Protein Expression in *Synechococcus* PCC 7002: a quantitative comparison of promoters and integration sites

ABSTRACT

Photosynthetic Cyanobacteria can be used as a chassis for different synthetic biology approaches. However, quantitative comparison of tools for engineering, such as those for heterologous gene expression, is often not available. Here, we report the comparative quantification of heterologous protein production in *Synechococcussp.*PCC 7002 regarding protein expression cassettes and locations of foreign gene integration using *sf*-GFP as a reporter. We used promoter cpc_{560} as reference because it was described as a "super strong" promoter. *sf*-GFP-expression constructs were integrated into neutral sites NS_1, NS_2, NS_3 and the extrachromosomal plasmid pAQ1. The latter induced a *sf*-GFP level of approximately 10-fold in comparison to a reference promotor expression. Protein-fusion with 6xHis increased *sf*-GFP as well as expression of *sf*-GFP fusion with ß subunit of phycocyanin.

Keywords: cyanobacteria, genomic integration, GFP, pAQ1, protein production, Synechococcus sp. PCC 7002

1. INTRODUCTION

Photoautotrophic prokaryotes, the Cyanobacteria, use sun energy to convert carbon dioxide into organic molecules by oxygenic photosynthesis. They show an up to 10-fold higher photosynthetic efficiency compared to higher plants. Together with emerging possibilities to manipulate these bacteria and available expression vectors, Cyanobacteria are currently evolving as an attractive protein production system[1].

Among these, *Synechococcussp*. PCC7002 is one of the suitable strains[2]. It is naturally competent and its nearly 3 Mbp genome is completely sequenced [3]. It can be transformed with high efficiency and integrates DNA by homologous recombination. In addition, it is fast growing, with minimal doubling times of 2.6 hours.

Promoter-efficiency is one of the key-factors for successful heterologous protein synthesis. For other Cyanobacteria-species such as *Synechocystissp.* PCC 6803, a couple of promoters were examined. Heterologous ones, such as *trc* and *lac*, or native promoters such as *cpc*, *rbc* and *psbA*2 were used in Cyanobacteria expression systems. The super strong promoter $cpc_{560}$was described in the cyanobacterium PCC6803 andtwo heterologous genes were expressed[4]. Besides this, a comparison of twelve native promoters in PCC6803 using the reporter protein eYFPwas published[5]. Constructs with cpc_{560} provided the highest values. For PCC7002, two orthogonal promoter libraries, one based on the above-mentioned cyanobacterial promoter *cpc* and the other on Escherichia coli promoter BBa_J23119, werecreated and evaluated[6]. A variety of IPTG inducible cassettes were subsequently developed and optimized by combining these promoter libraries. The resulting expression cassette showed superior performance compared to *trc* constructs and exhibited a 48-fold dynamic range. Another comparison of *cpc*, cpc_{560}, and *rbcL2*revealed that *cpc* and cpc_{560} were more effective than *rbcL2*[7].

Becausedifferent conditions were usedin the studies, it is difficult to judge which of these promoters is most efficient in *Synechococcus* compared to others. In addition, the integration

site in the three genome neutral sites or in plasmids might influence protein expression [8] as well as protein-fusion constructs with different proteins or a HIS-tag. To our knowledge, a quantitative and comparable evaluation of these variables is not available. We used Green Fluorescence Protein (GFP), more precisely *sf*-GFP, accumulation as a marker for protein expression due to the simplicity of GFP detection and quantification. "*sf*-GFP" is a genetically modified version of GFP that has greater stability, folding efficiency, and fluorescence than the original GFP [9]. We compared fluorescence intensities for six constructs with different integration sites and in addition, protein fusions.

As a result of our investigation, the protein expression data can be directly compared and assist to decide for a protein expression system in *Synechococcus sp.* PCC7002.

2. MATERIAL AND METHODS

2.1 CULTURE CONDITIONS

A^+ medium for PCC7002 cultivation was prepared according previous studies[10]. *Synechococcussp.* PCC7002 cell material was transferred from A^+ agar plates to Erlenmeyer flasks containing 20 ml of liquid A^+ media. Selective A^+ medium contained 100 µg / ml kanamycin or 100 µg / ml ampicillin. For fluorescence measurements, physiological characterization, or transformation, cultures were inoculated in 50 ml A^+ medium with an OD730nm of 0.1. Cultivation was continued to OD730nm 0.5 - 1. For *sf*-GFP quantification by fluorometer measurements cyanobacteria were grown in Erlenmeyer flasks at 30 °C, 100 µmol / m^2 white light, and 120 rpm in a Multitron II incubator, INFORS HT. In case of *sf*-GFP quantification by fluorescent microscopy, cyanobacteria were cultivated at 30 °C, 100 µmol / m^2 white light and 1% CO_2 in a 500 ml bubble column bioreactor.

