Molecular Systems Biology 5; Article number 305; doi:10.1038/msb.2009.63
Citation: *Molecular Systems Biology* 5:305
© 2009 EMBO and Macmillan Publishers Limited All rights reserved 1744-4292/09
www.molecularsystemsbiology.com

molecular systems biology

Evidence for a major role of antisense RNAs in cyanobacterial gene regulation

Jens Georg[1], Björn Voß[1], Ingeborg Scholz[1], Jan Mitschke[1], Annegret Wilde[2] and Wolfgang R Hess[1,*]

[1] Faculty of Biology and Freiburg Initiative in Systems Biology, University of Freiburg, Freiburg, Germany and [2] Justus-Liebig University Giessen, Institute of Microbiology and Molecular Biology, Heinrich-Buff-Ring, Giessen, Germany
* Corresponding author. Faculty of Biology and Freiburg Initiative in Systems Biology, University of Freiburg, Schänzlestr. 1, Freiburg 79104, Germany. Tel.: +49 761 2032796; Fax: +49 761 2036996; E-mail: wolfgang.hess@biologie.uni-freiburg.de

Received 22.12.08; accepted 3.8.09

Information on the numbers and functions of naturally occurring antisense RNAs (asRNAs) in eubacteria has thus far remained incomplete. Here, we screened the model cyanobacterium *Synechocystis* sp. PCC 6803 for asRNAs using four different methods. In the final data set, the number of known noncoding RNAs rose from 6 earlier identified to 60 and of asRNAs from 1 to 73 (28 were verified using at least three methods). Among these, there are many asRNAs to housekeeping, regulatory or metabolic genes, as well as to genes encoding electron transport proteins. Transferring cultures to high light, carbon-limited conditions or darkness influenced the expression levels of several asRNAs, suggesting their functional relevance. Examples include the asRNA to *rpl1*, which accumulates in a light-dependent manner and may be required for processing the L11 r-operon and the SyR7 noncoding RNA, which is antisense to the *murF* 5′ UTR, possibly modulating murein biosynthesis. Extrapolated to the whole genome, ∼10% of all genes in *Synechocystis* are influenced by asRNAs. Thus, chromosomally encoded asRNAs may have an important function in eubacterial regulatory networks.
Molecular Systems Biology **5**: 305; published online 15 September 2009; doi:10.1038/msb.2009.63
Subject Categories: functional genomics; RNA
Keywords: antisense RNA; cyanobacteria; microarray; noncoding RNA; *Synechocystis*

This is an open-access article distributed under the terms of the Creative Commons Attribution Licence, which permits distribution and reproduction in any medium, provided the original author and source are credited. This licence does not permit commercial exploitation or the creation of derivative works without specific permission.

Introduction

Bacteria, as well as eukaryotes, possess a significant number of regulatory RNAs. Eubacterial regulatory RNAs mainly control mRNA translation or decay, but some also bind proteins and thereby modify protein function (for reviews see Gottesman, 2004; Urban and Vogel, 2007). The majority of eubacterial regulatory RNAs are encoded at genomic locations far away from their target genes and exhibit only partial base complementarity to their mRNA targets. However, a small number of regulatory RNAs are transcribed from the reverse complementary strand of an annotated gene and hence these fully or partially overlap with their potential targets (*cis*-encoded regulatory RNAs). It was known early on that such natural antisense RNAs (asRNAs) control phage development and plasmid replication in bacteria (Wagner and Simons, 1994), yet recent work has made much more progress on *trans*-encoded regulatory RNAs. In several eukaryotic model organisms, it was found that the main transcriptional output from their genomes is noncoding RNA (ncRNA). Sense/antisense transcript pairs occur frequently in mammalian genomes (Katayama *et al*, 2005) and asRNAs were found opposite 1555 genes during high-resolution transcript screening of the yeast genome (David *et al*, 2006). It is now estimated that asRNAs or overlapping transcripts from adjacent transcriptional units exist for ∼22–26% of annotated genes in the human genome (Yelin *et al*, 2003; Chen *et al*, 2004; Zhang *et al*, 2006), for 14.9–29% of mouse genes (Okazaki *et al*, 2002; Kiyosawa *et al*, 2003; Katayama *et al*, 2005; Zhang *et al*, 2006), 15.4–16.8% of *Drosophila* genes (Zhang *et al*, 2006), and 8.9% of *Arabidopsis thaliana* genes (Jen *et al*, 2005; Wang *et al*, 2005).

Despite the earlier reported examples of antisense transcripts in prokaryotes, experimental evidence for a more general role of chromosomally encoded asRNAs in eubacteria has remained scarce. Using a tiled microarray and a protocol optimized for detection of sRNAs, two asRNAs to transposase genes, and three ncRNAs overlapping a substantial part of an mRNA or of another ncRNA were reported in *Caulobacter* (Landt *et al*, 2008). On the other hand, Selinger *et al* (2000) found a very high number of potential asRNAs in *Escherichia coli* by using Affymetrix microarrays with an inverted probe set capable of detecting antisense transcription. Although not corroborated by independent experiments, this array detected

antisense transcription for ~3000–4000 genes, suggesting that there is a low level of transcription virtually throughout the *E. coli* genome (Selinger *et al*, 2000). More recently, evidence for 127 putative asRNAs in *Vibrio cholera* was obtained through parallel sequencing (Liu *et al*, 2009) but these asRNAs were not further studied. There is only one publication describing the biocomputational prediction of asRNAs in bacteria (Yachie *et al*, 2006). On the basis of a combination of promoter and rho-independent terminator prediction, 87 ncRNA and 46 asRNA candidates were predicted for *E. coli*. Of these, eight ncRNAs and four asRNAs could be verified experimentally.

In cyanobacteria, evidence from earlier work indicated a function of chromosomal *cis*-encoded asRNAs in the regulation of gene expression. The asRNA IsrR in *Synechocystis* sp. PCC 6803 (from here: *Synechocystis*) regulates the accumulation of the *isiA* mRNA, and thereby controls the amount of IsiA protein and finally, protein–chromophore light harvesting complexes in cyanobacterial cells under iron limitation and redox stress (Duehring *et al*, 2006). A transcript complementary to the transcription factor *furA* mRNA was found in the filamentous cyanobacterium *Anabaena* PCC 7120. The *furA* asRNA originates by read-through from the adjacent gene alr1690 encoding a putative cell wall protein (Hernandez *et al*, 2005) and covers *furA* over its full length. Interrupting read-through from alr1690 resulted in an increased expression of FurA, thus the asRNA contributed in determining cellular levels of the protein. Other, less characterized, examples of asRNAs in cyanobacteria include a *cis*-encoded asRNA starting from the 3′ end of the gas vesicle gene *gvpB* and ending within the *gvpA* gene of the filamentous *Calothrix* PCC 7601 (Csiszar *et al*, 1987), and 24 asRNAs found by microarray hybridization in the marine unicellular *Prochlorococcus* MED4 (Steglich *et al*, 2008). In addition, there is a growing number of publications that hint at the impact of regulatory RNA in cyanobacteria without providing molecular details (Nakamura *et al*, 2007; Dienst *et al*, 2008; Voss *et al*, 2009).

Here, a computational search was implemented for the 3.6 Mb genome of *Synechocystis* to find such RNAs. To test the existence of predicted candidates efficiently, a tiling microarray was designed, in which all genome regions containing predicted regulatory RNAs were covered, together with a control set of the same size. Focusing on high scoring as well as on randomly selected candidates for asRNAs, 28 asRNAS were verified independently by 5′ RACE (rapid amplification of cDNA ends) and Northern blot analysis (Table I). Among the targets possibly influenced by these asRNAs are mRNAs for ribosomal proteins, mRNAs for enzymes of primary metabolism as well as for proteins that are involved in signal transduction and electron transfer.

Results

Large-scale analysis using a tiling microarray

A tiling microarray was developed, covering all genes and intergenic regions for which a terminator, and thus a candidate asRNA or ncRNA, was predicted. As a control set, probes were designed for genes and intergenic regions without a prediction, covering approximately the same total size. The resulting 102 739 probes amount to an accumulated length of 1 441 146 nt in tiled probes in both orientations, which represent ~40% of the chromosome. The arrays were hybridized in quadruplicates with pooled RNA from nine different conditions, such as exponential and stationary growth phase and different stress conditions (high light (HL), low light, 12 h incubation in the dark, iron and nitrogen depletion, heat and cold stress), to detect those transcripts, which are only induced under specific conditions. To avoid labeling artifacts from reverse transcription and second strand synthesis during cDNA synthesis (Perocchi *et al*, 2007), we labeled the RNA directly for microarray hybridization. Two additional microarrays were hybridized with genomic DNA and used for the normalization of signal intensities from individual probes as described by Huber *et al* (2006). The mapping of transcribed segments was carried out according to Huber *et al* (2006) yielding ~2500 transcript segments with arbitrary expression values from −5 to +10 (see Supplementary information 'Segmentation2500_final.pdf'). As evidence for low-level transcription of virtually every part of a bacterial genome has been provided (Selinger *et al*, 2000), we established a robust threshold at +1.0, leaving 646 transcript segments for closer inspection. As a positive control, IsrR (Duehring *et al*, 2006) was detected as one contiguous segment of the array (Figure 1 and Table I). The mapped 5′ end of IsrR is located 5 nt from the 5′ end of the transcript segment identified in the microarray, whereas its 3′ end is located 4 nt before the end of the last responding probe. These numbers yield a segment length of 186 nt compared with the fine-mapped asRNA length of 177 nt (Duehring *et al*, 2006), which is an excellent correlation for the chosen tiling factor. In the 20 kb genomic region, which also gives rise to the IsrR/*isiA* transcript pair, two further asRNAs were detected. The affected genes (as_sll1586 and as_ndhH) code for an unknown protein and NADH dehydrogenase subunit 7, respectively (Figure 1).

From the 646 transcript segments above the expression threshold of +1.0, 432 corresponded to mono-, di-, and multicistronic mRNAs, whereas 60 originated from intergenic regions and were considered ncRNAs and 73 at least partially overlap sense transcripts and therefore were designated asRNAs (see Supplementary Table S1 for details). We also detected transcripts, which likely represent short mRNAs (labeled 'new ORF' in Supplementary Table S1) and are not included in the numbers of the candidate asRNAs and ncRNAs, nor the segments representing putative 5′ and 3′ UTRs (Figure 2). In all, 28 asRNA candidates (Table I) and seven putative ncRNAs (Table II) were chosen for further analysis by Northern blot hybridization and 5′ RACE. Furthermore, we determined the distribution of medium-level-expressed segments (expression value from +0.99 to 0.0). This group contains 542 segments, among them 389 mRNA segments, 51 UTRs, 84 putative asRNAs and 18 putative ncRNAs (Figure 2).

