STRESS RESPONSE OF *SYNECHOCYSTIS* SP. PCC6803 TO DIFFERENT MUTANT STRAINS AND A SQUARE LIGHT REGIME.

M.BANCHERO & S.BANDSTRA

INTRODUCTION

The development of sustainable methods of production became fundamental to dealing with environmental issues. The utilization of microorganism for tackle these problems seems encouraging. One of the organisms that gained popularity as a experimental model is *Synechocystis* sp. PCC6803, considering its ability to fixate CO2. In order to improve the production of desired targets using *Synechocystis* sp. PCC6803 it is important to develop a deep understanding of the factors that produce stress and how the organism responds to this. Stress response takes up lot of responses and should be avoided as much as possible when developing green factories. Getting insight into stress response of *Synechocystis* sp. PCC6803 might help perfecting the development of green factories. The stress response of *Synechocystis* PCC6803 to different light intensities, temperature and high salt concentrations has been previously studied [1]. Of these three factors high light intensities and low temperature seemed to be limiting the growth rates of some substrains. These stress factors hinder the production of biomass and therefore their ability to produce side-products interesting for industrial purposes. Even though growth rates apparently are hindered by light stress, *Synechocystis* PCC6803 was proven to have high resistance to photoinhibition [2]. Furthermore, high stress conditions have been studied, but mild stress conditions could be more probable in industrial environments. Therefore, in this project we want to study the stress response of *Synechocystis* sp. PCC6803 to different mutant strains and a square light regime.

METHODS

Synechocystis PCC6803 was grown in multicultivators [3] which keeps the culture fresh and without nutrient restrictions. Consequently, the cells are kept in exponential phase and bubbled in 1% CO_2 and 99% O_2. The set-up of this experiment makes it possible to study light as the only limiting factor. Two light regimes were studied, namely a square light regime and a sinusoidal light regime. The light intensity peak for the light regime square is 64 µmol m^{-2} s^{-1} for the sinusoidal light regime 100 µmol m^{-2} s^{-1}, in this way the total intensity is the same for both regimes. The multicultivators were enabled at least 3 days before the sampling and continuous light was applied 2 days before sampling.

The sampling was performed over 9 different time points to cover the day/night cycle of 24 hours. The samples were taken from two multicultivators, each of them with 8 channels. 4 channels contained the ΔfumC strain which presents the knockout of fumarase, an enzyme that catalyzed the conversion of fumarate to malate and can therefore be a measure of anabolic activity. The other 4 channels contained the ΔfumCΔzwf strain which presents the additionally knock-out of Glucose-6-phosphate 1-dehydrogenase. Glucose-6-phosphate 1-dehydrogenase catalyzes the first step of Pentose phosphate pathway (PPP) and therefore a knock-out of this gene encourages a major flux towards the TCA cycle of glycogen during night time. This mutation therefore is an indication of the difference between catabolic and anabolic activity.

MICROBIOLOGY: EXPERIMENTAL DESIGN AND PROCEDURE

Glycogen concentrations were measured indirectly via the absorption of NADPH. This is possible when glycogen is first hydrolyzed by Amyloglucosidase into glucose. Then two consecutive reactions will result in the synthesis of NADPH, which can be measured. The relation NADPH to glucose is one to one, which allows us to calculate the glucose concentration and consequently the glycogen concentration. Furthermore, fumarate concentrations were measured with high performance liquid chromatography and biomass was obtained via OD_{730} measurements and CASY counting.

MICROBIOLOGY: DATA ANALYSIS

Growth Rates

The change in cell concentration (N) over time is linear when the natural logarithm is applied, when plotted the growth rate (μ) can be obtained from the slope.

$$\ln N_t = \mu t + \ln N_0$$

Yields

$$Y_{\frac{p}{x}} = \frac{[x]}{OD_{730}}$$

In this case [x] is the concentration of either fumarate or glycogen.

