Analysis of a photosynthetic cyanobacterium rich in internal membrane systems via gradient profiling by sequencing (Grad-seq)

Matthias Riediger, 1 Philipp Spät, 2 Raphael Bilger, 1 Karsten Voigt, 3 Boris Maček, 2 Wolfgang R. Hess 1, *

1 Genetics and Experimental Bioinformatics, Faculty of Biology, University of Freiburg, Schänzlestr. 1, 79104 Freiburg, Germany
2 Department of Quantitative Proteomics, Interfaculty Institute for Cell Biology, University of Tübingen, Auf der Morgenstelle 15, 72076 Tübingen, Germany
3 IT Administration, Institute of Biology 3, Faculty of Biology, University of Freiburg, Schänzlestr. 1, 79104 Freiburg, Germany

*Author for correspondence: wolfgang.hess@biologie.uni-freiburg.de

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Wolfgang R. Hess: wolfgang.hess@biologie.uni-freiburg.de

W.R.H. and M.R. designed the research and performed the experiments and bioinformatics analyses. P.S. and B.M. carried out the MS and the related analyses. M.R. developed the bioinformatics analysis workflow and database. R.B. performed validation experiments and K.V. set up the hosting webserver. W.R.H. and M.R. wrote the manuscript with input from all authors. W.R.H. and B.M. supervised and funded the project.

Abstract

Although regulatory small RNAs have been reported in photosynthetic cyanobacteria, the lack of clear RNA chaperones involved in their regulation poses a conundrum. Here, we analyzed the full complement of cellular RNAs and proteins using gradient profiling by sequencing (Grad-seq) in Synechocystis 6803. Complexes with overlapping subunits such as the CpcG1-type versus the CpcL-type phycobilisomes or the PsaK1 versus PsaK2 photosystem I pre(complexes) could be distinguished, supporting the high quality of this approach. Clustering of the in-gradient distribution profiles followed by several additional criteria yielded a short list of potential RNA chaperones that include an YlxR homolog and a cyanobacterial homolog of the KhpA/B complex. The data suggest previously undetected complexes between accessory proteins and CRISPR-Cas systems, such as a Csx1-Csm6 ribonucleolytic defense complex. Moreover, the exclusive association of either RpoZ or 6S RNA with the core RNA polymerase complex and the existence of a reservoir of inactive sigma–antisigma complexes is suggested. The Synechocystis Grad-seq resource is available online at https://sunshine.biologie.uni-freiburg.de/GradSeqExplorer/ providing a comprehensive resource for the functional assignment of RNA–protein complexes and multi-subunit protein complexes in a photosynthetic organism.

Introduction

Noncoding RNAs (ncRNAs) constitute a major component of the transcriptional output in all organisms (Cech and Steitz, 2014; Morris and Mattick, 2014). In photosynthetic cyanobacteria as well as in other bacteria, complex regulatory networks have been identified that include small regulatory ncRNAs (sRNAs) as major players in the posttranscriptional control of gene expression (Kopf and Hess, 2015; Wagner and Romby, 2015). While some sRNAs and cis-transcribed antisense RNAs (asRNAs) may function independently of proteins, the vast majority of ncRNAs requires interactions with specific RNA-binding proteins (RBPs;
Cyanobacteria form a single phylum of species with very different morphologies and highly diverse lifestyles. Their genome sizes vary by approximately one order of magnitude. Cyanobacteria are also the only bacteria that perform oxygenic photosynthesis. They are the presumed evolutionary ancestors of chloroplasts, are of paramount ecological importance as primary producers and enjoy high biotechnological interest (Hagemann and Hess, 2018; Vijay et al., 2019). In the unicellular model cyanobacterium Synechocystis sp. PCC 6803 (hereafter Synechocystis), three clustered regularly interspaced short palindromic repeat (CRISPR) systems (Scholz et al., 2013), a 6S RNA homolog (Heilmann et al., 2017), and hundreds of putative sRNAs have been identified (Kopf et al., 2014; Kopf and Hess, 2015). Several of these sRNAs are important post-transcriptional regulators. For instance, the sRNA PsrR1 (photosynthesis regulatory RNA1) limits the expression of several genes encoding photosynthesis proteins and pigment biosynthesis proteins during the acclimation of the photosynthetic machinery to high light conditions (Georg et al., 2014). Likewise, the sRNA IsaR1 (iron-stress activated RNA 1) contributes to the adjustment of iron-sulfur cluster biosynthesis, photosynthetic electron transport, and photosystem I (PSI) gene expression in the acclimation response to low iron (Georg et al., 2017). Moreover, two different types of ncRNAs, the sRNA NsiR4 (nitrogen stress-induced RNA 4) and the glutamine Type I riboswitch, were found to control cyanobacterial nitrogen assimilation in ways that differ considerably from the archetypical model developed for Escherichia coli (Klähn et al., 2015, 2018). Recently, the Synechocystis RBPs Rbp2 and Rbp3 (Ssr1480 and Slr0193) were shown to affect association of mRNAs encoding core subunits of both photosystems at the thylakoid membranes (Mahbub et al., 2020). However, despite the observed abundance of riboregulatory processes, no RBPs functionally similar to ProQ, CsrA, or Hfq have been identified in cyanobacteria to date. While there are no candidates for the former two, a structural homolog of Hfq is present in several cyanobacteria, including Synechocystis (Dienst et al., 2008; Bøggild et al., 2009). However, compared to the homologs in proteobacteria, cyanobacterial Hfq is truncated, does not bind RNA, and may function through protein–protein interactions (Schuergers et al., 2014).