Synechocystis transcripts expression levels

The 15 most highly accumulating mRNAs (see Supplementary Table S1) in our tiling microarray originate from an intron-located endonuclease gene (slr0915), the photosynthetic genes *psaAB* (slr1834/slr1835), *psbD2* (slr0927), *psbD* (sll0849),

© 2009 EMBO and Macmillan Publishers Limited

Table I Top-scoring antisense RNAs from the prediction and microarray analysis

Array segment		Mean		S	Annotation/reference	TSS	Northern	FC		
Start	Stop	asRNA	mRNA					Dark	HL	$-CO_2$
2269032	2269133	7.835	0.971	–	Internal as_slr0320 (t), cf. Figure 5	c2269144	100	–	–	–
1518029	1518214	7.003	5.878	+	Internal as_isiA (IsrR; Duehring *et al*, 2006) (t)	1518034	177	–	–	–
89735	89820	6.302	1.557	+	Internal as_sll1049, cf. Figure 5	neg.	90	–	–	–
2706752	2706984	5.14	1.51	–	as_rlpA (slr0423) (5′overlap), cf. Figure 5	c2706939	160	–	−3.04 ± 0.15	–
2859925	2860056	4.334	2.28	–	as_ndhF1 (slr0844) 2 segments (3′overlap), cf. Figure 5	c2860111 and c2860313	700	+ 3.21 ± 0.27	–	–
2859617	2859910	2.017								
924638	924780	4.201	4.365	+	as_rpl1 (sll1744) (5′overlap) (t), cf. Figure 8	924448	150	–	−3.84 ± 0.23	–
166860	167110	4.191	2.21	+	Internal as_sll0217 (t), cf. Figure 5	166849	250	–	–	–
1504253	1504333	4.01	0.447	+	Internal as_sll1586, cf. Figure 1	1504239	90	–	–	–
2163153	2163249	3.787	−0.308	–	Internal as_slr0408	c2163253	130	−2.37 ± 0.15	–	−2.06 ± 0.15
2823667	2824068	3.699	0.497	–	as_slr0580 (5′overlap), cf. Figure 5	c2823987	600	–	–	+ 1.95 ± 0.12
136748	136889	3.597	2.527	–	Internal as_infB (slr0744)	c136871	160	–	–	–
1992384	1992520	3.542	1.003	–	Internal as_pknA (slr1697), cf. Figure 5	c1992722	65	–	−2.0 ± 1.26	–
3465167	3465285	3.257	0.376	+	as_ppx (sll1546) (5′overlap), cf. Figure 5	neg.	250	–	−1.89 ± 0.27	+ 1.9 ± 0.05
3565927	3566261	3.17	4.747	–	Internal as_lepA (slr0604), cf. Figure 7	c3566241	380	–	–	–
3439416	3439581	2.99	−0.224	+	Internal as_sll0723, cf. Figure 5	3439412	170	–	–	–
865872	865924	2.907	2.099	+	hik3 (as_sll1124) (5′overlap)	865932	700	–	–	–
1510816	1511161	2.844	2.78	–	as_ndhH (slr0261) (3′overlap), cf. Figure 1	c1511138	220	–	–	–
1283031	1283175	2.705	0.015	+	as_sppA (sll1703) (5′overlap)	1283002	> 500	+ 1.90 ± 0.21	–	–
2512367	2512605	2.594	0.146	+	as_rfbA (sll0207) (3′overlap) (t), cf. Figure 5	2512327	550	–	–	–
1143957	1144210	2.253	0.506	–	as_slr0882 (3′overlap), cf. Figure 6	c1144439	450	–	+ 2.54 ± 0.37	–
3198747	3198980	2.179	1.656	–	as_hemE (slr0536) (5′overlap)	c3198959	500	–	–	–
695587	695865	2.162	0.759	+	Internal as_sll1289, cf. Figure 6	695567	250	–	–	–
1768770	1769050	1.853	0.518	+	as_ribA (sll1894) (3′overlap)	neg.	> 1000	–	−1.86 ± 0.44	–
198819	199396	1.54	0.816	–	Internal as_slr1102, cf. Figure 5	neg.	400	–	–	–
819499	819977	1.494	5.452	+	as_tktA (sll1070) (5′overlap), cf. Figure 7	819725	200	–	−2.08 ± 0.03	–
2216528	2217002	1.27	0.342	+	as_sll1864 (3′overlap) + orf	2215955	> 650	–	−4.33 ± 0.46	+ 2.4 ± 0.08
3207282	3207602	1.22	−0.164	+	as_sll0503 (3′overlap), cf. Figure 7	3207223	380	–	–	+ 2.72 ± 0.83
2422099	2422678	1.1	0.169	+	Internal as_sarA (sll0750), cf. Figure 5	2422099 and 2422045	350	–	–	–

The start and stop positions of hybridizing segments within the *Synechocystis* chromosome are based on tiling microarray data and are shown together with the average expression (mean) for the asRNA and for the respective mRNA calculated from the hybridization of RNA pooled from nine different conditions in quadruplicates. The orientation of the asRNA locus in the genome is given (S) and the annotation, including the classification as internal, 5′ or 3′ overlapping asRNA. 't' indicates that this asRNA was predicted based on a possible terminator structure. Precise 5′ ends of asRNAs were determined by 5′ RACE analysis (TSS). Size of major asRNA bands in Northern hybridizations are in nt. Moreover, fold changes (FC) are indicated under three different conditions from the expression arrays (if no FC: –). See supplementary Table S1 for the complete list of asRNAs.

© 2009 EMBO and Macmillan Publishers Limited

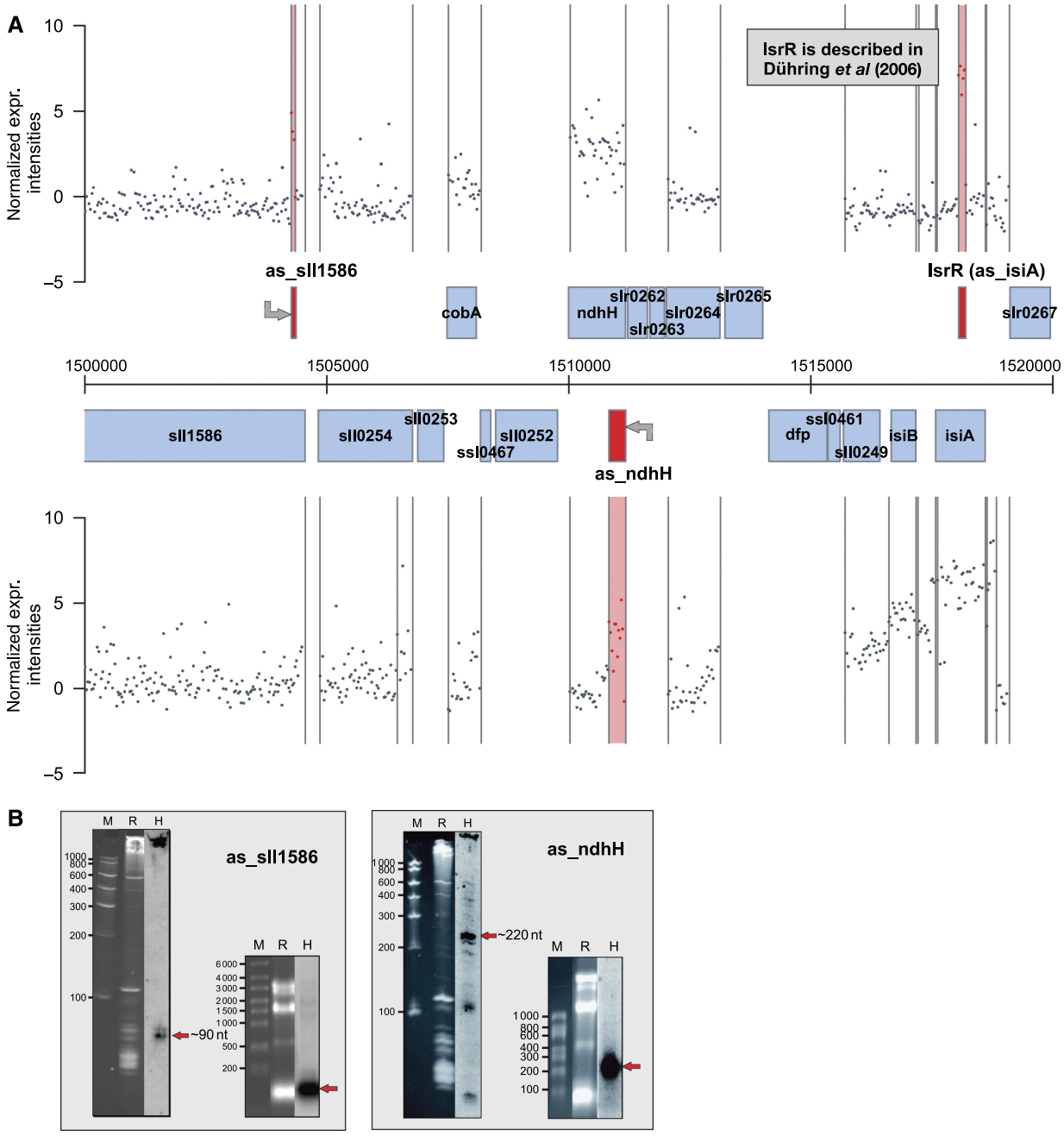


Figure 1 Example for verification of microarray-detected asRNAs in a 20 kb region of the *Synechocystis* genome, from coordinate 1 500 000–1 520 000. (**A**) Individual probes are indicated by dots, sets of probes with similar absolute expression levels were joined into contiguous segments, separated from each other and from regions not covered by the array by vertical lines (for the full data set see Supplementary information 'Segmentation2500_final.pdf'). Annotated protein-coding genes are represented by blue boxes. At least three clearly detectable asRNAs (segments in red) originate in this region: IsrR (Duehring *et al*, 2006), an ~90 nt asRNA to sll1586 and an asRNA to *ndhH* (slr0261). (**B**) Northern blot hybridizations based on high-resolution polyacrylamide gels and agarose gels. For each asRNA the hybridization (H), the corresponding lane in the RNA electrophoresis (R) and a molecular mass marker (M) is shown. As an additional experimental control, 5′ ends of the two new asRNAs were mapped by 5′ RACE to positions 1504239 (as_sll1586) and c1511138 (as_ndhH), providing a third line of evidence for the existence of these asRNAs (see also Table I).

psbT (smr0001), and *rbcL* (slr0009), the cell division cycle gene slr0374, the *groESL* operon (slr2075_slr2076), the genes slr0742, sll0524, sll0623, and slr1667, the RNA-binding protein A gene *rbpA* (sll0517), the molybdopterin biosynthesis gene *moeA* (slr0900), as well as the iron-stress-induced protein A gene *isiA* (sll0247). We found 14 ncRNAs and 4 asRNAs within the same range of expression levels. These asRNAs are opposite to *isiA*, slr0320, sll1121, and sll1049 (Supplementary Table S1). Finding stress-induced genes such as *isiA* among the top-expressed genes is not an artifact, but results from the fact that we hybridized pooled RNA samples from cultures grown under nine different conditions.

© 2009 EMBO and Macmillan Publishers Limited

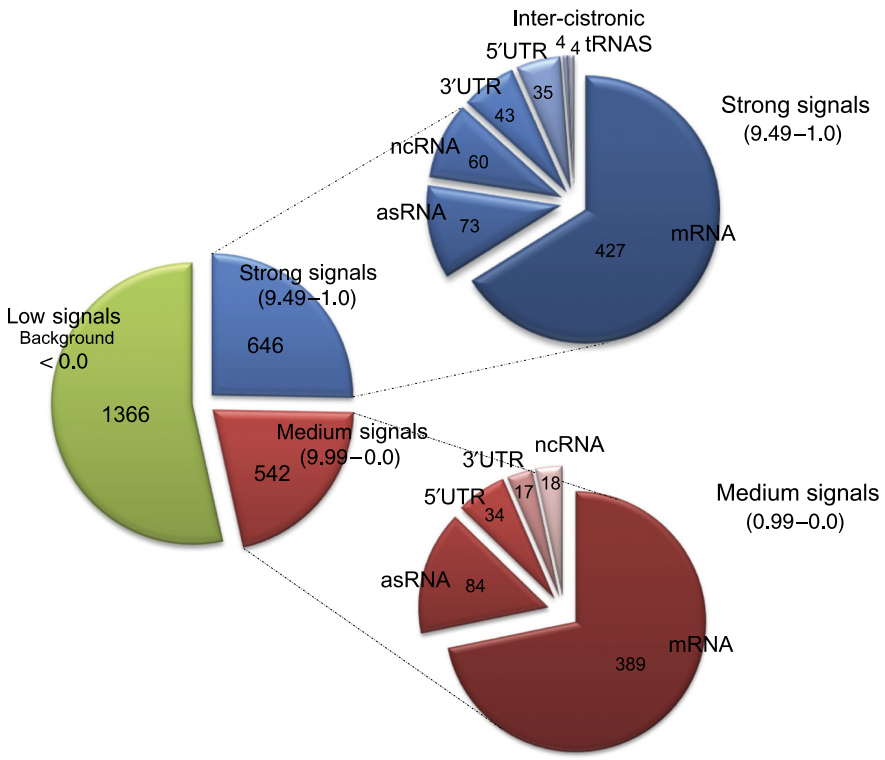


Figure 2 Composition of the population of high- and medium-scoring transcript segments. Distribution of the 646 segments with a mean expression value in the top third group of expression signals and 542 medium scoring segments among different classes of RNA molecules. For details of the annotation of these segments see Supplementary Table S1.