Production rates
Production rates ($Q_{p.day}$) were calculated for glycogen and fumarate during the day. [x] can be either fumarate or glycogen.

$$Q_{p.day} = \frac{[x]}{OD_{730} * t} = Y_{\frac{p}{x}} * \mu$$

Production rates ($Q_{p.night}$) were calculated for fumarate during the night:

$$Q_{p.night} = \frac{[fum(t_2)] - [fum(t_1)]}{Average(OD_{730}) * (t_2 - t_1)}$$

Consumption rate
The consumption rate (Q_s) were calculated only during the night:

$$Q_{\frac{s}{x}} = \frac{[gly(t_1)] - [gly(t_2)]}{Average(OD_{730}) * (t_2 - t_1)}$$

TRANSCRIPTOMICS-EXPERIMENTAL DESIGN AND PROCEDURE

RNA isolation was performed following the RNeasy Mini kit protocol. After the isolation, a DNA digestion kit was used followed by the measurement of RNA concentration using nanodrop. The RNAseq method was

used to sequence the mRNAs, this is a next generation technology high throughput method that allows the quantification of the cell its transcriptome under certain conditions.

TRANSCRIPTOMICS: DATA ANALYSIS

Quality control: Nanodrop and FastQC

The quality of the samples was asses by using nanodrop. The results were analyzed looking at the 260/280 and 260/230 ratios, indicating RNA/DNA and nucleotide purity, respectively.

Outlier removal and Normalization

Filtering for outliers was done for genes which maximum did not exceed 24 counts. After this, normalization is needed to make comparison between samples possible by dealing with technical bias. The normalization of the counts (c) in preparation of DESeq are calculated according to formula's below, calculations were performed in log space to account for underflow problems.

$$sf = e^{median(lin(c)) - mean(\ln(c))}$$

$$norm(c) = \frac{c}{sf}$$

DEseq Analysis

The differential expression of genes were analyzed using the R package DESeq [4□]. This methodology enables a more powerful quantitative analysis of RNAseq applying dispersion shrinkage to genes with similar expression levels.

PROTEOMICS: EXPERIMENTAL DESIGN AND PROCEDURE

The proteomes of the ΔfumCΔzwf strain *Synechocystis* sp. PCC6803 were analyzed using tandem mass spectrometry. Three biological replicates were taken at the end of the night and at midday. In order to perform the proteomic analysis, the proteins were isolated from the cells and the shotgun approach was followed [5]. The isolation of the proteins was followed by tryptic digestion, since peptides are more easy to identify. The pre-fractionation of the samples containing the digested peptides was performed using hydrophilic interaction liquid chromatography HILIC-HPLC, this is important for reducing the sample complexity in the MS/MS ion search [6]. Every biological replicate was fractionated four times, resulting in 24 samples in total.

PROTEOMICS: DATA ANALYSIS

In the data analysis both datasets for 2018 and 2019 were analyzed. The raw MS/MS data was processed with the MASCOT Distiller software following the MudPIT approach. Consequently, the peptides were identified using the MASCOT server. The identified peptides were then matched against a complete

Synechocystis sp. PCC6803 open reading frame translation database in Uniprot. In order to obtain protein quantification, the weighted average of the nitrogen isotopic ratios of light over heavy was used in order to obtain the isotopic ratio for all identified proteins.

Validation was performed by manually discarding abnormal L/H isotopic protein ratios and samples with counts below 30. The processed data was median normalized making the $_{14}N/_{15}N$ ratio equal to 1. The differential expression of proteins was found using the R package for reproducibility-optimized test statistic (ROTS) [[7]□]. This package provides a ranking of features that is the product of the optimization of a modified t-statistics. This package was used instead of DEseq, since DEseq is developed for counts data and does not handle the proteomics data very well