2.2 *sf*-GFP FLUORESCENCE QUANTIFICATION

"Freeze / Thaw" technique was used to disrupt the cells for fluorometer measurements. Briefly, the cells were frozen in liquid nitrogen and thawed at 37 °C for 5 minutes. Cell debris was sedimented at 14,800 rpm for 5 minutes and the supernatant analyzed with a Perkin Elmer LS50B fluorescence spectrometer. The excitation wavelength was 488 nm, while the emission wavelength was 510 nm. Slits were adjusted at 5 nm (excitation) and 10 nm (emission). The value at which *sf*-GFP production reached a maximum level was used for evaluation. This level was compared to the values of untransformed (WT) and strains transformed with pBK47_pcpc_{560}_NS_2_kanR_GFP. In pBK47_pcpc_{560}_NS_2_kanR_GFP the *sf*-GFP gene was expressed under control of cpc_{560}. The expression cassette was integrated in neutral site 2. All strains and plasmids used in this study are listed below (Table 1 and Table 2).

In addition, intact cells were evaluated by fluorescence microscopy (KEYENCE BZ-X800). The acquired images were analyzed by analyzer software (KEYENCE, "Hybrid Cell Count") with respect to cell count and cellular fluorescence quantification. Exposure time (1 sec.) and magnification (400x) was identic for all experiments.

2.3 VECTOR CONSTRUCTION

sf-GFP gene was cloned downstream cpc_{560} promoter. To investigate *sf*-GFP production at multiple genome integration sites, respective homologous upstream and downstream sequences were introduced. Successfully transformed clones were selected by antibiotic resistance marker, either kanamycin or ampicillin. Gene-integration was confirmed by PCR.

Construction of expression vectors was performed as depicted in Fig. 1. Oligonucleotides were used as listed (supplementary table). PCR fragments were cloned into linearized vector using In-Fusion Snap Assembly master mix (TaKaRa).

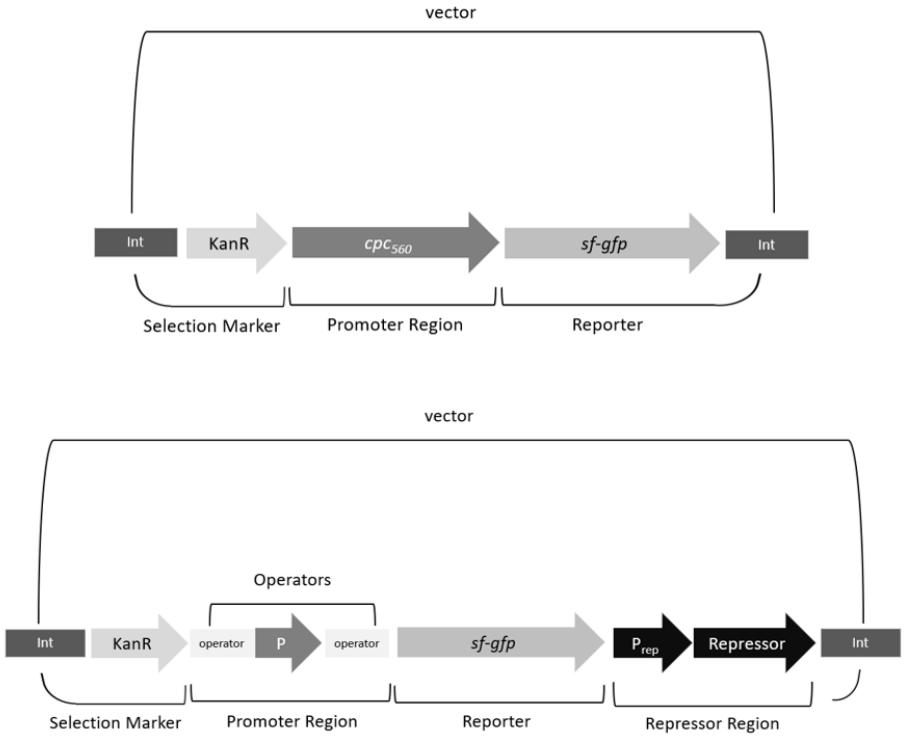

Fig. 1. Schematic illustration of plasmids: (A) Plasmids with the constitutive promoter cpc_{560}. (B) Plasmids with inducible promoters. Constructs p*cptOO* and p*tacO* (Table 1) were designed according to previous studies[6].