Table II Selected new or confirmed ncRNAs

Array segment		TSS	Annotation	Prediction	Mean	Strand	FC			References
Start	Stop						Dark	HL	$-CO_2$	
2960896	2960952	2960898	Yfr1	t, c	9.4183	+	–	–	–	Voss *et al* (2007)
1832218	1832334	1832234	SRP RNA ffs	None	9.3047	+	–	–	–	RFAM
3138669	3138773	c3138743	SyR5	t	8.7715	–	–	–	–	This study
2730501	2730626	2730523	Yfr2b	c	8.3092	+	+ 2.0 ± 0.02	–	–	Voss *et al* (2009)
1671897	1672056	1671919	SyR1	t, c	7.1891	+	–	+ 8.8 ± 0.59	+ 3.8 ± 0.1	Voss *et al* (2009)
1816523	1816625	c1816602	SyR6	t	6.9568	–	–	+ 3.0 ± 0.2	–	This study
1518643	1518856	c1518816 and c1518836	5′UTR *isiA* (sll0247) and ncRNA	None	6.6058	–	–	–	–	Duehring *et al* (2006) and this study
1431936	1431981	1431853	SyR2	t, c	5.9175	+	−2.6 ± 1.6	−7.0 ± 0.2	−3.9 ± 0.25	Voss *et al* (2009)
2512366	2512425	c2512423	SyR9	None	5.5599	–	–	−1.9 ± 0.17	–	This study
1748948	1749130	c1749138	SyR7	t	5.3804	–	–	−7.66 ± 1.85	–	This study
106687	106838	c106808	SyR8	t	3.9673	–	–	–	–	This study
727707	728258	c728041	SyR4	t, c	1.1912	–	–	–	+ 1.9 ± 0.22	This study
727492	728273	727885 and 728053	SyR3	t	0.4828	+	–	–	–	This study

The start and stop positions of hybridizing segments within the *Synechocystis* chromosome are based on tiling microarray data, TSS as determined by 5′ RACE or taken from the references. The list has been sorted according to the average expression signals in the tiling microarray experiment (mean). Prediction: 't' indicates prediction based on possible terminator, 'c', predicted in comparative analysis (Voss *et al*, 2009). The fold changes (FC) under three conditions were calculated from the expression microarray. See supplementary Table 'Synarray.xls' for the complete overview.

Assessing the reliability of the prediction strategy

The transcription of many bacterial genes, and thus also of ncRNAs and asRNAs, finishes at a rho-independent terminator, which can be computationally predicted (see Materials and methods). Our terminator prediction identified 713 putative transcripts within all non-annotated sequences (intergenic and antisense). Assuming an average transcript length of 300 nt, ~20% were completely intergenic (ncRNA candidates),

© 2009 EMBO and Macmillan Publishers Limited

whereas ~80% were antisense to an annotated gene. The iron stress regulated asRNA IsrR (Duehring *et al*, 2006), as well as the small ncRNAs Yfr1 (Axmann *et al*, 2005; Voss *et al*, 2007), SyR1, and SyR2 (Voss *et al*, 2009), were among the predicted transcripts, indicating the reliability of this procedure. To evaluate the performance of the prediction strategy further, we compared its outcome against the results from the tiling microarrays. As the segmentation procedure could be erroneous in itself, we took the following approach: for each predicted terminator, we computed the mean normalized expression of probes within four 100 nt long segments, starting from the 5′ end of the terminator. For expression cut-offs ranging from 0 to 9, the number of terminators passing it was computed. Two background sets (one antisense-only, and one freely distributed) of randomly chosen segments of size 100 nt were handled the same way. Altogether, the analyses showed that there is a clear tendency of regions close to predicted terminators to have a higher mean expression. This is even more pronounced in the antisense-only analyses (Supplementary Figure S1).

In absolute numbers, 11 out of 73 asRNAs and 27 out of 60 intergenic ncRNAs with a microarray expression level of at least +1, have been predicted here, based on the presence of a rho-independent terminator (Table II; Supplementary Table S1), including five ncRNAs reported earlier in a comparative genomics study (Voss *et al*, 2009). Examples for false-negatives include SyR9, the 5′ UTR of the *isiA* gene that accumulates in large quantities as an ~160 nt small RNA (Duehring *et al*, 2006) and *ffs*, the ncRNA of the signal recognition particle (Table II). If all 60 segments identified in the array were real ncRNAs, the true-positive rate of the terminator-based prediction for this class of RNA molecules would be ~45%. The higher true-positive rate for ncRNAs is reflected in their better terminator scores. In Figure 3, the free energy of the stem-loop (ΔG_S) and the hybridization energy of the DNA/RNA-hybrid (ΔG_H) in the transcribing RNA polymerase holoenzyme are plotted against each other for all predicted terminators. Good terminators are expected to have a strong hairpin (low ΔG_S) pushing the polymerase from the DNA to which it is relatively weakly bound (high ΔG_H). Overall, Figure 3 shows that predicted intergenic terminators are evenly distributed, whereas candidate antisense terminators tend to accumulate in the low scoring area (lower right corner). The same tendency is observed for the terminators of the independently verified asRNAs and ncRNAs. Intriguingly, the terminator for IsrR appears as one of the worst. These results suggest that terminators of antisense transcripts are not evolutionary optimized to the same degree as the other terminators, perhaps because of an influence of the coding strand sequence. Thus, antisense terminator predictions appear less sensitive and less specific than for terminators located in intergenic spacer regions.

Finding new intergenic ncRNAs

In total, our data revealed 60 segments that represent possible ncRNA genes within the total set of high scoring transcripts (Supplementary Table S1). Among these are the known ncRNAs Yfr1 (Voss *et al*, 2007), Yfr2b, SyR1, and SyR2 (Voss *et al*, 2009). Additionally, seven ncRNA candidates were verified in Northern blot experiments (Figure 4; Table II). These were named SyR for ***Sy***nechocystis nc***R***NA and originate from the intergenic spacers between genes *cpcB*–ssr2848 (SyR3 and SyR4), sll0048–sll0737 (SyR5), *rps1* (slr1984)–*dnaG* (sll1868) (SyR6), sll0208–*rfbA* (sll0207) (SyR9), sll1247–*murF* (slr1351) (SyR7), and *llal.2* (sll0790)–*ksgA* (SyR8). The length range of these ncRNAs is 80–350 nt (Figure 4), a typical size distribution for bacterial ncRNAs.

New asRNAs

An overview of 73 different candidate asRNAs detected in our array is provided in Supplementary Table S1. With an average expression value of 7.8, an asRNA to slr0320 was the most

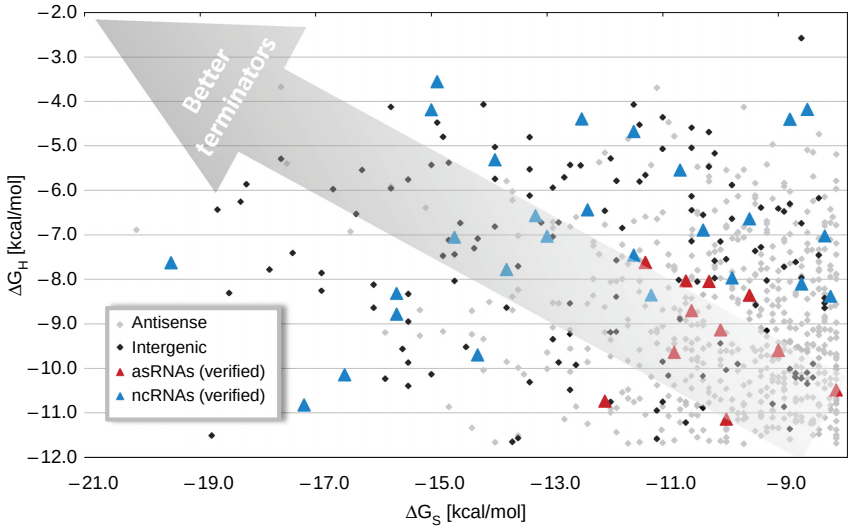


Figure 3 Distribution of terminator hairpin scores. Free energy of the terminator hairpin plotted against hybridization energy of the DNA/RNA-duplex. Good terminators are expected to appear in the upper left corner, having low values for the free energy of hairpins (ΔG_S) and high hybridization energy (ΔG_H).

© 2009 EMBO and Macmillan Publishers Limited

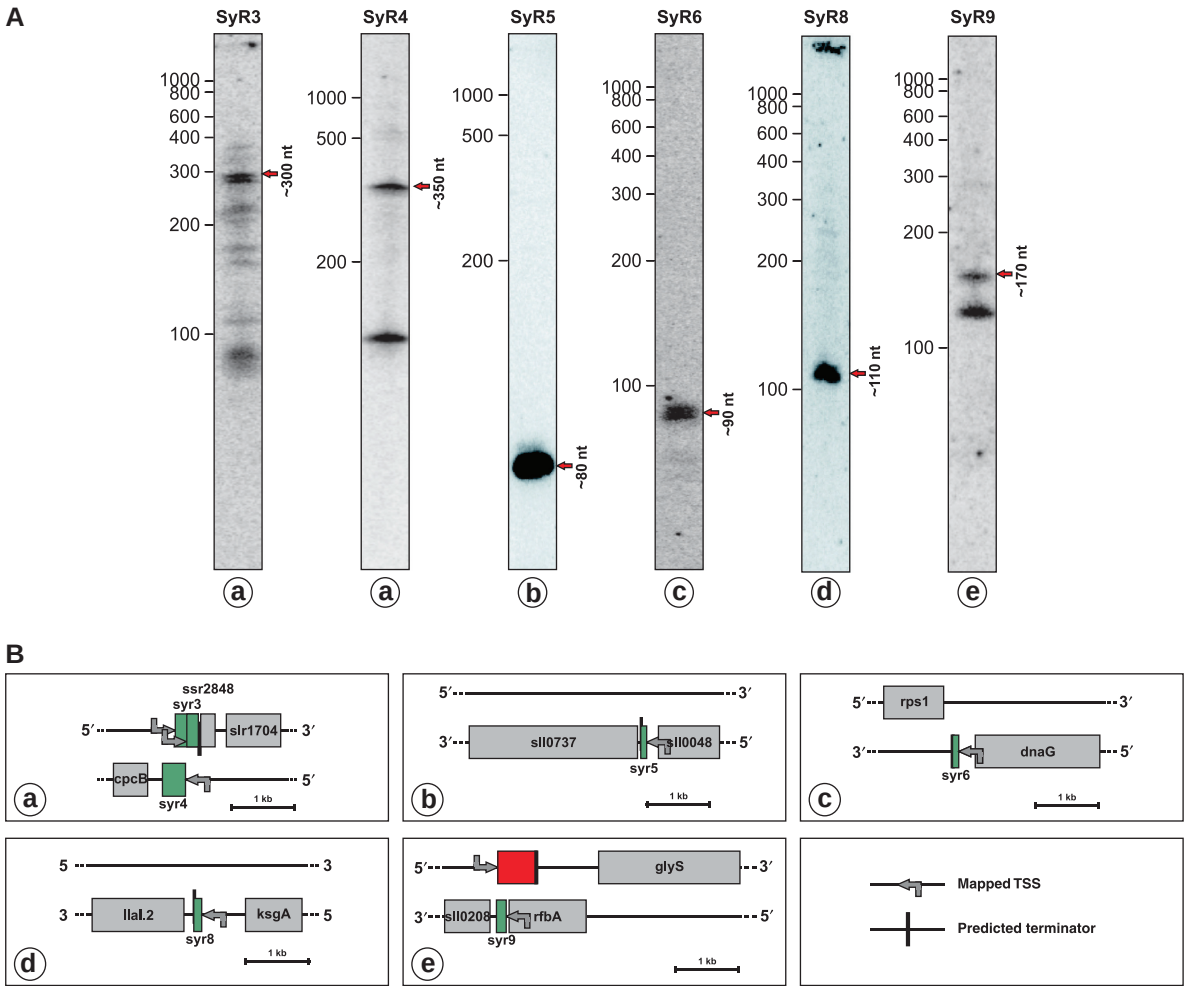


Figure 4 Detection of new ncRNAs by Northern blot hybridization. (**A**) For each ncRNA, the hybridization is shown resulting from separation on a high-resolution polyacrylamide gel. Red arrows indicate those signals corresponding to the segments in the tiling microarray in combination with the mapped TSS. (**B**) Representation of the ncRNA genes within the genome of *Synechocystis*. The forward and reverse strand is shown with confirmed ncRNA genes as green elements, protein-coding genes as grey boxes and asRNAs in red.

highly accumulated asRNA, followed by the asRNA IsrR (7.0), which served as the internal control (Duehring *et al*, 2006). The five next most highly expressed asRNAs have expression levels (4.2 to 6.3) similar to those of highly expressed protein-coding genes such as *amt1* (sll0108, 5.5) or *rbcL* (slr0009, 6.0).

