



UvA-DARE (Digital Academic Repository)

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

Aguirre von Wobeser, E.

Publication date
2010

[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*. [Thesis, fully internal, Universiteit van Amsterdam].

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, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

SAMENVATTING EN CONCLUSIES

1. Introductie

In dit proefschrift beschrijf ik mijn bijdrage aan de ontwikkeling van hedendaagse moleculair biologische methodes die fundamentele informatie kunnen leveren over hoe organismen zich aanpassen aan hun omgeving. Met name heb ik mij gericht op het mogelijk maken van ‘transcriptomics’ analyse die een kwalitatief beeld van de relatieve expressie van alle open reading frames (‘genen’) van een genoom kan verschaffen. De transcriptomics analyse heb ik toegepast op een aantal interessante groeicondities voor cyanobacteriën. Om de methode te kunnen gebruiken is het noodzakelijk dat het complete genoom van het betreffende organisme bekend is. De model cyanobacterie *Synechocystis* stam PCC 6803 was het allereerste fototrofe organisme dat volledig werd gesequenced. Daarnaast is al veel bekend van de fysiologie van dit organisme, zodat we hebben gekozen voor *Synechocystis* PCC 6803 om het nieuwe werkveld te gaan verkennen. Voor de uitvoering van transcriptomics zijn unieke gebruiksmiddelen nodig, dit zijn de zogenaamde ‘DNA-microarrays’. Hieronder bespreek ik het ontwerp van een microarray waarmee het ‘transcriptomics’ onderzoek in mijn proefschrift is uitgevoerd.

Bij transcriptomics experimenten worden omvangrijke datasets verkregen, waarbij vaak de transcriptie van veel genen onveranderd blijft, terwijl de transcriptie van sommige genen wordt versterkt en van sommige andere genen juist afneemt. Uit de vergelijking tussen een controle en test conditie kan in detail en voor alle genen tegelijk de reactie op een verandering in groeiomgeving worden vastgesteld. Met name de genen waarvan de expressie in een gegeven test daadwerkelijk verandert, zijn potentieel interessant. Door toepassing van transcriptomics kunnen de genen van *alle* betrokken metabole paden in feite in één keer worden overzien en dat in hun onderlinge relatie. De methode is daarmee equivalent met de uitvoering van vele duizenden polymerase chain reacties (PCR) in één keer. Essentiële voorwaarden voor het mogen trekken van harde conclusies zijn goede reproduceerbaarheid van de groeicondities (‘biological replicates’), betrouwbare uitvoering van de verschillende stappen in het array experiment en niet in de laatste plaats een goede keuze van statistische analyse methodes voor adequate data verwerking en het verkrijgen van betrouwbare conclusies.

In dit proefschrift heb ik een nieuw microarray platform ontwikkeld voor *Synechocystis* PCC 6803, alsmede een nieuwe normalisatie methode voor de statistische analyse van genen met hoge expressie. Het nieuwe microarray platform en de statistische analyses zijn in dit proefschrift gebruikt voor transcriptomics analyse bij het opleggen of opheffen van stikstof, licht en koolstoflimitatie. De ontwikkelde technieken en experimentele resultaten kunnen een belangrijke rol gaan spelen bij een beter begrip van de groei van cyanobacteriën onder verschillende milieucondities, eerst nog in het laboratorium en mogelijk later voor de beoordeling van de actuele groei en metabole status van cyanobacteriën in ecosystemen.

2. Microarray ontwerp

In een compleet geanalyseerd genoom is de positie en sequentie van alle open reading frames bekend. Een microarray wordt systematisch opgebouwd door kleine stukjes genoom op basis van de bekende sequentie te selecteren als probe. Daartoe worden duizenden verschillende probes die representatief zijn voor het gehele genoom in rijen en kolommen op het array oppervlak vastgehecht waarbij het soort probe in iedere array-cel bekend is. In de door ons ontwikkelde microarray hebben we gebruik gemaakt van zogenaamde lange oligonucleotide probes, met een lengte van 45 tot 60 nucleotiden per probe.

Bij het ontwerpen van zulke probes zijn verschillende eigenschappen van belang. Bij hoge kwaliteit probes zijn selectiviteit en afgestemde gevoeligheid essentieel. Probes voor een bepaald gen mogen alleen de transcripten van dat specifieke gen herkennen. De gevoeligheid van de probes om transcript te kunnen binden wordt bepaald door selecteerbare thermodynamische eigenschappen. De gevoeligheid van alle duizenden probes in een microarray dient zo homogeen mogelijk te zijn zodat naast specificiteit ook evenwichtige hybridisatie bij een vaste temperatuur verzekerd is. Aldus, heb ik voor alle open reading frames in het genoom met het software programma Array Designer systematisch gezocht naar unieke nucleotide volgordes, en een aantal andere randvoorwaarden gebruikt om te voldoen aan de eis van vergelijkbare thermodynamische eigenschappen. De thermodynamische eigenschappen omvatten de smelt-temperatuur van het probe-target paar alsmede de vrije energie van concurrerende secundaire structuren in de probes, zoals hairpin en dimeer hybridisatie. Er zijn nog andere parameters die de hybridisatie reactie tussen probe en target kunnen beïnvloeden waaronder de positie van de probe-sequentie binnen het gen, de lengte van de probe en de lengte van de target. Voor ieder gen zijn op deze wijze tenminste twee en veelal drie probes gemaakt, soms zelfs meer (maximaal zes probes/gen). Deze opzet maakt het mogelijk om de kwaliteit van de verkregen data te toetsen.