Table 1. Strains used in this study.

Strain	Description	Source
E. coli HST08	E. coli strain used for molecular cloning	Takara Bio Europe
WT	Wild type; a marine cyanobacterium	Kachel& Mack, 2020 [11]
NS_1	NS_1 genome integration and cpc_{560} promoter	This study
NS_2	NS_2 genome integration and cpc_{560} promoter	This study
NS_3	NS_3 genome integration and cpc_{560} promoter	This study
pAQ1	pAQ1 genome integration and cpc_{560} promoter	This study
ß-SU FUSION	NS_2 genome integration and gene expression in fusion with Phycocyanin ß-subunit using the cpc_{560} promoter	This study
HIS FUSION	NS_2 genome integration and cpc_{560} promoter, but using HIS-tagged *sf*-GFP	This study
p*cptOO*	NS_2 genome integration and IPTG-inducible *cpt* promoter	This study
p*tacO*	NS_2 genome integration into NS-2 and IPTG-inducible *tac* promoter	This study
p*trcO*	NS_2 genome integration and theophylline-inducible *trc* promoter	This study

Table 2. Plasmids used and constructed in this study.

Plasmid	Source
pBK47_pcpc_{560}_NS_1_kanR_GFP	This study
pBK47_pcpc_{560}_NS_2_kanR_GFP	Kachel& Mack, 2020[11]
pBK47_pcpc_{560}_NS_3_kanR_GFP	This study
pBK47_pcpc_{560}_pAQ1_kanR_GFP	This study
pBK47_p*cpc*B_NS_2_kanR_ßSU_GFP	This study
pBK47_pcpc_{560}_NS_2_kanR_GFP_HIS	This study
pBK47_p*cptOO*_NS_2_kanR_GFP	This study
pBK47_p*tacO*_NS_2_kanR_GFP	This study
pBK47_p*trcO*_NS_2_kanR_GFP	This study

3. RESULTS AND DISCUSSION

3.1 *sf*-GFP fluorescence detection

Initially we measured intact cells for fluorescence quantification. However, there was only a limited quantitative relation between *sf*-GFP fluorescence detected in a fluorimeter and number of *sf*-GFP expressing cells. Above a culture density of OD 2, no correlation between fluorescence signal intensity and cell amount was detected (Fig. 2A). After cell disruption by Freeze / Thaw technique, a linear relation between *sf*-GFP fluorescence and culture density was observed (Fig. 2B).