We chose 28 asRNAs for independent verification by Northern blot analysis and 5′ RACE. The Northern data can be broadly divided into clear signals and more complex patterns observed for a subset of asRNAs, which may result from either co-degradation or co-processing with their corresponding mRNAs. Prominent examples are asRNAs to the flavoprotein gene sll0217, *rlpA*, slr0580, and *ndhF1* (Figure 5). Possible false-positives are to be expected predominantly among those 13 asRNA candidates whose existence was suggested only by one or two strongly responding oligonucleotides in the microarray (Supplementary Table S1). Indeed, in Northern hybridizations, two of these candidates (as_sll1121 and as_slr1552) showed in addition to a signal at ~90 nt a high molecular weight smear indicating potential cross-hybridization (not shown). No 5′ RACE signals were obtained for as_sll1049, as_*ppx*, as_slr1102, and as_*ribA*; in the latter case probably because of read-through from the adjacent gene slr1964. With as_ndhF1 and as_slr0882 we noticed two more examples for asRNAs, which included a short open reading frame as part of the transcript (Figures 5 and 6B). Nevertheless, these were counted here as asRNAs, as both have a substantial overlap with the respective mRNA 3′ ends and as long overlapping transcription has been shown to be effective in cyanobacteria (Hernandez *et al*, 2005).

Verification and characterization of newly found asRNAs by transcriptome microarrays

A novel transcriptome microarray was designed as an efficient tool for the verification and examination of possible regulation of the newly found asRNAs and ncRNAs. This array includes

© 2009 EMBO and Macmillan Publishers Limited

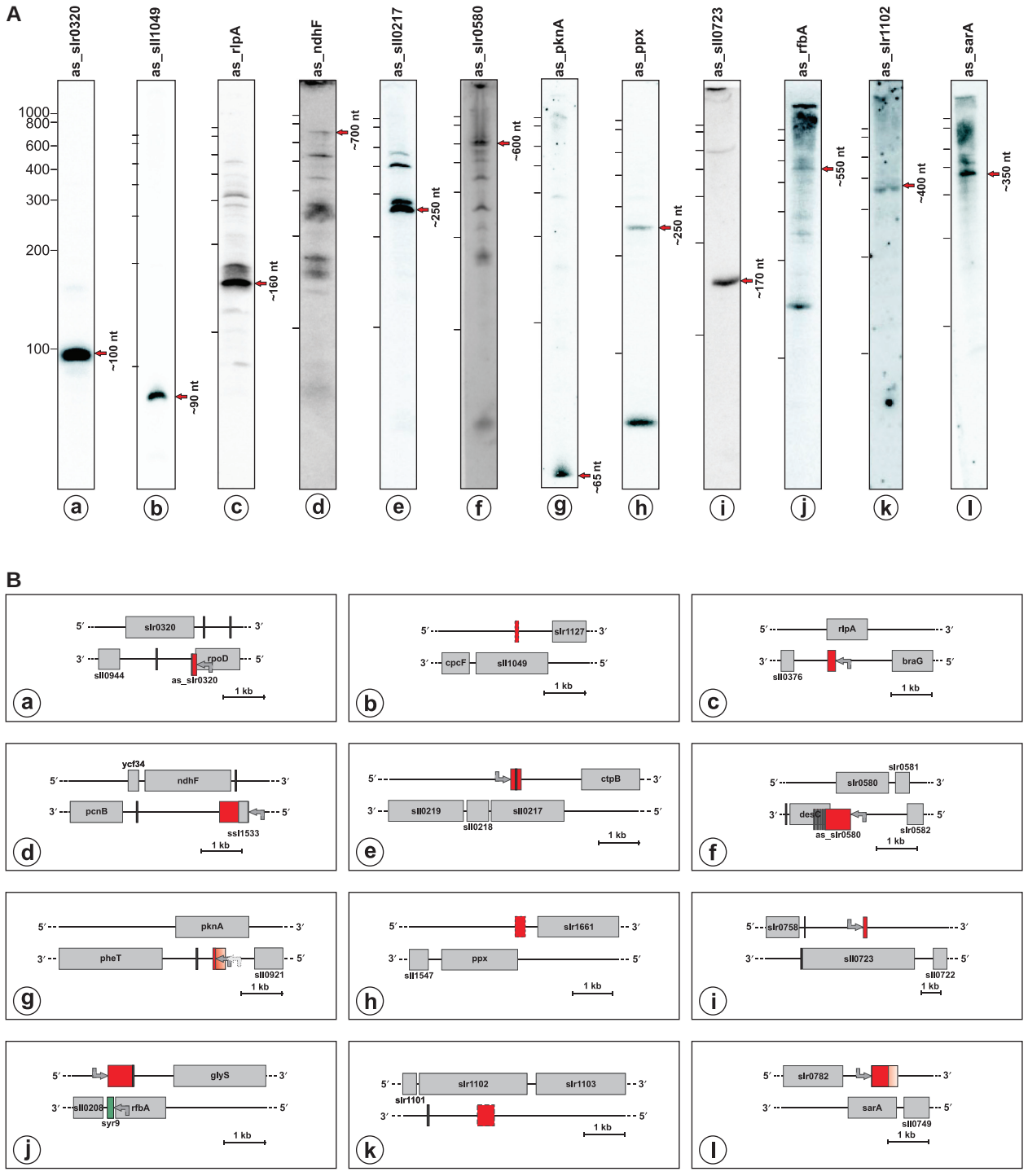


Figure 5 Selected examples of novel asRNAs. Validation of computational prediction and microarray analysis by representative Northern blot experiments and 5′ mapping. (**A**) For each of the 12 tested asRNAs the hybridization is shown. The positions of bands of a molecular mass marker are indicated by short bars. Red arrows indicate major products corresponding to microarray segments in combination with the mapped TSS. (**B**) Schematic drawing showing newly found asRNAs in red boxes (major signals in Northern blot, when possible mapped to the genome by microarray segmentation data) and light red boxes (weaker signals in Northern blot), intergenic spacer-located genes for ncRNAs in green. Predicted terminators are indicated by black vertical lines, mapped TSS by grey arrows, broken boxes indicate 5′ ends were not mapped. The origin of as_rfbA was mapped far into the sll0208_*rfbA* intergenic spacer. In this region it overlaps with yet another transcript, the small ncRNA SyR9 that accumulates as a doublet of ∼150/170 nt (Figure 4).

probe sets for all protein-coding genes as well as for all other transcripts, which we identified in the course of this study. Cultures were treated with three different stress conditions, which are highly relevant for a photosynthetic organism, namely HL, darkness and CO_2 depletion. The fold changes (FCs) in expression levels were measured for all ncRNAs,

© 2009 EMBO and Macmillan Publishers Limited

asRNAs and their cognate mRNAs in triplicates and can be found in Tables I and II. For six selected asRNA/mRNA pairs and for the SyR7 ncRNA, we confirmed the changes in expression levels by Northern blot hybridization (Figures 6 and 7). Here, SyR7 was included as a particularly interesting example. SyR7 appears as a *bona fide* ncRNA as it is intergenic over its full length. However, its transcriptional start site (TSS) was mapped to the reverse complementary strand only 6 nt upstream of the *murF* start codon. Consequently, we wondered as to whether SyR7 would overlap with the 5′ UTR of the *murF* gene (Malakhov *et al*, 1995). A TSS 206 nt upstream of *murF* was mapped, which indicates that this is indeed the case. This TSS has very recently been confirmed by another group (Hedger *et al*, 2009). Interestingly, we found the expression level of SyR7 more than 20 times higher than that of the *murF* mRNA under three different conditions. However, on a shift to HL, the SyR7/*murF* ratio declined dramatically to ~ 1 (Figure 6A). This change may be in part because of activation of the P1 promoter activated by the cAMP-dependent transcription factor SyCRP1 (Hedger *et al*, 2009), contributing to a slight increase in *murF* mRNA concentrations, by a factor of 2.34 ± 0.56 (Figure 6A). The light effect on the SyR7 steady state level was, however, much more pronounced as it declined to only 15% of its initial value (Figure 6A). Therefore, it is tempting to assume that under HL the *de novo* synthesis of MurF is required for an acceleration of cell wall biosynthesis, and that the main control is exerted at the level of *murF* translation through SyR7 whose expression becomes repressed in HL.

Characteristic changes were also obtained for the other asRNA/mRNA pairs studied in more detail. A situation inverse to SyR7 was observed with as_slr0882, which increases dramatically under HL and almost disappears in darkness (Figure 6B), whereas the slr0882 mRNA accumulation does not change significantly. The concentration of as_sll1289 is about equal to its cognate mRNA in darkness, whereas under all other conditions the amount of as_sll1289 appears higher than that of the sll1289 mRNA (Figure 6C). The concentrations of as_lepA appear higher than those of its mRNA under all conditions (Figure 7A). The internal asRNA to *lepA* may affect protein biosynthesis as *lepA* encodes ribosomal back translocase; a protein only recently recognized as a third essential bacterial elongation factor (Qin *et al*, 2006). Another situation is provided by as_sll0503, which appears in small amounts under control conditions and in darkness. On a shift to HL, the expression levels increased for both the mRNA sll0503 and its asRNA as_sll0503. However, if CO_2 was depleted only as_sll0503 went up, thus the resulting asRNA/mRNA ratio shifted to ~ 4 (Figure 7B). The amount of as_tktA is always considerably lower compared with the *tktA* mRNA. Yet, we observed characteristic changes within the hybridizing pattern of three narrowly spaced transcript bands (Figure 7C).

as_rpl1: a possible role in discoordinating gene expression

The ratio between as_rpl1 and the *rpl1* mRNA is close to 1 under all tested conditions, except under HL, where it declines to 0.2 (Figure 8). This asRNA overlaps with the 5′ end of ribosomal protein 1 (*rpl1*) mRNA, which belongs to the L11 ribosomal protein operon. In this operon, the adjacent genes *rpl1* and *rpl11* were found in microarray studies to become upregulated under HL in *Synechocystis* (Hihara *et al*, 2001). On the other hand, this operon is one of the best studied r-operons in *E. coli* and Rpl1 in particular has been characterized as a feedback translational regulator of overwhelming regulatory relevance for this operon (Yates *et al*, 1980; Branlant *et al*, 1981; Lindahl and Zengel, 1986). Thus, as_rpl1 could be involved in the specific processing of the precursor mRNA or in specific translational regulation of *rpl1* mRNA. Therefore, the impact of HL on the expression of as_rpl1 was studied in more detail. Using three different strand-specific probes, we monitored the expression of *rpl1*, *rpl11* and as_rpl1 during a shift from intermediate light to HL. As both mRNA probes detected a (weak) ~ 5300 nt and an ~ 3000 nt band, these probably represent the unprocessed operon precursor RNA and a specific processed fragment (Figure 8). Additionally, a single main product appeared for each probe: ~ 1800 nt for *rpl1* and ~ 1100 nt for *rpl11*. These transcript species probably represent the respective monocistronic mRNAs. Thus, a specific processing of the polycistronic precursor mRNA within the *rpl1-rpl11* intergenic spacer is indicated. As expected, a significant increase in their amount can be readily observed for both mRNAs after just 15 min in HL. Such an increase is not observed for as_rpl1. On the contrary, the amount of this asRNA decreases, albeit with a time-delay, as the minimum was observed 60 min after a shift to HL. It appears that the *rpl11-rpl1* mRNA precursor is converted into the monocistronic mRNA species in a time-delayed manner at the expense of as_rpl1.