RESULTS

METABOLOMICS

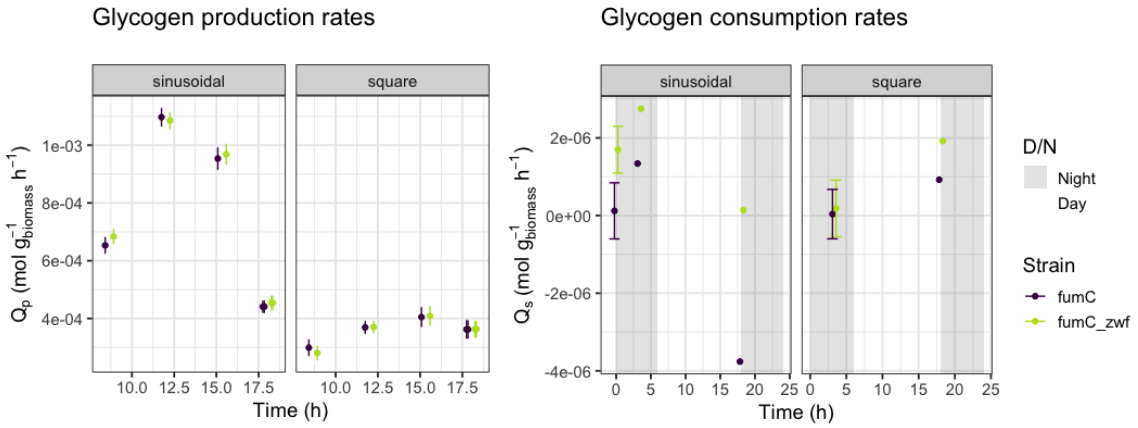

Figure1: Glycogen production and consumption rates per strain and for different light regimes.

The glycogen production rates increase during the day for both regime types. For the sinusoidal light regime, the production rate is at its highest around midday and for the square light regime the production rate displays a small delay and peaks around after midday. The overall production rates for square light regime are a lot lower than sinusoidal light regime. This lower production rates and observed delay is probably because the cells under square light regime are exposed to more stress and might be using their resources for stress responses instead of glycogen production. The mutations for both strains do not affect the glycogen synthesis pathway, resulting in no visible difference for glycogen production rates between strains. The consumption rates are shown in **Figure1** and display very erratic results, leading to the conclusion that these probably should not be used for interpretation.

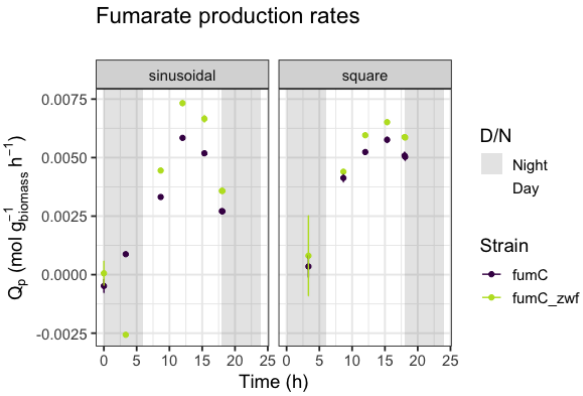

Figure2: Fumarate production rates per strain and for different light regimes

During the day fumarate is primarily produced via urea and purine metabolism and during the night by means of the TCA cycle. The ΔfumCΔzwf achieves higher fumarate production rates during the day, which is somewhat unexpected. The knockdown of the zwf-gene should not affect the urea and purine metabolism and this should be the same for both strains. This difference could be explained by that the TCA cycle is still a little bit active during the day, resulting in more glycogen being directed to the TCA cycle for the ΔfumCΔzwf strain. Differences could also be caused by measurement inaccuracies or data analysis mistakes. During the night the fumarate production rates are lower than during the day, which is in contrast to what would be expected. Again for the night production rates it is probably best to not use these results for interpretation.