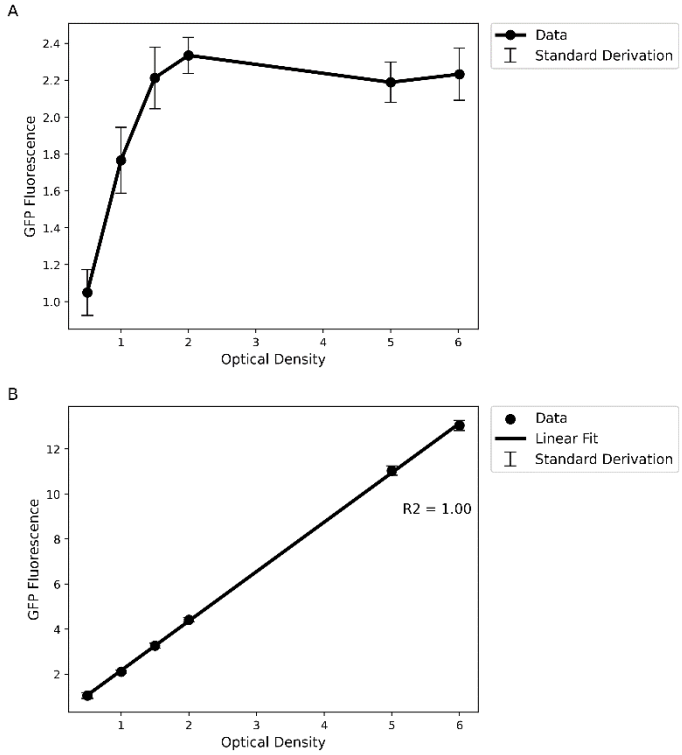

Fig. 2. *sf*-GFP fluorescence of NS_2 by fluorometer fluorescence analysis. After transformation, cells were grown on selective A^+ agar plates for 8 days, transferred to selective liquid medium for another 10 days and diluted to the given OD (n=5). Error bars represent standard deviation. (A) intact cells, (B) disrupted cells.

3.2 *sf*-GFP fluorescence measurements

Evaluation of fluorescence microscopy data from NS_2 showed that the *sf*-GFP level per cell increased and decreased within 10 days after cultivation onset (Fig. 3). Maximal levels were observed between 72 hrs. and 120 hrs. after culture start. The three maximal fluorescence values of each culture were averaged for comparison.

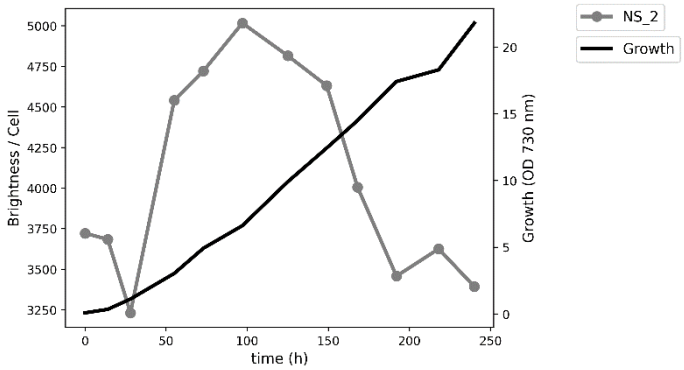

Fig. 3. *sf*-GFP fluorescence of reference culture NS_2 during culture growth. Grey (left ordinate): *sf*-GFP fluorescence as a function of time following inoculation (n=5000 cells). Black (right ordinate): Culture density.

3.3 *sf*-GFP fluorescence Visualization

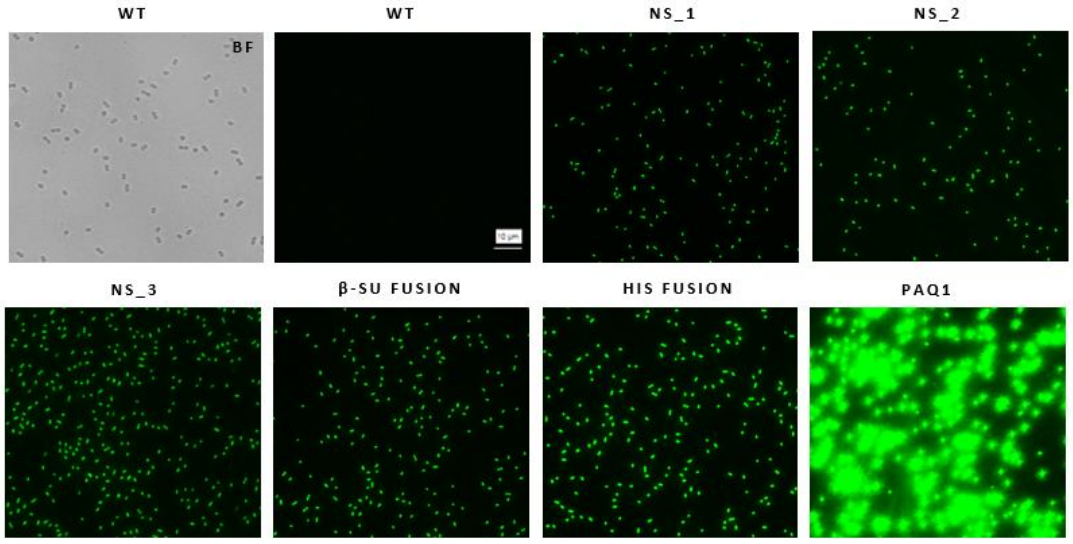

Fig. 4. Microscopic fluorescence images of different *Synechococcus* cells at maximum fluorescence levels except BF – brightfield. The images were captured using Keyence BZ-X800 fluorescence microscope.