Discussion

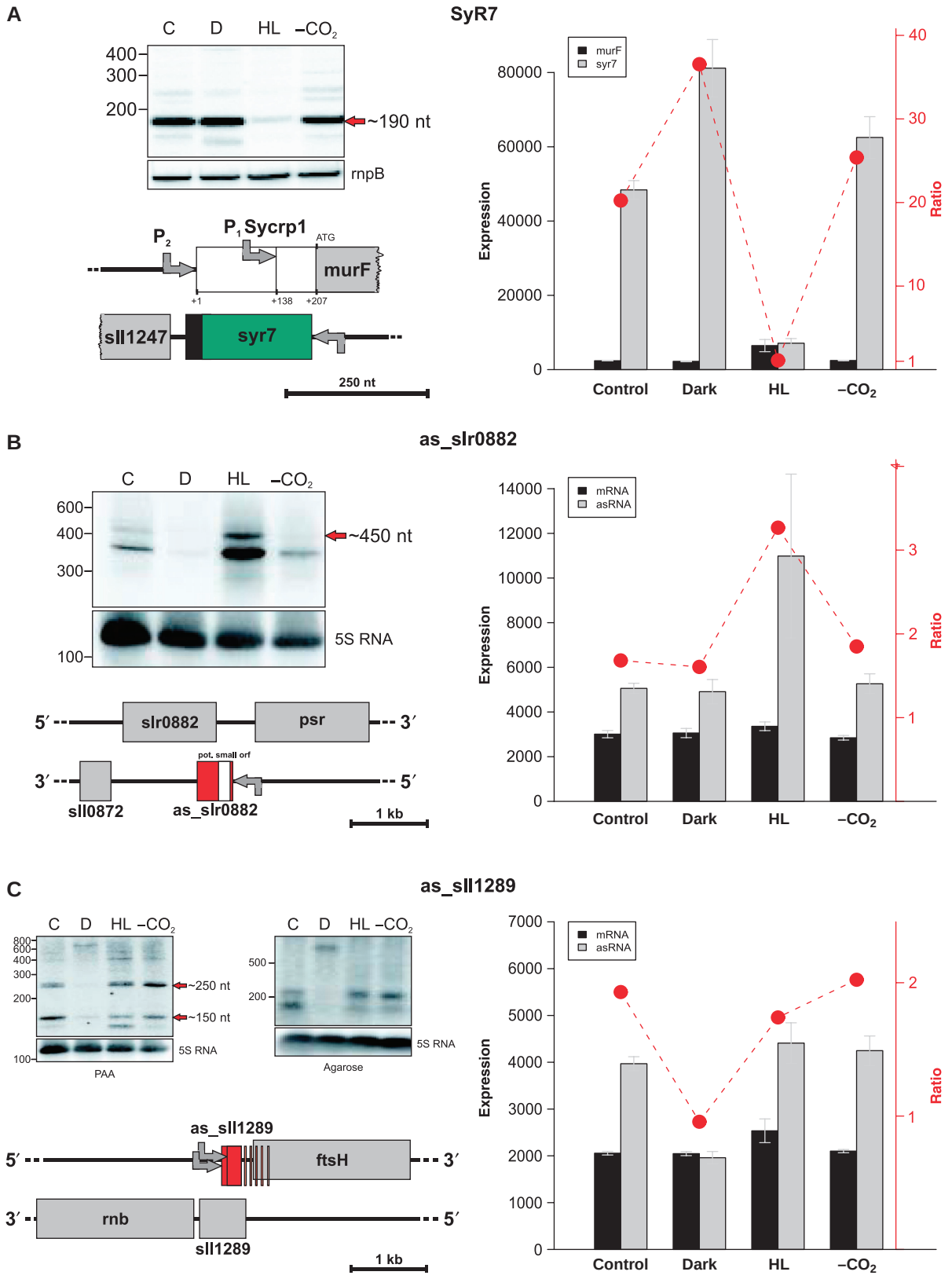
Identification of eubacterial asRNAs

Despite early reports on asRNAs in bacteria and phages (Wagner and Simons, 1994) a systematic screening for asRNAs in bacteria is missing. Here, we present a partial transcriptome analysis in the cyanobacterial model organism *Synechocystis*, combined with extensive verification, and provide first functional insight into the role of asRNAs. There are three main technical problems in dealing with antisense transcription in bacteria: (i) the general lack of robust algorithms to predict them; (ii) the high risk of measuring experimental artifacts generated during cDNA synthesis in microarray analyses (Perocchi *et al*, 2007); and (iii) a low level of transcription reported to occur virtually throughout the entire genome (Selinger *et al*, 2000), making it difficult to differentiate asRNA with a regulatory function from transcriptional noise.

Here, we have tried to overcome all three obstacles by (i) rigorously interrogating all predictions made in a computational approach using tiled microarrays. To overcome the problem of unintended second strand synthesis (ii) we labeled RNA samples directly before their hybridization on the microarray, and finally (iii) we focused predominantly on very highly expressed asRNAs.

© 2009 EMBO and Macmillan Publishers Limited

Computational screens have been used successfully for the prediction of ncRNAs in various eubacteria, but very rarely for finding asRNAs. Yachie *et al* (2006) presented a strategy that also predicts asRNAs, based on sequence patterns, nucleotide biases, and higher-order base relations, as they, for example, occur through basepairing in structured RNA molecules. This



© 2009 EMBO and Macmillan Publishers Limited

is reasonable for (intergenic) ncRNA prediction, yet it is less suitable for a prediction focusing on asRNAs, as these function mainly by complementarity rather than specific sequence and/or structure features.

Here, we found a correlation between the prediction and the actual presence of a terminator. However, based on the array results, the number of false-negative predictions turned out to be high. The predicted terminators come with the following parameters: free energy of the stem-loop ΔG_S, hybridization energy ΔG_H and a poly-U scoring. Comparing any single parameter or combination of parameters with the actual presence of a transcript did not indicate any particular correlation. The poor performance of the prediction for antisense transcripts may be explained by the existence of alternative termination signals (involving proteins similar to Rho, or RNA–RNA interaction (Stork *et al*, 2007)), or a lack of specific termination because of functional peculiarities, such as transcriptional interference (Sneppen *et al*, 2005). Moreover, the accumulation of asRNAs with secondary 3′ ends resulting from co-degradation or co-processing of asRNAs with their cognate mRNAs could, in some cases, also provide an explanation. Further work is required to differentiate between these possibilities.

Total number of asRNAs

Here, we found 73 candidates for *cis*-asRNAs and 60 free-standing genes for putative ncRNAs which all had an average expression of more than $+1.0$. With regard to mRNAs, such an expression threshold of $+1.0$ corresponded to the top third of the most-strongly expressed genes. The false-positive rate appears low in this candidate set. False-positives would be expected predominantly among those 13 asRNA candidates (18% of all) represented only by one or two probes in the microarray; however, further testing did not support this view. Nevertheless, if we conservatively assume a false-positive rate of 5% and a true-positive rate of 95% for the array-selected candidate asRNAs, 69 of the 73 asRNA candidates can be expected to exist. On the other hand, focusing on one third of the most-strongly accumulating transcripts leaves two thirds of the segments to be investigated. In fact, there is strong evidence to suggest that also less highly expressed asRNAs exist in *Synechocystis*. We selected exemplarily three possible asRNAs for the genes *uvrA*, *dnaX*, and *accA*, which were predicted based on the possible presence of a terminator but not found during autosegmentation of the array data. Their expression levels were also below the threshold of $+1.0$. These candidate asRNAs were detectable in Northern hybridizations (Supplementary Figure S2) and verified in 5′ RACE experiments. These three weakly expressed asRNAs accumulate to levels that correspond to the amounts of their respective mRNAs. The stoichiometric ratio between an asRNA and its respective mRNA is probably more important than the absolute accumulation level of the asRNA. Therefore, it appears valid to assume that an equal number of 69 asRNAs exists among the medium-expressed third of all transcripts as in the top third, suggesting 138 asRNAs from 40% of the bacterial chromosome. Extrapolated to the whole genome the resulting number of more than 300 chromosomally encoded asRNAs does not appear unlikely for a bacterial cell. Recently, evidence for 127 asRNAs was found by parallel sequencing in *Vibrio cholera* (Liu *et al*, 2009). Chromosomally encoded *cis*-asRNAs in *Synechocystis* are much more frequent than originally thought and seem to outnumber intergenic ncRNAs. With this conservative approximation taken into account, asRNAs may affect 8–10% of all genes in *Synechocystis*, a number that lies within the range of asRNAs in eukaryotic genomes.

Possible mechanisms of asRNA functions

If nearly every tenth open reading frame has an asRNA encoded on the opposite DNA strand, very complex regulatory circuits would be possible (Levine *et al*, 2007; Shimoni *et al*, 2007). We detected a large variety among the asRNAs in our study. The asRNAs can be classified by their transcript level, the mRNA/asRNA ratio, and their position relative to their corresponding open reading frame. Functionally, it makes a difference if an asRNA overlaps a 5′ or 3′ end of its cognate mRNA or if it is fully internal. For this reason, we differentiated the asRNAs into these three classes according to mapping data from 5′ RACE, the lengths of hybridizing fragments in Northern blots, and by array hybridization. From the set of 28 asRNAs confirmed by multiple methods (Table I) 13 were internal, 8 were 5′ overlapping, and 7 were 3′ overlapping. Together with other factors, such as half-life, length or expression patterns (induced, transient, constitutive), a multitude of functions and mechanisms appear possible. Some of these are discussed below, but more experimental effort is necessary to investigate the individual functions of *Synechocystis* asRNAs.

It is well established that asRNAs and their *cis*-targets can form RNA–RNA duplexes, which are degraded by dsRNA-specific RNases (Hernandez *et al*, 2005; Duehring *et al*, 2006; Darfeuille *et al*, 2007; Kawano *et al*, 2007; Fozo *et al*, 2008). Hence, antisense transcription is a powerful natural tool in repressing gene expression. There is a growing number of examples, which support the idea of bacterial asRNAs serving as novel types of transcriptional terminators such as the 427 nt asRNA RNAβ in *Vibrio anguillarum* (Stork *et al*, 2007) to achieve discoordinated expression of different operon segments. Obviously, the most likely candidates for such

Figure 6 Quantitative analysis of expression microarray data and their verification. For each panel, a Northern blot is shown reproducing the results obtained for the individual small RNA in the expression microarray analysis. RNA was analyzed from cultures kept under control conditions (C), darkness for 1 h (D), high light for 30 min (HL), or depletion for CO_2 for 6 h ($-CO_2$). As a control for equal loading either 5S ribosomal RNA or the RNase P RNA (*rnpB*) was hybridized. The diagram shows the average of normalized probe set signal intensities from three biological replicates and two technical replicates each. The ratios of asRNA/mRNA signal intensities are indicated by filled circles. (**A**) Analysis of the SyR7 ncRNA, which is overlapping the *murF* 5′ UTR. Two TSS for *murF*, P1, and P2 are indicated. (**B**) Analysis of the asRNA to gene slr0882. (**C**) Analysis of the asRNA to gene sll1289. The second Northern hybridization from a low-resolution gel confirms a ~600 nt long asRNA under the dark condition.