TRANSCRIPTOMICS

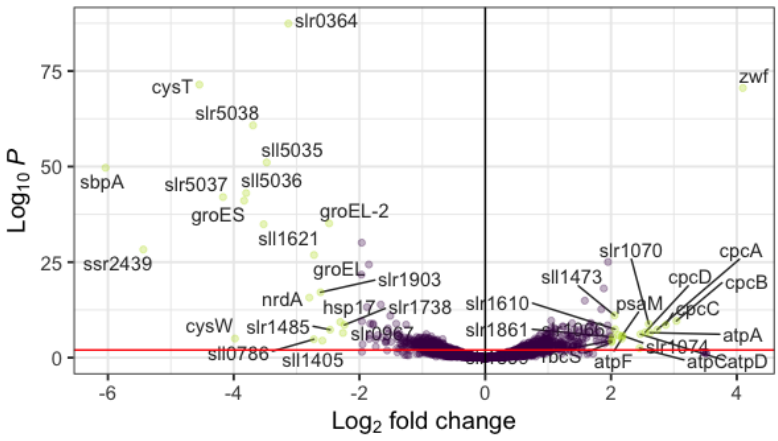

Figure3: transcriptomics compared for two mutant strains **ΔfumC and ΔfumCΔzwf** are compared.

It seems that heat shock genes are more expressed for the ΔfumCΔzwf strain in comparison with ΔfumC strain. The groES, groEL-2, groEL genes all belong to a heat shock regulon present in multiple organisms [8□]. Another heat stress associated gene that is upregulated is the hsp17 gene. This gene works as a thermosensor and controls photosynthesis activity under stress conditions [9□]. Other genes that are more expressed for the ΔfumCΔzwf strain are involved in other processes as membrane transport, transcriptional regulation, Quinone reductases and photosystems.

What is remarkable is that the conditions used here are not expected to be stressful, following the normal sinusoidal light regime. For the ΔfumC strain the genes responsible for making up phycobilisomes and the ATP-ase complexes are more expressed.

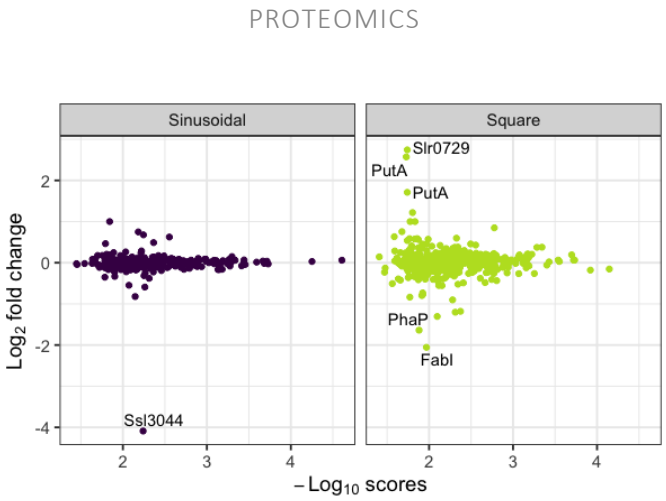

Figure4: Fold change versus protein abundance

Protein levels are less dynamic than transcript levels, due to higher costs for protein production and longer lifetimes. From this it would be expected that proteins are not really differentially expressed, especially for the timescale of 24 hours. **Figure4** gives an indication that there are some differentially expressed proteins especially for the square light regime. The cells under square light regime are under more light stress than normal, which has been known to result in more differentially expressed proteins. However, these proteins that seem interesting also all fall in the lower protein abundance regions (1.5-2.5). For the sinusoidal light regime there is one protein Ssl3044, a hydrogenase component, that has a very large negative fold change. This indicates that this protein is more upregulated during the night. In order to assess if these proteins are differentially expressed a statistical analysis is needed.