3.4 Comparison of neutral integration sites NS_1, NS_2, NS_3 and pAQ1

sf-GFP fluorescence levels of *Synechococcus* with NS_1, NS_2, NS_3 or pAQ1 integration was monitored for a 10-day cultivation period (Fig. 5). The maximum fluorescence level (n=5000 / construct) by integration into NS_1 was 4683, 4830 in NS_2 and 4707 in NS_3, respectively.

The maximum fluorescence level (n = 5000 / construct) was calculated as mean of the three maximum values and was 4683 by integration in NS_1, 4830 in NS_2 and 4707 in NS_3, respectively. Maximum *sf*-GFP fluorescence after pAQ1 integration was 46,927. Distribution of brightness per cell frequencies at maximum fluorescence level by NS_1, NS_2, NS_3 or pAQ1 is given in Fig. 6. The distribution curves of *sf*-GFP fluorescence for the different constructs display similar shapes, with pAQ1 cells exhibiting significantly higher brightness than the other strains. Since plasmids in general could be present in a cell with multiple

copies, this is not surprising [12]. The higher fluctuations in the cell count of pAQ1 cells suggest greater heterogeneity in *sf*-GFP expression within the cell population. Regarding the almost sigmoidal distribution of signal intensity per cell it might be worth to select for strains with enhanced expression in a set of transformed cells.However, despite the observed heterogeneity, pAQ1 integration, on average, resulted in a nearly 10-fold higher *sf*-GFP fluorescence than respective gene integration into the genome.

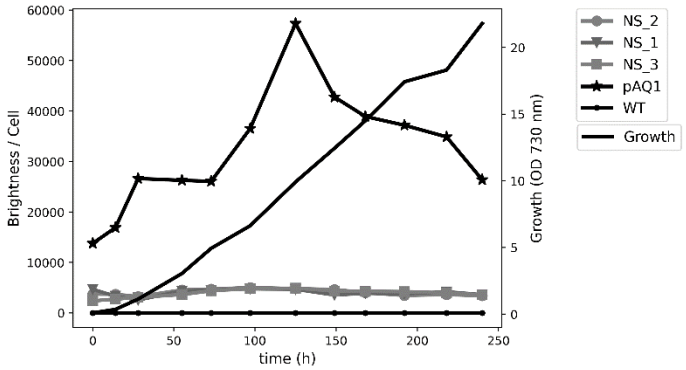

Fig. 5. *sf*-GFP fluorescence levels of wild type and transformants quantified by fluorescence microscopy. *sf*-GFP expression cassette was integrated in NS_1, NS_2, NS_3 or pAQ1, respectively. *sf*-GFP fluorescence levels (left ordinate) as a function of time after inoculation (n = 5000 cells). Culture density (right ordinate).