© 2009 EMBO and Macmillan Publishers Limited

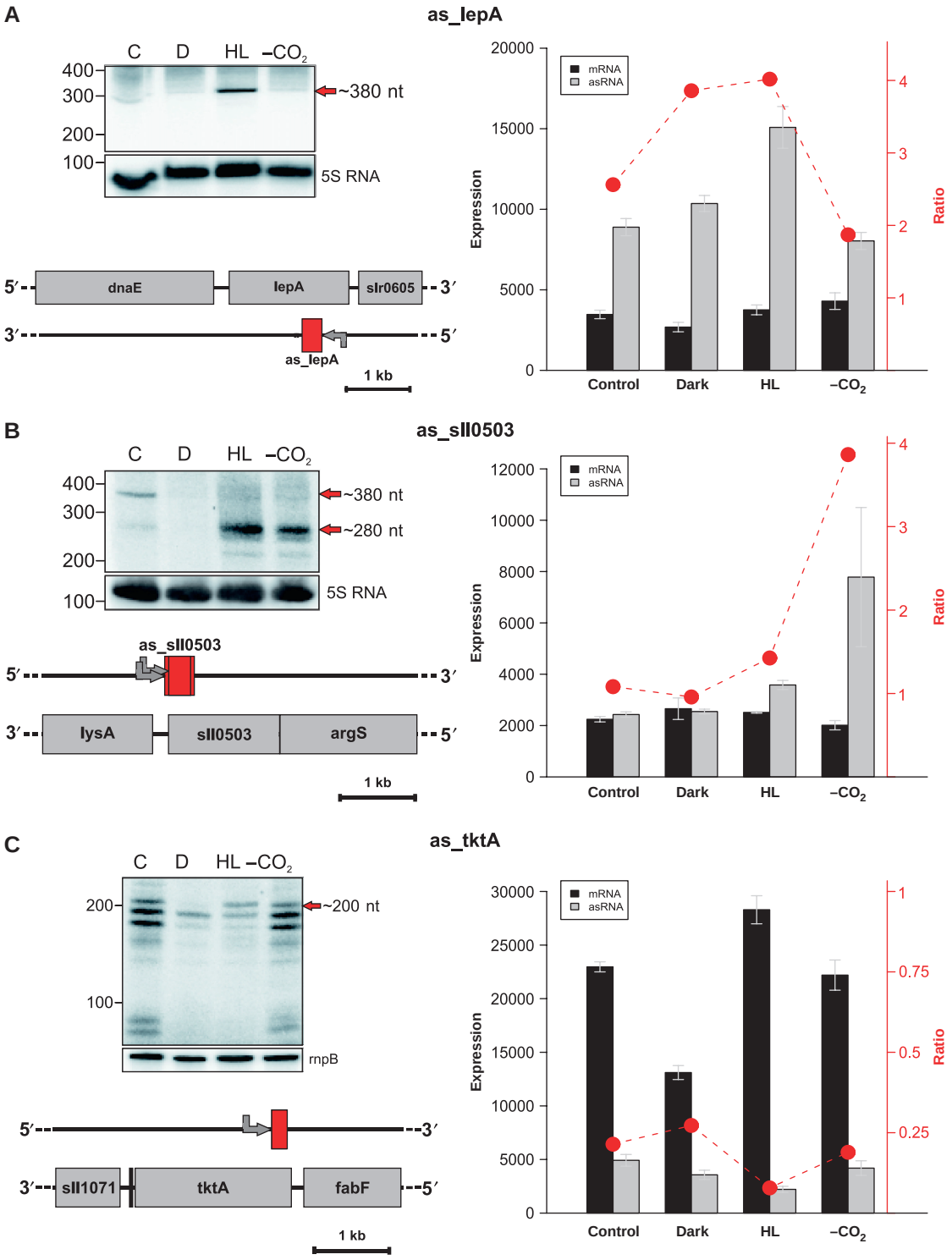


Figure 7 Microarray analysis and verification of three asRNA/mRNA pairs. All panels are arranged and labeled as in Figure 6. (**A**) Analysis of the asRNA to gene *lepA*. (**B**) Analysis of the asRNA to gene sll0503. (**C**) Analysis of the asRNA to *tktA*.

termination and processing events are asRNAs overlapping the 3′ ends of their target mRNAs. Another such candidate is as_rpl1 (Figure 8), which spans a whole intergenic spacer. Rpl1 acts as an important feedback regulator in *E. coli* (Yates *et al*, 1980; Branlant *et al*, 1981; Lindahl and Zengel, 1986), whereas Rpl1 was shown to be required for the autogenous control of the L11–L1 operon (Cole and Nomura, 1986). It was not shown whether this feedback regulation would also be

© 2009 EMBO and Macmillan Publishers Limited

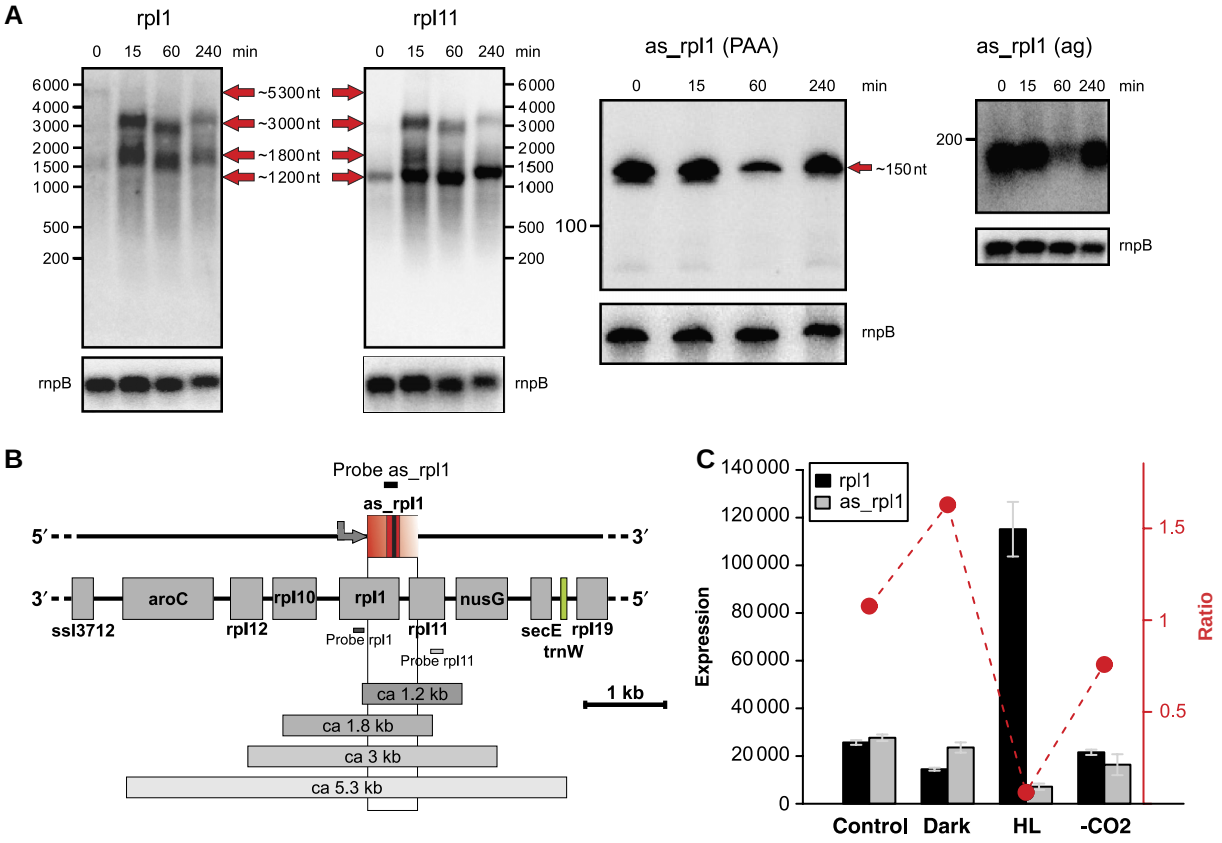


Figure 8 A possible role of as_rpl1 in discoordinating gene expression within the L11 r-operon under high light. (**A**) Northern blots showing the accumulation of as_rpl1 during a shift from standard to high light conditions (50 to 500 μmol of photons m^{-2} s^{-1}), together with its cognate mRNA *rpl1*, the *rpl11* mRNA and several bands that correspond to precursor and putative processing intermediates. Samples were collected 0, 15, 60, and 240 min after the light shift. Gels were hybridized with a probe for *rnpB* (coding for the RNA subunit of RNAse P) to correct for slight differences in the loaded amounts of RNA. For as_rpl1 two different hybridizations are shown, from separation of RNA in a polyacrylamide gel (PAA) and in an agarose gel (ag). (**B**) Organization of the *Synechocystis* L11 r-operon. The location of regions complementary to the transcript probes is given, together with the putative identity of hybridizing transcript species from part A. (**C**) Quantitative analysis of expression microarray data and their verification for as_rpl1 and the *rpl1* mRNA. The panel is arranged and labeled as described in the legend to Figure 6.

sufficient for this control. In fact, studies of the S10 ribosomal protein operon suggested that, at least in *E. coli*, additional regulatory processes are required to coordinate the synthesis of ribosomal proteins with cell growth rate (Lindahl and Zengel, 1990). In *Synechocystis*, we noticed a transient decrease in the amount of as_rpl1 (Figure 8) during the activation of operon transcription while undertaking a light upshift experiment. This observation points toward a possible consumption of as_rpl1 during the adaptation process, compatible with both a regulatory function as well as with mRNA maturation.

Another possible level of regulation includes asRNAs, which directly modulate transcriptional activity. There is strong evidence to suggest that divergently located promoters can interfere with each other (Prescott and Proudfoot, 2002) and work with *E. coli* showed that the length of transcripts generated from the divergently located promoter (Sneppen *et al*, 2005) is one important factor for this interaction. We noticed that the average length of asRNAs tends to be longer than that of ncRNAs. According to literature, the latter are typically 50–250 nt in length (Vogel and Papenfort (2006) and see Supplementary Figure S4 in Shi *et al* (2009)). Here, we observed ~180 nt as the average ncRNA length (Figure 4; Table II), with a maximum of 350 nt in case of SyR4 (Figure 4). In contrast, the lengths of the asRNAs, as confirmed by Northern blots, range here from 65 to 700 nt (Figures 1, 5 and 6; Table I), with many asRNAs longer than 300 nt lending support to the idea that some of them may have a function in transcriptional interference. A recent example of the transcriptional interference mechanism is an asRNA in *Clostridium acetobutylicum*, which can be up to 1000 nt long. This asRNA is involved in the sulfur-dependent expression of the *ubiG* operon (Andre *et al*, 2008).

We found several asRNAs extending into the 5′ UTR region of their mRNA targets and some of them probably terminate beyond the TSS of the mRNA on the reverse complementary strand. It is well established that initiation of degradation through RNase E requires free 5′ ends (Mackie, 1998). Therefore, the selective stabilization of transcripts by masking of endonuclease (RNase E) recognition sites appears to be another important function of natural asRNAs. Moreover, such 5′ overlapping asRNAs are prime candidates for providing translational regulation by extending into the regions for interaction with the ribosome, regulating rather translation than RNA stability (Darfeuille *et al*, 2007; Kawano *et al*, 2007; Fozo *et al*, 2008).

© 2009 EMBO and Macmillan Publishers Limited

Biological relevance of asRNAs

The substantial amounts of different asRNAs in *Synechocystis* raise the question of their biological benefit for the organism. One known role of bacterial asRNAs is to act as the antidote to mRNAs coding for toxic peptides (Kawano *et al*, 2007; Fozo *et al*, 2008) or transposons (Sittka *et al*, 2008). Systematic searches for toxin–antitoxin systems have revealed an abundance in free-living prokaryotes, including *Synechocystis* (Pandey and Gerdes, 2005). But what is the relevance of the majority of the asRNAs detected here? Their appearance is not restricted to a specific functional class of genes (such as regulation, primary metabolism, transcription, translation, DNA repair, etc.). Furthermore, their expression level, which is in part very high (IsrR, as_sll1049, as_slr0320) and otherwise covers the whole range of mRNA expression levels, indicates a vital function.

A bacterial cell has several means of achieving gene regulation. There are regulatory proteins as well as RNA-based elements, for example, riboswitches or ncRNAs. Although one regulatory protein per gene is clearly impossible and not very sophisticated, the concept of asRNA theoretically allows the system to have an individual regulator for every single element at a very low cost. Moreover, mathematical modeling of sRNA-based gene regulation has revealed a particular niche for regulatory RNA in allowing cells to transition quickly yet reliably between distinct states, consistent with the widespread appearance of bacterial sRNAs in stress regulatory networks (Mehta *et al*, 2008). In addressing this possibility, we examined the expression of all asRNAs and ncRNAs found in this study in a genome-wide expression microarray under four different conditions and verified the results for seven of them in more detail. In several of the newly found asRNAs, we discovered the expression to be strongly affected by some of these conditions, resulting in distinct and characteristic changes in the ratios between asRNAs and their cognate mRNAs. These changes provide circumstantial evidence for a functional role of the newly found asRNAs in regulatory networks.