A recently developed method uses gradient profiling by sequencing (Grad-seq) to directly detect groups of RNAs and comigrating proteins. In this approach, whole-cell lysates are fractionated on a sucrose density gradient by ultracentrifugation. The fractions are subjected to mass spectrometry (MS)-based proteome measurements and transcriptome deep-sequencing (RNA-seq) analyses, hence guiding the potential discovery of new globally acting RBPs and
polypeptides acting together in large multiprotein complexes (Smirnov et al., 2016). This has been prolific in the discovery of ProQ as a previously overlooked major sRNA-binding protein in the bacterial pathogen Salmonella enterica (Smirnov et al., 2016; Westermann et al., 2019) and other enteric bacteria (Melamed et al., 2020), in the discovery of FopA as a new member of the emerging family of FinO/ProQ-like RBPs (Gerovac et al., 2020), in the characterization of the seemingly noncoding RNA RyEG as an mRNA that encodes a small toxic protein (Hör et al., 2020) and in identifying the unexpected involvement of exoribonuclease activity in the stabilization and activation of sRNAs in the Gram-positive pathogen Streptococcus pneumonia (Hör et al., 2020).

While there are reports on the composition of protein complexes (Xu et al., 2020) and the association of these complexes with the thylakoid or cell membrane systems in the model Synechocystis (Baers et al., 2019), no information on RNA/protein complexes is currently available for any cyanobacterium. Here, we report the results of Grad-seq analysis applied to Synechocystis, identifying the sedimentation profiles of thousands of transcripts and proteins simultaneously. We applied a hierarchical clustering approach and provide a database that supports further analyses of this dataset, based on multiple filter criteria. In contrast to previous Grad-Seq studies that initially focused on the sedimentation profiles of ncRNAs followed by further experimental analysis, we directed our efforts to the comigration analysis of proteins and RNAs. We computed a support vector machine (SVM) score for the prediction of RBPs among the cosedimenting proteins using RNApred (Kumar et al., 2011). In addition, we assumed that relevant RBPs would be more widely phylogenetically conserved and thus tested for the presence of putative homologs in a set of 57 different cyanobacteria, Arabidopsis (Arabidopsis thaliana), E. coli, and Salmonella enterica and evaluated the degree of synteny between 34 different cyanobacteria. This strategy allowed the direct delineation of RBPs or other proteins of interest. These results provide a rich RNA/protein complexome resource in a photosynthetic cyanobacterium.

**Results**

Grad-seq analysis resolves the proteome and transcriptome of a photosynthetic bacterium

We cultivated triplicate cultures of Synechocystis under moderate light intensities (50 μE) in BG11 medium. We prepared and fractionated whole-cell lysates by density gradient centrifugation, as described (Smirnov et al., 2016) but using n-Dodecyl β-D-maltoside (DDM) as membrane solubilizer and building the gradients from sucrose. The gradients showed the clear color separation profiles from the different pigment-containing complexes (Figure 1A). After separation, we analyzed the eluted fractions from each gradient via MS and RNA-seq for the identification of their protein and RNA composition, respectively. The overall in-gradient distribution of RNA profiles (Supplemental Figure 1A) and protein profiles (Supplemental Figure 1B) showed a strong correlation (median RNA-seq R = 0.76, median MS R = 0.85) between replicates and thus confirmed good reproducibility. The fractionated proteins and RNAs were also visualized using conventional denaturing polyacrylamide gels (Figure 1A and B).

Following the identification of proteins and RNAs by MS and RNA-seq, we determined the sedimentation profile of 2,394 proteins and 4,251 different transcripts detected in all three replicates (Figure 1C). These transcripts consisted of 2,968 mRNAs, 530 separate 5′ untranslated regions (UTRs), 359 3′ UTRs, 140 ncRNAs, 139 asRNAs, 65 transposase-associated RNAs, 42 transfer RNAs (tRNAs), 6 ribosomal RNAs (rRNAs), and 3 types of CRISPR RNAs (crRNAs). A total of 3,559 annotated protein-coding genes were previously defined in this Synechocystis strain (Trautmann et al., 2012); hence, we detected 67.3% of all annotated proteins in our experimental conditions. When including 122 more recently identified additional genes, we detected 65% of the 3,681 total proteins. The 4,251 different detected transcripts corresponded to 2,544 previously defined transcriptional units (TUs) out of a total of 4,091 (Kopf et al., 2014), indicating that we detected 62.2% of all TUs. In our previous definition of TUs, we had assigned multiple genes and their UTRs to one and the same TU if coverage by RNA-seq reads did not indicate interrupting terminators or internal transcriptional start sites. However, aiming for maximum sensitivity, we kept all genes and UTRs separate here. The percentages of 67.3% detected proteins and of 62.2% detected TUs are quite exhaustive, as the samples were from a single culture condition, while a subset of genes will be expressed only under specific conditions (Kopf et al., 2014). An overview of fraction complexity is given in Supplemental Data Set 1 and the detailed distribution of all detected proteins and transcripts can be found in Supplemental Data Set 2. We also provide a comprehensive resource, available at https://sunshine.biologie.uni-freiburg.de/GradSeqExplorer/, to scan for all detected proteins and RNAs from this dataset.