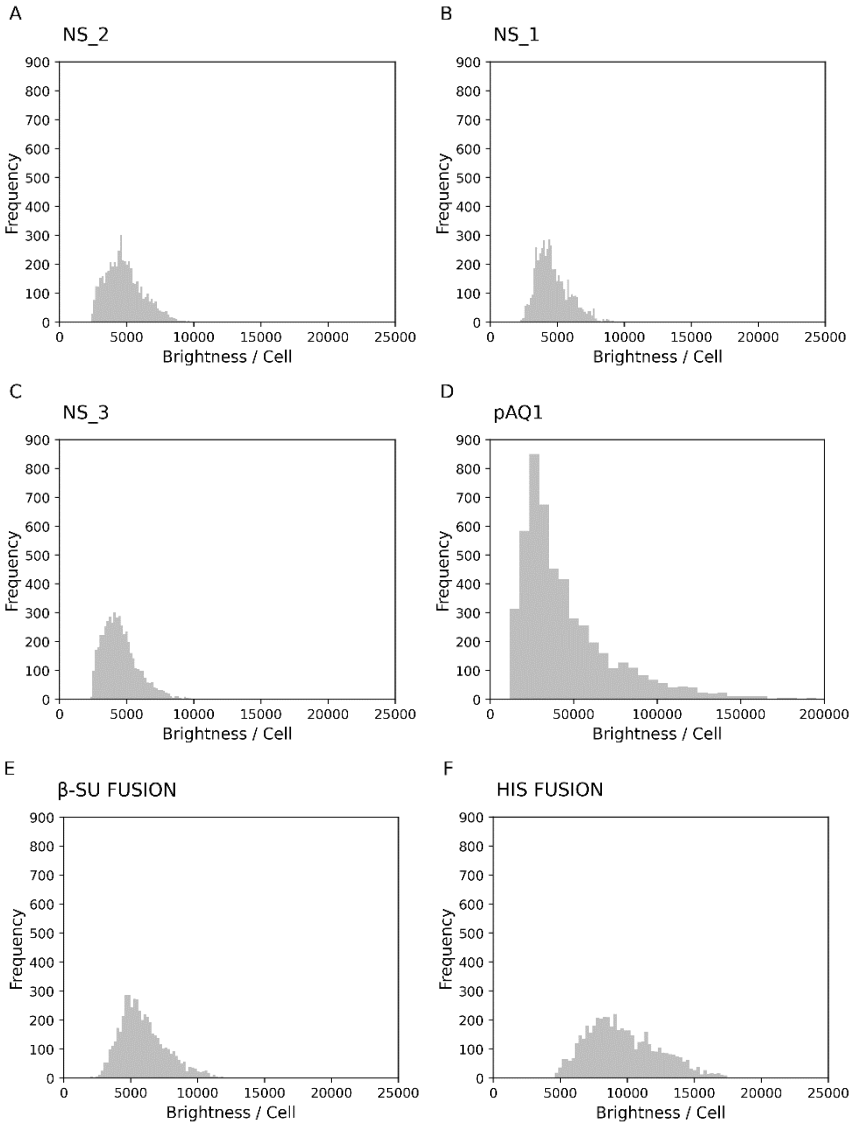

Fig. 6. Distribution of brightness per cell frequencies at maximum fluorescence levels. *sf*-GFP expression cassette was integrated in NS_2 (A), NS_1 (B) NS_3 (C) or pAQ1 (D), respectively. For ß-SU FUSION (E) and HIS FUSION (F) the cassette was also integrated in NS 2, but the *sf*-GFP gene was expressed in fusion with either the phycocyanin ß-subunit or the 6xHIS tag. For each value, the corresponding cell count is shown.

3.5 Effect of protein-fusion on expression levels

Likewise, we investigated if the fluorescence levels were affected by protein fusion. 6xHIS as a tag is often used as an anchor to isolate the expressed protein. It was fused to *sf*-GFP C-terminus. Using the same promoter and integration site, fluorescence levels of NS_2 and HIS FUSION were compared (Fig. 7). Our data indicate that protein amount was about doubled as judged by comparative *sf*-GFP fluorescence with or without 6xHIS tag. Fusion to phycocyanin β subunit was used as an effective protein expression strategy as well [13]. Combination of the corresponding gene with that of *sf*-GFP and integration in NS_2 increase

sf-GFP-fluorescence by a factor of 1.2.It remains to be determined if this is a general effect, i.e., for any protein, or is just specific for *sf*-GFP.

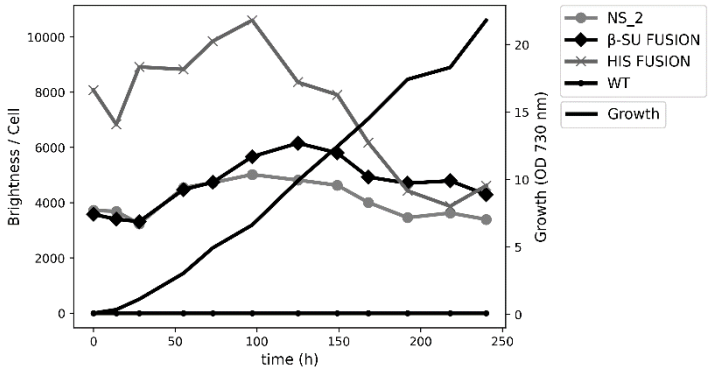

Fig. 7. *sf*-GFP fluorescence levels of PCC7002 wild type and transformants quantified by fluorescence microscopy. *sf*-GFP expression cassette was integrated in NS_2. The *sf*-GFP gene was expressed in fusion with coding sequence for a 6xHIS tag or that for phycocyanin β-subunit. *sf*-GFP fluorescence levels (left ordinate) as a function of time after inoculation (n = 5000 cells). Culture density (right ordinate).

