



### Matthias Kopf

2006–2012 Diplomstudium der Biologie an der Universität Freiburg. 2012–2015 Promotion an der Universität Freiburg in der Arbeitsgruppe von Prof. Dr. W. R. Hess. Seit 2015 Software Requirements Engineer bei der Molecular Health GmbH in Heidelberg.

DOI: 10.1007/s12268-016-0690-3  
© Springer-Verlag 2016

■ Vor über 20 Jahren wurde das Genom von *Synechocystis* spec. PCC 6803 (*Synechocystis* 6803) sequenziert. Mittels bioinformatischer Vorhersagen wurden damals über 3.000 Protein-codierende Gene identifiziert. Gut einem Drittel konnte durch Sequenzähnlichkeit zu bereits charakterisierten Proteinen eine Funktion zugewiesen werden. Das Genom von *Synechocystis* 6803, das zu 87 Prozent aus Protein-codierenden Genen, tRNA und rRNA besteht, besitzt demnach eine sehr hohe Gendichte [1]. Seitdem wurde jedoch zunehmend die Bedeutung nicht-codierender RNAs (ncRNAs) und kleiner Proteine bekannt, die in der ursprünglichen Genomannotation nicht beachtet wurden. Es fehlt noch immer für etwa die Hälfte der Gene in *Synechocystis* 6803 eine funktionelle Annotation.

Diesen noch offenen Fragestellungen wurde in Form einer bioinformatischen Analyse von Genom- und Transkriptomdaten nachgegangen. Da diese Analysen stark von Homologieinformationen aus verwandten Organismen profitieren, wurden die gleichen Daten auch für den von uns neu sequenzierten Schwesterstamm *Synechocystis* spec. PCC 6714 (*Synechocystis* 6714) erhoben und gemeinsam analysiert. Durch eine Kombination von klassischen RNA-Sequenzdaten (RNA-Seq) und auf intakte 5'-Enden angereicherten RNA-Sequenzdaten (dRNA-Seq) aus zehn unterschiedlichen Wachstumsbedingungen konnten wir sowohl transkriptionelle Einheiten

## VAAM-Promotionspreis 2016

# Genom- und Transkriptomanalyse in *Synechocystis*

MATTHIAS KOPF

MOLECULAR HEALTH GMBH, HEIDELBERG

und deren Transkriptionsstartstellen kartieren als auch die zugehörigen Expressionsprofile analysieren.

Im Modellorganismus *Synechocystis* 6803 identifizierten wir so insgesamt 371 potenzielle *trans*-agierende kleine ncRNAs (sRNAs) [2], von denen 60 Prozent in *Synechocystis* 6714 konserviert und 26 Prozent unter einer spezifischen Bedingung induziert sind [3]. Durch eingehende Analysen der die sRNAs umgebenden Transkriptionslandschaft identifizierten wir 13 Vorkommen einer bisher nicht beschriebenen chimären sRNA/mRNA-Klasse, die wir aufgrund der potenziell transkriptionsaktivierenden Rolle der sRNA-Komponente *Actuaton* taufen [3]. In einem *Actuaton* ist eine sRNA-Komponente vor einem Protein-codierenden Gen ins Genom integriert und kann durch seine Präsenz maßgeblich die Expressionsstärke des nachfolgenden Gens beeinflussen. Zudem wurden mittels vergleichender bioinformatischer Analysen insgesamt 21 bisher nicht annotierte offene Leserahmen (ORFs) identifiziert, die mit hoher Wahrscheinlichkeit für kleine Proteine codieren. Einige dieser ORFs, beispielsweise *nsiR6* und *hliR1*, sind unter bestimmten Stressbedingungen stark induziert und stellen interessante Kandidaten für eine tiefer gehende Charakterisierung dar (**Abb. 1**). Die unter den analysierten Bedingungen transkribierten Genombereiche nehmen 97 Prozent des Genoms von *Synechocystis* 6803 ein. Damit ist das Genomarrangement, wenn man nicht-codierende Elemente und kleine Proteine miteinbezieht, sogar noch deutlicher kompakter als ursprünglich angenommen.