Om te bepalen welke thermodynamische eigenschappen van de probes van belang zijn, is met behulp van multiple regressie bepaald wat het effect was van ontwerp parameters van de probes op de waargenomen hybridisatie signalen in een microarray experiment. We vonden hierbij dat alle ontwerp parameters een zeer significant effect op de hybridisatie efficiëntie hadden. Een hogere smelt-temperatuur leidde tot een versterkt hybridisatie signaal, terwijl langere probes, meer hairpin and dimeer structuren, langere gen sequenties, en de positionering van probes richting het 3' einde van de gen sequentie leidden tot een significant lager hybridisatie signaal. Gebruikelijk is dat er wordt gecorrigeerd voor dergelijke verschillen in hybridisatie efficiëntie door de resultaten te vergelijken met een standaard, welke de controle wordt genoemd. Om voor verschillen in hybridisatie efficiëntie van de probes te corrigeren is analyse van de log-ratios van een controle monster en een test monster dat een behandeling heeft ondergaan een gebruikelijke methode. Tot onze verrassing bleek dat sommige ontwerp parameters ook een klein maar significant effect hadden op de log-ratio waarde, wat aangeeft dat het ontwerp van de probes ook nog van enige invloed is op de vaststelling van veranderingen in gen expressie.

Er kan worden geconcludeerd dat een doeltreffend ontwerp voor microarrays zou moeten streven naar minimale variatie in smelttemperatuur tussen probes, de voorkeur

zou moeten geven aan probes aan het 5' einde van de gen sequentie, en hairpin en dimeer hybridisatie dient te vermijden. Deze probe eigenschappen zijn essentieel voor succesvolle microarray experimenten.

3. Normalisatie van microarray data

Microarray data worden beïnvloed door een aantal artefacten die ontstaan door verschillen in de RNA hoeveelheid per monster, de verhouding van messenger RNA ten opzichte van ribosomaal RNA, de efficiëntie van de inbouw van kleurstof tijdens reverse transcriptie, enzovoorts. Om voor deze artefacten te corrigeren is een aantal normalisatie technieken beschikbaar. Een wijdverbreide methode voor normalisatie is het expliciet analyseren van de effecten van kleurstoffen, verschillen tussen arrays, enzovoorts, met behulp van een ANOVA (Kerr & Churchill, 2001; Wolfinger *et al.*, 2001). Een andere veel gebruikte methode is de omzetting van de verdeling van de ruwe hybridisatie signalen van de verschillende microarrays naar eenzelfde target verdeling. Kwantiel methoden (zoals qspline) passen dit idee toe, waarbij de oorspronkelijke rangorde van de ruwe data behouden blijft (Workman *et al.*, 2002; Bolstad, 2003). Uit de analyses bleek echter dat normalisatie met kwantiel methodes te sterk kan zijn en aanleiding kan geven tot geforceerde data fitting. In het bijzonder geldt dit voor genen in de staarten van de verdeling waar qspline normalisatie kan leiden tot ongewenst sterke transformaties van datapunten zonder dat dit ondersteund wordt door de feitelijke data. Zo vonden wij dat de kwantiel methode bij stikstofgebrek matig presteerde. Vele genen met een hoge expressie in de controle hebben juist een heel lage expressie in stikstof beperkte monsters, en vice versa. Als gevolg daarvan neigt normalisatie op basis van de kwantiel methodes ernaar om genen waarvan de expressie eigenlijk niet wordt verlaagd toch een hogere expressie te geven. Dit artefact ontstaat omdat deze methode beoogd om in het totaalplaatje dezelfde target verdeling te behouden.

Vanwege dit artefact heb ik een nieuwe methode ontwikkeld die geschikt is voor genen met een hoge expressie. De methode is gebaseerd op de 'Generalized Extreme Value' (GEV) verdeling, een wiskundige verdeling die een goede beschrijving gaf van alle microarray data die we ter beschikking hebben. De GEV verdeling maakt parametrische transformatie mogelijk die in tegenstelling tot kwantiel methodes de data niet distribueert op basis van hun oorspronkelijke rangorde, maar op basis van de oorspronkelijk gemeten waarden. De GEV distributie blijkt zeer flexibel en was beter bruikbaar voor genen met een hoge expressie dan qspline normalisatie.

4. Veranderingen in gen expressie bij limitatie van stikstof, licht of koolstof

Het microarray ontwerp en de bio-informatica technieken die we in dit proefschrift hebben ontwikkeld, zijn vervolgens gebruikt voor onderzoek naar de expressie van alle open reading frames (genen) in stikstof-gelimiteerde batch cultures en continu cultures van *Synechocystis* PCC 6803. De resultaten van de batch cultuur stemden goed overeen met de resultaten van eerder gepubliceerd onderzoek met kort-durende batch experimenten (Osanaï *et al.*, 2006). Deze overeenstemming laat zien dat resultaten verkregen met totaal verschillende microarray platforms heel goed reproduceerbaar zijn.

