



UvA-DARE (Digital Academic Repository)

**Genome-wide expression analysis of environmental stress in the cyanobacterium
Synechocystis PCC 6803**

Aguirre von Wobeser, E.

[Link to publication](#)

Citation for published version (APA):

Aguirre von Wobeser, E. (2010). *Genome-wide expression analysis of environmental stress in the cyanobacterium Synechocystis PCC 6803*.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

SUMMARY AND CONCLUSIONS

1. Introduction

In this thesis, I have designed a microarray platform covering the entire genome of the cyanobacterium *Synechocystis* PCC 6803. The design included the selection of high-quality oligonucleotide probes with similar thermodynamic properties and the development of a new normalization procedure suitable for the analysis of highly expressed genes. The new microarray platform was applied to study the whole-genome response of freshwater cyanobacteria to nitrogen limitation, light limitation, and inorganic carbon limitation.

2. Microarray design

In order to choose optimal oligonucleotide sequences for microarray analysis, several characteristics of the candidate probes need to be considered. An essential property of probes representing one gene in a microarray design is their lack of similarity to all other genes of the genome. Probes that lack similar stretches in other parts of the genome confer specificity to a probe design. Furthermore, the thermodynamic properties of a probe-target hybrid determine the sensitivity of the probe, and a design with homogeneous sensitivity is desirable. These thermodynamic properties include the melting temperature of the probe-target hybrid and the free energies of competing secondary structures of the probes, like hairpin and dimer hybridization. Other parameters that might affect the hybridization reaction are the position of the probe in the gene, the probe length and the length of the target gene.

In order to determine whether the thermodynamic properties of probes are indeed relevant to the performance of microarrays, we tested whether these design parameters had significant effects on the observed hybridization signals of a microarray experiment, using multiple linear regression. We found that all design parameters had highly significant effects on the hybridization intensities. Higher melting temperature increased the hybridization signal, whereas longer probes, more hairpin and dimer structures, longer gene sequences, and the positioning of probes towards the 3' end of the gene sequences significantly decreased the signal. Differences in hybridization intensity between treatment and control samples are often expressed as log ratios. The log ratio between treatment and control was also significantly affected by some of the design parameters of the probes, although effects on the log ratio were relatively small.

We conclude that a careful microarray design strategy should aim at probes with similar thermodynamic properties across the entire microarray. An adequate microarray design minimizes the variation in melting temperature between probes, favors probes closer to the 5' end of the sequence, and avoids cross-hybridization. These probe properties are considered essential for successful microarray experiments.

3. Normalization of microarray data

Microarray data are affected by a number of experimental artifacts that arise from differences in the amount of RNA in each sample, the proportion of mRNA to ribosomal and transfer RNA, the efficiency of incorporation of the dyes during labeling, etc. To remove those artifacts, a number of normalization techniques are available. One common approach is to model the effects of dyes, arrays, and so on, explicitly using ANOVA (Kerr and Churchill, 2001; Wolfinger *et al.*, 2001). Another widely used normalization method is to adjust all data sets to the same distribution. Quantile-based methods apply this idea by adjusting the data to a common distribution, while keeping the original ranking of the data (Workman *et al.*, 2002; Bolstad, 2003). However, we observed that forcing of data to the same target distribution in quantile-based methods can be too strict, and might result in over-fitting. Especially for genes at the extremes of the distribution, quantile-based transformations can alter the relation between data-points of different samples in a way that is not necessarily justified by the data.

Therefore, we developed a new normalization method that can handle highly expressed genes. This method is based on the Generalized Extreme Value (GEV) distribution, which fitted adequately to all microarray data sets we have available. Similar to quantile-based normalization, GEV-based normalization results in data sets that fit the same distribution. Contrary to quantile-based methods, however, the GEV distribution allows a parametric transformation of the data. It does not force the data sets to produce exactly the same target distribution, but preserves information contained in the distribution of the raw data. The Generalized Extreme Value distribution was shown to be extremely flexible, and was more suitable for the analysis of highly expressed genes than quantile-based approaches such as *qspline* normalization.