Beyond *Synechocystis*

In a systematic screening for cyanobacterial ncRNAs in four strains of marine *Prochlorococcus/Synechococcus*, seven different ncRNAs were identified based on comparative genome analysis (Axmann *et al*, 2005). More recently, we used high coverage whole genome microarrays to screen genome wide for the presence of ncRNAs in *Prochlorococcus* MED4 (Steglich *et al*, 2008). This complements the earlier analysis of Axmann *et al* (2005) in the identification of 14 novel ncRNAs and 24 possible asRNAs (Steglich *et al*, 2008), although these were not characterized in detail. Considering *Prochlorococcus* MED4 is the cyanobacterium with the most streamlined genome (Strehl *et al*, 1999; Rocap *et al*, 2003; Hess, 2004) and given the paucity of such analyses for this class of bacteria as a whole, the number of asRNAs detected here in a related unicellular cyanobacterium is astonishing. *Synechocystis* or even cyanobacteria as a whole may not be so exceptional in this respect. Recent publications have presented a growing number of asRNAs in a wide variety of bacteria such as *Calothrix* (Csiszar *et al*, 1987), *Anabaena* sp. PCC7120 (Hernandez *et al*, 2005), *Vibrio anguillarum* (Stork *et al*, 2007), *Vibrio cholera* (Liu *et al*, 2009), *Caulobacter crescentus* (Landt *et al*, 2008), *Clostridium acetobutylicum* (Andre *et al*, 2008), *Streptomyces coelicolor* (Swiercz *et al*, 2008), *Bacillus subtilis* (Eiamphungporn and Helmann, 2009), and *Salmonella* (Sittka *et al*, 2008). A closer look at *E. coli* supports this view: first, albeit not studied in detail, an *E. coli* tiling array detected antisense transcription (Selinger *et al*, 2000). Second, Vogel *et al* (2003a, b) and Kawano *et al* (2005) detected asRNAs in RNomics experiments. Third, a bioinformatic approach predicted 46 asRNAs from which four were verified (Yachie *et al*, 2006). Finally, the five QUAD1 or Sib RNAs in *E. coli* lie antisense to short open reading frames coding for toxic oligopeptids (Fozo *et al*, 2008). Taking into account, that most of the approaches to systematically detect ncRNAs, discriminate against asRNAs, for example by size exclusion of the relatively big asRNAs (<65 nt (Kawano *et al*, 2005), <50 nt (Swiercz *et al*, 2008), 50–500 nt (Vogel *et al*, 2003b)), the focus on Hfq-bound RNAs (Sittka *et al*, 2008), or on intergenic regions (Landt *et al*, 2008), the actual number of asRNAs in *E. coli* and other bacteria is undoubtedly underestimated. Therefore, a potentially high number of bacterial asRNAs still remaining to be discovered could dramatically increase the regulatory capacity, flexibility and redundancy. It is very likely that chromosomally encoded asRNAs constitute an important component of another, not yet fully appreciated, level of gene regulation in bacteria.

Materials and methods

Bacterial strains and growth conditions

Synechocystis sp. PCC 6803 used in this study (originally from S. Shestakov, Moscow State University, Russia) was propagated on BG11 (Rippka *et al*, 1979) 1% (w/v) agar (Bacto agar, Difco) plates. Liquid cultures of *Synechocystis* 6803 were grown at 30 °C in BG11 (20 mM TES pH 7.6) medium under continuous illumination with white light of 50 μmol of photons $m^{-2} s^{-1}$ and a continuous stream of air. Different growth and stress conditions were applied to exponentially growing *Synechocystis* cultures (OD_{750} 0.6–0.8) to allow virtually all kinds of RNAs to be expressed. For HL stress, light intensity was shifted from 50 to 500 μmol of photons $m^{-2} s^{-1}$, samples were collected 30, 60, and 120 min after the shift. For low light conditions, light intensity was shifted from 100 to 10 μmol of photons $m^{-2} s^{-1}$, samples were collected 30, 60, and 120 min after the shift. For iron and nitrogen stress, cells were collected by centrifugation and washed twice with iron-free (replacing ammonium iron (III) citrate with di-ammonium hydrogen citrate) or nitrogen-free (omitting sodium nitrate from the medium) BG11 medium. Resulting pellets were then resuspended in their respective medium. For iron stress, cells were harvested after 20 and 45 h, for nitrogen stress after 12.5 and 20 h. Heat and cold stress were applied by a temperature shift from 30 to 42 °C or 15 °C, respectively. For heat stress, sample collection occurred after 20 and 60 min, for cold stress after 30 and 120 min. Another culture was harvested after 12 h incubation in the dark. For stationary phase cells, a culture was harvested at OD_{750} of 3.5. Exponentially growing cells were harvested at OD_{750} 0.56. The cultures for the expression microarray were grown at control conditions (OD 0.6 at 750 nm; 50 μmol photons $m^{-2} s^{-1}$), or transferred to dark for 1 h, depleted for CO_2 for 6 h by transferring to carbon-free BG11 (BG11 w/o $NaCO_2$, pH 7.0) without aeration after washing once in carbon-free BG11, or transferred to HL (500 μmol of photons $m^{-2} s^{-1}$) for 30 min.

© 2009 EMBO and Macmillan Publishers Limited

RNA extraction and analysis

Synechocystis 6803 cells were collected by rapid filtration (Pall Supor 800 Filter, 0.8 μm). Filters with cells were dissolved in 1 ml TRIzol (Invitrogen) per 40 ml culture, immediately frozen in liquid nitrogen and incubated 15 min at 65 °C in a water bath. Further RNA isolation followed the manufacturer's protocol.

Northern blot analysis and 5′ RACE

High resolution Northern blots were prepared from the separation of 10 to 25 μg of total RNA on 10% urea-polyacrylamide gels as described by Steglich *et al* (2008). Blots for RNAs with higher molecular weight were prepared from the separation of 5 to 10 μg of total RNA on 1.5% denaturing agarose gels. Hybridization conditions were described by Steglich *et al* (2008). 5′ RACE was performed as described in Steglich *et al* (2008). The sequences of all oligonucleotides used in this study for the preparation of transcript probes and 5′ RACE are listed in Supplementary Table S2.

Microarray hybridization

For RNA hybridizations, the RNA mix was labeled directly, without cDNA synthesis in 5 μg aliquots with the Kreatech 'ULS labeling kit for Agilent gene expression arrays' with Cy3 or Cy5 according to the manufacturer's protocol. Fragmentation and hybridization was performed following the manufacturer's instructions for Agilent one color microarrays with 3 to 5.5 μg of labeled RNA. DNA was fragmented by 3 h incubation at 95°C in H_2O and Cy3 labeled with the Kreatech kit mentioned above. Hybridization was performed similar to RNA hybridization, without the fragmentation step in the Agilent protocol. For DNA hybridization, 0.5 to 3.8 μg of labeled DNA were used. For the expression microarray, we directly labeled 2 μg RNA using the Cy3 labeling kit mentioned above. Hybridization was done with 1.5 μg RNA per array according to the Agilent protocol for 4 × 44k single color microarrays. Each stress condition was hybridized in triplicates. The data for both types of microarrays have been deposited in the GEO database under the accession numbers GSE16162 and GSE14410.

Transcript prediction

In general, a transcribed region of a genome is characterized by a TSS and a region of termination. A TSS can be identified by its preceding promoter region and the nucleotide identity (preferably A or G; Vogel *et al* (2003a)). Preliminary studies showed that the current standard method for TSS prediction based on a position-specific scoring matrix as developed by Vogel *et al* (2003a) alone is statistically not significant for *ab initio* transcript prediction and also does not improve significance in combination with terminator prediction. For this reason, we only made use of terminator prediction, described in the following. For termination of transcription, two possibilities exist: rho-dependent and -independent termination. Only the latter can be identified on the sequence level, as it shows a characteristic GC-rich hairpin in front of a T/U-rich region, the so-called T-tail. The T-tail can be further divided into the proximal (first five bases) and the distal part (the four bases after the proximal part). With the help of RNAll (Wan and Xu, 2005) such intrinsic terminators were predicted and subjected to a post-filtering step with the following rules: (1) at least four G–C or G–U pairs; (2) at most 2 nt spacer between stem and T-tail; (3) at least three 'T/U's in the proximal part; (4) no more than one 'G' in the proximal part; (5) a 'T' at position 2 or 3 in the proximal part; (6) at most three purines or three cytosines in the distal part; (7) at least 4 'T's in proximal and distal part together; (8) no multiloops and at most 1 bulged nucleotide; and (9) free energy of the stem-loop at most −8.0 kcal/mol. Rules 1–7 were taken from Lesnik *et al* (2001) and rules 8 and 9 were defined by ourselves. Calculation of the free energy was performed using RNAshapes (Steffen *et al*, 2006) as RNAll provides a heuristic structure prediction, leading to artifacts in the subsequent energy calculation by efn2 (Mathews *et al*, 2004).

Design of microarrays

The design of probes for the tiling microarray was based on the terminator predictions. To each prediction, the sequence of the corresponding gene or intergenic region was extracted in both orientations and redundancy was removed. Neighboring genes were concatenated with their intergenic spacer, to get antisense transcripts overlapping two genes. This resulted in 646 (480 antisense + 166 intergenic) sequences with a total length of 691 759 nt. As controls, we aimed at a similar number of genes and intergenic regions yielding a similar amount of bases. We selected 474 genes with a total length of 698 590 nt and 158 intergenic regions comprising 50 797 nt. This sums up to a total of 632 control sequences holding 749 387 nt. Altogether, the sequences on the array covered 1 441 146 nt. The probe design included generating overlapping sequences of length 50 with an offset of 28 nt, trimming of sequences to get a Tm as close as possible to 72°C with a minimum length of 25 nt, checking redundancy of trimmed sequence within the genome and the plasmids pcB.2.4, pSYSG, pSYSX, pSYSM, and pSYSA and discarding sequences with multiple perfect or 1-mismatch hits or Tm out of 70–74°C. This procedure resulted in 102 739 probes fitting on a 2 × 104K-Agilent custom array together with control probes from mouse actin gene.

The expression microarray holds probe sets for all annotated genes from the chromosome (NC_000911) as well as the seven plasmids (pSYSA: NC_005230, pSYSG: NC_005231, pSYSM: NC_005229, pSYSM: NC_005232, pCA2.4, pCB2.4, pCC5.2 available at http://genome.kazusa.or.jp/cyanobase/Synechocystis/) and, additionally, each genomic region corresponding to an expressed segment seen with the tiling microarray. On average, 3 to 5 probes per transcript were designed using the Agilent eArray system (https://earray.chem.agilent.com/earray/). The chosen design criteria were 'best distribution method', Tm 80°C and a length between 45 and 60 nt, resulting in 20 293 probes. These probes were manufactured in a 44K Agilent custom microarray format with an internal duplication of all probes, hence providing an internal obligatory technical replicate. Descriptions of the array design and probe sequences for both microarrays have been deposited in the GEO database under the accession numbers GSE16162 and GSE14410.

Data normalization, transcript mapping, and identification of antisense transcripts

The procedure of transcript mapping on data from the tiling microarray was performed as described in Huber *et al* (2006). To be able to make use of the author-provided software (R-package tiling_array) we had to design virtual probes for the genomic regions not covered by the probes on the microarray and assigned to them the arbitrarily chosen normalized expression value of −20.0. This is possible without affecting the segmentation algorithm, as the latter is optimizing the sum of the summed up residuals, that is the squared difference of an individual probe to the mean of all probes in the segment, over all segments. Segments containing solely virtual probes have a mean of −20.0 and as each probe has the same expression value, the contribution of such virtual-only segments is 0.0 and thereby does not affect the overall optimization. To find the optimal segmentation, the algorithm needs to be given an expected number of segments. To calculate this number, we considered 646 regions based on predictions and 632 regions as controls, making a total of 1278 genomic regions. As a region always implies 'empty' regions surrounding it we get 2 × 1278=2556 regions. Overall, this gives an estimate of ∼2500 segments per strand.

Data extraction from transcriptome microarray. Spot intensities were extracted with the 'Agilent Feature Extraction Software 10.5.1.1' (Protocol: GE1_105_Dec08), for further processing we used the R-package 'limma'. The median spot intensities were quantile normalized and the contrasts between control and stress conditions were extracted using the linear model provided by limma. The *P*-values were calculated with Benjamini–Hochberg adjustment. Only probes with an adjusted *P*-value <0.05 were used for further calculations. All probes of one feature were unified in a probe set for calculation of FC and mean expression. To test the experimental variability, we determined the average in-group FCs between the normalized triplicates, the

© 2009 EMBO and Macmillan Publishers Limited

borders for a significance level of 0.05 are −0.34 and 0.31 for the control (all log2 values), −0.72 and 0.54 for the sample from dark, −0.92 and 0.85 for HL and −0.64 and 0.46 for CO_2 depletion. Thus, FC′s greater than ±0.9 (log2) were listed as differentially expressed. The mean expression is the mean of all quantile normalized median probe intensities of one probe set. For the calculation of asRNA/mRNA ratios, the mean expression of the asRNA was divided by the mean expression of the corresponding mRNA.

ORF analysis

Candidate asRNAs and ncRNAs were scanned for conserved ORFs. Initially, ORFs with possible start codons (ATG, GTG, TTG, and ATT) and a minimum length of 45 nt were predicted. Conservation was checked using TBLASTN against the NCBI *nr* database.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft Focus program 'Sensory and regulatory RNAs in Prokaryotes' SPP1258 (project HE 2544/4-1 to WRH and WI 2014/3-1 to AW), the graduate school 'Signal systems in plant model organisms' (to JG) and by the BMBF—Freiburg Initiative in Systems Biology, project 0313921 (WRH).

