–512.

Davidi, D., Liebermeister, W., Bar-Even, A., Flamholz, A., Tummler, K., Barenholz, U., Goldenfeld, M., Shlomi, T., Noor, E., and Milo, R. (2016). Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements. Proc Natl Acad Sci USA *113*, 3401–3406.

Davis, J.H., Rubin, A.J., and Sauer, R.T. (2011). Design, construction and characterization of a set of insulated bacterial promoters. Nucleic Acids Res. *39*, 1131–1141.

Deisseroth, K. (2015). Optogenetics : 10 years of microbial opsins in neuroscience. *18*, 1213–1225.

DeLong, E.F., and Beja, O. (2010). The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. PLoS Biol *8*, e1000359.

Diender, M., Stams, A.J.M., and Sousa, D.Z. (2015). Pathways and Bioenergetics of Anaerobic Carbon Monoxide Fermentation. Front. Microbiol. *6*, 1–18.

Dittmar, K.A., Sorensen, M.A., Elf, J., Ehrenberg, M., Pan, T., Sørensen, M. a, Elf, J., Ehrenberg, M., Pan, T., Sorensen, M.A., et al. (2005). Selective charging of tRNA isoacceptors induced by amino-acid starvation. EMBO Rep. *6*, 151–157.

Dittmar, K.A., Goodenbour, J.M., and Pan, T. (2006). Tissue-specific differences in human transfer RNA expression. PLoS Genet *2*, e221.

Drew, D., Slotboom, D.-J., Friso, G., Reda, T., Genevaux, P., Rapp, M., Meindl-Beinker, N.M., Lambert, W., Lerch, M., Daley, D.O., et al. (2005). A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. Protein Sci. *14*, 2011–2017.

Drew, D., Lerch, M., Kunji, E., Slotboom, D., and Gier, J. De (2006). Optimization of membrane protein overexpression and purification using GFP fusions. Nat. Methods *3*, 303–313.

Drew, D.E., Heijne, G. Von, and Gier, J.L. De (2001). Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*. *507*, 220–224.

Drummond, D.A., and Wilke, C.O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell *134*, 341–352.

Ducat, D.C., and Silver, P.A. (2012). Improving carbon fixation pathways. Curr. Opin. Chem. Biol. *16*, 337–344.

Ducat, D.C., Avelar-Rivas, J.A., Way, J.C., and Silvera, P.A. (2012). Rerouting carbon flux to enhance photosynthetic productivity. Appl. Environ. Microbiol. *78*, 2660–2668.

Durão, P., Aigner, H., Nagy, P., Mueller-Cajar, O., Hartl, F.U., and Hayer-Hartl, M. (2015). Opposing effects of folding and assembly chaperones on evolvability of Rubisco. Nat Chem Biol *11*, 148–155.

Dürre, P., and Eikmanns, B.J. (2015). C1-carbon sources for chemical and fuel production by microbial gas fermentation. Curr. Opin. Biotechnol. *35*, 63–72.

Ebrahim, A., Lerman, J.A., Palsson, B.Ø., and Hyduke, D.R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst. Biol. *7*.

Elena, C., Ravasi, P., Castelli, M.E., Peirú, S., and Menzella, H.G. (2014). Expression of codon optimized genes in microbial systems: Current industrial applications and perspectives. Front. Microbiol. *5*, 1–8.

Elf, J., and Ehrenberg, M. (2005). What makes ribosome-mediated transcriptional attenuation sensitive to amino acid limitation? PLoS Comput Biol *1*, e2.

Elf, J., Nilsson, D., Tenson, T., and Ehrenberg, M. (2003). Selective charging of tRNA isoacceptors explains patterns of codon usage. Science *300*, 1718–1722.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One *3*.

Engqvist, M.K.M., McIsaac, R.S., Dollinger, P., Flytzanis, N.C., Abrams, M., Schor, S., and Arnold, F.H. (2015). Directed Evolution of *Gloeobacter violaceus* Rhodopsin Spectral Properties. J. Mol. Biol. *427*, 205–220.

Erb, T.J. (2011). Carboxylases in natural and synthetic microbial pathways. Appl. Environ. Microbiol. *77*, 8466–8477.

Erb, T.J., Berg, I.A., Brecht, V., Müller, M., Fuchs, G., and Alber, B.E. (2007). Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. Proc. Natl. Acad. Sci. U. S. A. *104*, 10631–10636.

Erb, T.J., Brecht, V., Fuchs, G., Müller, M., and Alber, B.E. (2009). Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase. Proc. Natl. Acad. Sci. U. S. A. *106*, 8871–8876.

Eremina, N.S., Yampolskaya, T.A., Altman, I.B., Mashko, S. V, and Stoynova, N. V (2010). Overexpression of ydbK-encoding putative pyruvate synthase improves L-valine production and aerobic growth on ethanol media by an *Escherichia coli* strain carrying an oxygen-resistant alcohol dehydrogenase. J. Microb. Biochem. Technol. *2*, 77–83.

Eriksson, H. (2009). Intracellular vesicles induced by monotopic membrane protein in *Escherichia coli*. Stockholm University.

Eriksson, H.M., Wessman, P., Ge, C., Edwards, K., Wieslander, A., and Wieslander, Å. (2009). Massive formation of intracellular membrane vesicles in *Escherichia coli* by a monotopic membrane-bound lipid glycosyltransferase. J. Biol. Chem. *284*, 33904–33914.

Ernst, O.P., Lodowski, D.T., Elstner, M., Hegemann, P., Brown, L.S., and Kandori, H. (2014). Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. Chem. Rev. *114*, 126–163.

Estelmann, S., Hügler, M., Eisenreich, W., Werner, K., Berg, I. a., Ramos-Vera, W.H., Say, R.F., Kockelkorn, D., Gad'on, N., and Fuchs, G. (2011). Labeling and enzyme studies of the central carbon metabolism in *Metallosphaera sedula*. J. Bacteriol. *193*, 1191–1200.

Farasat, I., Kushwaha, M., Collens, J., Easterbrook, M., Guido, M., and Salis, H.M. (2014). Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. Mol. Syst. Biol. *10*, 731.

Fast, A.G., and Papoutsakis, E.T. (2012). Stoichiometric and energetic analyses of non-photosynthetic CO_2-fixation pathways to support synthetic biology strategies for production of fuels and chemicals. Curr. Opin. Chem. Eng. *1*, 380–395.

Fast, A.G., Schmidt, E.D., Jones, S.W., and Tracy, B.P. (2015). Acetogenic mixotrophy: novel options for yield improvement in biofuels and biochemicals production. Curr. Opin. Biotechnol. *33*, 60–72.

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., and Palsson, B.Ø. (2007). A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. *3*.

Feist, A.M., Nagarajan, H., Rotaru, A.-E., Tremblay, P.-L., Zhang, T., Nevin, K.P., Lovley, D.R., and Zengler, K. (2014). Constraint-based modeling of carbon fixation and the energetics of electron transfer in Geobacter metallireducens. PLoS Comput. Biol. *10*, e1003575.

Feng, J., Liu, H.-C., Chu, J.-F., Zhou, P.-J., Tang, J.-A., and Liu, S.-J. (2006). Genetic cloning and functional expression in *Escherichia coli* of an archaerhodopsin gene from *Halorubrum xinjiangense*. Extremophiles *10*, 29–33.

Fenno, L., Yizhar, O., and Deisseroth, K. (2011). The development and application of optogenetics. Annu. Rev. Neurosci. *34*, 389–412.

Flamholz, A., Noor, E., Bar-Even, A., and Milo, R. (2012). EQuilibrator - The biochemical thermodynamics calculator. Nucleic Acids Res. *40*, D770-775.

Fluman, N., Navon, S., Bibi, E., and Pilpel, Y. (2014). mRNA-programmed translation pauses in the targeting of *E. coli* membrane proteins. eLife *3*, e03440.

Frenkel-Morgenstern, M., Danon, T., Christian, T., Igarashi, T., Cohen, L., Hou, Y.M., and Jensen, L.J. (2012). Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. Mol Syst Biol *8*, 572.

Friedrich, T., Geibel, S., Kalmbach, R., Chizhov, I., Ataka, K., Heberle, J., Engelhard, M., and Bamberg, E. (2002). Proteorhodopsin is a light-driven proton pump with variable vectoriality. J. Mol. Biol. *321*, 821–838.

Frigaard, N.-U., Martinez, A., Mincer, T.J., and DeLong, E.F. (2006). Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. Nature *439*, 847–850.

Fu, H.-Y., Lin, Y.-C., Chang, Y.-N., Tseng, H., Huang, C.-C., Liu, K.-C., Huang, C.-S., Su, C.-W., Weng, R.R., Lee, Y.-Y., et al. (2010). A novel six-rhodopsin system in a single archaeon. J. Bacteriol. *192*, 5866–5873.

Fuchs, G. (2011). Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? Annu. Rev. Microbiol. *65*, 631–658.

Fuhrman, J.A., Schwalbach, M.S., and Stingl, U. (2008). Proteorhodopsins: an array of physiological roles? Nat. Rev. Microbiol. *6*, 488–494.

Ganapathy, S., Bécheau, O., Venselaar, H., Frölich, S., Steen, J.B. van der, Chen, Q., Radwan, S., Lugtenburg, J., Hellingwerf, K.J.J., Groot, H.J.M. de, et al. (2015). Modulation of spectral properties and pump activity of proteorhodopsins by retinal analogs. Biochem. J. *467*, 333–343.

Ganapathy, S., Venselaar, H., Chen, Q., Groot, J.M. De, Hellingwerf, K.J., and Grip, W.J. De (2017). Retinal-based Proton Pumping in the Near Infra-red Retinal-based Proton Pumping in the Near Infra-red. J. Am. Chem. Soc. DOI: 10.1021/jacs.6b11366.

Gao, X., Sun, T., Pei, G., Chen, L., and Zhang, W. (2016). Cyanobacterial chassis engineering for enhancing production of biofuels and chemicals. Appl. Microbiol. Biotechnol. *100*, 3401–3413.

Gardin, J., Yeasmin, R., Yurovsky, A., Cai, Y., Skiena, S., and Futcher, B. (2014). Measurement of average decoding rates of the 61 sense codons in vivo. eLife *3*, 1–20.

Geertsma, E.R., Groeneveld, M., Slotboom, D.-J., and Poolman, B. (2008). Quality control of overexpressed membrane proteins. Proc. Natl. Acad. Sci. U. S. A. *105*, 5722–5727.

Gerhardt, A., Çinkaya, I., Linder, D., Huisman, G., and Buckel, W. (2000). Fermentation of 4-aminobutyrate by *Clostridium aminobutyricum*: cloning of two genes involved in the formation and dehydration of 4-hydroxybutyryl-CoA. Arch. Microbiol. *174*, 189–199.

Gerosa, L., Haverkorn Van Rijsewijk, B.R.B., Christodoulou, D., Kochanowski, K., Schmidt, T.S.B., Noor, E., and Sauer, U. (2015). Pseudo-transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data. Cell Syst. *1*, 270–282.

Giacalone, M., Gentile, A., Lovitt, B., Berkley, N., Gunderson, C., and Surber, M. (2006). Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system. Biotechniques *40*, 355–364.

Gialama, D., Kostelidou, K., Michou, M., Delivoria, D.C., Kolisis, F.N., and Skretas, G. (2016). Development of *Escherichia coli* strains that withstand membrane protein-induced toxicity and achieve high-level recombinant membrane protein production. ACS Synth. Biol.

Gibson, D.G. (2014). Programming biological operating systems: genome design, assembly and activation. Nat. Methods *11*, 521–526.

Gimpel, J.A., Henríquez, V., and Mayfield, S.P. (2015). In Metabolic Engineering of Eukaryotic Microalgae: Potential and Challenges Come with Great Diversity. Front. Microbiol. *6*, 1376.

Gimpel, J. a., Nour-Eldin, H.H., Scranton, M. a, Li, D., and Mayfield, S.P. (2016). Refactoring the six-gene Photosystem II core in the chloroplast of the green algae *Chlamydomonas reinhardtii*. ACS Synth. Biol. *5*, 589–596.

Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sorensen, K.D., et al. (2014). A dual program for translation regulation in cellular proliferation and differentiation. Cell *158*, 1281–1292.

Giovannoni, S.J., Bibbs, L., Cho, J.C., Stapels, M.D., Desiderio, R., Vergin, K.L., Rappé, M.S., Laney, S., Wilhelm, L.J., Tripp, H.J., et al. (2005). Proteorhodopsin in the ubiquitous marine bacterium SAR11. Nature *438*, 82–85.

Godinic-Mikulcic, V., Jaric, J., Greber, B.J., Franke, V., Hodnik, V., Anderluh, G., Ban, N., and Weygand-Durasevic, I. (2014). Archaeal aminoacyl-tRNA synthetases interact with the ribosome to recycle tRNAs. Nucleic Acids Res *42*, 5191–5201.

Gokarn, R.R., Eiteman, M.A., and Altman, E. (1998). Expression of pyruvate carboxylase enhances succinate production in *Escherichia coli* without affecting glucose uptake. Biotechnol. Lett. *20*, 795–798.

Goldbeck, C.P., Jensen, H.M., TerAvest, M. a, Beedle, N., Appling, Y., Hepler, M., Cambray, G., Mutalik, V., Angenent, L.T., and Ajo-Franklin, C.M. (2013). Tuning promoter strengths for improved synthesis and function of electron conduits in *Escherichia coli*. ACS Synth. Biol. *2*, 150–159.

Goltermann, L., Borch Jensen, M., and Bentin, T. (2011). Tuning protein expression using synonymous codon libraries targeted to the 5′ mRNA coding region. Protein Eng. Des. Sel. *24*, 123–129.

Gómez-Consarnau, L., González, J.M., Coll-Lladó, M., Gourdon, P., Pascher, T., Neutze, R., Pedrós-Alió, C., Pinhassi, J., Gomez-Consarnau, L., Gonzalez, J.M., et al. (2007). Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. Nature *445*, 210–213.

Gómez-Consarnau, L., González, J.M., Riedel, T., Jaenicke, S., Wagner-Döbler, I., Sañudo-Wilhelmy, S.A., and Fuhrman, J.A. (2015). Proteorhodopsin light-enhanced growth linked to vitamin-B1 acquisition in marine Flavobacteria. ISME J. 1 11.

Gong, F., Liu, G., Zhai, X., Zhou, J., Cai, Z., and Li, Y. (2015). Quantitative analysis of an engineered CO_2-fixing *Escherichia coli* reveals great potential of heterotrophic CO_2 fixation. Biotechnol. Biofuels *8*.

Gonzalez-Esquer, C.R., Shubitowski, T.B., and Kerfeld, C.A. (2015). Streamlined Construction of the Cyanobacterial CO2-Fixing Organelle via Protein Domain Fusions for Use in Plant Synthetic Biology. Plant Cell *27*, 2637–2644.

Goodman, D.B., Church, G.M., and Kosuri, S. (2013). Causes and effects of N-terminal codon bias in bacterial genes. Science *342*, 475–479.

Gould, N., Hendy, O., and Papamichail, D. (2014). Computational tools and algorithms for designing customized synthetic genes. Front. Bioengeneering Biotechnol. *2*, 41.

Gourdon, P., Alfredsson, A., Pedersen, A., Malmerberg, E., Nyblom, M., Widell, M., Berntsson, R., Pinhassi, J., Braiman, M., Hansson, Ö., et al. (2008). Optimized in vitro and in vivo expression of proteorhodopsin: A seven-transmembrane proton pump. Protein Expr. Purif. *58*, 103–113.

Gouy, M., and Gautier, C. (1982). Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res *10*, 7055–7074.