3.6 Inducible promoter

In addition, we examined three inducible promoters. This includes the IPTG-inducible *tac*O and *cptOO* and the theophylline-inducible *trc*. Expression was induced by 1 mM and 2 mM in case of theophylline and 1 mM, 2 mM, and 5 mM for IPTG induction. Nonetheless, *sf*-GFP production with both inducible promoter constructs remained below the fluorescence level of reference NS_2. Application of inducible promoters reached not more than 0.5-fold of the reference fluorescence level (Fig. 8).

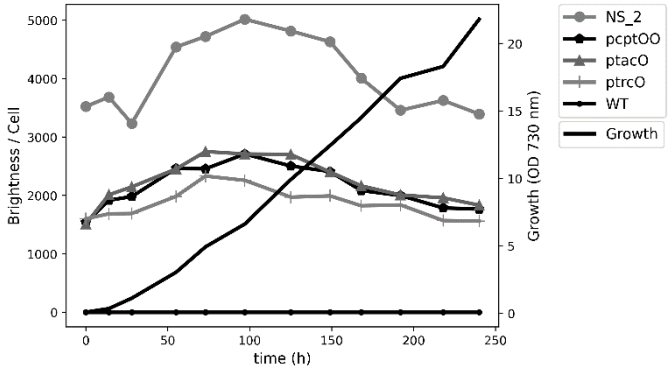

Fig. 8. *sf*-GFP fluorescence levels, quantified by fluorescence microscopy, with *sf*-GFP expression cassette integrated in NS_2. Promoter cpc_{560} (circles), theophylline-inducible promoter *trc* (plusses), IPTG-inducible promoter *cptOO* (pentagons) or IPTG-inducible promoter *tacO* (triangles). *sf*-GFP fluorescence levels (left ordinate) as a function of time after inoculation (n = 5000 cells). Culture density (right ordinate).

3.7 Comparison of *sf*-GFP quantification by fluorescence microscopy and fluorimetry for all constructs

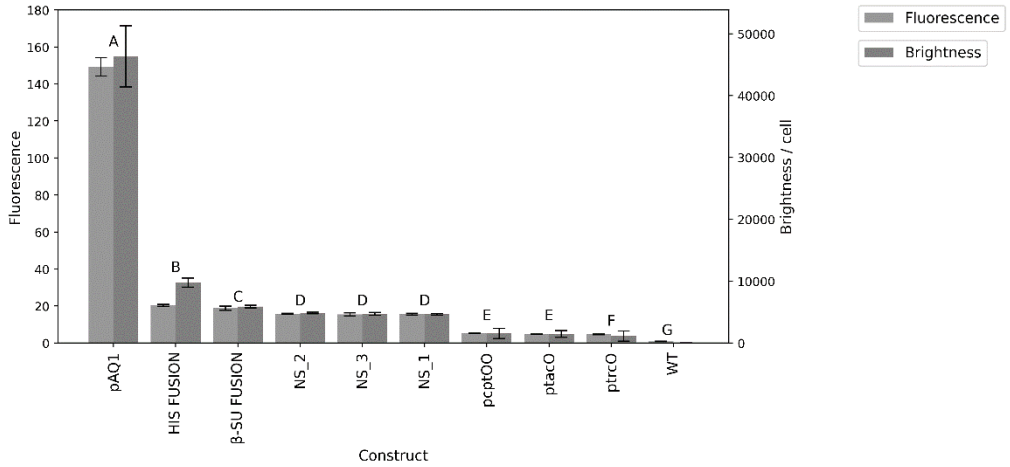

Fig. 9. Average maximal fluorescence levels and maximum brightness per cell by different constructs. Fluorescence values were averaged (n = 5000).

Data are divided into significance groups A, B, C, D, E, F and G (ANOVA). P = .05
Mean ± S.E.M = Mean values ± Standard error of means of twelve experiments

sf-GFP fluorescence of clones with expression cassette inserted into NS_1, NS_2, or NS_3 do not differ significantly (Fig. 9, *P=.16* for NS_1 and NS_2, *P=.77* for NS_1 and NS_3, *P=.45* for NS_2 and NS_3). The 1.2-fold increase in maximal fluorescence levels when using the β subunit fusion construct (C, *P=.004*) and the doubling of these when fusing the *sf*-GFP to 6xHIS tag (B, *P<.001*) are significant. pAQ1 integration results in a 10-fold increase in maximum fluorescence (*P<.001)*. In this case, we did, however, observe individual cells with even higher values. Therefore, if the goal is to achieve high expression rates, an additional selection for these clones with extra high expression levels could be reasonable.