Hierarchical clustering splits the Grad-seq dataset into protein- and RNA-dominated main branches

Based on their distribution along the gradient, we assigned proteins and transcripts to 17 different clusters using the dynamic tree cut algorithm from the Weighted correlation network analysis (WGCNA) R package (Langfelder and Horvath, 2008, 2012; Figure 2A, Supplemental Figure 2A). A striking result from this approach was the division of the entire dataset into a protein-dominated branch whose members mainly sedimented in the lower molecular fractions, and an RNA-dominated branch whose members belonged predominantly to the higher molecular fractions (Supplemental Figure 2B and C). The protein-dominated branch, represented by Clusters 1–9, consisted of soluble proteins, small- and medium-sized protein complexes, and tRNAs in the lower molecular Fractions 2–6 (~10%–20%...
sucrose). Approximately 80% of all detected proteins were classified into this branch. The RNA-dominated branch, represented by Clusters 10–17, included most transcripts and large protein complexes, predominantly in the higher molecular Fractions 6–18 (20%–40% sucrose), and comprised ~90% of all detected transcripts (Figure 2B). As expected, the sedimentation properties of different proteins varied largely and reflected the molecular weight or the size of the associated protein complex. Based on calibration with known complexes, proteins peaking in Fractions 6 and
Figure 2 Composition of cluster profiles, functional, and phylogenetic analysis. A, Profiles of Clusters 1–17 (numbers to the right) along the different Fractions 1–18 (numbers on the X-axis). The gray Clusters 1–9 represent the protein-dominated branch, while the colored Clusters 10–17 represent the RNA-dominated branch (pink background, see Supplemental Figure 2). The red Cluster 10 represents the main RNA cluster, yellow Clusters 11–12 represent small clusters between the RNA clusters, green Clusters 13–15 represent other RNA clusters containing transposase-associated RNAs and housekeeping ncRNAs, such as 6S RNA, tmRNA, or RNase P, while purple Clusters 16–17 represent the main ribosomal RNA clusters. B, Pie charts illustrating the relative content in proteins and different RNA species of each cluster. The circle sizes correlate with the number of their respective constituents (scale in lower left corner). C, Pie charts illustrating the association of proteins in each cluster with a functional category (Kanehisa
above were associated with larger complexes of over 250 kDa in size (Supplemental Figure 3A). By contrast, the majority of RNAs exhibited a largely similar sedimentation pattern, accumulating mainly in Clusters 10, 13, and 14. Notable exceptions were the RNase P RNA, transfer-messenger RNA (tmRNA), 6S RNA as well as 5S, 16S, and 23S rRNA, which sedimented toward higher buoyancies. These results are in accordance with previous observations (Smirnov et al., 2016) that RNA sedimentation profiles are largely independent of transcript length but rely more on their corresponding binding protein(s) (Supplemental Figure 3B).

The proteins of our dataset were categorized into functional groups (Figure 2C), as defined by the Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa and Goto, 2000; Kanehisa, 2019; Kanehisa et al., 2019). We then assessed the existence of a likely ortholog for all proteins in selected organisms (Supplemental Table 1). We chose a set of 56 different cyanobacterial proteomes to judge protein conservation in cyanobacteria only. We also compared our proteome as a reference for bacteria with characterized sRNA chaperones. This analysis led to the classification of proteins into five categories: "Synechocystis only," "cyanobacteria only," "bacteria," "cyanobacteria and Arabidopsis," and "all." The dominance of proteins of unknown function in the "Synechocystis only" category and their proportion of over 50% in the "cyanobacteria only" category illustrate that these organisms are still largely unexplored. Only 20% of the Synechocystis proteome was assigned to clusters from the RNA-dominated branch (Clusters 10–17 with maximum peaks in Fractions 6–18). These clusters included proteins that are part of large, well-known complexes, such as the majority of PS subunits, Rubisco, RNA Polymerase (RNAP), most ribosomal proteins, and a small number of largely undescribed proteins, depicting promising candidates for interaction with such complexes or potential RBPs, especially if not conserved in E. coli, S. enterica, or Arabidopsis (Figure 2C).

Many transcripts show bimodal sedimentation profiles
Most transcripts were assigned to one of the three Clusters 10, 13, or 14. The transcripts belonging to Clusters 13 and 14 had largely similar sedimentation profiles and were analyzed together, while the sedimentation profiles of Cluster 10 transcripts clearly differed. The transcripts from Cluster 10 mainly occurred in Fractions 5–7 but with a second occurrence above Fraction 8, often peaking with Fraction 9. Therefore, we grouped Cluster 10 transcripts according to their main peaks, yielding three subgroups (Figure 3A). In these subgroups, the main peaks of mRNAs, their associated asRNAs, and, to the extent of our knowledge, their respective regulatory ncRNAs, largely overlapped. Hence, mRNA-asRNA and mRNAsRNA hybrids may exist in these fractions. However, the secondary peaks for mRNA occurrences did not overlap with asRNA or ncRNA peaks. Hence, these mRNAs cofractionated with their putative respective RNA regulators in the primary peak fractions and fractionated separately in the secondary peak fractions.

We grouped transcripts from Clusters 13 and 14 in the same manner, as a function of the number of peaks and their corresponding fraction(s). Transposase mRNAs are a major constituent of these clusters. Notably, only two annotated ncRNAs, Ncr0560 and Ncr1310, were included in Cluster 14, exhibiting a single peak in Fraction 9. Ncr0560 originates from a transcriptional start site internal to gene slr2062 encoding the IS200/IS605 family element transposase accessory protein TnpB and therefore is closely connected to transposases. However, most transcripts attached to Clusters 13 and 14 occurred in the narrow range of Fractions 9 and 10, cosedimenting with RNAP subunits, but also with secondary peaks before Fraction 8, often in Fraction 4 where they cosedimented with the 30S ribosomal protein S1 homologs Rps1A and Rps1B (Figures 3B and 4A).