Grantham, R., Gautier, C., Gouy, M., Mercier, R., and Pave, A. (1980). Codon catalog usage and the genome hypothesis. Nucleic Acids Res *8*, r49–r62.

Grayson, K.J., Faries, K.M., Huang, X., Qian, P., Dilbeck, P., Martin, E.C., Hitchcock, A., Vasilev, C., Yuen, J.M., Niedzwiedzki, D.M., et al. (2017). Augmenting light coverage for photosynthesis through YFP-enhanced charge separation at the *Rhodobacter sphaeroides* reaction centre. Nat. Commun. *8*, 13972.

Greene, D.N., Whitney, S.M., and Matsumura, I. (2007). Artificially evolved *Synechococcus* PCC6301 Rubisco variants exhibit improvements in folding and catalytic efficiency. Biochem. J. *404*, 517–524.

Gromadski, K.B., Daviter, T., and Rodnina, M. V. (2006). A uniform response to mismatches in codon-anticodon complexes ensures ribosomal fidelity. Mol. Cell *21*, 369–377.

Grosjean, H., de Crecy-Lagard, V., and Marck, C. (2010). Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. FEBS Lett *584*, 252–264.

Grossman, T.H., Kawasaki, E.S., Punreddy, S.R., and Osburne, M.S. (1998). Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. Gene *209*, 95–103.

Guadalupe-Medina, V., Wisselink, W.H., Luttik, M.A., de Hulster, E., Daran, J.-M., Pronk, J.T., and van Maris, A.J. (2013). Carbon dioxide fixation by Calvin-Cycle enzymes improves ethanol yield in yeast. Biotechnol. Biofuels *6*, 125.

Gustafsson, C., Govindarajan, S., and Minshull, J. (2004). Codon bias and heterologous protein expression. Trends Biotechnol *22*, 346–353.

Gustafsson, C., Minshull, J., Govindarajan, S., Ness, J., Villalobos, A., Welch, M., Drive, O.B., Suite, A., and Park, M. (2012). Engineering genes for predictable protein expression. Proteins Expr. Purif. *83*, 37–46.

Gutman, G.A., and Hatfield, G.W. (1989). Nonrandom utilization of codon pairs in *Escherichia coli* . Proc Natl Acad Sci U S A *86*, 3699–3703.

Hal Alper, Curt Fischer, Elke Nevoigt, and G.S. (2005). Tuning genetic control through promoter engineering. *103*.

Han, X., Chow, B.Y., Zhou, H., Klapoetke, N.C., Chuong, A., Rajimehr, R., Yang, A., Baratta, M. V, Winkle, J., Desimone, R., et al. (2011). A high-light sensitivity optical neural silencer: development and application to optogenetic control of non-human primate cortex. Front. Syst. Neurosci.

Hawkins, A.S., Han, Y., Lian, H., Loder, A.J., Menon, A.L., Iwuchukwu, I.J., Keller, M., Leuko, T.T., Adams, M.W.W., and Kelly, R.M. (2011). Extremely Thermophilic Routes to Microbial Electrofuels. ACS Catal. *1*, 1043–1050.

Hawkins, A.S., McTernan, P.M., Lian, H., Kelly, R.M., and Adams, M.W.W. (2013). Biological conversion of carbon dioxide and hydrogen into liquid fuels and industrial chemicals. Curr. Opin. Biotechnol. *24*, 376–384.

Hershberg, R., and Petrov, D.A. (2008). Selection on codon bias. Annu. Rev. Genet. *42*, 287–299.

Hildebrandt, V., Fendler, K., Heberle, J., Hoffmann, A., Bamberg, E., and Büldt, G. (1993). Bacteriorhodopsin expressed in *Schizosaccharomyces pombe* pumps protons through the plasma membrane. Proc. Natl. Acad. Sci. U. S. A. *90*, 3578–3582.

Hitchcock, A., Jackson, P.J., Chidgey, J.W., Dickman, M.J., Hunter, C.N., and Canniffe, D.P. (2016). Biosynthesis of chlorophyll *a* in a purple bacterial phototroph and assembly into a plant chlorophyll-protein complex. ACS Synth. Biol. *5*, 948–954.

Hjelm, A. (2015). Optimizing membrane and secretory protein production in Gram-negative bacteria. Stockholm University.

Hjelm, A., Schlegel, S., Baumgarten, T., Klepsch, M., Wickström, D., Drew, D., and Gier, J.-W. de (2013). Optimizing E. coli -Based Membrane Protein Production Using Lemo21(DE3) and GFP-Fusions. In Membrane Biogenesis: Methods and Protocols, D. Rapaport, and J.M. Herrmann, eds. (Berlin: Springer Science & Business Media), pp. 381 400.

Hoffmann, A., Hildebrandt, V., Heberle, J., and Büldt, G. (1994). Photoactive mitochondria: In vivo transfer of a light-driven proton pump into the inner mitochondrial membrane of *Schizosaccharomyces pombe*. Proc. Natl. Acad. Sci. U. S. A. *91*, 9367–9371.

Hoffmeister, S., Gerdom, M., Bengelsdorf, F.R., Linder, S., Flüchter, S., Öztürk, H., Blümke, W., May, A., Fischer, R.-J., Bahl, H., et al. (2016). Acetone production with metabolically engineered strains of *Acetobacterium woodii*. Metab. Eng. *36*, 37–47.

Hohmann-Marriott, M.F., and Blankenship, R.E. (2012). The photosynthetic world. In Photosynthesis: Plastid Biology, Energy Conversion and Carbon Assimilation, Advances in Photosynthesis and Respiration, J.J. Eaton-Rye, B.C. Tripathy, and T.D. Sharkey, eds. (Dordrecht: Springer Netherlands), pp. 3–32.

Hsieh, J.M., Besserer, G.M., Madej, M.G., Bui, H.Q., Kwon, S., and Abramson, J. (2010). Bridging the gap: A GFP-based strategy for overexpression and purification of membrane proteins with intra and extracellular C-termini. Protein Sci. *19*, 868–880.

Hu, P., Chakraborty, S., Kumar, A., Woolston, B.M., Liu, H., Emerson, D., and Stephanopoulos, G. (2016). Integrated Bioprocess for Conversion of Gaseous Substrates to Liquids. Proc. Natl. Acad. Sci. *113*, 14–19.

Huber, H., Gallenberger, M., Jahn, U., Eylert, E., Berg, I. a, Kockelkorn, D., Eisenreich, W., and Fuchs, G. (2008). A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum *Ignicoccus hospitalis*. Proc. Natl. Acad. Sci. U. S. A. *105*, 7851–7856.

Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol *2*, 13–34.

Imam, S., Fitzgerald, C.M., Cook, E.M., Donohue, T.J., and Noguera, D.R. (2015). Quantifying the effects of light intensity on bioproduction and maintenance energy during photosynthetic growth of *Rhodobacter sphaeroides*. Photosynth. Res. *123*, 167–182.

Imasheva, E.S., Balashov, S.P., Choi, A.R., Jung, K.-H., and Lanyi, J.K. (2009). Reconstitution of *Gloeobacter violaceus* rhodopsin with a light-harvesting carotenoid antenna. Biochemistry *48*, 10948–10955.

Imasheva, E.S., Balashov, S.P., Wang, J.M., and Lanyi, J.K. (2011). Removal and reconstitution of the carotenoid antenna of xanthorhodopsin. J. Membr. Biol. *239*, 95–104.

Ingolia, N.T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. Nat. Rev. Genet. *15*, 205–213.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Inoue, K., Ono, H., Abe-Yoshizumi, R., Yoshizawa, S., Ito, H., Kogure, K., and Kandori, H. (2013). A light-driven sodium ion pump in marine bacteria. Nat. Commun. *4*.

Inui, M., Suda, M., Kimura, S., Yasuda, K., Suzuki, H., Toda, H., Yamamoto, S., Okino, S., Suzuki, N., and Yukawa, H. (2008). Expression of *Clostridium acetobutylicum* butanol synthetic genes in *Escherichia coli*. Appl. Microbiol. Biotechnol. *77*, 1305–1316.

Ipekoğlu, E.M., Göçmen, K., Öz, M.T., Gürgan, M., and Yücel, M. (2016). Cloning and heterologous expression of chlorophyll a synthase in Rhodobacter sphaeroides. J. Basic Microbiol. 1–7.

Ivanovsky, R.N., Krasilnikova, E.N., and Fal, Y.I. (1993). A pathway of the autotrophic CO_2 fixation in *Chloroflexus aurantiacus*. Arch. Microbiol. *3*, 257–264.

Iwaki, T., Haranoh, K., Inoue, N., Kojima, K., Satoh, R., Nishino, T., Wada, S., Ihara, H., Tsuyama, S., Kobayashi, H., et al. (2006). Expression of foreign type I ribulose-1,5-bisphosphate carboxylase/ oxygenase (EC 4.1.1.39) stimulates photosynthesis in cyanobacterium *Synechococcus* PCC7942 cells. Photosynth. Res. *88*, 287–297.

Jain, S., Caforio, A., and Driessen, A.J.M. (2014). Biosynthesis of archaeal membrane ether lipids. Front. Microbiol. *5*, 641.

Jang, H.-J., Yoon, S.-H., Ryu, H.-K., Kim, J.-H., Wang, C.-L., Kim, J.-Y., Oh, D.-K., and Kim, S.-W. (2011). Retinoid production using metabolically engineered *Escherichia coli* with a two-phase culture system. Microb. Cell Fact. *10*.

Janke, C., Scholz, F., Becker-Baldus, J., Glaubitz, C., Wood, P.G., Bamberg, E., Wachtveitl, J., and Bamann, C. (2013). Photocycle and vectorial proton transfer in a rhodopsin from the eukaryote *Oxyrrhis marina*. Biochemistry *52*, 2750–2763.

Jaschke, P.R., Saer, R.G., Noll, S., and Beatty, J.T. (2011). Modification of the Genome of *Rhodobacter sphaeroides* and Construction of Synthetic Operons. In Methods in Enzymology, (Elsevier Inc.), pp. 519–538.

Jensen, H.M., Albers, A.E., Malley, K.R., Londer, Y.Y., Cohen, B.E., Helms, B.A., Weigele, P., Groves, J.T., and Ajo-Franklin, C.M. (2010). Engineering of a synthetic electron conduit in living cells. Proc. Natl. Acad. Sci. U. S. A. *107*, 19213–19218.

Jeschek, M., Gerngross, D., and Panke, S. (2016). Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort. Nat. Commun. *7*, 11163.

Johnson, E.T., and Schmidt-Dannert, C. (2008). Light-energy conversion in engineered microorganisms. Trends Biotechnol. *26*, 682–689.

Johnson, E.T., Baron, D.B., Naranjo, B., Bond, D.R., Schmidt-Dannert, C., and Gralnick, J.A. (2010). Enhancement of survival and electricity production in an engineered bacterium by light-driven proton pumping. Appl. Environ. Microbiol. *76*, 4123–4129.

Kalyuzhnaya, M.G., Puri, A.W., and Lidstrom, M.E. (2015). Metabolic engineering in methanotrophic bacteria. Metab. Eng. *29*, 142–152.

Kamennaya, N.A., Ahn, S., Park, H., Bartal, R., Sasaki, K.A., Holman, H.Y., and Jansson, C. (2015). Installing extra bicarbonate transporters in the cyanobacterium *Synechocystis* sp. PCC6803 enhances biomass production. Metab. Eng. *29*, 76–85.

Kamo, N., Hashiba, T., Kikukawa, T., Araiso, T., Ihara, K., and Nara, T. (2006). A light-driven proton pump from *Haloterrigena turkmenica*: Functional expression in *Escherichia coli* membrane and coupling with a H^+ co-transporter. Biochem. Biophys. Res. Commun. *341*, 285–290.

Kanaya, S., Yamada, Y., Kudo, Y., and Ikemura, T. (1999). Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene *238*, 143–155.

Karnik, S., Doi, T., Molday, R., and Khorana, H.G. (1990). Expression of the archaebacterial bacterio-opsin gene with and without signal sequences in *Escherichia coli*: the expressed proteins are located in the membrane but bind retinal poorly. Proc. Natl. Acad. Sci. U. S. A. *87*, 8955–8959.

Karnik, S.S., Nassal, M., Doi, T., Jay, E., Sgaramella, V., and Khorana, H.G. (1987). Structure-function studies on bacteriorhodopsin. II. Improved expression of the bacterio-opsin gene in *Escherichia coli*. J. Biol. Chem. *262*, 9255–9263.

Kashket, E.R. (1985). The proton motive force in bacteria: a critical assessment of methods. Annu. Rev. Microbiol. *39*, 219–242.

Kawanabe, A., Furutani, Y., Jung, K.-H., and Kandori, H. (2009). Engineering an inward proton transport from a bacterial sensor rhodopsin. J. Am. Chem. Soc. *131*, 16439–16444.

Kebeish, R., Niessen, M., Thiruveedhi, K., Bari, R., Hirsch, H.-J., Rosenkranz, R., Stäbler, N., Schönfeld, B., Kreuzaler, F., and Peterhänsel, C. (2007). Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in *Arabidopsis thaliana*. Nat. Biotechnol. *25*, 593–599.

Keller, M.W., Schut, G.J., Lipscomb, G.L., Menon, A.L., Iwuchukwu, I.J., Leuko, T.T., Thorgersen, M.P., Nixon, W.J., Hawkins, A.S., Kelly, R.M., et al. (2013). Exploiting microbial hyperthermophilicity to produce an industrial chemical, using hydrogen and carbon dioxide. Proc. Natl. Acad. Sci. U. S. A. *110*, 5840–5845.

Kelly, C.L., Liu, Z., Yoshihara, A., Jenkinson, S.F., Wormald, M.R., Otero, J.M., Estévez, A.M., Kato, A., Marqvorsen, M.H.S., Fleet, G.W.J., et al. (2016). Synthetic chemical inducers and genetic decoupling enable orthogonal control of the rhaBAD promoter. ACS Synth. Biol. acssynbio.6b00030.

Keniya, M., Holmes, A., M Niimi, E.L., Gillet, J., Gottesman, M., and Cannon, R. (2014). Drug resistance is conferred on the model yeast *Saccharomyces cerevisiae* 1 by expression of the melanoma-associated human ABC transporter ABCB5. Mol. Pharm. *11*, 3452–3452.

Kerfeld, C.A. (2015). Plug-and-play for improving primary productivity. Am. J. Bot. *102*, 1949–1950.

Kernan, T., Majumdar, S., Li, X., Guan, J., West, A.C., and Banta, S. (2016). Engineering the iron-oxidizing chemolithoautotroph *Acidithiobacillus ferrooxidans* for biochemical production. Biotechnol. Bioeng. *113*, 189–197.

Khade, P.K., and Joseph, S. (2011). Messenger RNA interactions in the decoding center control the rate of translocation. Nat. Struct. Mol. Biol. *18*, 1300–1302.

Khaengraeng, R., and Reed, R.H. (2005). Oxygen and photoinactivation of *Escherichia coli* in UVA and sunlight. J. Appl. Microbiol. *99*, 39–50.

Khan, N.E., Myers, J.A., Tuerk, A.L., and Curtis, W.R. (2014). A process economic assessment of hydrocarbon biofuels production using chemoautotrophic organisms. Bioresour. Technol. *172*, 201–211.

Kim, H.S., Ernst, J. a, Brown, C., Bostrom, J., Fuh, G., Lee, C. V, Huang, A., Vandlen, R.L., and Yansura, D.G. (2012a). Translation levels control multi-spanning membrane protein expression. PLoS One *7*, e35844.