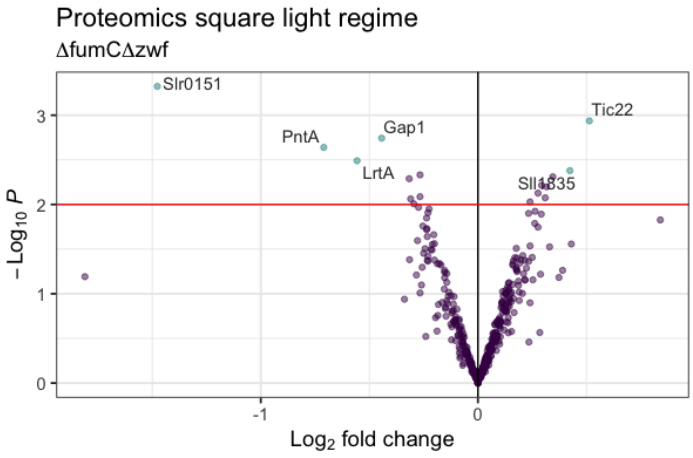

Figure5: volcano plot for proteomics square light regime for the ΔfumCΔzwf strain. Only 5 out of 17 significant proteins were.

In **Figure5** the results are in line with what we would expect from Figure4 with more differentially expressed proteins for the square light regime when compared to the sinusoidal light regime. There are 17 proteins for the square light regime that have p-values below α(=0.01) and are therefore differentially expressed. This means that the observed change in protein levels between the two conditions, day and night, is significant.

DIFFERENTIALLY EXPRESSED PROTEINS SQUARE LIGHT REGIME

The LrtA protein was found to be differentially expressed during the night. The protein is a ribosome hibernation factor that promotes the dimerization of 70S ribosomes into inactive 100S ribosomes. In this way translation is silenced which can be seen as a stress response. What is interesting about this result is that proteins involved in ribosomal synthesis and translation pathways displayed higher counts compared to sinusoidal light regime. This means probably more ribosomes being synthesize but at the same time they are getting deactivated by the hibernation factors. How this can be beneficial for the cell is hard to grasp.

The protein Gap1 is a Glyceraldehyde-3-phosphate dehydrogenase and was found to be differentially expressed during the night. This enzyme catalyzes the conversion of glyceraldehyde 3-phosphate (G3P) to 1,3-bisphosphoglycerate (BPG) which is the first step of glycolysis. It makes sense that this protein is expressed more during the night, but it is peculiar that this is differentially expressed for the square light regime and not for the sinusoidal light regime.

The protein Slr0151 is a tetratricopeptide repeat(TPR) motif that is present in a wide spectrum of ligand-binding proteins. The main function is expected to be protein-protein interaction and are known to have an important role in the assembly of multiprotein complexes [11]. In literature it is mentioned that TPR is critical for the correct assembly of the Hsp70-Hsp90 chaperone complex. It could be that this result is also an indication for a stress response, however it does not make sense that this happens during the night.

STRESS RESPONSE DURING THE DAY

The Tic22 protein is evolutionary conserved fold among plants, algae and cyanobacteria [12]. In cyanobacteria it acts as a chaperone for proteins located in the inner thylakoid membrane. This is key in keeping the proteins in an unfolded state when transported to the inner membrane [13]. The Tic22 protein is differentially expressed in the square light regime. This means that more proteins that make up the photosystem complex are being transported and placed into the membrane. Why more photosystems are needed might be explained by that they somehow they got damaged by the mild light stress conditions and need to be replaced. It could also be explained by that the cells are just trying to maximize its photosynthesis to the unexpected higher light intensities for the square light regime.