Grad-seq analysis allows the detection of ribonucleoprotein complexes
We detected 53 ribosomal or ribosome-associated proteins and their ribosomal RNAs (rRNAs) in the fractions containing the complexes with the highest sedimentation coefficients. A preference for the entire set of ribosomal components was restricted to Fraction 18 (Figure 4A). Notable exceptions were the 50S ribosomal protein L11 (Rpl11, encoded by slr1743), Rpl12 (slr1746 gene), Rpl16 (slr1805 gene), Rps2 (slr1260 gene), and Rps10 (slr1101 gene), pointing to their association with other complexes or transcripts. Moreover, Ycf65 encoded by slr0923, previously

Figure 2 (Continued)
and Goto, 2000; Kanehisa, 2019; Kanehisa et al., 2019). Proteins not included in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database or annotated as “Function unknown” or “General function prediction only” were manually annotated as “Function unknown.” The phylogenetic distribution of likely orthologs was determined based on the domclust algorithm in the Microbial Genome Database (Uchiyama et al., 2019) and 59 selected genomes (Supplemental Table 1). As some examples, RNAP subunits, most ribosomal and photosystem subunits or the Rubisco cluster together based on their sedimentation profiles and group together based on their phylogenetic occurrence. Largely uncharacterized proteins assigned to the RNA-dominated branch and not conserved in Arabidopsis, Salmonella or E. coli are promising candidates as potential cyanobacterial RNA chaperones (highlighted by the dashed rectangle). The circle sizes correlate with their respective constituents (scale to the right lower end). The distribution of homologs in the selected reference organisms is indicated by red (absence) or green (presence) rectangles at the bottom of the panel. For the visualization of sedimentation velocity versus molecular weight, see Supplemental Figure 3.
reported as a plastid-specific ribosomal protein in some eu-
karyotic algae (Turmel et al., 2009) and land plants
(Yamaguchi and Subramanian, 2003), did not associate with
other ribosomal proteins, and nor were the Rps1A and
Rps1B proteins encoded by slr1356 and slr1984, respectively
(Figure 4A). Rather, Rps1A and Rps1B may play a role in the
Shine–Dalgarno-independent initiation of translation in cya-
nobacteria (Mutsuda and Sugiura, 2006; Nakagawa et al.,
2010). Their occurrence in Fractions 3–5 (Rps1A) or 3–10
(Rps1B) indicates a wide set of associations to different tran-
sic transcripts, while their lack of association with the ribosomal
subunits likely is due to the buffer conditions used here.

CRISPR-Cas systems are particularly interesting macromo-
lecular complexes that consist of both protein and RNA
components. CRISPR-Cas systems confer a form of acquired
immunity against invading nucleic acids in many different
bacteria and archaea and can be classified into greatly vary-
ing subtypes that belong to two major classes: multisubunit
effector complexes in Class 1 and single-protein effectors in
Class 2 (Makarova et al., 2020). Despite the high level of di-
vergence among different systems, CRISPR-Cas systems usu-
ally consist of repeat-spacer arrays and the associated
protein-coding cas genes.

Synechocystis encodes three separate and complete
CRISPR-Cas systems that are highly expressed and active in
interference assays (Scholz et al., 2013; Reimann et al., 2017;
Behler et al., 2018; Kieper et al., 2018; Scholz et al., 2019).

According to the presence of signature genes, these three
systems have been classified as subtypes I-D, III-D, and III-Bv
(Scholz et al., 2013; Behler et al., 2018).

The gradient profiles of the crRNAs of the three CRISPR
systems of Synechocystis varied vastly from one another. While CRISPR1 RNAs (cr1RNAs) were mainly detected in Fraction 9 (Cluster 14), CRISPR3 RNAs (cr3RNAs) peaked in Fraction 10 (Figure 4B). However, the detected CRISPR2-
associated proteins colocalized together with the cr2RNAs
in low molecular Fractions 4–6, at the same sedimentation
level as tRNAs. The detected CRISPR1-associated proteins
were scattered throughout the gradient, with only the Cas7/
Csc2 protein (encoded by slr7012) cofractionating with
cr1RNAs in higher molecular Fractions 9–11. This tight asso-
ciation with cr1RNAs is consistent with the known function
of this protein in binding the variable crRNA spacer se-
quence (Hrle et al., 2014). All structural proteins associated
with the CRISPR3 system (proteins Cmr1, Cmr5, Cmr4,
Cmr3, and Cmr2/Cas10 encoded by the genes sll7085 to
sll7087, sll7089, and sll7090) fractionated together and with
the CRISPR3-associated crRNAs, with a peak in Fraction 10. This
cofractionation indicates that the interference complex
remained largely intact during the preparation. However,
these proteins also exhibited a second peak in the lower
Fractions F3–F4 without associated cr3RNA or the Cas10
protein (encoded by slr7090), hence constituting a poten-
tially nonfunctional subcomplex or resulting from partial
**Figure 4** In-gradient distribution of ribonucleoprotein complexes in *Synechocystis*. Heatmap representation of standardized relative abundances (z-score) for each detected protein and RNA in different shades of gray (proteins) or red (RNA) alongside the fractions of the sucrose density gradient. Cluster affiliations are shown on the right (cluster color code as in Figure 2 and Supplemental Figure 2). A, Ribosomal RNAs and proteins. The fractions of co-occurring 16S rRNA and 30S ribosomal proteins as well as 5S, 23S rRNA, and 50S ribosomal proteins are indicated by a red dashed box. The names of ribosome-associated proteins with diverging gradient profiles are indicated in red, such as Ycf65, Rps1A, Rps1B, Rps10 (30S subunit), Rpl12, Rpl11, and Ssl0438 (50S subunit). B, Crispr RNAs and associated proteins. The fractions of respective cRNAs and co-occurring CRISPR proteins are indicated by red dashed boxes. C, RNAP subunits, sigma factors, and 6S RNA, essential housekeeping RNAs and their protein binding partners, as well as selected proteins involved in RNA metabolism. The fractions of co-occurring RNAP subunits, Sigma factors, and 6S RNA, as well as the tmRNA, RNase P RNA, and SRP RNA co-occurring with their respective cognate binding proteins, are indicated by red dashed boxes. Sigma factors and proteins involved in RNA metabolism with peaks above Fraction 5 are marked in red.
disassembly of the complex during lysis preparation or centrifugation (Figure 4B). Interestingly, the slr7080 gene product, a Csx3 accessory protein, cofractionated with these two CRISPR-Cas complexes. Cas accessory genes encode diverse groups of proteins that co-occur only sporadically with CRISPR-Cas systems and are not essential for their functionality (Shmakov et al., 2018; Shah et al., 2019). Slr7080 has been characterized as a Csx3-AAA protein (Shah et al., 2019), a distant member of the Cas Rossmann fold (CARF) protein family (Makarova et al., 2014). CARF proteins sense ligands such as cyclic oligoadenylate (cOA) signaling molecules that can be synthesized by the polymerase domain of Cas10 upon recognition of its target (Kazlauskienė et al., 2017; Niewoehnner et al., 2017; Han et al., 2018). The Archaeoglobus fulgidus Csx3 protein AfCxs3 is a ring nuclease that degrades the CRISPR signaling molecule cOA4 (Athukoralage et al., 2020). However, Csx3/Slr7080 is substantially larger than AfCxs3 (310 compared to 110 amino acids) and possesses an additional AAA ATPase domain in its C-terminal region (Shah et al., 2019). Our data suggest that Csx3/Slr7080 is expressed relatively abundantly and that its product Slr7080/Csx3 is likely physically linked to both the cr3RNA-free subcomplex and the complete functional interference complex.