Kim, J., Jo, B.H., Jo, Y., and Cha, H.J. (2012b). Improved production of biohydrogen in light-powered *Escherichia coli* by co-expression of proteorhodopsin and heterologous hydrogenase. Microb. Cell Fact.

Kim, S.J., Yoon, J.S., Shishido, H., Yang, Z., Rooney, L.A., Barral, J.M., and Skach, W.R. (2015). Translational tuning optimizes nascent protein folding in cells. Science *348*, 444–448.

Kim, S.Y., Waschuk, S.A., Brown, L.S., and Jung, K.-H. (2008). Screening and characterization of proteorhodopsin color-tuning mutations in *Escherichia coli* with endogenous retinal synthesis. Biochim. Biophys. Acta, Bioenerg. *1777*, 504–513.

Kim, Y.-S., Kim, N.-H., Yeom, S.-J., Kim, S.-W., and Oh, D.-K. (2009). In Vitro Characterization of a Recombinant Blh Protein from an Uncultured Marine Bacterium as a β-Carotene 15,15'-Dioxygenase. J. Biol. Chem. *284*, 15781–15793.

Kimchi-Sarfaty, C., Oh, J.M., Kim, I.W., Sauna, Z.E., Calcagno, A.M., Ambudkar, S. V, and Gottesman, M.M. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science *315*, 525–528.

Kirchman, D.L., and Hanson, T.E. (2013). Bioenergetics of photoheterotrophic bacteria in the oceans. Environ. Microbiol. Rep. *5*, 188–199.

Kirst, H., Formighieri, C., and Melis, A. (2014). Maximizing photosynthetic efficiency and culture productivity in cyanobacteria upon minimizing the phycobilisome light-harvesting antenna size. Biochim. Biophys. Acta - Bioenerg. *1837*, 1653–1664.

Klare, J.P.P., Chizhov, I., and Engelhard, M. (2008). Microbial rhodopsins: scaffolds for ion pumps, channels, and sensors. In Bioenergetics, G. Schäfer, and H. Penefsky, eds. (Springer Berlin Heidelberg), pp. 73-122.

Klyszejko, A.L., Shastri, S., Mari, S.A., Grubmüller, H., Muller, D.J., and Glaubitz, C. (2008). Folding and assembly of proteorhodopsin. J. Mol. Biol. *376*, 35–41.

Knight, R.D., Freeland, S.J., and Landweber, L.F. (2001). A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. Genome Biol *2*, 10.

Komar, A.A. (2009). A pause for thought along the co-translational folding pathway. Trends Biochem. Sci. *34*, 16–24.

Konneke, M., Schubert, D.M., Brown, P.C., Hugler, M., Standfest, S., Schwander, T., Schada von Borzyskowski, L., Erb, T.J., Stahl, D.A., and Berg, I.A. (2014). Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO_2 fixation. Proc. Natl. Acad. Sci. *111*, 8239–8244.

Köpke, M., Held, C., Hujer, S., Liesegang, H., Wiezer, A., Wollherr, A., Ehrenreich, A., Liebl, W., Gottschalk, G., Dürre, P., et al. (2010). *Clostridium ljungdahlii* represents a microbial production platform based on syngas. Proc. Natl. Acad. Sci. U. S. A. *107*, 13087–13092.

Kosuri, S., and Church, G.M. (2014). Large-scale de novo DNA synthesis: technologies and applications. Nat. Methods *11*, 499–507.

Kosuri, S., Goodman, D.B., Cambray, G., Mutalik, V.K., Gao, Y., Arkin, A.P., Endy, D., and Church, G.M. (2013). Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. *110*, 14024–14029.

Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene *361*, 13–37.

Kromdijk, J., Głowacka, K., Leonelli, L., Gabilly, S.T., Iwai, M., Niyogi, K.K., Long, S.P., Jordan, A.M., Yoon, C.H., Jung, K.W., et al. (2016). Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. Science *354*, 857–861.

Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009a). Coding-sequence determinants of gene expression in *Escherichia coli*. Science *324*, 255–258.

Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009b). Coding-sequence determinants of gene expression in *Escherichia coli*. Science 255–258.

Kuniyoshi, T.M., Severino, D., Balan, A., Schenberg, A.C.G., and Hallenbeck, P.C. (2015). Heterologous expression of proteorhodopsin enhances H_2 production in *Escherichia coli* when endogenous Hyd-4 is overexpressed. J. Biotechnol. *206*, 52–57.

Lajoie, M.J., Kosuri, S., Mosberg, J.A., Gregg, C.J., Zhang, D., and Church, G.M. (2013). Probing the limits of genetic recoding in essential genes. Science *342*, 361–363.

Lange, B.M., Rujan, T., Martin, W., and Croteau, R. (2000). Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. Proc. Natl. Acad. Sci. U. S. A. *97*, 13172–13177.

Larkum, A.W.D., Ross, I.L., Kruse, O., and Hankamer, B. (2012). Selection, breeding and engineering of microalgae for bioenergy and biofuel production. Trends Biotechnol. *30*, 198–205.

Latif, H., Zeidan, A.A., Nielsen, A.T., and Zengler, K. (2014). Trash to treasure: production of biofuels and commodity chemicals via syngas fermenting microorganisms. Curr. Opin. Biotechnol. *27*, 79–87.

Leal, M.R.L. V., Walter, A.S., and Seabra, J.E.A. (2012). Sugarcane as an energy source. Biomass Convers. Biorefinery *3*, 17–26.

Lee, K.A., and Jung, K.H. (2011). ATP regeneration system using *E. coli* ATP synthase and gloeobacter rhodopsin and its stability. J. Nanosci. Nanotechnol. *11*, 4261–4264.

Lee, C., Kang, H.J., Hjelm, A., Qureshi, A.A., Nji, E., Choudhury, H., Beis, K., de Gier, J.-W., and Drew, D. (2014). MemStar: a one-shot *Escherichia coli*-based approach for high-level bacterial membrane protein production. FEBS Lett. *588*, 3761–3769.

Lee, T.S., Krupa, R. a, Zhang, F., Hajimorad, M., Holtz, W.J., Prasad, N., Lee, S.K., and Keasling, J.D. (2011). BglBrick vectors and datasheets: A synthetic biology platform for gene expression. J. Biol. Eng. *5*, 12.

Levin-Karp, A., Barenholz, U., Bareia, T., Dayagi, M., Zelcbuch, L., Antonovsky, N., Noor, E., and Milo, R. (2013). Quantifying translational coupling in *E. coli* synthetic operons using RBS modulation and fluorescent reporters. ACS Synth. Biol. *2*, 327–336.

Li, G.-W.W., Oh, E., and Weissman, J.S. (2012a). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature *484*, 538–541.

Li, G.W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell *157*, 624–635.

Li, H., Opgenorth, P.H., Wernick, D.G., Rogers, S., Wu, T.-Y., Higashide, W., Malati, P., Huo, Y.-X., Cho, K.M., and Liao, J.C. (2012b). Integrated Electromicrobial Conversion of CO_2 to Higher Alcohols. Science *335*, 1596–1596.

Liang, F., and Lindblad, P. (2016). Effects of overexpressing photosynthetic carbon flux control enzymes in the cyanobacterium *Synechocystis* PCC 6803. Metab. Eng. *38*, 56–64.

Liao, J.C., Mi, L., Pontrelli, S., and Luo, S. (2016). Fuelling the future: microbial engineering for the production of sustainable biofuels. Nat. Rev. Microbiol. *14*, 288–304.

Liebermeister, W., Noor, E., Flamholz, A., Davidi, D., Bernhardt, J., and Milo, R. (2014). Visual account of protein investment in cellular functions. Proc. Natl. Acad. Sci. U. S. A. *111*, 8488–8493.

Liebeton, K., Lengefeld, J., and Eck, J. (2014). The nucleotide composition of the spacer sequence influences the expression yield of heterologously expressed genes in *Bacillus subtilis*. J. Biotechnol. *191*, 214–220.

Lintig, J. Von, Vogt, K., and von Lintig, J. (2000). Filling the gap in vitamin A research. J. Biol. Chem. *275*, 11915–11920.

Liu, L.-N. (2016). Distribution and dynamics of electron transport complexes in cyanobacterial thylakoid membranes. Biochim. Biophys. Acta - Bioenerg. *1857*, 256–265.

Liu, C.L., and Mortenson, L.E. (1984). Formate Dehydrogenase of *Clostridium pasteurianum*. J. Bacteriol. *159*, 375–380.

Liu, R.S.H., and Asato, A.E. (2003). Tuning the color and excited state properties of the azulenic chromophore: NIR absorbing pigments and materials. J. Photochem. Photobiol. C Photochem. Rev. *4*, 179–194.

Liu, C., Gallagher, J.J., Sakimoto, K.K., Nichols, E.M., Chang, C.J., Chang, M.C.Y., and Yang, P. (2015). Nanowire–Bacteria Hybrids for Unassisted Solar Carbon Dioxide Fixation to Value-Added Chemicals. Nano Lett. *15*, 3634–3639.

Liu, C., Colon, B.C., Ziesack, M., Silver, P.A., Nocera, D.G., Colón, B.C., Ziesack, M., Silver, P.A., Nocera, D.G., Colon, B.C., et al. (2016). Water splitting-biosynthetic system with CO_2 reduction efficiencies exceeding photosynthesis. Science *352*, 1210–1213.

Liu, J.K., O'Brien, E.J., Lerman, J. a, Zengler, K., Palsson, B.O., and Feist, A.M. (2014). Reconstruction and modeling protein translocation and compartmentalization in *Escherichia coli* at the genome-scale. BMC Syst. Biol. *8*, 110.

Long, S.P., Marshall-Colon, A., and Zhu, X.G. (2015). Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. Cell *161*, 56–66.

Lovley, D.R., and Nevin, K.P. (2013). Electrobiocommodities: powering microbial production of fuels and commodity chemicals from carbon dioxide with electricity. Curr. Opin. Biotechnol. *24*, 385–390.

Lu, Q., and Jiao, F. (2016). Electrochemical CO_2 reduction: Electrocatalyst, reaction mechanism, and process engineering. Nano Energy.

Lu, J., Brigham, C.J., Gai, C.S., and Sinskey, A.J. (2012). Studies on the production of branched-chain alcohols in engineered *Ralstonia eutropha*. Appl. Microbiol. Biotechnol. *96*, 283–297.

Lundstrom, K. (2007). Structural genomics and drug discovery. J. Cell. Mol. Med. *11*, 224–238.

Mackenzie, T.D.B.B., Burns, R. a., and Campbell, D. a. (2004). Carbon status constrains light acclimation in the cyanobacterium *Synechococcus elongatus*. Plant Physiol. *136*, 3301–3312.

Maertens, B., Spriestersbach, A., von Groll, U., Roth, U., Kubicek, J., Gerrits, M., Graf, M., Liss, M., Daubert, D., Wagner, R., et al. (2010). Gene optimization mechanisms: a multi-gene study reveals a high success rate of full-length human proteins expressed in *Escherichia coli*. Protein Sci *19*, 1312–1326.

Mahadevan, R., and Schilling, C.H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng. *5*, 264–276.

Mainguet, S.E., Gronenberg, L.S., Wong, S.S., and Liao, J.C. (2013). A reverse glyoxylate shunt to build a non-native route from C4 to C2 in *Escherichia coli*. Metab. Eng. *16*, 116–127.

Makoff, A.J., and Smallwood, A.E. (1990). The use of two-cistron constructions in improving the expression of a heterologous gene in *E.coli*. Nucleic Acids Res. *18*, 1711–1718.

Malvankar, N.S., and Lovley, D.R. (2014). Microbial nanowires for bioenergy applications. Curr. Opin. Biotechnol. *27*, 88–95.

Malyshev, D.A., Dhami, K., Lavergne, T., Chen, T., Dai, N., Foster, J.M., Correa Jr., I.R., and Romesberg, F.E. (2014). A semi-synthetic organism with an expanded genetic alphabet. Nature *509*, 385–388.

de Marco, A. (2013). Recombinant polypeptide production in *E. coli*: towards a rational approach to improve the yields of functional proteins. Microb. Cell Fact. *12*, 101.

Marino, J., Hohl, M., Seeger, M. a, Zerbe, O., and Geertsma, E.R. (2015). Bicistronic mRNAs to Enhance Membrane Protein Overexpression. J. Mol. Biol. *427*, 943–954.

Martinez, A., Bradley, A.S., Waldbauer, J.R., Summons, R.E., and DeLong, E.F. (2007). Proteorhodopsin photosystem gene expression enables photophosphorylation in a heterologous host. Proc. Natl. Acad. Sci. U. S. A. *104*, 5590–5595.

Mattis, J., Tye, K.M., Ferenczi, E. a, Ramakrishnan, C., O'Shea, D.J., Prakash, R., Gunaydin, L. a, Hyun, M., Fenno, L.E., Gradinaru, V., et al. (2012). Principles for applying optogenetic tools derived from direct comparative analysis of microbial opsins. Nat. Methods *9*, 159–172.

Mattozzi, M.D., Ziesack, M., Voges, M.J., Silver, P.A., and Way, J.C. (2013). Expression of the sub-pathways of the *Chloroflexus aurantiacus* 3-hydroxypropionate carbon fixation bicycle in *E. coli*: Toward horizontal transfer of autotrophic growth. Metab. Eng. *16*, 130–139.

Maurino, V.G., and Weber, A.P.M. (2013). Engineering photosynthesis in plants and synthetic microorganisms. J. Exp. Bot. *64*, 743–751.

McCarren, J., and DeLong, E.F. (2007). Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. Environ. Microbiol. *9*, 846–858.

McEwen, J.T., Kanno, M., and Atsumi, S. (2016). 2,3 Butanediol production in an obligate photoautotrophic cyanobacterium in dark conditions via diverse sugar consumption. Metab. Eng. *36*, 28–36.

Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2009). BioNumbers The database of key numbers in molecular and cell biology. Nucleic Acids Res. *38*, 750–753.

Miranda, M.R.M., Choi, A.R., Shi, L., Bezerra, A.G., Jung, K.-H., and Brown, L.S. (2009). The photocycle and proton translocation pathway in a cyanobacterial ion-pumping rhodopsin. Biophys. J. *96*, 1471–1481.

Miroux, B., and Walker, J.E. (1996). Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J. Mol. Biol. *260*, 289–298.

Mirzadeh, K., Martinez, V., Toddo, S., Guntur, S., Herrgard, M., Elofsson, A., Nørholm, M.H.H., and Daley, D.O. (2015). Enhanced protein production in *Escherichia coli* by optimization of cloning scars at the vector:coding sequence junction. ACS Synth. Biol. *4*, 959–965.

Misawa, N., Nakagawa, M., Kobayashi, K., Yamano, S., Izawa, Y., and Harashima, K. (1990). Elucidation of the *Erwinia uredovora* Carotenoid Biosynthetic Pathway by Functional Analysis of Gene Products Expressed in *Escherichia coli*. *172*, 6704–6712.

Mogi, T., Stern, L.J., Marti, T., Chao, B.H., and Khorana, H.G. (1988). Aspartic acid substitutions affect proton translocation by bacteriorhodopsin. Proc. Natl. Acad. Sci. U. S. A. *85*, 4148–4152.

Molitor, B., Richter, H., Martin, M.E., Jensen, R.O., Juminaga, A., Mihalcea, C., and Angenent, L.T. (2016). Carbon recovery by fermentation of CO-rich off gases – turning steel mills into biorefineries. Bioresour. Technol. *215*, 386–396.

Mueller-Cajar, O., and Whitney, S.M. (2008). Evolving improved *Synechococcus* Rubisco functional expression in *Escherichia coli*. Biochem. J. *414*, 205–214.