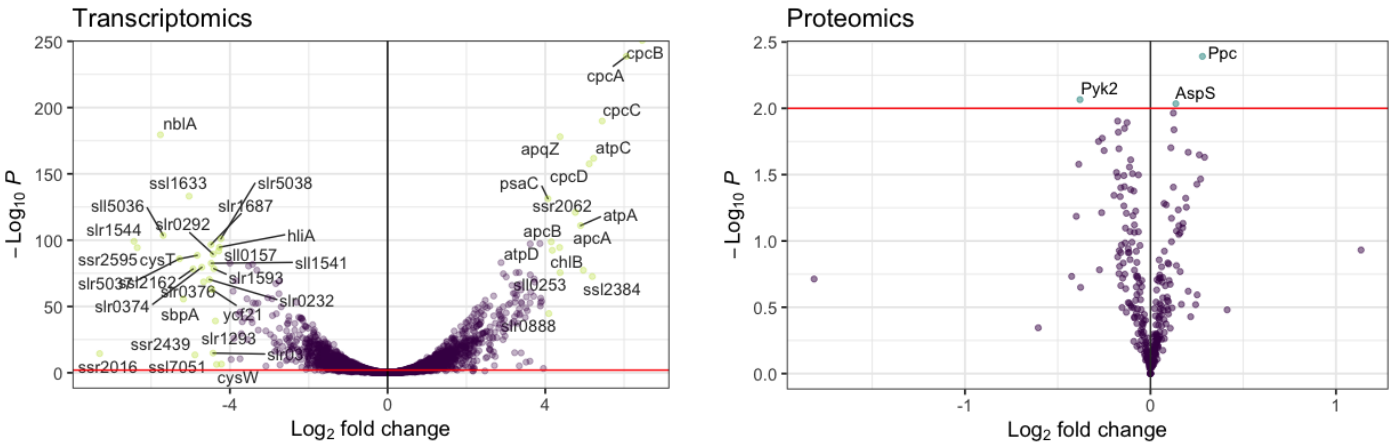

Figure6: Comparison between transcriptomics and proteomics for the ΔfumCΔzwf strain under sinusoidal light regime.

The protein phosphoenolpyruvate carboxylase (Ppc) is very important for CO2 fixation in cyanobacteria. This protein catalyzes phosphoenolpyruvate adding bicarbonate to give oxaloacetate [14] During the day is expected to observed an increase in the rate of C02 fixation.

The most obvious difference is that for transcriptomics differentially expression is a lot more present. This is in line with the known dynamic behavior of transcriptomics. As mentioned before, protein production is costly and among other reasons explains the low amount of differentially expressed proteins under these semi-normal conditions, double mutations aside.

For the three differentially expressed proteins it is expected that the fold change for corresponding genes in the transcriptomics results are multiplied by approximately 2.3 [14]. Surprisingly the Pyk2 protein, that is involved in the glycolysis by catalyzing D-glyceraldehyde 3-phosphate to pyruvate, is not found in the transcriptomics data. This could indicate low immeasurable amounts of transcript or the occurrence of a technical error during sequencing. The ppc gene, responsible for CO_2 fixation, has a lower fold change (-0.12) in the transcriptomics results than its corresponding protein (0.28) and is differentially expressed during night instead of day. In addition, the p-value for this gene is very high, showing little significance. The aspS gene is very significant (3.38e-22) and has a ~14 times increase in fold change when compared to the protein instead of the expected 2.3 times increase.

CONCLUSION

On the metabolic level we observed significantly lower glycogen production rates under square light regime when compared to the sinusoidal light regime. The square light regime is known to cause more stress for the cells and is the only differential factor between the different light regimes. This could mean that resources elsewise used for glycogen production are now needed for dealing with stressful conditions.

The double mutation compared with the single mutation under sinusoidal light regime shows that the ΔfumCΔzwf strain expresses more heat shock genes indicating that they are experiencing more stress. This is somewhat surprising taking into account that the temperature and light conditions are the same.

Differential expression of proteins during a short timescale of 24 hours is uncommon for normal conditions. However, under stress conditions differential expression is more common, which is in line with our results. Most of the proteins found to be differentially expressed cannot directly be linked to a stress response, but for the lrtA, Slr0151 and Tic22 proteins there is a relation to stress response.