Zusätzlich zur Identifizierung bisher unentdeckter Genomelemente unterstützt die hier durchgeführte vergleichende Transkriptomanalyse unter verschiedenen Wachstumsbedingungen auch maßgeblich die funktionelle Genomanalyse, da sie insbesondere die Auswahl interessanter Gene für eine umfangreiche Charakterisierung erheblich vereinfacht und somit eine wichtige Ressource für experimentell arbeitende Wissenschaftler darstellt.

### Danksagung

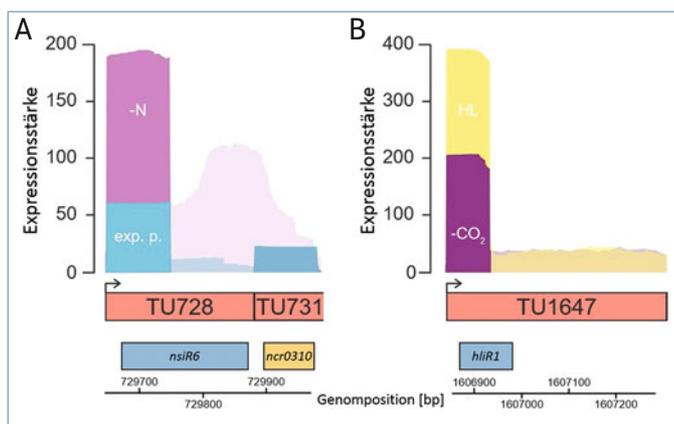
Ich danke allen ehemaligen Kollegen und Kooperationspartnern für die tolle Zusammenarbeit, insbesondere Stephan Klähn und Björn Voß, mit denen ich an vielen Projekten erfolgreich gearbeitet habe. Auch meinem Doktorvater Wolfgang Hess, der mich während all der Jahre immer hervorragend betreut und gefördert hat, gilt mein besonderer Dank. ■

### Literatur

- [1] Kaneko T, Sato S, Kotani H et al. (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res* 3:109–136
- [2] Kopf M, Klähn S, Scholz I et al. (2014) Comparative analysis of the primary transcriptome of *Synechocystis* sp. PCC 6803. *DNA Res* 21:527–539
- [3] Kopf M, Klähn S, Scholz I et al. (2015) Variations in the non-coding transcriptome as a driver of inter-strain divergence and physiological adaptation in bacteria. *Sci Rep* 5:9560

### Korrespondenzadresse:

Dr. Matthias Kopf  
Molecular Health GmbH  
Kurfürsten-Anlage 21  
D-69115 Heidelberg  
Tel.: 06221-43851-2337  
matthias.kopf@molecularhealth.com



◀ **Abb. 1:** Innerhalb der ursprünglich als sRNA vorhergesagten Gene *nsiR6* (A) und *hliR1* (B) wurden mittels vergleichender Analysen zwei konservierte kleine Proteine identifiziert, die spezifisch unter Stickstoffmangel (*nsiR6*) und Starklicht (*hliR1*) induziert sind. Zusätzlich zu diesen Bedingungen ist jeweils die Wachstumsbedingung mit der zweithöchsten Expressionsstärke angezeigt. Die abgebildeten Bedingungen sind Stickstoffmangel (-N), exponentielle Phase (exp. p.), Starklicht (HL) und CO<sub>2</sub>-Mangel (-CO<sub>2</sub>). Die Expressionsstärke ist als Quadratwurzel der normalisierten Sequenzierentiefe angegeben, und die x-Achse zeigt die Genomposition der dargestellten Region an. Die Expressionsdaten der ersten 100 Nukleotide sind nach vorhergesagten Transkriptionsstartstellen farblich hervorgehoben.