Mueller-Cajar, O., Morell, M., and Whitney, S.M. (2007). Directed evolution of Rubisco in *Escherichia coli* reveals a specificity-determining hydrogen bond in the form II enzyme. Biochemistry *46*, 14067–14074.

Mukai, T., Hoshi, H., Ohtake, K., Takahashi, M., Yamaguchi, A., Hayashi, A., Yokoyama, S., and Sakamoto, K. (2015). Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon. Sci. Rep. *5*, 9699.

Müller, J.E.N., Meyer, F., Litsanov, B., Kiefer, P., Potthoff, E., Heux, S., Quax, W.J., Wendisch, V.F., Brautaset, T., Portais, J.-C., et al. (2015). Engineering *Escherichia coli* for methanol conversion. Metab. Eng. *28*, 190–201.

Mutalik, V.K., Guimaraes, J.C., Cambray, G., Mai, Q.-A., Christoffersen, M.J., Martin, L., Yu, A., Lam, C., Rodriguez, C., Bennett, G., et al. (2013a). Quantitative estimation of activity and quality for collections of functional genetic elements. Nat. Methods *10*, 347–353.

Mutalik, V.K., Guimaraes, J.C., Cambray, G., Lam, C., Christoffersen, M.J., Mai, Q.-A., Tran, A.B., Paull, M., Keasling, J.D., Arkin, A.P., et al. (2013b). Precise and reliable gene expression via standard transcription and translation initiation elements. Nat. Methods *10*, 354–360.

Na, Y.-A., Lee, J.-Y., Bang, W.-J., Lee, H.J., Choi, S.-I., Kwon, S.-K., Jung, K.-H., Kim, J.F., and Kim, P. (2015). Growth retardation of *Escherichia coli* by artificial increase of intracellular ATP. J. Ind. Microbiol. Biotechnol. *42*.

Nakahigashi, K., Toya, Y., Ishii, N., Soga, T., Hasegawa, M., Watanabe, H., Takai, Y., Honma, M., Mori, H., and Tomita, M. (2009). Systematic phenome analysis of *Escherichia coli* multiple-knockout mutants reveals hidden reactions in central carbon metabolism. Mol. Syst. Biol. *5*, 1–14.

Nakamura, Y., Kaneko, T., Sato, S., Mimuro, M., Miyashita, H., Tsuchiya, T., Sasamoto, S., Watanabe, A., Kawashima, K., Kishida, Y., et al. (2003). Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. DNA Res. *10*, 137–145.

Nakayama, T., Yonekura, S., and Yonei, S. (2013). Escherichia coli pyruvate : flavodoxin oxidoreductase, YdbK - regulation of expression and biological roles in protection against oxidative stress. Genes Genet. Syst. *88*, 175–188.

Nissley, D.A., Sharma, A.K., Ahmed, N., Friedrich, U., Kramer, G., Bukau, B., and O'Brien, E.P. (2016). Accurate prediction of cellular co-translational folding indicates proteins can switch from post- to co-translational folding. Nat. Commun. *7*, 10341 |.

Nogales, J., Gudmundsson, S., Knight, E.M., Palsson, B.O., and Thiele, I. (2013). Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. Proc. Natl. Acad. Sci. U. S. A. *109*, 2678–2683.

Noor, E., Bar-Even, A., Flamholz, A., Lubling, Y., Davidi, D., and Milo, R. (2012). An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. Bioinformatics *28*, 2037–2044.

Noor, E., Haraldsdóttir, H.S., Milo, R., and Fleming, R.M.T. (2013). Consistent estimation of Gibbs energy using component contributions. PLoS Comput. Biol. *9*, e1003098.

Noor, E., Bar-Even, A., Flamholz, A., Reznik, E., Liebermeister, W., and Milo, R. (2014). Pathway thermodynamics highlights kinetic obstacles in central metabolism. PLoS Comput. Biol. *10*.

Nørholm, M.H.H., Light, S., Virkki, M.T.I., Elofsson, A., von Heijne, G., and Daley, D.O. (2012). Manipulating the genetic code for membrane protein production: what have we learnt so far? Biochim. Biophys. Acta *1818*, 1091–1096.

Nørholm, M.H.H., Toddo, S., Virkki, M.T.I., Light, S., von Heijne, G., and Daley, D.O. (2013). Improved production of membrane proteins in Escherichia coli by selective codon substitutions. FEBS Lett. *587*, 2352–2358.

Novoa, E.M., Pavon-Eternod, M., Pan, T., and Ribas de Pouplana, L. (2012). A role for tRNA modifications in genome structure and codon usage. Cell *149*, 202–213.

Nybo, S.E., Khan, N.E., Woolston, B.M., and Curtis, W.R. (2015). Metabolic engineering in chemolithoautotrophic hosts for the production of fuels and chemicals. Metab. Eng. *30*, 105–120.

Oesterhelt, D., and Stoeckenius, W. (1971). Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. Nature *233*, 149–152.

Oliver, J.W.K., Machado, I.M.P., Yoneda, H., and Atsumi, S. (2013). Cyanobacterial conversion of carbon dioxide to 2,3-butanediol. Proc. Natl. Acad. Sci. U. S. A. *110*, 1249–1254.

Ort, D.R., Merchant, S.S., Alric, J., Barkan, A., Blankenship, R.E., Bock, R., Croce, R., Hanson, M.R., Hibberd, J.M., Long, S.P., et al. (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. Proc. Natl. Acad. Sci. *112*, 8529–8536.

Orth, J.D., Fleming, R.M.T., and Palsson, B.Ø. (2009). EcoSal-*Escherichia coli* and *Salmonella*: cellular and molecular biology (ASM Press, Washington DC).

Orth, J.D., Thiele, I., and Palsson, B.Ø. (2010). What is flux balance analysis? Nat. Biotechnol. *28*, 245–248.

Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., and Palsson, B.Ø. (2011). A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism. Mol. Syst. Biol.

Ortiz-Marquez, J.C.F., Nascimento, M. Do, Angeles Dublan, M. de los, and Curatti, L. (2012). Association with an ammonium-excreting bacterium allows diazotrophic culture of oil-rich eukaryotic microalgae. Appl. Environ. Microbiol. *78*, 2345–2352.

Pandit, A. V, and Mahadevan, R. (2011). In silico characterization of microbial electrosynthesis for metabolic engineering of biochemicals. Microb. Cell Fact. *10*, 1–14.

Parikh, M.R., Greene, D.N., Woods, K.K., and Matsumura, I. (2006). Directed evolution of RuBisCO hypermorphs through genetic selection in engineered *E. coli*. Protein Eng. Des. Sel. *19*, 113–119.

Pechmann, S., and Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nat. Struct. Mol. Biol. *20*, 237–243.

Pechmann, S., Chartron, J.W., and Frydman, J. (2014). Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. Nat. Struct. Mol. Biol. *21*, 1–9.

Peck, R.F., Echavarri-Erasun, C., Johnson, E.A., Ng, W. V, Kennedy, S.P., Hood, L., DasSarma, S., and Krebs, M.P. (2001). brp and blh are required for synthesis of the retinal cofactor of bacteriorhodopsin in *Halobacterium salinarum*. J. Biol. Chem. *276*, 5739–5744.

Peter, D.M., Schada von Borzyskowski, L., Kiefer, P., Christen, P., Vorholt, J. a., and Erb, T.J. (2015). Screening and Engineering the Synthetic Potential of Carboxylating Reductases from Central Metabolism and Polyketide Biosynthesis. Angew. Chemie Int. Ed. *54*, 1–6.

Pfleger, B.F., Pitera, D.J., Smolke, C.D., and Keasling, J.D. (2006). Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat. Biotechnol. *24*, 1027–1032.

Pinchuk, G.E., Hill, E.A., Geydebrekht, O. V, De Ingeniis, J., Zhang, X., Osterman, A., Scott, J.H., Reed, S.B., Romine, M.F., Konopka, A.E., et al. (2010). Constraint-based model of Shewanella oneidensis MR-1 metabolism: a tool for data analysis and hypothesis generation. PLoS Comput. Biol. *6*, e1000822.

Pop, C., Rouskin, S., Ingolia, N.T., Han, L., Phizicky, E.M., Weissman, J.S., and Koller, D. (2014). Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. Mol. Syst. Biol. *10*, 1–15.

Portnoy, V.A., Bezdan, D., and Zengler, K. (2011). Adaptive laboratory evolution - harnessing the power of biology for metabolic engineering. Curr. Opin. Biotechnol. *22*, 590–594.

Portnoy, V. a, Scott, D. a, Lewis, N.E., Tarasova, Y., Osterman, A.L., and Palsson, B.Ø. (2010). Deletion of genes encoding cytochrome oxidases and quinol monooxygenase blocks the aerobic-anaerobic shift in *Escherichia coli* K-12 MG1655. Appl. Environ. Microbiol. *76*, 6529–6540.

Purvis, I.J., Bettany, A.J., Santiago, T.C., Coggins, J.R., Duncan, K., Eason, R., and Brown, A.J. (1987). The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. J Mol Biol *193*, 413–417.

Pushkarev, A., and Béjà, O. (2016). Functional metagenomic screen reveals new and diverse microbial rhodopsins. ISME J. *10*, 2331–2335.

Qian, W., Yang, J.R., Pearson, N.M., Maclean, C., and Zhang, J. (2012). Balanced codon usage optimizes eukaryotic translational efficiency. PLoS Genet. *8*, e1002603.

Quax, T.E., Wolf, Y.I., Koehorst, J.J., Wurtzel, O., van der Oost, R., Ran, W., Blombach, F., Makarova, K.S., Brouns, S.J., Forster, A.C., et al. (2013). Differential translation tunes uneven production of operon-encoded proteins. Cell Rep *4*, 938–944.

Quax, T.E.F., Claassens, N.J., Söll, D., and van der Oost, J. (2015). Codon Bias as a Means to Fine-Tune Gene Expression. Mol. Cell *59*, 149–161.

Raab, D., Graf, M., Notka, F., Schödl, T., and Wagner, R. (2010). The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. Syst. Synth. Biol. *4*, 215–225.

Rabaey, K., and Rozendal, R.A. (2010). Microbial electrosynthesis - revisiting the electrical route for microbial production. Nat. Rev. Microbiol. *8*, 706–716.

Rabaey, K., Girguis, P., and Nielsen, L.K. (2011). Metabolic and practical considerations on microbial electrosynthesis. Curr. Opin. Biotechnol. *22*, 371–377.

Ramey, C.J., Barón-Sola, Á., Aucoin, H.R., and Boyle, N.R. (2015). Genome Engineering in Cyanobacteria: Where we are and where we need to go. ACS Synth. Biol. *4*, 1186–1196.

Ran, W., and Higgs, P.G. (2012). Contributions of speed and accuracy to translational selection in bacteria. PLoS One *7*, e51652.

Reed, J.L., Vo, T.D., Schilling, C.H., and Palsson, B.Ø. (2003). An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). Genome Biol. *4*, R54.

Reeve, B., Hargest, T., Gilbert, C., and Ellis, T. (2014). Predicting Translation Initiation Rates for Designing Synthetic Biology. Front. Bioeng. Biotechnol. *2*, 1–6.

dos Reis, M., Savva, R., and Wernisch, L. (2004). Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res *32*, 5036–5044.

Riedel, T., Gómez-Consarnau, L., Tomasch, J., Martin, M., Jarek, M., González, J.M., Spring, S., Rohlfs, M., Brinkhoff, T., Cypionka, H., et al. (2013). Genomics and Physiology of a Marine Flavobacterium Encoding a Proteorhodopsin and a Xanthorhodopsin-Like Protein. PLoS One *8*, e57487.

Rosenbaum, M.A., and Henrich, A.W. (2014). Engineering microbial electrocatalysis for chemical and fuel production. Curr. Opin. Biotechnol. *29*, 93–98.

Rousset, M., and Liebgott, P. (2014). Engineering Hydrogenases for H_2 Production: Bolts and Goals. In Microbial BioEnergy: Hydrogen Production, Advances in Photosynthesis and Respiration, pp. 43–77.

Rovner, A.J., Haimovich, A.D., Katz, S.R., Li, Z., Grome, M.W., Gassaway, B.M., Amiram, M., Patel, J.R., Gallagher, R.R., Rinehart, J., et al. (2015). Recoded organisms engineered to depend on synthetic amino acids. Nature *518*, 89–93.

Sabehi, G., Loy, A., Jung, K.-H., Partha, R., Spudich, J.L., Isaacson, T., Hirschberg, J., Wagner, M., and Béjà, O. (2005). New insights into metabolic properties of marine bacteria encoding proteorhodopsins. PLoS Biol *3*, e273.

Sakimoto, K.K., Wong, A.B., and Yang, P. (2016). Self-photosensitization of nonphotosynthetic bacteria for solar-to-chemical production. Science *351*, 74–77.

Salis, H.M. (2011). The Ribosome Binding Site Calculator. In Methods in Enzymology, (Elsevier Inc.), pp. 19–42.

Salis, H.M., Mirsky, E. a, and Voigt, C. a (2009). Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol. *27*, 946–950.

Sambrook, J., Fritsch, E., and Maniatis, T. (1989). Molecular Cloning : A laboratory manual (New York: Cold Spring Harbor Laboratory).

Sarduy, E.S., Muñoz, A.C., Trejo, S.A., and Chavéz Planes, M.D.L.A. (2012). High-level expression of Falcipain-2 in *Escherichia coli* by codon optimization and auto-induction. Protein Expr. Purif. *83*, 59–69.

Sauer, U., Canonaco, F., Heri, S., Perrenoud, A., and Fischer, E. (2004). The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. J. Biol. Chem. *279*, 6613–6619.

Saunders, R., and Deane, C.M. (2010). Synonymous codon usage influences the local protein structure observed. Nucleic Acids Res. *38*, 6719–6728.

Savakis, P., and Hellingwerf, K.J. (2015). Engineering cyanobacteria for direct biofuel production from CO*2*. Curr. Opin. Biotechnol. *33*, 8–14.

Scaife, M.A., Burja, A.M., and Wright, P.C. (2009). Characterization of cyanobacterial β-carotene ketolase and hydroxylase genes in *Escherichia coli*, and their application for astaxanthin biosynthesis. Biotechnol. Bioeng. *103*, 944–955.

Schiel-Bengelsdorf, B., and Dürre, P. (2012). Pathway engineering and synthetic biology using acetogens. FEBS Lett. *586*, 2191–2198.

Schlegel, S., Löfblom, J., Lee, C., Hjelm, A., Klepsch, M., Strous, M., Drew, D., Slotboom, D.J., and de Gier, J.-W. (2012). Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21(DE3). J. Mol. Biol. *423*, 648–659.

Schlegel, S., Hjelm, A., Baumgarten, T., Vikström, D., and de Gier, J.-W. (2014). Bacterial-based membrane protein production. Biochim. Biophys. Acta *1843*, 1739–1749.

Schlegel, S., Genevaux, P., and de Gier, J.-W. (2015). De-convoluting the Genetic Adaptations of *E. coli* C41(DE3) in Real Time Reveals How Alleviating Protein Production Stress Improves Yields. Cell Rep. *10*, 1758–1766.

Schlegel, S., Genevaux, P., and de Gier, J.-W. (2016a). Isolating *Escherichia coli* strains for recombinant protein production. Cell. Mol. Life Sci. 1–18.

Schlegel, S., Genevaux, P., and de Gier, J.-W. (2016b). Isolating Escherichia coli strains for recombinant protein production. Cell. Mol. Life Sci. 1–18.

Schmidl, S.R., Sheth, R.U., Wu, A., and Tabor, J.J. (2014). Refactoring and Optimization of Light-Switchable Escherichia coli Two-Component Systems. ACS Synth. Biol. *3*, 820–831.