When considering differential expression there seems to be little correspondence between transcriptomics and proteomics. It should be noted however that there were only three differentially expressed proteins to compare with the transcriptomics results, making this conclusion preliminary. For further research, comparison between transcriptomics and proteomics for the square light regime might be more interesting to study.

Applying different mutations and different environmental conditions to *Synechocystis* should be taken into consideration when developing green factories, since inducing the stress response can be very un-beneficial.

REFERENCES

[1] T. Zavřel, P. Očenášová, and J. Červený, "Phenotypic characterization of Synechocystis sp. PCC 6803 substrains reveals differences in sensitivity to abiotic stress," *PLoS One*, vol. 12, no. 12, p. e0189130, Dec. 2017.

[2] A. Cordara *et al.*, "Analysis of the light intensity dependence of the growth of Synechocystis and of the light distribution in a photobioreactor energized by 635 nm light.," *PeerJ*, vol. 6, p. e5256, 2018.

[3] V. Bryson and W. Szybalski, "Microbial Selection.," *Science*, vol. 116, no. 3003, pp. 45–51, Jul. 1952.

[4] S. Anders and W. Huber, "Differential expression analysis for sequence count data," *Genome Biol.*, vol. 11, no. 10, p. R106, Oct. 2010.

[5] Y. Zhang, B. R. Fonslow, B. Shan, M.-C. Baek, and J. R. Yates, "Protein Analysis by Shotgun/Bottom-up Proteomics," *Chem. Rev.*, vol. 113, no. 4, pp. 2343–2394, Apr. 2013.

[6] D. Bensaddek, A. Nicolas, and A. I. Lamond, "Evaluating the use of HILIC in large-scale, multi dimensional proteomics: Horses for courses?," *Int. J. Mass Spectrom.*, vol. 391, pp. 105–114, Nov. 2015.

[7] T. Suomi, F. Seyednasrollah, M. K. Jaakkola, T. Faux, and L. L. Elo, "ROTS: An R package for reproducibility-optimized statistical testing," *PLOS Comput. Biol.*, vol. 13, no. 5, p. e1005562, May 2017.

[8] Sean M Hemmingsen *et al.*, "Homologous plant and bacterial proteins chaperone oligomeric protein assembly," 1988.

[9] J. Kortmann, S. Sczodrok, J. Rinnenthal, H. Schwalbe, and F. Narberhaus, "Translation on demand by a simple RNA-based thermosensor.," *Nucleic Acids Res.*, vol. 39, no. 7, pp. 2855–68, Apr. 2011.

[10] N. Zeytuni and R. Zarivach, "Structural and Functional Discussion of the Tetra-Trico-Peptide Repeat, a Protein Interaction Module," *Structure*, vol. 20, no. 3, pp. 397–405, Mar. 2012.

[11] C. Scheufler *et al.*, "Structure of TPR Domain–Peptide Complexes," *Cell*, vol. 101, no. 2, pp. 199–210, Apr. 2000.

[12] J. Tripp *et al.*, "Structure and Conservation of the Periplasmic Targeting Factor Tic22 Protein from Plants and Cyanobacteria," *J. Biol. Chem.*, vol. 287, no. 29, pp. 24164–24173, Jul. 2012.

[13] S. Glaser *et al.*, "Tic22 is an essential chaperone required for protein import into the apicoplast.," *J. Biol. Chem.*, vol. 287, no. 47, pp. 39505–12, Nov. 2012.

[14] J. R. Waldbauer, S. Rodrigue, M. L. Coleman, and S. W. Chisholm, "Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle.," *PLoS One*, vol. 7, no. 8, p. e43432, 2012.

[15] M. Takeya, M. Y. Hirai, and T. Osanai, "Allosteric Inhibition of Phosphoenolpyruvate Carboxylases is Determined by a Single Amino Acid Residue in Cyanobacteria," *Sci. Rep.*, vol. 7, no. 1, p. 41080, Dec. 2017.