Conflict of interest

The authors declare that they have no conflict of interest.

References

Andre G, Even S, Putzer H, Burguiere P, Croux C, Danchin A, Martin-Verstraete I, Soutourina O (2008) S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Res* **36:** 5955–5969

Axmann IM, Kensche P, Vogel J, Kohl S, Herzel H, Hess WR (2005) Identification of cyanobacterial non-coding RNAs by comparative genome analysis. *Genome Biol* **6:** R73

Branlant C, Krol A, Machatt A, Ebel JP (1981) The secondary structure of the protein L1 binding region of ribosomal 23S RNA. Homologies with putative secondary structures of the L11 mRNA and of a region of mitochondrial 16S rRNA. *Nucleic Acids Res* **9:** 293–307

Chen JJ, Sun M, Kent WJ, Huang XQ, Xie HQ, Wang WQ, Zhou GL, Shi RZ, Rowley JD (2004) Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acids Res* **32:** 4812–4820

Cole JR, Nomura M (1986) Translational regulation is responsible for growth-rate dependent and stringent control of the synthesis of ribosomal proteins L11 and L1 in *Escherichia coli*. *Proc Natl Acad Sci USA* **83:** 4129–4133

Csiszar K, Houmard J, Damerval T, Tandeau de Marsac N (1987) Transcriptional analysis of the cyanobacterial *gvpABC* operon in differentiated cells: occurrence of an antisense RNA complementary to three overlapping transcripts. *Gene* **60:** 29–37

Darfeuille F, Unoson C, Vogel J, Wagner EG (2007) An antisense RNA inhibits translation by competing with standby ribosomes. *Mol Cell* **26:** 381–392

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW, Steinmetz LM (2006) A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci USA* **103:** 5320–5325

Dienst D, Dühring U, Mollenkopf H, Vogel J, Golecki J, Hess WR, Wilde A (2008) The RNA chaperone Hfq is essential for cell motility of the cyanobacterium *Synechocystis* PCC 6803. *Microbiology* **154:** 3134–3143

Duehring U, Axmann IM, Hess WR, Wilde A (2006) An internal antisense RNA regulates expression of the photosynthesis gene *isiA*. *Proc Natl Acad Sci USA* **103:** 7054–7058

Eiamphungporn W, Helmann JD (2009) Extracytoplasmic function sigma factors regulate expression of the *Bacillus subtilis yabE* gene via a cis-acting antisense RNA. *J Bacteriol* **191:** 1101–1105

Fozo EM, Kawano M, Fontaine F, Kaya Y, Mendieta KS, Jones KL, Ocampo A, Rudd KE, Storz G (2008) Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. *Mol Microbiol* **70:** 1076–1093

Gottesman S (2004) The small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu Rev Microbiol* **58:** 303–328

Hedger J, Holmquist PC, Leigh KA, Saraff K, Pomykal C, Summers ML (2009) Illumination stimulates cAMP receptor protein-dependent transcriptional activation from regulatory regions containing class I and class II promoter elements in *Synechocystis* sp. PCC 6803. *Microbiology*, epub ahead of print

Hernandez JA, Muro-Pastor AM, Flores E, Bes MT, Peleato ML, Fillat MF (2005) Identification of a *furA cis* antisense RNA in the cyanobacterium *Anabaena* sp. PCC 7120. *J Mol Biol* **355:** 325–334

Hess WR (2004) Genome analysis of marine photosynthetic microbes and their global role. *Curr Opin Biotechnol* **15:** 191–198

Hihara Y, Kamei A, Kanehisa M, Kaplan A, Ikeuchi M (2001) DNA microarray analysis of cyanobacterial gene expression during acclimation to high light. *Plant Cell* **13:** 793–806

Huber W, Toedling J, Steinmetz LM (2006) Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics* **22:** 1963–1970

Jen CH, Michalopoulos I, Westhead DR, Meyer P (2005) Natural antisense transcripts with coding capacity in *Arabidopsis* may have a regulatory role that is not linked to double-stranded RNA degradation. *Genome Biol* **6:** R51

Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, Nishida H, Yap CC, Suzuki M, Kawai J, Suzuki H, Carninci P, Hayashizaki Y, Wells C, Frith M, Ravasi T, Pang KC, Hallinan J, Mattick J, Hume DA *et al* (2005) Antisense transcription in the mammalian transcriptome. *Science* **309:** 1564–1566

Kawano M, Aravind L, Storz G (2007) An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Mol Microbiol* **64:** 738–754

Kawano M, Reynolds AA, Miranda-Rios J, Storz G (2005) Detection of 5′- and 3′-UTR-derived small RNAs and cis-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res* **33:** 1040–1050

Kiyosawa H, Yamanaka I, Osato N, Kondo S, Hayashizaki Y (2003) Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. *Genome Res* **13:** 1324–1334

Landt SG, Abeliuk E, McGrath PT, Lesley JA, McAdams HH, Shapiro L (2008) Small non-coding RNAs in *Caulobacter crescentus*. *Mol Microbiol* **68:** 600–614

Lesnik EA, Sampath R, Levene HB, Henderson TJ, McNeil JA, Ecker DJ (2001) Prediction of rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Res* **29:** 3583–3594

Levine E, Zhang Z, Kuhlman T, Hwa T (2007) Quantitative characteristics of gene regulation by small RNA. *PLOS Biol* **5:** e229

Lindahl L, Zengel JM (1986) Ribosomal genes in *Escherichia coli*. *Annu Rev Genet* **20:** 297–326

Lindahl L, Zengel JM (1990) Autogenous control is not sufficient to ensure steady-state growth rate-dependent regulation of the S10 ribosomal protein operon of *Escherichia coli*. *J Bacteriol* **172:** 305–309

Liu JM, Livny J, Lawrence MS, Kimball MD, Waldor MK, Camilli A (2009) Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res* **37:** e46

© 2009 EMBO and Macmillan Publishers Limited

Mackie GA (1998) Ribonuclease E is a 5′-end-dependent endonuclease. *Nature* **395:** 720–723

Malakhov MP, Los DA, Wada H, Semenenko VE, Murata N (1995) Characterization of the *murF* gene of the cyanobacterium *Synechocystis* sp. PCC 6803. *Microbiology* **141:** 163–169

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci USA* **101:** 7287–7292

Mehta P, Goyal S, Wingreen NS (2008) A quantitative comparison of sRNA-based and protein-based gene regulation. *Mol Sys Biol* **4:** 221

Nakamura T, Naito K, Yokota N, Sugita C, Sugita M (2007) A cyanobacterial non-coding RNA, Yfr1, is required for growth under multiple stress conditions. *Plant Cell Physiol* **48:** 1309–1318

Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full length cDNAs. *Nature* **420:** 563–573

Pandey DP, Gerdes K (2005) Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res* **33:** 966–976

Perocchi F, Xu Z, Clauder-Muenster S, Steinmetz LM (2007) Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res* **35:** e128

Prescott EM, Proudfoot NJ (2002) Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci USA* **99:** 8796–8801

Qin Y, Polacek N, Vesper O, Staub E, Einfeldt E, Wilson DN, Nierhaus KH (2006) The highly conserved LepA is a ribosomal elongation factor that back-translocates the ribosome. *Cell* **127:** 721–733

Rippka R, Deruelles J, Waterbury JB, Herdmann M, Stanier RY (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol* **111:** 1–61

Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, Arellano A, Coleman M, Hauser L, Hess WR, Johnson ZI, Land M, Lindell D, Post AF, Regala W, Shah M, Shaw SL, Steglich C, Sullivan MB, Ting CS *et al* (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424:** 1042–1047

Selinger DW, Cheung KJ, Mei R, Johansson EM, Richmond CS, Blattner FR, Lockhart DJ, Church GM (2000) RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nat Biotechnol* **18:** 1262–1268

Shi Y, Tyson GW, DeLong EF (2009) Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459:** 266–269

Shimoni Y, Friedlander G, Hetzroni G, Niv G, Altuvia S, Biham O, Margalit H (2007) Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol Sys Biol* **3:** 138

Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JC, Vogel J (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* **4:** e1000163

Sneppen K, Dodd IB, Shearwin KE, Palmer AC, Schubert RA, Callen BP, Egan JB (2005) A mathematical model for transcriptional interference by RNA polymerase traffic in *Escherichia coli*. *J Mol Biol* **346:** 399–409

Steffen P, Voss B, Rehmsmeier M, Reeder J, Giegerich R (2006) RNAshapes: an integrated RNA analysis package based on abstract shapes. *Bioinformatics (Oxford, England)* **22:** 500–503

Steglich C, Futschik ME, Lindell D, Voss B, Chisholm SW, Hess WR (2008) The challenge of regulation in a minimal phototroph: Non-coding RNAs in *Prochlorococcus*. *PLoS Genet* **4:** e1000173

Stork M, Di Lorenzo M, Welch TJ, Crosa JH (2007) Transcription termination within the iron transport-biosynthesis operon of *Vibrio anguillarum* requires an antisense RNA. *J Bacteriol* **189:** 3479–3488

Strehl B, Holtzendorff J, Partensky F, Hess WR (1999) A small and compact genome in the marine cyanobacterium *Prochlorococcus marinus* CCMP 1375: lack of an intron in the gene for tRNA(Leu)(UAA) and a single copy of the rRNA operon. *FEMS Microbiol Lett* **181:** 261–266

Swiercz JP, Hindra, Bobek J, Bobek J, Haiser HJ, Di Berardo C, Tjaden B, Elliot MA (2008) Small non-coding RNAs in *Streptomyces coelicolor*. *Nucleic Acids Res* **36:** 7240–7251

Urban JH, Vogel J (2007) Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. *Nucleic Acids Res* **35:** 1018–1037

Vogel J, Axmann IM, Herzel H, Hess WR (2003a) Experimental and computational analysis of transcriptional start sites in the cyanobacterium *Prochlorococcus* MED4. *Nucleic Acids Res* **31:** 2890–2899

Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jager JG, Huttenhofer A, Wagner EG (2003b) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res* **31:** 6435–6443

Vogel J, Papenfort K (2006) Small non-coding RNAs and the bacterial outer membrane. *Curr Opin Microbiol* **9:** 605–611

Voss B, Georg J, Schön V, Ude S, Hess WR (2009) Biocomputational prediction of non-coding RNAs in model cyanobacteria. *BMC Genomics* **10:** 123

Voss B, Gierga G, Axmann IM, Hess WR (2007) A motif-based search identifies the ortholog of the small RNA Yfr1 in all major lineages of cyanobacteria. *BMC Genomics* **8:** 375

Wagner EG, Simons RW (1994) Antisense RNA control in bacteria, phages, and plasmids. *Ann Rev Microbiol* **48:** 713–742

Wan X-F, Xu D (2005) Intrinsic terminator prediction and its application in *Synechococcus* sp. WH 8102. *Internat J Comp Sci Technol* **20:** 465–482

Wang XJ, Gaasterland T, Chua NH (2005) Genome-wide prediction and identification of cis-natural antisense transcripts in *Arabidopsis thaliana*. *Genome Biol* **6:** R30

Yachie N, Numata K, Saito R, Kanai A, Tomita M (2006) Prediction of non-coding and antisense RNA genes in *Escherichia coli* with Gapped Markov Model. *Gene* **372:** 171–181

Yates JL, Arfsten AE, Nomura M (1980) *In vitro* expression of *Escherichia coli* ribosomal protein genes: autogenous inhibition of translation. *Proc Natl Acad Sci USA* **77:** 1837–1841

Yelin R, Dahary D, Sorek R, Levanon EY, Goldstein O, Shoshan A, Diber A, Biton S, Tamir Y, Khosravi RR, Nemzer S, Pinner E, Walach S, Bernstein J, Savitsky K, Rotman G (2003) Widespread occurrence of antisense transcription in the human genome. *Nature Biotechnol* **21:** 379–386

Zhang Y, Liu XS, Liu QR, Wei L (2006) Genome-wide *in silico* identification and analysis of *cis* natural antisense transcripts (*cis*-NATs) in ten species. *Nucleic Acids Res* **34:** 3465–3475

SOME RIGHTS RESERVED

Molecular Systems Biology is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*.

This article is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Licence.