Schomburg, I., Chang, A., Placzek, S., Söhngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., et al. (2013). BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic Acids Res. *41*, D764-72.

Schuchmann, K., and Müller, V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. Nat. Rev. Microbiol. *12*, 809–821.

Schwander, T., Schada von Borzyskowski, L., Burgener, S., Cortina, N.S., and Erb, T.J. (2016). A synthetic pathway for the fixation of carbon dioxide in vitro. Science *354*, 900–904.

Shah, P., Ding, Y., Niemczyk, M., Kudla, G., and Plotkin, J.B. (2013). Rate-limiting steps in yeast protein translation. Cell *153*, 1589–1601.

Shao, Z.Q., Zhang, Y.M., Feng, X.Y., Wang, B., and Chen, J.Q. (2012). Synonymous codon ordering: a subtle but prevalent strategy of bacteria to improve translational efficiency. PLoS One *7*, e33547.

Sharma, A.K., Spudich, J.L., and Doolittle, W.F. (2006). Microbial rhodopsins: functional versatility and genetic mobility. Trends Microbiol. *14*, 463–469.

Sharp, P.M., and Li, W.H. (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evol. *24*, 28–38.

Sharp, P.M., and Li, W.H. (1987). The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res *15*, 1281–1295.

Shastri, A.A., and Morgan, J.A. (2005). Flux balance analysis of photoautotrophic metabolism. Biotechnol. Prog. *21*, 1617–1626.

Shih, P.M., Zarzycki, J., Niyogi, K.K., and Kerfeld, C.A. (2014). Introduction of a synthetic CO_2-fixing photorespiratory bypass into a cyanobacterium. J. Biol. Chem. *289*, 9493–9500.

Siegel, J.B., Smith, A.L., Poust, S., Wargacki, A.J., Bar-Even, A., Louw, C., Shen, B.W., Eiben, C.B., Tran, H.M., Noor, E., et al. (2015). Computational protein design enables a novel one-carbon assimilation pathway. Proc. Natl. Acad. Sci. *112*, 3704–3709.

Siu, K., Chen, R.P., Sun, Q., Chen, L., Tsai, S., and Chen, W. (2015). Synthetic scaffolds for pathway enhancement. Curr. Opin. Biotechnol. *36*, 98–106.

Slamovits, C.H., Okamoto, N., Burri, L., James, E.R., and Keeling, P.J. (2011). A bacterial proteorhodopsin proton pump in marine eukaryotes. Nat. Commun. *2*, 1–6.

Smanski, M.J., Bhatia, S., Zhao, D., Park, Y., B A Woodruff, L., Giannoukos, G., Ciulla, D., Busby, M., Calderon, J., Nicol, R., et al. (2014). Functional optimization of gene clusters by combinatorial design and assembly. Nat. Biotechnol. *32*, 1241–1249.

Smanski, M.J., Zhou, H., Claesen, J., Shen, B., Fischbach, M.A., and Voigt, C.A. (2016). Synthetic biology to access and expand nature's chemical diversity. Nat. Rev. Microbiol. *14*, 135–149.

Smith, M.J., and Francis, M.B. (2016). A designed *A. vinelandii* - *S. elongatus* coculture for chemical photoproduction from air, water, phosphate and trace metals. ACS Synth. Biol. *5*, 955–961.

Snijder, H.J.A., and Hakulinen, J. (2016). Membrane Protein Production in *E. coli* for Applications in Drug Discovery. In Advanced Technologies for Protein Complex Production and Characterization, pp. 59–77.

Soll, D., Jones, D.S., Ohtsuka, E., Faulkner, R.D., Lohrmann, R., Hayatsu, H., and Khorana, H.G. (1966). Specificity of sRNA for recognition of codons as studied by the ribosomal binding technique. J Mol Biol *19*, 556–573.

Spencer, P.S., Siller, E., Anderson, J.F., and Barral, J.M. (2012). Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. J Mol Biol *422*, 328–335.

Stadler, M., and Fire, A. (2011). Wobble base-pairing slows in vivo translation elongation in metazoans. RNA *17*, 2063–2073.

Steindler, L., Schwalbach, M.S., Smith, D.P., Chan, F., and Giovannoni, S.J. (2011). Energy starved *Candidatus* Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. PLoS One *6*, e19725.

Stoletzki, N., and Eyre-Walker, A. (2007). Synonymous codon usage in *Escherichia coli*: Selection for translational accuracy. Mol. Biol. Evol. *24*, 374–381.

Stolzenberger, J., Lindner, S.N., Persicke, M., Brautaset, T., and Wendisch, V.F. (2013). Characterization of fructose 1,6-bisphosphatase and sedoheptulose 1,7-bisphosphatase from the facultative ribulose monophosphate cycle methylotroph *Bacillus methanolicus*. J. Bacteriol. *195*, 5112–5122.

Stone, K.M., Voska, J., Kinnebrew, M., Pavlova, A., Junk, M.J.N., and Han, S. (2013). Structural insight into proteorhodopsin oligomers. Biophys. J. *104*, 472–481.

Straub, M., Demler, M., Weuster-Botz, D., and Dürre, P. (2014). Selective enhancement of autotrophic acetate production with genetically modified *Acetobacterium woodii*. J. Biotechnol. *178*, 67–72.

Supek, F., and Smuc, T. (2010). On relevance of codon usage to expression of synthetic and natural genes in *Escherichia coli*. Genetics *185*, 1129–1134.

Sydow, A., Krieg, T., Mayer, F., Schrader, J., and Holtmann, D. (2014). Electroactive bacteria - molecular mechanisms and genetic tools. Appl. Microbiol. Biotechnol. *98*, 8481–8495.

Takyar, S., Hickerson, R.P., and Noller, H.F. (2005). mRNA helicase activity of the ribosome. Cell *120*, 49–58.

Tats, A., Tenson, T., and Remm, M. (2008). Preferred and avoided codon pairs in three domains of life. BMC Genomics *9*.

TerAvest, M.A., Zajdel, T.J., and Ajo-Franklin, C.M. (2014). The Mtr Pathway of *Shewanella oneidensis* MR-1 Couples Substrate Utilization to Current Production in *Escherichia coli*. ChemElectroChem *1*, 1874–1879.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. *25*, 4876–4882.

Tian, B., and Hua, Y. (2010). Carotenoid biosynthesis in extremophilic Deinococcus-Thermus bacteria. Trends Microbiol. *18*, 512–520.

Tikh, I., and Schmidt-Dannert, C. (2013). Towards engineered light–energy conversion in nonphotosynthetic microorganisms. In Synthetic Biology, H. Zhao, ed. (Boston: Academic Press), pp. 303–316.

Tikh, I.B., Quin, M.B., and Schmidt-Dannert, C. (2014a). A tale of two reductases: extending the bacteriochlorophyll biosynthetic pathway in *E. coli*. PLoS One *9*, e89734.

Tikh, I.B., Held, M., and Schmidt-Dannert, C. (2014b). BioBrick(TM) compatible vector system for protein expression in *Rhodobacter sphaeroides*. Appl. Microbiol. Biotechnol. *98*, 3111–3119.

Tishkov, V.I., and Popov, V.O. (2006). Protein engineering of formate dehydrogenase. Biomol. Eng. *23*, 89–110.

Torella, J.P., Gagliardi, C.J., Chen, J.S., Bediako, D.K., Colon, B., Way, J.C., Silver, P.A., and Nocera, D.G. (2015). Efficient solar-to-fuels production from a hybrid microbial-water-splitting catalyst system. Proc Natl Acad Sci U S A *112*, 2337–2342.

Tsukamoto, T., Inoue, K., Kandori, H., and Sudo, Y. (2013a). Thermal and spectroscopic characterization of a proton pumping rhodopsin from an extreme thermophile. J. Biol. Chem. *288*, 21581–21592.

Tsukamoto, T., Kikukawa, T., Kurata, T., Jung, K.-H., Kamo, N., and Demura, M. (2013b). Salt bridge in the conserved His-Asp cluster in Gloeobacter rhodopsin contributes to trimer formation. FEBS Lett. *587*, 322–327.

Tsukamoto, T., Demura, M., and Sudo, Y. (2014). Irreversible Trimer to Monomer Transition of Thermophilic Rhodopsin upon Thermal Stimulation. J. Phys. Chem. B *118*, 12383–12394.

Tsukamoto, T., Mizutani, K., Hasegawa, T., Takahashi, M., Honda, N., Hashimoto, N., Shimono, K., Yamashita, K., Yamamoto, M., Miyauchi, S., et al. (2016). X-ray Crystallographic Structure of Thermophilic Rhodopsin: Implications for High Thermal Stability and Optogenetic Function. J. Biol. Chem. *291*, 12223–12232.

Tsunoda, S.P., Ewers, D., Gazzarrini, S., Moroni, A., Gradmann, D., and Hegemann, P. (2006). H^+-pumping rhodopsin from the marine alga *Acetabularia*. Biophys. J. *91*, 1471–1479.

Tuller, T., and Zur, H. (2014). Multiple roles of the coding sequence 5′ end in gene expression regulation. Nucleic Acids Res. *43*, 13–28.

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell *141*, 344–354.

Tulloch, F., Atkinson, N.J., Evans, D.J., Ryan, M.D., and Simmonds, P. (2014). RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. eLife *3*, e04531.

Ueki, T., Nevin, K.P., Woodard, T.L., and Lovley, D.R. (2014). Converting carbon dioxide to butyrate with an engineered strain of *Clostridium ljungdahlii*. MBio *5*, 19–23.

Varaljay, V.A., Satagopan, S., North, J.A., Witte, B., Dourado, M.N., Anantharaman, K., Arbing, M.A., Hoeft McCann, S., Oremland, R.S., Banfield, J.F., et al. (2015). Functional metagenomic selection of RubisCO from uncultivated bacteria. Environ. Microbiol. *18*, 1187–1199.

Vazquez-Albacete, D., Cavaleiro, A.M., Christensen, U., Seppälä, S., Møller, B.L., and Nørholm, M.H.H. (2016). An expression tag toolbox for microbial production of membrane bound plant cytochromes P450.

Vignais, P.M., and Billoud, B. (2007). Occurrence, classification, and biological function of hydrogenases: An overview. Chem. Rev. *107*, 4206–4272.

Vollmers, J., Voget, S., Dietrich, S., Gollnow, K., Smits, M., Meyer, K., Brinkhoff, T., Simon, M., and Daniel, R. (2013). Poles apart: arctic and antarctic Octadecabacter strains share high genome plasticity and a new type of xanthorhodopsin. PLoS One *8*, e63422.

Volpers, M., Claassens, N.J., Noor, E., Oost, J. van der, Vos, W.M. de, Kengen, S.W.M., and Santos, V.A.P.M. dos (2016). Integrated *in silico* analysis of pathway designs supporting synthetic photo-electro-autotrophy. PLoS One DOI:10.1371/journal.pone.0157851.

De Vries, S.P.W., Van Hijum, S.A.F.T., Schueler, W., Riesbeck, K., Hays, J.P., Hermans, P.W.M., and Bootsma, H.J. (2010). Genome analysis of Moraxella catarrhalis strain RH4, a human respiratory tract pathogen. J. Bacteriol. *192*, 3574–3583.

Vuoristo, K.S., Mars, A.E., van Loon, S., Orsi, E., Eggink, G., Sanders, J.P.M., and Weusthuis, R. a. (2015). Heterologous expression of *Mus musculus* immunoresponsive gene 1 (irg1) in *Escherichia coli* results in itaconate production. Front. Microbiol. *6*, 1–6.

Wagner, G., and Hope, A.B. (1976). Proton transport in *Halobacterium halobium*. Funct. Plant Biol. *3*, 665–676.

Wagner, S., Klepsch, M.M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., van Wijk, K.J., Slotboom, D.J., Persson, J.O., et al. (2008). Tuning *Escherichia coli* for membrane protein overexpression. Proc. Natl. Acad. Sci. U. S. A. *105*, 14371–14376.

Wallin, E., and von Heijne, G. (1998). Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Sci. *7*, 1029–1038.

Walter, J.M., Greenfield, D., Bustamante, C., and Liphardt, J. (2007). Light-powering *Escherichia coli* with proteorhodopsin. Proc. Natl. Acad. Sci. U. S. A. *104*, 2408–2412.

Walter, J.M., Greenfield, D., and Liphardt, J. (2010). Potential of light-harvesting proton pumps for bioenergy applications. Curr. Opin. Biotechnol. *21*, 265–270.

Wan, N., Abernathy, M., Tang, J.K.-H., Tang, Y.J., and You, L. (2015). Cyanobacterial photo-driven mixotrophic metabolism and its advantages for biosynthesis. Front. Chem. Sci. Eng. *9*, 308–316.

Wang, B., Wang, J., Zhang, W., and Meldrum, D.R. (2012). Application of synthetic biology in cyanobacteria and algae. Front. Microbiol. *3*, 344.

Wang, Q., Wu, C., Chen, T., Chen, X., and Zhao, X. (2006). Expression of galactose permease and pyruvate carboxylase in *Escherichia coli* ptsG mutant increases the growth rate and succinate yield under anaerobic conditions. Biotechnol. Lett. *28*, 89–93.

Wang, S., Huang, H., Moll, J., and Thauer, R.K. (2010). NADP+ reduction with reduced ferredoxin and NADP+ reduction with NADH are coupled via an electron-bifurcating enzyme complex in *Clostridium kluyveri*. J. Bacteriol. *192*, 5115–5123.

Wang, S., Huang, H., Kahnt, J., and Thauer, R.K. (2013). *Clostridium acidurici* electron-bifurcating formate dehydrogenase. Appl. Environ. Microbiol. *79*, 6176–6179.

Wang, Y., Li, Y., Xu, T., Shi, Z., and Wu, Q. (2015). Experimental evidence for growth advantage and metabolic shift stimulated by photophosphorylation of proteorhodopsin expressed in *Escherichia coli* at anaerobic condition. Biotechnol. Bioeng. *112*, 947–956.

Waschuk, S. a, Bezerra, A.G., Shi, L., and Brown, L.S. (2005). Leptosphaeria rhodopsin: bacteriorhodopsin-like proton pump from a eukaryote. Proc. Natl. Acad. Sci. U. S. A. *102*, 6879–6883.

Way, J.C., Collins, J.J., Keasling, J.D., and Silver, P.A. (2014). Integrating biological redesign: Where synthetic biology came from and where it needs to go. Cell *157*, 151–161.

Wegener, K.M., Singh, A.K., Jacobs, J.M., Elvitigala, T., Welsh, E.A., Keren, N., Gritsenko, M.A., Ghosh, B.K., Camp, D.G., Smith, R.D., et al. (2010). Global proteomics reveal an atypical strategy for carbon/nitrogen assimilation by a cyanobacterium under diverse environmental perturbations. Mol. Cell. Proteomics *9*, 2678–2689.

Wegerer, A., Sun, T., and Altenbuchner, J. (2008). Optimization of an *E. coli* L-rhamnose-inducible expression vector: test of various genetic module combinations. BMC Biotechnol. *8*, 2.

Welch, M., Villalobos, A., Gustafsson, C., and Minshull, J. (2009a). You're one in a googol: optimizing genes for protein expression. J. R. Soc. Interface *6 Suppl 4*, S467-76.

Welch, M., Govindarajan, S., Ness, J.E., Villalobos, A., Gurney, A., Minshull, J., and Gustafsson, C. (2009b). Design parameters to control synthetic gene expression in *Escherichia coli*. PLoS One *4*, e7002.