We found compelling evidence for an even more intriguing complex of Cas accessory proteins for the CRISPR2 system, in which the four proteins Csx1/Slr7061, Csm6/Sll7062, Sll7070, and Sll7073 overlapped with the fractionation profile of cr2RNAs (Figure 4B). Csx1/Slr7061 was bioinformatically classified as a CARF-HEPN (Higher Eukaryotes and Prokaryotes Nucleotide-binding) domain protein and Csm6/Sll7062 as a CARF-RelE domain-containing protein (Shah et al., 2019). The protein structurally most closely related to Csx1/Slr7061 is the Sulfolobus islandicus Csx1 protein, an RNase that becomes allosterically activated upon cOA4 binding (Molina et al., 2019). Csm6 is a cOA-dependent CRISPR ribonuclease as well (Garcia-Doval et al., 2020); hence, our data suggest the presence of a ribonucleolytic defense complex associated with cr2RNAs. The other two proteins, Sll7070 and Sll7073, are entirely uncharacterized, but the genomic localization of their corresponding genes that frame the CRISPR2 cas1 cas2 adaptation module suggests a functionality closely associated with the CRISPR2 system. Prior to this work, neither the accumulation of Csx1/Slr7061, Csm6/Sll7062, Sll7070, and Sll7073 proteins nor their likely physical association with cr2RNAs was known. Hence, these proteins are suggested as highly interesting candidates for future experimental analyses.

The cyanobacterial RNAP is unique to splitting of the β′ subunit into an N-terminal γ (RpoC1) and a C-terminal β′ (RpoC2) subunit (Schneider and Hasekorn, 1988) as well as an insertion of >600 amino acids nterminal RpoC2 (Iyer et al., 2004). Moreover, the interplay between the core subunits, multiple sigma factors (9 in Synechocystis, SigA to SigL) and inhibitory 6S RNA is only partially understood (Heilmann et al., 2017). We observed that all RNAP core subunits colocalized with the sigma factors SigB and SigC, as well as 6S RNA, SigE, and SigA in higher molecular fractions, indicating the parallel presence of multiple forms of the RNAP holoenzyme. By contrast, the bulk of the RNAP ω subunit (RpoZ), as well as SigF and SigG, together with three detected anti-sigma factors, were all distributed only into the low molecular fractions, suggesting the presence of a reservoir of inactive binary complexes (Figure 4C). Furthermore, we noticed the mutually exclusive occurrence of RpoZ and the 6S RNA together with the RNAP core subunits (see F8 in Figure 4C). This observation may explain the observation that cyanobacterial RpoZ facilitates the association of the primary sigma factor SigA with the RNAP core, thus stimulating transcription of highly expressed genes (Gunnelius et al., 2014). In E. coli, 6S binds the β/β′ subunits (Wassarman and Storz, 2000). Hence, RpoZ may affect this binding by interacting with the γ and β′ subunits in Synechocystis (Gunnelius et al., 2014).

RNase P and the tmRNA (SsrA) showed similar sedimentation patterns as the RNAP holoenzyme and 6S RNA, while the signal recognition particle (SRP) mostly clustered together with the bulk of mRNAs (Figure 4C). Several other factors, which were annotated with different functionalities in RNA metabolism, were associated with larger complexes. Such factors were the DEAD-box RNA helicase ChhR, a putative SNF2 helicase encoded by sll1366, polynucleotide phosphorylase (PNPase), enolase, RNase J, and RNase E. The latter enzyme is central for RNA metabolism in bacteria because it is involved in numerous functions, such as the posttranscriptional regulation of gene expression, maturation of transcripts, or initiation of mRNA degradation (Bandrya and Luisi, 2018). Examples in Synechocystis include the posttranscriptional regulation of gene expression with the sRNA PsrR1 as specificity factor (Georg et al., 2014), operon discoordination of the rimO-crhR dicistron (Rosana et al., 2020) or the maturation of functionally active crRNAs from the CRISPR3 system (Behler et al., 2018). Indeed, RNase E was detected in several fractions, peaking together with ribosomal proteins, co-occurring with the bulk of mRNAs as well as cr3RNAs (Figure 4C). Cyanobacterial RNase E forms a complex with PNPase through an oligo-Arg repeat (Zhang et al., 2014), consistent with their co-occurrence in Fraction 10 and their separate occurrences in other fractions (Figure 4C), indicating that the PNPase-RNase E complex is likely only one of several different combinations in which RNase E exists in the cell.