4. Gene expression responses to nitrogen, light and carbon limitation

The microarray design and the bioinformatics techniques introduced in this thesis were applied to the study of whole-genome gene expression in nitrogen-starved *Synechocystis* PCC 6803 batch cultures and in nitrogen-limited continuous cultures. The results from starvation studies using batch cultures agreed well with previous reports using shorter times scales (Osanai *et al.*, 2006), indicating a remarkable consistency between *Synechocystis* PCC 6803 microarray platforms.

In continuous culture we studied gene expression patterns during the transition from nitrogen-limited growth to light-limited growth. This revealed a large number of genes that were downregulated during nitrogen-limited growth but upregulated during light-limited growth. This included many genes involved in photosynthesis, and also genes related to ATP synthetase, P uptake and polyphosphate formation. Conversely, other genes were upregulated during nitrogen-limited growth but downregulated during light-limited growth. This included several genes involved in nitrogen uptake and assimilation. These mirror images in gene expression were consistent with observed changes in cellular physiology and photosynthesis.

Comparison of the batch culture and continuous culture experiments showed that many, but not all of the genes responding to nitrogen starvation in batch culture responded to nitrogen limitation in continuous culture as well. In batch culture, nitrogen

concentrations are completely exhausted, cells become nitrogen starved, and ultimately cease growth in the stationary phase. In contrast, in continuous culture, cell growth is sustained through a low but continuous supply of nitrogen into the culture vessel. Indeed, several genes involved in cell growth were repressed only temporarily in our continuous-culture experiment, when population growth almost ceased shortly after the reduction in nitrogen supply. However, the expression of these genes recovered when cell growth resumed. These included, for instance, genes encoding RNA polymerase, ribosomal proteins, the core phycobilisome chromophore allophycocyanin, the photosystem I core subunits PsaA and PsaB, and the cell division protein FtsH. Hence, genes that are principally involved in cell growth must remain active in nitrogen-limited continuous culture, whereas genes responding to nitrogen limitation in batch culture indicate the cellular response to complete nitrogen starvation including responses caused by arrested cell growth.

Carbon limitation was studied in separate experiments. Carbon constitutes the largest part of the cellular biomass. Carbon limitation resulted in a strong upregulation of genes encoding inducible high-affinity carbon uptake systems. Surprisingly, however, almost all genes for carboxysomal proteins were downregulated. Carboxysomes are microcompartments within cyanobacterial cells that contain the enzyme Rubisco essential for carbon fixation. The downregulation of carboxysomal proteins might be compensated by the observed extension of the half-life time of the large subunit of the Rubisco protein. In our experiments, the well-documented tight connection between C and N metabolism was confirmed by the observed down-regulation of genes involved in nitrogen assimilation when carbon became limiting (Eisenhut *et al.*, 2007).

Summarizing, in this thesis I developed a new long-oligonucleotide microarray to study the whole-genome response of a model cyanobacterium to different environmental conditions. The results illustrate the power of this approach. In particular, we observed concerted changes in the expression of many functionally related genes in response to nitrogen, carbon and light limitation.

5. References

- Bolstad BM, Irizarri RA, Åstrand M, Speed T. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19:185-193.
- Eisenhut M, Aguirre von Wobeser E, Jonas L, Schubert HJ, Ibelings BW, Bauwe H, Matthijs HCP, Hagemann M. 2007. Long-term response toward inorganic carbon limitation in wild type and glycolate turnover mutants of the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Plant Physiology* 144:1946-1959.
- Kerr MK, Churchill GA. 2001. Statistical design and the analysis of gene expression microarray data. *Genetics Research* 77:123-128.
- Osanai T, Imamura S, Asayama M, Makoto Shirai M, Iwane Suzuki I, Murata N, Tanaka K. 2006. Nitrogen induction of sugar catabolic gene expression in *Synechocystis* sp. PCC 6803. *DNA Research* 13: 185-195.

- Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C, Paules RS. 2001. Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology* 8:625-637.
- Workman C, Jensen LJ, Jarmer H, Berka R, Gautier L, Nielsen HB, Saxild HH, Nielsen C, Brunak S, Knudsen S. 2002. A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biology* 3(9):research0048.