Weusthuis, R.A., Lamot, I., van der Oost, J., and Sanders, J.P.M. (2011). Microbial production of bulk chemicals: development of anaerobic processes. Trends Biotechnol. *29*, 153–158.

Whitaker, W.B., Sandoval, N.R., Bennett, R.K., Fast, A.G., and Papoutsakis, E.T. (2015). Synthetic methylotrophy: engineering the production of biofuels and chemicals based on the biology of aerobic methanol utilization. Curr. Opin. Biotechnol. *33*, 165–175.

Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J.J., van der Oost, J., Doudna, J. a., and Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. Nature *477*, 486–489.

Wijffels, R.H., and Barbosa, M.J. (2010). An outlook on microalgal biofuels. Science *329*, 796–799.

Wijffels, R.H., Kruse, O., and Hellingwerf, K.J. (2013). Potential of industrial biotechnology with cyanobacteria and eukaryotic microalgae. Curr. Opin. Biotechnol. *24*, 405–413.

Woolstenhulme, C.J., Parajuli, S., Healey, D.W., Valverde, D.P., Petersen, E.N., Starosta, A.L., Guydosh, N.R., Johnson, W.E., Wilson, D.N., and Buskirk, A.R. (2013). Nascent peptides that block protein synthesis in bacteria. Proc Natl Acad Sci U S A *110*, E878-87.

Work, V.H., D'Adamo, S., Radakovits, R., Jinkerson, R.E., and Posewitz, M.C. (2012). Improving photosynthesis and metabolic networks for the competitive production of phototroph-derived biofuels. Curr. Opin. Biotechnol. *23*, 290–297.

Xu, Y., Ma, P., Shah, P., Rokas, A., Liu, Y., and Johnson, C.H. (2013). Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. Nature *495*, 116–120.

Yamamoto, M., Arai, H., Ishii, M., and Igarashi, Y. (2003). Characterization of two different 2-oxoglutarate:ferredoxin oxidoreductases from *Hydrogenobacter thermophilus* TK-6. Biochem. Biophys. Res. Commun. *312*, 1297–1302.

Ye, V.M., and Bhatia, S.K. (2012). Pathway engineering strategies for production of beneficial carotenoids in microbial hosts. *34*, 1405–1414.

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J.D., Osterhout, R.E., Stephen, R., et al. (2011). Metabolic engineering of *Escherichia coli* for direct production of 1, 4-butanediol. Nat. Chem. Biol. *7*, 445–452.

Yishai, O., Lindner, S.N., Gonzalez de la Cruz, J., Tenenboim, H., and Bar-Even, A. (2016). The formate bio-economy. Curr. Opin. Chem. Biol. *35*, 1–9.

Yoshitsugu, M., Yamada, J., and Kandori, H. (2009). Color-changing mutation in the E-F loop of proteorhodopsin. Biochemistry *48*, 4324–4330.

Yoshizawa, S., Kawanabe, A., Ito, H., Kandori, H., and Kogure, K. (2012). Diversity and functional analysis of proteorhodopsin in marine Flavobacteria. Environ. Microbiol. *14*, 1240–1248.

Young, C.L., Britton, Z.T., and Robinson, A.S. (2012). Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. Biotechnol. J. *7*, 620–634.

Zhang, F., Vierock, J., Yizhar, O., Fenno, L.E., Tsunoda, S., Kianianmomeni, A., Prigge, M., Berndt, A., Cushman, J., Polle, J., et al. (2011). The microbial opsin family of optogenetic tools. Cell *147*, 1446–1457.

Zhang, G., Hubalewska, M., and Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat. Struct. Mol. Biol. *16*, 274–280.

Zhang, Y.M., Shao, Z.Q., Yang, L.T., Sun, X.Q., Mao, Y.F., Chen, J.Q., and Wang, B. (2013). Non-random arrangement of synonymous codons in archaea coding sequences. Genomics *101*, 362–367.

Zhang, Z., Kuipers, G., Niemiec, Ł., Baumgarten, T., Slotboom, D.J., de Gier, J.-W., and Hjelm, A. (2015). High-level production of membrane proteins in *E. coli* BL21(DE3) by omitting the inducer IPTG. Microb. Cell Fact. *14*, 142.

Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J.M., Sachs, M.S., and Liu, Y. (2013). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature *495*, 111–115.

Zhou, P., Xu, X.-W., Wu, M., Huang, W.-D., and Oren, A. (2009a). Isolation and functional expression of the bop gene from *Halobiforma lacisalsi*. Microbiol. Res. *164*, 553–559.

Zhou, T., Weems, M., and Wilke, C.O. (2009b). Translationally optimal codons associate with structurally sensitive sites in proteins. Mol. Biol. Evol. *26*, 1571–1580.

Zhu, X.-G., Long, S.P., and Ort, D.R. (2008). What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? Curr. Opin. Biotechnol. *19*, 153–159.

Zhu, X.G., Long, S.P., and Ort, D.R. (2010). Improving photosynthetic efficiency for greater yield. Annu Rev Plant Biol *61*, 235–261.

Zhuang, Z.-Y., and Li, S.-Y. (2013). Rubisco-based engineered *Escherichia coli* for in situ carbon dioxide recycling. Bioresour. Technol. *150*, 79–88.

Zobel, S., Benedetti, I., Eisenbach, L., de Lorenzo, V., Wierckx, N., and Blank, L.M. (2015). A Tn7-based device for calibrated heterologous gene expression in *Pseudomonas putida*. ACS Synth. Biol. 150702003749009.

Van Zyl, L.J., Taylor, M.P., Eley, K., Tuffin, M., and Cowan, D.A. (2014). Engineering pyruvate decarboxylase-mediated ethanol production in the thermophilic host *Geobacillus thermoglucosidasius*. Appl. Microbiol. Biotechnol. *98*, 1247–1259.

Nederlandse samenvatting

Microbiële en plant fotosynthese is een interessant proces om toe te passen voor biologische productie van chemicaliën en brandstoffen, aan de andere kant hebben fotosynthetische processen een vrij lage efficiëntie. Vandaar is het bestuderen van fotosynthetische processen en het aanpassen van fotosystemen van groot belang. Het verkennen van het aanpassen van fotosystemen in micro-organismen is het overkoepelende doel van dit proefschrift. Zoals beschreven in **Hoofdstuk 1**, kunnen natuurlijke fotosystemen worden aangepast of getransplanteerd ten behoeve van efficiëntere omzetting van zonlicht in biochemische energievormen. Voor het herontwerpen van biologische fotosynthese zijn ambitieuze voorstellen gedaan, om deze voorstellen te realiseren zijn zowel synthetische biologie en systeembiologie belangrijke vakgebieden. Om te verkennen hoe deze vakgebieden kunnen worden toegepast hiervoor, concentreert het werk in dit proefschrift op de transplantatie van simpele fotosystemen (proton-pompende rhodopsines; PPR's) in het celmembraan van de niet-fotosynthetische, heterotrofe modelbacterie *Escherichia coli*. Hiervoor hebben we *in silico* analyses uitgevoerd, bijvoorbeeld met metabole en thermodynamische modellen (Hoofdstuk 3), en hebben we daarnaast verschillende experimentele studies uitgevoerd (Hoofdstukken 4,6 en 7) voor de transplantatie van die PPR fotosystemen en ook andere membraaneiwitten. Deze studies hebben verschillende uitdagingen, beperkingen maar ook allerlei oplossingen aan het licht gebracht.

Hoofdstuk 2 beschrijft details over eerder werk rondom het transplanteren van PPR fotosystemen in verschillende organismen, en de fysiologische impact van deze systemen op de organismes. Gebaseerd op deze evaluatie worden enkele suggesties gedaan om de transplantatie van PPR fotosystemen verder te verbeteren en te benutten. De relatieve lage capaciteit van PPR's om protonen te pompen, en ook de relatief beperkte absorptie van het lichtspectrum worden geïdentificeerd als belangrijke beperkingen voor het toepassen van PPR's als energie-genererende fotosystemen. Opties worden besproken om de fysiologische capaciteit van PPR fotosystemen te verbeteren, zoals het verhogen van het aantal PPR systemen in het celmembraan, het aanpassen of selecteren van PPR's die sneller protonen kunnen pompen, of het werken met PPR systemen die een breder deel van het lichtspectrum kunnen benutten, zoals dubbel-pigment PPR systemen.

In **Hoofdstuk 3** wordt een systematische, geïntegreerde analyse gemaakt voor anaeroob, foto-elektro-autotroof synthetisch metabolisme in *E. coli*, bestaande uit (i) een PPR fotosysteem voor ATP regeneratie, (ii) een elektronenopname route, en (iii) een natuurlijke of synthetische CO_2-fixatie route. Zogenaamd 'constraint-based' metabolisch modelleren van het centraal metabolisme van *E. coli* is hiervoor gebruikt, dit is gecombineerd met kinetische en thermodynamische analyses van metabole routes. De ontwerpen voor de foto-elektro-autotrofe metabole routes blijken volgens de voorspellingen van de modellen een beperkt potentieel te hebben voor anaerobe groei van *E. coli* vanwege de relatief lage ATP-productiecapaciteit van de PPR fotosystemen. En ook de relatief hoge ATP-vraag voor celonderhoud speelde hierbij een grote rol. Als de ATP-regeneratie door PPR's kan worden verhoogd en/of ATP nodig voor celonderhoud kan worden verminderd, voorspelt het metabole model dat efficiënte CO_2-fixatie kan worden gerealiseerd, bijvoorbeeld met de natuurlijke reductieve citroenzuurcyclus of via de synthetische pyurvaat-synthase-pyruvaat-carboxylase-glyoxylaat bi-cyclus. Beide zuurstofgevoelige CO_2-fixatie routes zijn erg ATP-efficiënt en hebben daarbovenop volgens de voorspelling een voordelige reactiekinetiek. Dit laatste resulteert in een geschatte, lage vraag naar enzymen voor deze routes in de cel. Verder is de genoemde synthetische bi-cyclus thermodynamisch erg aantrekkelijk onder de geanalyseerde omstandigheden. Over het algemeen laat deze analyse zien hoe nuttig *in silico* analyses kunnen zijn om eventuele problemen en oplossingen te vinden in ingewikkelde ontwerpen voor autotroof en fotosynthetisch metabolisme, als een solide basis voor latere experimentele realisatie.

Om een belangrijk probleem van de PPR fotosystemen aan te pakken, hun lage functionele productieniveau in het celmembraan, is de heterologe productie in *E. coli* van het proton-pompende rhodopsine van de bacterie *Gloeobacter violaceus* (GR) en *Thermus thermophilus* JL18 (TR) gemeten en geoptimaliseerd in **Hoofdstuk 4.** Hoge constitutieve productie van beide PPR eiwitten is bereikt door het goed afstemmen van transcriptie en translatie. Maximum productieniveaus van ongeveer 50.000 PPR's/cel ofwel 7 mg PPR/L cultuur zijn bereikt, wat dicht tegen het maximaal haalbare voor *E. coli* aan lijkt te zitten. Naast het altijd voor PPR's benodigde retinal-pigment, heeft het GR systeem de mogelijkheid om een antenne pigment te binden en gebruiken, het echinenone-pigment. Na optimalisatie van de heterologe pigment productieroutes voor

retinal en echinenone afzonderlijk, konden zowel voldoende hoeveelheden van ofwel retinal of echinenone pigment worden gemaakt voor binding met de PPR fotosystemen in *E. coli*. Vervolgens bleek dat binding van echinenone met GR het absorptiespectrum van dit systeem in *E. coli* verbreedt, waardoor ook blauw licht kan worden geabsorbeerd. Daarnaast vonden we in absorptiespectra dat het TR systeem mogelijk ook dat echinenone-piment kan binden, wijzend op een mogelijke functie hiervan als antenne pigment voor TR. Voor het GR-systeem konden toegenomen ATP productie of andere fysiologische voordelen van dit systeem in blauw licht in *E. coli* nog niet goed worden aangetoond. Optimalisatie van de vertakte metabole route om tegelijk zowel retinal als echinenone te maken is geprobeerd met een set van rationeel ontworpen, variabele ribosomale bindingsplaatsen van verschillende translatie initiatie sterktes. Dit onderzoek toont aan dat er hoogstwaarschijnlijk een kinetische knelpunt is voor de laatste omzetting van beta-carotene in echinenone door het enzym CrtO, in vergelijking met de efficiëntere omzetting door Blh van beta-carotene in retinal. Daarom moeten de productieniveaus of enzymkinetiek op het knooppunt in de vertakte route verder worden verbeterd voor gebalanceerde productie van beide pigmenten. In het algemeen zijn de hier beschreven aanpakken voor het verbeteren van de productie van PPR fotosystemen en bijbehorende pigmenten in *E. coli* waarschijnlijk breed toepasbaar voor de productie van complexere fotosystemen en andere systemen.

In **Hoofdstuk 5** wordt een bijtijds overzicht gegeven hoe het gebruik van codons in genen het functionele productieniveau van eiwitten kan bepalen. In alle domeinen van het leven zijn voorkeuren voor bepaalde synonieme codons gevonden op meerdere niveaus: genoom, regulon, operon en ook binnen genen. Het feit dat alle bekende organismes een incomplete set tRNAs bezitten, impliceert dat codongebruik een algemeen mechanisme zou kunnen zijn voor het afstemmen van de translatiesnelheid. Naast een voorkeur voor het gebruiken van bepaalde synonieme codons, zijn andere varianten van codonvoorkeur beschreven, zoals de voorkeur voor bepaalde codonparen. Hoewel translatie-initiatie de belangrijkste stap is voor het controleren van eiwitproductie, is het breed geaccepteerd dat codongebruik, in het bijzonder in regio's verder weg vanaf het start codon, bij kan dragen aan het afstemmen van de translatie-elongatie snelheid. Bovendien speelt codongebruik een belangrijke rol in het controlleren van een veelheid aan cellulaire processen,

variërend van eiwitvouwing (intra-gen codonvoorkeur) tot relatieve verschillen productieniveaus voor verschillende soorten eiwitten (inter-gen codonvoorkeur). In natuurlijke, geëvolueerde systemen is een optimale set codons geselecteerd die zorgt voor goede modulatie van translatie-elongatiesnelheid, die -in elk geval voor bepaalde eiwitten- van belang is voor correcte co-translationele vouwing, en daarmee voor functionele eiwitproductie. Modulatie van die translatie-elongatiesnelheden hangt af van een complexe combinatie van factoren, bijvoorbeeld de secundaire structuur van het transcript (meer of minder RNA-haarspelden), het 'codonlandschap' (frequente en zeldzamere codons) en voor bacteriën ook ribosomale bindingsplaats-achtige motieven waarop de ribosomen pauzeren. Ondanks deze inzichten is het intrigerend dat meer dan een halve eeuw na het ontcijferen van de genetische code, we nog steeds niet goed in staat zijn om een goedwerkende synthetische DNA-code te ontwerpen voor de heterologe expressie van elk gewenst gen. Het belangrijkste probleem is de complexe combinatie van factoren gerelateerd aan codongebruik die translatie-initiatie en translatie-elongatie, kunnen beïnvloeden. Samenvattend staat het ontwerpen van synthetische genen voor heterologe expressie nog in de kinderschoenen, ondanks de beschikbaarheid van allerlei algoritmes hiervoor blijft het veelal een zaak van 'trial-and-error'.