These results illustrate the sedimentation profiles of native ribonucleoprotein complexes in a sucrose density gradient, supplying valuable information about the compositional structure of such complexes in a cyanobacterial model organism. In contrast to the Grad-seq analysis performed with Salmonella (Smirnov et al., 2016), we detected the RBP Hfq in the lightest fractions, therefore not cosedimenting with any RNAs or known ribonucleoprotein complexes (Figure 4C), confirming the different functionality of this protein in cyanobacteria.
Sedimentation profiles of mature and intermediate protein complexes associated with the thylakoid membrane illustrate the complexity of a photosynthetic lifestyle

Focusing on protein complexes of the cyanobacterial thylakoid membrane helps to understand the factors that maintain its photosynthetic machinery. Most core components of these complexes cosedimented, indicating the presence of largely intact PSI, PSII, and cytochrome-b6f core complexes (Figure 5A), NADH-dehydrogenase (NDH-1), and cytochrome-c-oxidase complexes as well as F0F1-ATP synthase (Supplemental Figure 4A–C). The detection of several small proteins, such as the NDH-1 subunits NdhP encoded by sml0013 (Schwarz et al., 2013; Zhang et al., 2014a), NdhN, NdhH, NdhJ, and NdhO (He and Mi, 2016), cosedimenting with other NDH-1 proteins, supported the high quality of the data. The soluble electron carriers plastocyanin (PetE), cytochrome c6 (Petl), and ferredoxins (PetF, Fdx) were mainly located between the lowest molecular Fractions 2 and 3 and were not associated with any complex, with the exception of Fdx7 (Supplemental Figure 4D).

The occurrence of the PSII and PSI core complexes was in agreement with the measured absorption spectra for fully assembled complexes in Fractions 6 and 7 (Figure 5B). The three membrane-extrinsic PSII subunits PsbO, PsbV, and PsbU, located on the luminal side of the thylakoid membrane, were disconnected and cosedimented in Fraction 3 (Figure 5A). We identified several other PSII accessory proteins in lower fractions as well, including Psb27 (Figure 5A) and most PSII assembly factors (reviewed by Nickelsen and Rengstl, 2013; Supplemental Figure 4E). A PSII subcomplex lacking PSII core reaction center polypeptides but containing Psb27 was described as a PSII repair cycle chlorophyll-protein complex (Weisz et al., 2019); therefore, corresponding complexes in transition between disassembled and assembled PSII were likely present in Fractions 4–5. This hypothesis is supported by the measured absorption spectra, which mainly showed the presence of carotenoids and absence of chlorophylls (Figure 5B). Detected PSII assembly factors cosedimenting with and potentially associated with PSII subcomplexes included Slr1471, PratA/Slr2048, CtpA/Slr0008, and Ycf48/Slr2034, which are involved in D1 integration, processing, and stabilization in the early stages of reaction center assembly; Psb27/Slr1645 and Slr0606, which are involved in CP43 assembly and integration during the later stages of reaction center assembly and PSII repair cycle; and Slr1768, Sll1390, and Sll1414, which are involved in D1 turnover as part of the PSII repair cycle (Nickelsen and Rengstl, 2013). Interestingly, few PSI assembly factors showed a more complex sedimentation pattern with additional peaks in higher molecular fractions. Slr1768 exhibited a second peak in Fraction 11 and Slr2013 in Fraction 18, while Slr0933, the homolog of the PSI assembly factor PHOTOSYNTHESIS-AFFECTED MUTANT 68 (PAM68) in Arabidopsis (Rengstl et al., 2013; Rast et al., 2016), solely sedimented in higher molecular Fraction 11 and from Fractions 14–18, leaving their exact role in PSII assembly in *Synechocystis* open for discussion.