In continu cultuur bestudeerden we gen expressie patronen gedurende de overgang van stikstof-gelimiteerde naar licht-gelimiteerde condities. Dit leverde een groot aantal genen op die een aanzienlijk lagere expressie hadden tijdens stikstof-gelimiteerde groei dan tijdens licht-gelimiteerde groei; veel van deze genen waren betrokken bij de fotosynthese. Maar er waren ook genen die juist hoger tot expressie kwamen tijdens stikstof limitatie en minder tijdens licht limitatie, waaronder veel genen die betrokken zijn bij stikstof opname en assimilatie. Deze tegenovergestelde patronen in gen expressie waren in overeenstemming met de gemeten veranderingen in cel fysiologie en fotosynthese eigenschappen van de cellen.

Vergelijking van de resultaten in batch cultuur en continu cultuur liet zien dat vele genen die in batch cultuur gereguleerd werden ook op gelijke wijze in continu cultuur reageerden. Dit gold echter niet voor alle genen. In batch cultuur wordt de concentratie van beschikbaar stikstof volledig uitgeput waarna de cultuur een stationaire fase bereikt. In continu cultuur daarentegen wordt de groei gehandhaafd in een steady state, waarvoor altijd enige stikstof aanwezig blijft in het medium. Tijdens overgang van een stikstof gelimiteerde steady state naar een steady state waarbij licht de groei beperkende factor wordt zijn er aanpassingen in gen transcriptie die in aanzet gelijk zijn voor continu en batch cultuur. Echter, in batch cultuur leidt limitatie tot een stationaire fase waarbij de celgroei stopt, terwijl in continu cultuur een nieuwe steady state wordt bereikt waarbij cellen hun metabolisme dusdanig hebben aangepast dat ze kunnen blijven groeien. Inderdaad vonden we dat verschillende genen die bij celgroei betrokken zijn in continu cultuur een hogere expressie vertoonden dan in batch cultuur. Voorbeelden zijn RNA polymerases, ribosomale eiwitten, allofycocyanine eiwitten die behoren tot de kern van het fycobilisoom, de fotosysteem 1 subunits PsaA en PsaB, en het celdelings eiwit FtsH. Daarentegen was de expressie van genen betrokken bij de overgang van cellen naar de stationaire fase veel lager in continu cultuur dan in batch cultuur. Daarom suggereren onze resultaten dat genen die alleen in continu cultuur reageren op stikstof limitatie cellulaire reacties aangeven die succesvolle groei bij lage stikstof concentratie mogelijk maken, terwijl genen die alleen in batch cultuur geregeerde expressie tonen een indicatie geven van cellulaire processen bij volledige stikstofuitputting en stopzetting van de groei.

Licht limitatie en de veranderingen in gen expressie als gevolg daarvan werden in de context van het stikstoflimitatie experiment bestudeerd. Eerder was ook al in een batch-cultuur experiment met verzadigende stikstof condities een licht limitatie status vastgesteld. Typische aanpassingen aan licht limitatie waren een verhoogde expressie van genen betrokken bij ATP synthetase, P-opname en polyfosfaatvorming, terwijl de genen betrokken bij stikstof en koolstofmetabolisme juist werden gerepresseerd. Opmerkelijk was dat de genen voor fotosysteem 2 tijdens licht limitatie minder tot expressie kwamen dan die voor fotosysteem 1.

Koolstof limitatie werd in aparte experimenten bestudeerd. Koolstof is de belangrijkste bouwsteen in cellen. Bij koolstof gebrek vonden we een sterke toename van verschillende hoge affiniteit transport systemen voor koolstof. Echter, verrassender wijze bleek dat alle genen betrokken bij carboxysoom synthese gerepresseerd werden. Dit kan mogelijk samenhangen met de waargenomen verlenging van de halfwaardetijd van de levensduur van de grote subunit van het RuBisCo eiwit. De bekende sterke koppeling tussen C en N metabolisme werd bevestigd in onze resultaten, waarin we

vonden dat koolstof limitatie leidt tot repressie van genen die betrokken zijn bij de assimilatie van stikstof verbindingen (Eisenhut *et al.*, 2007).

Samenvattend heb ik in dit proefschrift de ontwikkeling van een nieuwe DNA microarray beschreven dat is gebaseerd op lange oligonucleotide probes. Het ontwerp werd gebruikt voor het vaststellen van de aanpassing van de expressie van alle genen op het genoom van een model cyanobacterie aan veranderende omgevingscondities. De resultaten laten zien dat de methode veel mogelijkheden biedt om aanpassingen van het complete metabolisme te bestuderen. Hieruit kwam naar voren dat er veel samenhang is in de aanpassing van verschillende genproducten die betrokken zijn bij stikstof, licht en koolstoflimitatie.

5. Referenties

- 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): research 0048.