In **Hoofdstuk 6** is het effect van verschillende codonalgoritmes (optimalisatie en harmonisatie) experimenteel getest voor de heterologe productie van membraaneiwitten. Naast het effect van deze algoritmes op eiwitproductie werd ook het gecombineerde effect van transcriptionele afstelling in *E. coli* LEMO21(DE3) getest. De overexpresssie van zes verschillende membraan-geïntegreerde eiwitten, waaronder 4 PPR varianten (uit bacteriën, archaea en eukaryoten), werd getest. Voor al deze membraaneiwitten werd de eiwitproductie vergeleken voor die van de (i) originele wild-type codonvariant, (ii) een geoptimaliseerde codonvariant van een commercieel algoritme en (iii) een geharmoniseerde codonvariant. Transcriptionele afstelling speelt een hoofdrol bij het verbeteren van de productie van alle geteste eiwitten. Voor de verschillende PPR varianten hebben de geteste codonvarianten nauwelijks invloed op de bereikte productieniveaus. Voor de andere twee geteste membraaneiwitten geven sommige codonvarianten significant verbeterde productie. Voor beide eiwitten gaf de geoptimaliseerde codonvariant lagere productie van de wild-type codonvariant, terwijl de geharmoniseerde

codonvariant significant verbeterde productie gaf, gelijk of zelfs hoger dan voor de wild-type variant.

In **Hoofdstuk 7** wordt aangetoond dat een nieuw systeem voor afstelling op translationeel niveau gebruikt kan worden voor het succesvol verbeteren van de productie van verschillende membraaneiwitten, gebaseerd op eerdere bevindingen uit Hoofdstuk 4. Het toegepaste systeem is het recentelijk ontwikkelde 'Bicistronic Design' (BCD) systeem, dat is gebaseerd op de translationele koppeling van een kort 'leider' gen met het gen dat in dit geval het gewenste membraaneiwit codeert. Translatie initiatie van het tweede gen wordt gecontroleerd door variabele ribosoom bindingsplaatsen met verschillende sterktes. Een set van 22 verschillende RBS-varianten worden voor deze membraaneiwit genen geplaatst, deze zijn geschikt voor het precies, gen-context onafhankelijk, afstellen van genexpressie. De hier gebruikte methode is gebaseerd op een makkelijke DNA-assemblage aanpak, gevolgd door het screenen van hoog producerende stammen in een microtiter plaat. Voor alle vier de membraaneiwitten die getest worden in deze studie geeft dit BCD systeem 3 tot 7 keer hogere eiwitproductieniveaus dan twee onlangs ontwikkelde methodes voor het fijn afstellen van membraaneiwitproductie. De geteste aanpak geeft hoge constitutitieve productie, zonder de noodzaak voor het toevoegen van inducerende componenten. Dit systeem kan waarschijnlijk toegepast worden voor zowel efficiënte membraaneiwitpurificaties als voor synthetische biologie projecten met membraaneiwitten.

In **Hoofdstuk 8** worden een breed overzicht en perspectieven gegeven over de mogelijkheden van microbiële autotrofie en de productie van waardevolle chemicaliën uit CO_2. Zowel fotoautotrofe en chemolithoautotrofe productieplatforms worden bediscussieerd, en recente ontwikkelingen in het verbeteren van de efficiëntie en productiemogelijkheden van dit soort platforms worden besproken. Gerealiseerde transplantaties van fotosystemen, maar ook van CO_2-fixatie routes en elektronenopname routes worden besproken. Verder wordt er een overzicht gegeven van nieuwe *in silico* en experimentele gereedschappen om systemen gerelateerd aan autotrofie in te bouwen en aan te passen in verschillende autotrofe en heterotrofe organismes, met inbegrip van gereedschappen toegepast in dit proefschrift. Verder worden veelbelovende voorbeelden besproken van hybride productie system van

autotrofe en heterotrofe organismen, bio-inorganische hybrides van autotrofe organismen met elektrocatalysatoren of licht-omzettende halfgeleider materialen. Toekomstige mogelijkheden worden besproken om efficiëntere productieplatformen te realiseren, zoals een geïntegreerde synthetische biologie aanpak die gebaseerd is op een combinatie van een rationele ontwerp samen met methodes gebaseerd op laboratoriumevolutie.

About the author

Photo: Jonne Seijdel, Code43.nl

Nicolaas (Nico) Joannes Hubertus Petrus Claassens was born on February 10th 1989 in Geleen, The Netherlands. He studied Environmental Sciences and Biotechnology at Wageningen University, from which he obtained his BSc (2010) and MSc (2012) degree. During his MSc study he performed research on Methanogenic Microbial Electrolysis Cells in the sub-department of Environmental Technology, Wageningen University. In 2012, he spent 5 months at the Department of Biotechnology at the Norwegian University of Science and Technology in Trondheim, where he characterized the XylS-*Pm* system for recombinant gene expression in *Corynebacterium glutamicum* and *Rhodobacter capsulatus*.

In September 2012 Nico started his PhD research at Wageningen University in the Laboratory of Microbiology, under supervision of Prof. John van der Oost and Prof. Willem de Vos in a collaboration with Prof. Vitor Martins Dos Santos (Laboratory of Systems and Synthetic Biology). His work focussed on taking steps towards more sustainable, efficient conversion of light, and also CO_2 into chemicals and fuels by microorganisms and improved production of membrane proteins, mainly focussing on *Escherichia coli* as a host organism. His research included the engineering of rhodopsin photosystems and various other membrane proteins in *Escherichia coli*. Hereto he applied and explored state-of-the-art synthetic biology techniques and systems biology modelling tools. Most of this work has been described in this thesis. In addition, Nico has been involved in supervising the Wageningen UR student iGEM team 2014: BananaGuard, which received the 2nd prize in this international synthetic biology competition in Boston. He has also been active in the organisation of Science Café Wageningen for several years. Since 2016 he has been a blogger for 'Faces of Science', supported by KNAW and Kennislink, providing him the opportunity to share his passion for his PhD research and innovations in synthetic biology with a broader audience.

List of publications

Claassens, N.J., Finger Bou, M., van Velden, S.E.J., Vallvé Odena, M., Chordia, S., Martins dos Santos, V.A.P., de Vos, W.M., van der Oost, J. Tuning the functional production of rhodopsin photosystems with echinenone antennae in *Escherichia coli*. *Manuscript in preparation*.

Claassens, N.J., Siliakus, M.F., Spaans, S.K.S, Creutzburg, S., Nijsse, B., Quax, T.E.F., van der Oost, J. Improving heterologous membrane protein production in *Escherichia coli* by combining transcriptional tuning and codon usage algorithms. *Manuscript in preparation*.

Claassens, N.J., Scholten, B., Muis, F., de Groot, J., de Vos, W.M., van der Oost, J. High functional membrane protein production in *Escherichia coli* by tuning with translational coupling elements. *Manuscript in preparation*.

Claassens, N.J. (2017). A warm welcome for alternative CO_2 fixation pathways in microbial biotechnology. *Microbial Biotechnology 10*, 31–34.

Claassens, N.J., Sousa, D.Z., Santos, V.A.P.M. dos, de Vos, W.M., and van der Oost, J. (2016). Harnessing the power of microbial autotrophy. *Nature Reviews Microbiology 14*, 692–706

Ouwerkerk, J.P.*, van der Ark, K.C.H.*, Davids, M., Claassens, N.J., Robert Finestra, T., de Vos, W.M., and Belzer, C. (2016). Adaptation of Akkermansia muciniphila to the oxic-anoxic interface of the mucus layer. *Applied and Environmental Microbiology* DOI:10.1128/AEM.01641-16

Volpers, M.*, Claassens, N.J.*, Noor, E., Oost, J. van der, Vos, W.M. de, Kengen, S.W.M., and Santos, V.A.P.M. dos (2016). Integrated *in silico* analysis of pathway designs supporting synthetic photo-electro-autotrophy. *PLoS One* DOI:10.1371/journal.pone.0157851.

Quax, T.E.F.*, Claassens, N.J.*, Söll, D., and van der Oost, J. (2015). Codon bias as a means to fine-tune gene expression. *Molecular Cell 59*, 149–161.

Claassens, N.J., Volpers, M., Martins dos Santos, V.A.P., van der Oost, J., de Vos, W.M. (2013). Potential of proton-pumping rhodopsins: engineering photosystems into microorganisms. *Trends in Biotechnology 31*, 633–642.
*contributed equally

Overview of completed training activities

Discipline-specific activities

Meetings & conferences

- Lorentz Workshop: Photosynthesis Plug and Play, Leiden (NL), 2016**
- Dutch Biotechnology Conference (NBC), Wageningen (NL), 2016**
- BioSB NL, Lunteren (NL), 2016**
- EUSynBioS Conference, London (UK), 2016**
- KNVM Microbial Biotechnology 2.0, Delft (NL), 2015**
- 2nd Synthetic Biology Congress, London (UK), 2015*
- Retreat Wageningen Centre of Systems Biology, Hengelo (Geld.) (NL), 2015**
- International Conference on Metabolic Engineering in Bacteria, Amsterdam (NL), 2015**
- NVBMB Fall symposium: Controlling biology with light, Groningen (NL), 2014
- KNVM Microbial Biotechnology 1.0, Utrecht (NL), 2014
- iGEM Giant Jamboree 2015, Boston (US), 2014
- WCSB Symposium Systems Biology for Food, Feed, and Health, Wageningen (NL), 2014*
- Metabolic engineering X conference, Vancouver (CA), 2014*
- Retreat Wageningen Centre of Systems Biology, Twello (NL), 2014**
- EMBO-EMBL Symposium: New Approaches and Concepts in Microbiology, Heidelberg (DE), 2013*
- NWO-ALW meeting, Lunteren (NL), 2013*
- WCSB Symposium Systems Biology for Food, Feed, and Health, Wageningen (NL), 2013*
- NBV-KNCV Symposium Chemistry of Life – Engineering Biology, Nijmegen (NL), 2013
- Retreat Wageningen Centre of Systems Biology, Gaanderen (NL), 2013**

*poster presentation, **oral presentation

Courses

- EMBO Synthetic Biology in Action, Heidelberg (DE), 2015
- NCSB Tutorial: Network Inference from Data, Wageningen (NL), 2013
- NCSB Tutorial: Basics of Parameter Estimation, Wageningen (NL), 2013

General courses

- Grant writing course (Yellow Research), Amsterdam (NL), 2016
- Abode InDesign, Wageningen (NL), 2015
- Speed reading, Wageningen (NL), 2015
- Thesis supervision and organisation, Wageningen (NL), 2013
- Project and Time Management, Wageningen (NL), 2013
- Scienitific Writing, Wageningen (NL), 2013
- PhD week VLAG, Baarlo (NL), 2012

Optionals

- PhD trip MIB-SSB, California (US), 2015
- Organising member PhD trip committee, 2014-2015
- PhD trip MIB-SSB, Boston area (US) and Toronto area (CA), 2013
- Bio-information technology MSc course, Wageningen (NL), 2012
- Metabolic engineering tools meeting (organising member 2015-2016)
- Microbiology Seminars (organising member 2014)
- Bacterial Genetics group meetings
- PhD meetings Laboratory of Microbiology

Acknowledgements - Dankwoord

It's done, my PhD thesis has been written after numerous DNA gels, modified *E. coli* strains, analyzed Excel sheets and presented Powerpoint slides. I have learned a lot during the past 4+ years. I enjoyed the freedom to develop and do my own research in middle of exciting developments in synthetic and molecular microbiology. Seemingly unavoidable frustrations were of course also often part of this trajectory. But after all I have learned and done, the most valuable part of this trajectory were many amazing people that were around to interact with, enjoy time with and get support and ideas from, both related and unrelated to this work. I am fortunate that there were so many that meant a lot to my work and especially to me personally over the past years!

Allereerst John, jij bood mij in 2012 de fantastische kans om in BacGen te komen werken. Vanaf het begin bood jij mij veel vrijheid en gaf je me ondersteuning en veel vertrouwen om deze uitdaging aan te gaan. Regelmatig brainstormden we er heerlijk op los, en soms moest je me zelfs onderbreken: de "mag ik nu ook wat zeggen, Nico" in een van onze eerste gesprekken zal ik niet snel vergeten. Maar gelukkig hield ik ook af en toe mijn mond om te kunnen luisteren naar jouw goede, wilde ideeën en enthousiasme. Jouw begeleiding was een heerlijke basis om mijn onderzoek op te bouwen. Het is bijzonder hoe jij BacGen hebt opgebouwd tot een diverse groep met spannende onderzoekslijnen en al die getalenteerde mensen.

Willem, vanaf het begin van dit project wist jij ook hele fijn ondersteuning te bieden aan mijn werk, te midden van jouw overvolle werkschema. Je gaf regelmatig en vooral ook heel snel jouw input voor dit project, en je steunde ook op fantastische wijze mijn persoonlijke wetenschappelijke ontwikkeling. Ik ben dankbaar voor alle kansen die je me hebt gegeven om mij als jonge wetenschappers te mogen ontwikkelen binnen het door jou zo succesvol gerunde lab. Voor jouw efficiëntie en energie heb ik veel bewondering en daar heb ik ook veel van kunnen leren, al hoop ik dat ik je spreektempo niet op topsnelheid heb overgenomen.

Vitor, dank voor jouw bijdrages aan dit werk, jouw netwerk en visies op systeembiologie en synthetische biologie waren een goede aanvulling voor dit onderzoek. Jij hebt me goed uitgedaagd om zowel experimentele als *in silico* aspecten van het onderzoek te integreren. En ondanks dat dit niet altijd

zo simpel bleek, heeft dit geleid tot een *in silico* publicatie en hoofdstuk in deze thesis, en zeker ook mijn persoonlijke interesse voor de mogelijkheden van modelleren verder aangewakkerd.

Michael, together we started on a for us both relatively new challenge to explore the metabolism and potential of *E. coli* with rhodopsin photosystems using metabolic models. I am very happy with all the modelling work you did, chapter 3 would have been absolutely impossible for me to perform without your hard work. It was a great pleasure discussing and working with you on all the results and getting our first research paper published together.

Ruben, na Michaels vertrek, nam jij de rol als modelleur over in het 'rhodopsine-team'. En met jouw modelleerexpertise, aangevuld met het goede werk van onze gezamenlijke student Henri, gaat dit ook nog tot een mooi verhaal leiden.

Diana, Ruud and Servé, thanks for further enlightening me in microbial physiology and metabolism aspects, another scientific love of me, for which I got your help and insights for my projects at several times!

Also many thanks to several collaborators, literally around the world. Jan-Willem de Gier in Stockholm, thanks for sharing insights and help on the membrane protein production challenges. To our ETH Basel collaborators, Sven Panke, Markus Jeschek and Daniel Gerngross, thanks for all your support on the RedLibs algorithm. Across the ocean at NCBI, Yuri Wolf thanks for sharing many codon tables and your help for the harmonization project. At Yale, the eminent Professor Dieter Söll, it was a big honor to have you on board for our codon review. Back to Europe in ETH Zürich, Elad Noor, I am very happy that you were willing to share your brilliant knowledge and insights on how to model thermodynamics and metabolism for chapter 3! And in my hopefully future home town Potsdam, Arren, thanks already for our inspiring exchange of ideas on how to improve C1-metabolism and looking forward to work with you in the future. Closer to home in Leiden, Vidya and Wim de Grip, even though we faced some challenges and the data did not make it into this thesis, thanks for all your support on the proton-pumping assays and congrats on your great achievements in the color-tuning of rhodopsins!

Many of the numerous DNA gels, *E. coli* strains, Excel sheets and much more work would never have existed with the help of all BSc and MSc thesis students that

worked with me over the past years. I was happy that each of you showed great enthusiasm to work on my appealing, but in real lab life also often challenging and sometimes frustrating projects. Shreyans, Bas, Robert, Max, Sem, Jonas, Bart, Marta, Fernando, Frederieke, Henri and Suzan, you all put in a lot of your time into these projects. I am very grateful for this! I am happy some of you already entered into PhD positions yourselves and I am confident that all you have motivation and capacities to do great work in the future!