Figure 5 Heatmap showing the in-gradient distribution of the subunits of major protein complexes involved in photosynthesis. A, The sucrose density gradient (w/v 10%–40%) showed a characteristic color pattern (left) corresponding to the pigments bound to distinct macromolecular complexes detected in the same sedimentation range (grayscale heatmap). The in-gradient distribution of detected photosystem subunits and phycobiliproteins is given in the heatmap with standardized relative abundances (z-score). The profiles of CpcG1 and CpcG2 are shown separately to illustrate the difference between the G-type PBS (CpcG1 as rod-core linker) and L-type PBS (CpcG2 as rod-core linker) types (Kondo et al., 2005; Liu et al., 2019). PsbP, refers to the protein Sll1418 that has been characterized as a PsbP-like protein associated with PSI (Ishikawa et al., 2002; Summerfield et al., 2005; Sveshnikov et al., 2007). Below the heatmap, complex subunits are abbreviated with the last letter of their encoding gene name. Subunits of PSI, e.g., psaA, of the cytochrome-b6f complex, e.g., petA, subunits of PSII, e.g., psbA, subunits of PBS, e.g., phycocyanin cpcA, or allophycocyanin apcA. Green coloring indicates membership to the core complex, while light green refers to subcomplexes, accessorial parts of the complex, or assembly factors, showing gradient profiles differing from the corresponding core complex. For the in-gradient distribution of other membrane-associated complexes and proteins associated with photosynthesis see Supplemental Figure 4. B, Absorption spectra taken from the indicated pigmented fractions. The dashed lines indicate the average absorption spectrum of all pigmented fractions.
Furthermore, our data illustrated the roles of PsaL and PsaK in PSI trimer formation. PsaL and PsaK are the last subunits to be incorporated into PSI in the trimer assembly process, with the incorporation of PsaK2 mainly involved in state transition and PsaK1 in the constitutive state (Fujimori et al., 2005; Dühring et al., 2007), while PsaL is required for PSI trimer formation (Chitnis and Chitnis, 1993). The intact PSI monomer was easily detected, but the PSI trimer peak was only weakly detectable, since the majority of PSI appeared to occur as monomers, possibly a consequence of our sample preparation procedure. However, the peak appeared more clearly when the proteomics data were normalized to constant protein amounts per fraction (Supplemental Figure 4G and 4H). Matching previous findings (Fujimori et al., 2005; Dühring et al., 2007), our data showed the cosedimentation of PsaL and PsaK1 with the fully assembled PSI monomer as well as the trimer, while the majority of PsaK2 was only partially incorporated into the PSI monomer together with most assembly factors. Similar to previous observations of multiple peaks of some PSII assembly factors, Ycf4/Sll0226 exhibited a second peak; VESICLE-INDUCING PROTEIN IN PLASTIDS 1 (VIPP1/Sll0617) exhibited two more peaks in higher molecular fractions (Supplemental Figure 4F).

The sedimentation profiles of phycobiliproteins pointed to another aspect of the complex dynamics of the photosynthetic apparatus. Two types of phycobilisomes (PBS) were previously described. The main phycobilisome (CpcG1-PBS type) is characterized by CpcG1 as a rod-core linker; the smaller CpcL-PBS type is characterized by CpcG2 as its rod core linker (Kondo et al., 2005; Liu et al., 2019). The distinction between these two types according to their protein composition illustrated the different sedimentation profiles of CpcG2-PBS and CpcG1-PBS within the pigmented sedimentation range (Figure 5). The association of CpcG2 with PsaK was suggested to function as a PSI antenna involved in state transitions under high light conditions (Kondo et al., 2005; Liu et al., 2019). As part of an NDH1-PSI-CpcG2-PBS supercomplex, CpcG2 is involved in facilitating efficient cyclic electron transport (Gao et al., 2016). Interestingly, ferredoxin-NADP⁺ reductase (FNR, encoded by PetH) and PsaK1 colocalized with CpcG1-PBS. The association of FNR with PBS has been reported previously (Kondo et al., 2005; Liu et al., 2019), while the interaction of PsaK1 and CpcG1-PBS, analogous to the interaction of PsaK2 and CpcG2-PBS, had not yet been observed until now.

**Smaller complexes and protein–protein interactions in toxin-antitoxin systems and regulation**

In addition to the analysis of ribonucleoprotein complexes and membrane-bound protein complexes, the dataset also provided clues about smaller complexes comprising a limited number of interacting proteins.

One class of such complexes with frequently only two proteins is the toxin–antitoxin (TA) systems. TA systems consist of a stable toxic protein and its unstable cognate antitoxin. In the case of Type II TA systems, both components are small proteins (Schuster and Bertram, 2013). *Synechocystis* was previously predicted to code for 69 Type II TA systems (Kopfmann et al., 2016). While the majority of their encoding genes are borne on plasmids, consistent with their function in plasmid maintenance, 22 predicted TA systems are encoded on the main chromosome. These must have functions distinct from postsegregational killing but are severely understudied. We detected at least four such TA systems: the pairs Slr0770/Slr0771 and Ssl0258/Ssl0259 in Cluster 2; Ssl2138/Sll1092 in Cluster 5, and Ssl2245/Sll1130 in Cluster 8 (Figure 6A). Of these, the pairs from Cluster 2 have not been studied, while the functionality of the pair Ssl2138/Sll1092 as a TA system was experimentally validated (Ning et al., 2013). Some evidence indicates that the pair Ssl2245/Sll1130 is involved in a wider regulatory context (Srikumar et al., 2017). Interestingly, Ssl2245/Sll1130 was assigned to Cluster 8 suggests that they may be part of a larger complex. Hence, our data not only support the existence of TA systems but also show their active expression and support their supposed mode of action due to their cofractionation.

Other smaller protein complexes have regulatory functions. Here, we chose proteins interacting with components of the nitrogen regulatory system for illustration. The GntR-like transcriptional regulator PlmA is in an inactive state when bound to the signaling protein PiII bound to PiI-interacting protein X (PipX; Labella et al., 2016) but is
involved in phycobilisome degradation when activated under nitrogen starvation (Sato et al., 2008). The formation of the PII-PipX complex also controls the activity of the transcriptional master regulator for nitrogen starvation, NtcA (for review see Labella et al., 2020a). Under nitrogen replete conditions, PII forms a complex with either N-acetylglutamate kinase (NAGK, encoded by argB), which stabilizes it and therefore enhances Arg biosynthesis. Furthermore, PII binds about 85% of all available PipX to sequester it away from NtcA and thus inhibit its regulatory properties. The remaining 15% of PipX is bound by PlmA to keep it inactive. Under low nitrogen conditions, PII is sequestered by 2OG, and NtcA-dependent gene expression is activated upon binding to PipX (Llacer et al., 2010; Forcada-Nadal et al., 2018). Our data illustrated the co-occurrence of the PII-PipX complex and all mentioned proteins interacting with its single components (Figure 6B). Another protein interacting with PII during high nitrogen conditions was the membrane protein PamA, implicated in the control of nitrogen- and sugar metabolism-related genes (Osanai et al., 2005). Notably, this was the only previously mentioned proteins that displayed peaks outside the range of the PII-PipX complex, pointing to its potential association with another, larger complex (Figure 6B). Interestingly, the single occurrence of PipX in Fraction 9 (Figure 6B) might be linked to its recently proposed new functionality as an interactor with the translation machinery (Cantos et al., 2019).