4. CONCLUSION

Using *sf*-GFP-fluorescence as an indicator for the level of protein expression and application of comparable conditions, we could assess the effectivity of the promoters and insertion sites under investigation. We have chosen promotors that are known to allow high expression levels as well as inducible ones. Gene expression from plasmid inserted sequences is most effective. However, it is interesting to note that it is increased by 10-fold on average or even higher in selected cells. The increase in protein production by addition of small amino acid sequences such as HIS-tag or the ß-subunit fragment used herein could be of importance for the design of experiment. On the other hand, the location of genomic integration seems to play a minor role with this regard and the *sf*-GFP protein level in strains with genes under inducible promotors is lower in comparison to constitutive promotor expression.

With this investigation, we could provide valuable information for heterologous gene expression in *Synechococcussp.* PC7002 and support the design of projects aiming to overexpress proteins in this organism.

References

1. Li S, Sun T, Xu C, Chen L,Zhang W. Development and optimization of genetic toolboxes for a fast-growing cyanobacterium Synechococcus elongatus UTEX 2973. Metabolic engineering. 2018; 48:163–174.
2. Li Y, Horsman M, Wu N, Lan C Q,Dubois-Calero N. Biofuels from microalgae. Biotechnology progress. 2008; 24:815–820.
3. Frank J. Causation and Consequences of Polyploidy in Cyanobacteria, Synechococcus sp. PCC7002. University of Colorado, 2022.
4. Zhou J, Zhang H, Meng H, Zhu Y, Bao G, Zhang Y, Li Y,Ma Y. Discovery of a super-strong promoter enables efficient production of heterologous proteins in cyanobacteria. Scientific reports. 2014; 4:4500.
5. Liu D,Pakrasi H B. Exploring native genetic elements as plug-in tools for synthetic biology in the cyanobacterium Synechocystis sp. PCC 6803. Microbial cell factories. 2018; 17:48.
6. Markley A L, Begemann M B, Clarke R E, Gordon G C,Pfleger B F. Synthetic biology toolbox for controlling gene expression in the cyanobacterium Synechococcus sp. strain PCC 7002. ACS synthetic biology. 2015; 4:595–603.
7. Gupta J K,Srivastava S. The Effect of Promoter and RBS Combination on the Growth and Glycogen Productivity of Sodium-Dependent Bicarbonate Transporter (SbtA) Overexpressing Synechococcus sp. PCC 7002 Cells. Frontiers in microbiology. 2021; 12
8. Xu Y, Alvey R M, Byrne P O, Graham J E, Shen G,Bryant D A. Expression of genes in cyanobacteria: adaptation of endogenous plasmids as platforms for high-level gene expression in Synechococcus sp. PCC 7002. Methods in molecular biology (Clifton, N.J.). 2011; 684:273–293.
9. Pédelacq J-D, Cabantous S, Tran T, Terwilliger T C,Waldo G S. Engineering and characterization of a superfolder green fluorescent protein. Nature biotechnology. 2006; 24:79–88.
10. Stevens S E, Patterson COP,Myers J. The production of hydrogen peroxide by blue-green algae: A survey. Journal of Phycology. 1973; 9:427–430.
11. Kachel B,Mack M. Engineering of Synechococcus sp. strain PCC 7002 for the photoautotrophic production of light-sensitive riboflavin (vitamin B2). Metabolic engineering. 2020; 62:275–286.
12. Nagy C, Thiel K, Mulaku E, Mustila H, Tamagnini P, Aro E-M, Pacheco C C,Kallio P. Comparison of alternative integration sites in the chromosome and the native plasmids of the cyanobacterium Synechocystis sp. PCC 6803 in respect to expression efficiency and copy number. Microbial cell factories. 2021; 20:130.
13. Betterle N, Hidalgo Martinez D,Melis A. Cyanobacterial Production of Biopharmaceutical and Biotherapeutic Proteins. Frontiers in plant science. 2020; 11:237.