Bart, dank voor je programmeerhulp bij het codon-harmoniseren! Marcel Janssen, dank voor advies bij mijn lichtopstellingen.

Martine en de andere betrokken bij het Faces of Science project van Kennislink en KNAW, dank voor de goede support om blogs van mij en andere promovendi zo goed voor de dag te laten komen. Faces of Science is een gaaf project dat voor mij weer nieuwe invalshoeken heeft geboden en deuren heeft geopend. Phelim, dank voor het doen van een mooi en inspirerend bio-art project samen!

Much of the work in this thesis was performed in office 0.045 at the Dreijen on the 'hill' were I could enjoy nice company of amazing office mates, both in 'early' mornings and real late evenings sometimes. Bas, dank voor alle gesprekken en gelul over alle kleine en grote 'ups' en 'downs' van onze onderzoeksprojecten, wetenschap en allerlei andere boeiende dingen. Raymond, het was goed om jou naast mij te hebben om m'n ego een beetje in toom te houden met jouw rake humor en fanatieke practical jokes. Natuurlijk dank voor alle gefaalde pogingen om mij te vergiftigen met creatieve toevoegingen in mijn thee. En ik ben blij dat je na een tijdje heel ver weg weer terug bent gekomen naar BacGen om hier als getalenteerde wetenschapper je werk voort te zetten. Tessa, jij was een fantastische toevoeging als wijs persoon en wetenschapper aan ons kantoor, met soms wat al te puberaal gedragende mannen. Ook al nam je verdacht vaak regenachtig weer mee naar Wageningen, maar gelukkig ook snoep. Ik ben blij dat we samen wat stappen hebben kunnen zetten om die onbegrijpelijke taal van codons wat beter op een rij te krijgen, resulterend in een prachtig review. Melvin, na het vertrek van Tessa vulde jij die vacante plek voor een redelijk verstandige promovendus en fijne kantoorgenoot prima in. Onze projecten overlapten elkaar op een paar vlakken en daarmee konden we goed onze ideeën en frustraties delen bij onze pogingen om *E. coli* en z'n membraan succesvol te verbouwen. En daarbij hebben we ook nog samen kunnen werken om die ingewikkelde codontaal wat

beter in te zetten bij dat soort werk. Franklin, I have great respect for all your ideas and devotion for science. Despite less successful attempts to train each other in Portuguese and Dutch, I had great fun with you in the office. Brenda, it was great fun having you and I will never forget your enthusiasm when watching from our office window the first, directly melting snowflakes in the only 'real' winter in your life so far. Luckily you, together with Leo, shared fun and Brazilian warmth with us!

And then mid-2016 we finally moved downhill to the campus, where I ended up in Helix in 6.073 with a great view and much more important another great collection of colleagues. Teunke, dankzij jouw begonnen mijn maandagen ondanks ons traplooppact prima met even heerlijk bijkletsen over onze weekenden enzo. Ik heb genoten van jouw interesse in de mensen om je heen en je fantastische doorzettingsvermogen en zie uit naar nog een paar maanden samen in kantoor! Hanne, dank voor alle gezelligheid en de leerzame blik in het leven van een duizendpotende, enthousiaste postdoc. En terloops heb je ook nog mijn Vlaamse taalvaardigheden en kledingsmaak met een klein beetje succes verbeterd. Tim, wij begonnen ooit tegelijk als overburen op de Dreijen en eindigen beide onze promotietrajecten in hetzelfde kantoor in het Helix. Fijn hoe jij het kantoor verstrekt met CRISPR-kennis, het spotten van mooie zonsondergangen en de verassende Duitse humor, die toch echt bestaat. Lucia, dinsdagen waren altijd nog een beetje gezelliger op kantoor met jou en het was altijd fijn om met jou, zowel face-to-face als in vele tientallen mails, te sparren over die onbegrijpelijke codontaal. En dan onze jongste kantoortelg, Thijs, het is leuk een frisse en kritische nieuwe promovendus erbij te hebben in ons kantoor op het uitdagende onderwerp van betere eiwitproductie/codongebruik, alle succes de komende jaren!

Alex, even though I barely escaped some violent attacks from your drone and plastic hammer, it is a lot of fun to be your office neighbor, and to keep track of each other's stupidities every now and then. Thanks for joining me on countless food expeditions to AH To Go. And of course for that other great expedition to California, which was great fun organizing and enjoying together! Thanks for all of this and more! And I am very happy to have you as one of my paranymphs!

Talking about paranymphs, Becca, I am sure in my 4+ years there was nobody I spent more time chatting with at the lab bench than you, even though you were only around for a little over 1 year in Wageningen. Your cheerful personality and never-ending great stories made many boring pipetting sessions a big pleasure.

And I think that focusing on a Kiwi accent and simultaneously focusing on filling up a 96-well plate was definitely a great lesson in multitasking. And so it's great to have you back on my side for my defense!

And then there is so many more colleagues that made my stay at Microbiology so good and much more fun, countless lunchbreaks, birthday cakes, Friday (and other day) drinks, poker nights, one memorable snow ball fight and many chats, advices and support! Special thanks to all not-yet-mentioned (former) BacGen colleagues Tijn, Elleke, Tom, Daan, Stan, Servé, Yifan, Edze, Marnix, Richard, Mark, Peter, Sjoerd, Jorrit, Ioannis, Wen, Prarthana, Joyshree, Jeroen, Lione, Jochem, Patrick, Sebastian and Enrico, many thanks!

Anja, Carolien, Mirella, Philippe en meester-manusje-van-alles Wim, dank voor een hele hoop praktische zaken en al jullie inzet en toewijding om microbiologie op fantastische wijze draaiende te houden!

Voor talloze bestellingen, autoclaafladingen, geduldig oplossen van al mijn veroorzaakte HPLC-issues en hulp bij velerlei technische vragen en zoektochten in het lab, veel dank aan Tom vdW, Ton, Tom S, Steven, Sjon, Merlijn, en Monika en alle andere analisten van MIB en SSB.

Gosse en Caroline, dank voor jullie tomeloze enthousiasme bij het onderwijs bij mijn eerste kennismaking met microbiologie tijdens mijn studie.

Teunke, Noora, Corina, Floor, Bas, Vicente and Javier thanks for the great trip to the US and Canadian east coast in 2013, a great trip with high-level science and good fun. And thanks to several guys for joining my first exploration of the North American continent, before the science part started Melvin joined me to explore amazing NY, and afterwards I got stuck with Bas, Tijn, Daan and Jasper in the snow in May, with countless moose, some beavers and one small bear in Algonquin Provincial Park in Ontario.

Yue, Anna, (congrats both by the way for finishing your PhDs at a jealous-making high speed!), Ruben, Maarten, Alex, Jasper and Kees thanks for together organizing a great PhD trip to California in 2015. Johanna, LooWee, Ying, Lara, Klaudyna, Ioannis, Alex K, Alex U, and Michi thanks for adding a great two weeks of exploring

natural wonders during an epic road trip from San Diego up to Yellowstone (and luckily road-killing only one rabbit along the way).
Petra and Romy thanks for organizing some nice Microbiology Seminars together.

Kiira thanks for a great exploration of the Canadian east coast and great times, also together with Elleke, during the Metabolic Engineering Conference in Vancouver. And as a nice follow-up we set up the interdepartmental metabolic engineering meetings, which I enjoyed and are being nicely continued by Pauline and Jeroen.

Rob, Ruben, Stamatios, Kees, Marnix, and Christian, it was a pleasure to supervise the iGEM 2014 team with you. Especially thanks to all the enthusiastic students in the team Marlène, Tjaša, Teresa, Max, Walter, Michiel, Rik, Bob, Miquel, Kevin, Jeremy and of course the small, great team captain, and now my great colleague Wen! You all made my banana carnival costume gather world fame during the Jamboree in Boston before taking the second prize and some other great awards to Wageningen, and of course thanks for trying to rescue my favorite fruit with cool synthetic biology ideas.

Master team captain Tom and all other runners of the Veluweloop teams in the past years, thanks for the great running spirit and giving me the motivation and fun to explore a new sport!

Carrie, dank voor kantoordoorhaalavondjes en -weekendjes om dit boek af te krijgen, spinsessies en superleuk dat we naast floorballteamies ook nog even collega's gaan worden.

Emmy, dank voor je enthousiaste voortzetting van mijn cursus Vlaamse taalvaardigheid enzo ;-), maar bovenal voor je bijzondere steun en inspiratie tijdens de laatste fase van mijn promotietraject!

And then there is so many other people in the diverse and vibrant MIB-SSB family I owe my thanks to for great help and great times. At countless moments and for countless things you were very nice and helpful, keep that spirit! Hereby I want to thank you all!

En dan naast al die geweldige collega's waren er nog een hoop andere dierbare mensen in mijn omgeving in de afgelopen vier jaar en natuurlijk al daarvoor, en ook in de toekomst hoop ik met jullie nog veel moois te beleven.

Stefan, dank voor mijn langstdurende vriendschap, op nog vele jaren en mooie treinavonturen enzo! Alle dierbare milieustudievrienden, inclusief de inmiddels verzamelde gezellige aanhang, Sander & Dehorah, Marina & Tim, Barbara, Anouk & Douwe, Marlies & Jan, Tanya, Joep & Anneloes, Ralph, Maarten & Binh, dank voor allerlei gezellige spellenavonden, verhuizingen, bruiloften, wandelingen, tripjes en nog zo veel meer!

Wieneke & Bas en Jouke, ben blij dat ik ondanks mijn afdwaling van de waterzuiveringskunde toch nog een waterzuiveringsclubje heb met jullie om de lachspieren te blijven trainen.

Dank aan Lieneke, Anneloes, Sander, Jochem en vele andere oud-Haarweg-103-ers voor een jarenlange gezellige thuisbasis, ook nog tijdens de eerste jaren van mijn semi-burger bestaan als phd-student.

Gelukkig zijn afgelopen jaren maar weinig woensdagavonden voorbijgegaan zonder floorball, samen eten, daarna zwoegen met of zonder wat push-ups en af en toe scoren, en daarna natuurlijk nog de kantine in voor gezelligheid, gigamix en goede wist-je-datjes. Anne, Luc, Stijn, Francine, Robin, Bram, Anouk, Christiaan, (H)Anna, (H)anna, (euro!), Martine, Jochem S, Carrie, Eefje, David, Sigrid, Tess, Elvy, Abhi, Jim, Paul, Jochem B, Hetty, Justin, Katharina, Lysanne en vele andere (oud)-stickers dank voor al dit sportieve plezier en meer!

Mijn echte enthousiasme voor biologie en Wageningen begon eigenlijk op te bloeien na mijn deelnames aan de Nationale Biologie Olympiade in 2006 en 2007. Oud-olympisten vormen voor mij een bijzondere vriendengroep met onder andere Eva, Marije, Mojca, Diederik, Roel, Nienke, Julian, Inge, Michiel, Leonie C, Leonie vS, Bert, Remie, Anna, Jim, Jord, Oskar, Lianne, Anne-Maud, Imma, Rob, Simeon en vele andere om samen KOENs, vakanties, fietstochten, een aanstaande molzoektocht en zoveel meer mee te beleven. Lezen jullie mijn proefschrift trouwens maar niet te goed, want dan zullen jullie opmerken dat ik RubisCO een beetje afkraak :-P.

En voorbij het studentvakbondsidealisme is er ook nog een hoop gezelligheid en inspiratie met de Flauweculco: Marlies & Jan, Jouke, Evelien, Andy, Niels & Frazen & Bastiaan, Karmijn, Ewoud & Rhodé, op naar nog vele kerstdinertjes en weekendjes vol gezelligheid, goede discussies en flauwekul!

Het organiseren van Science Café Wageningen was vooral door het diverse team een van de meest inspirerende activiteiten buiten mijn werk de afgelopen jaren. Veel dank aan Pim, Dolf, Neli, Gerlinde, Jelle, Margo, Jochem, Nick, Karin, Kristin, Jeroen, Siri, Pola, Kees en andere (oud)-teamleden!

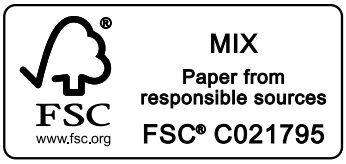
Ook veel dank voor de steun en interesse van mijn lieve familieleden en 'aanhang', ooms Pierre, Aart, Willy, neef Jasper, tantes Lieske en Bertie, en Els & Paul. Altijd leuk om mails van jullie te ontvangen, bijvoorbeeld naar aanleiding van nieuwe blogs.

Mijn lieve zus Marieke & 'schoonbroer' Robin, dank voor de fijne band die we samen hebben en hoop dat we nog heel lang onze heerlijk diverse levens kunnen blijven delen. Ik ben een zeer trotse broer van zo'n creatieve zus die een fantastische cover heeft getoverd voor dit proefschrift. Mam, pap, jullie zijn er al mijn hele leven op formidabele wijze voor mij. Al die jaren al voortdurende steun, aandacht en veel interesse voor alles waarmee ik me bezighoud. Jullie zijn absoluut onmisbaar geweest voor alles wat ik heb mogen meemaken en bereiken tot dusverre, op noch veule jaore um saame te geneete!

About the cover
The background of the cover symbolizes light, the source of energy for photosystems, especially blue and green light, which were of specific relevance for the rhodopsin photosystems in this thesis. The hexagonal pattern in the background is formed by a photo of the LED-light panel used for some experimental work. On the front, the tuner symbolizes all the genetic tuning performed in this work. The metallic tuner represents the tuning mechanisms that are engineered, encoded in genetic information, inside the cell. The black half circle on tuner represents a visual symbol used for ribosome binding sites, a core genetic part applied for tuning in this thesis. The black half circle points to the optimal point of tuning on the indicator circle, an intense red point, similar to the intense red colour of *E. coli* cells producing a high number of rhodopsins. The indicator circle also symbolizes the membrane covering a cell; the genetic tuning inside the cell results in more membrane proteins and photosystems on the outside, in the cell membrane.

The research described in this thesis was financially supported by the IP/OP program Systems Biology of Wageningen University and Research and by the NWO Spinoza Grant to Prof W. M. de Vos.

Cover design by Marieke Claassens.
Lay-out by Ferdinand van Nispen, *my*-thesis.nl.
Printed by GVO drukkers & vormgevers B.V. Ede, on recycled, FSC-certified paper.



Propositions

1. Engineering a *Gloeobacter* rhodopsin photosystem in *Escherichia coli* provides limited physiological advantages but provides insights that are widely applicable. (this thesis)

2. Constitutive expression, tuned by translation initiation elements, is a better strategy than inducible expression systems for high membrane protein production. (this thesis)

3. Chemolithoautotrophic microorganisms should be prioritized above photoautotrophic microorganisms as future cell factories.

4. The seemingly universal language of codons is a complex myriad of dialects, making optimal gene design impossible.

5. A global standard for DNA assembly is unrealistic and unnecessary.
   (Casini et al., 2015, Nature Reviews Molecular Cell Biology)

6. Carbon dioxide emissions from aviation should be compensated by funding research on jet biofuels rather than by planting trees.

7. Democratic decision-making processes without public voting should urgently be explored.

8. The perfect proof of unanticipated societal impacts of fundamental science is Jennifer Lopez producing a TV series entitled CRISPR. (Science News, Oct 2016)

Belonging to the thesis entitled:

'Tuning for light and more: engineering phototrophy and membrane proteins in *Escherichia coli*'

Nicolaas J.H.P. Claassens

Wageningen, March 17th 2017