Classification of sedimentation zones linked to various steps in protein biosynthesis assists in the identification of cyanobacterial RNA chaperones

Most transcripts accumulated in a narrow gradient range, equivalent to large protein complexes of 300 kDa in size, similar to the PS monomers. Intriguingly, this section was largely free of known RNA-interacting complexes, indicating the presence of one or more RNA chaperones responsible for the characteristic sedimentation pattern. We classified the transcripts from Cluster 10 as Group 1 RNAs. Other transcripts assigned to Clusters 13 and 14 were classified as Group 2 RNAs (Figure 7).

Most RNA-interacting complexes involved in protein biosynthesis sedimented deeper, accumulating in higher molecular fractions. These proteins included the RNAP holoenzyme, 6S RNA, RNase P, tmRNA (encoded by ssrA), several other proteins functionally linked to RNA degradation and a subset of ribosomal proteins. SmpB, the cognate RBP for tmRNA, occurred in Fraction 2, indicating a substantial amount of free protein. However, the second SmpB peak in Fractions 7 and 8 overlapped with tmRNA, indicative of the formation of a complex (Figure 4C). The co-occurrence of tmRNA and Rps1A in Fractions 8 and 10 (Figure 4A) is consistent with reports of the direct physical interaction between Rps1 and tmRNA in E. coli (Wower et al., 2000).

A large set of transcripts, including 16S rRNA, peaked in the same range, indicating the location of a large transcription-translation complex still binding to RNAs (Figure 7). Proteins that sedimented in these fractions were another group of interesting candidates with potential involvement in RNA metabolism. Interestingly, although CRISPR1, CRISPR3, and most transposase mRNAs largely cosedimented with those complexes, their involvement in specific regulatory aspects is not known.

Based on these findings, we performed further analysis to characterize the proteins cosedimenting with transcripts and to determine interesting candidates for future work (Supplemental Figure 5). We primarily focused on 19% (451/2,394) of all detected proteins, which were assigned to RNA-dominated clusters in the higher molecular fractions (cluster ≥ 10), to which the essential complexes involved in protein biosynthesis also belonged. Specifically, 27% (124/451) of the proteins that were annotated as having unknown functions were of high interest. Of those, a total of 41% (51/124) were conserved in at least 50% of selected cyanobacterial genomes but not in E. coli, S. enterica, or Arabidopsis (Supplemental Table 1). Hence, these 51 proteins were conserved, but largely undescribed and exhibited sedimentation profiles linking them to RNA metabolism. To aid our selection further, we computed a SVM score for the prediction of RBPs from their amino acid sequence using the RNApred webserver (Kumar et al., 2011). We performed this prediction for all proteins and included the SVM score in the database associated with this article. Approximately 27% of the resulting 51 proteins fulfilled the SVM score criteria for probably and very probable RBPs, which were set according to the RNApred performance for Synechocystis proteins (Supplemental Figure 6), narrowing the list of RNA chaperone candidates to only 14 proteins (0.67%) of the initially 2,394 detected proteins (Table 1).

Manual curation of this shortlist of 14 proteins that had passed the criteria of appropriate cluster assignment, unknown functional categorization, minimum phylogenetic conservation, and minimum SVM score resulted in the identification of three proteins that were unlikely to be involved with nucleic acids due to their previously described functions. These proteins were SII1515, the glutamine synthetase-inactivating factor IF17 (Garcia-Dominguez et al., 1999); SII0169, a CsoD1 ortholog that is part of the β-carboxysome, containing two bacterial microcompartment (BMC) domains (Kinney et al., 2011); and SII1939, an FtsZ-interacting Ftn6 homolog involved in cell division (Marbouty et al., 2009). Nevertheless, we kept them in the list because a small number of bifunctional proteins exist in bacteria that are active in metabolism and in controlling gene expression, including mechanisms in which RNA is bound (Commichau and Stülke, 2008).

Another four proteins might interact with nucleic acids but less so with RNA. These proteins were SII2874, a homolog of RemA, an essential regulator of biofilm formation in Bacillus subtilis (Winkelman et al., 2009, 2013); SII0639, a DHH phosphoesterase domain-containing protein with similarities to a RecJ domain protein, nucleotidyltransferase/poly(A) polymerase or exopolyphosphatase-like enzyme;
Figure 7 Sedimentation zones functionally linked to gene expression. Diagram illustrating the sedimentation profiles of intact macromolecular protein complexes (left) alongside the sucrose density gradient, matching the observed pigment pattern. The overall protein and RNA distribution along the gradient are depicted on the right side of the gradient diagram. The majority of proteins (~80%) occurs in the lower molecular fractions (below F6, Clusters 1–9), indicating their association with small complexes or occurrence as soluble proteins. Proteins that fall into that group and are involved in gene expression are the majority of transcription factors and two component systems, anti-Sigma factors, the Sigma factors SigG and SigF, several proteins involved in RNA degradation and tRNA metabolism. The minority of proteins (~20%) occurs in the higher molecular fractions (above F5, Clusters 10–17), indicating association with large complexes. Proteins that fall into that group and are involved in protein biosynthesis are RNA polymerase (RNAP), the Sigma factors SigC, SigB, SigE, and SigA, several proteins and RNA–protein complexes involved in RNA metabolism, such as RNase D, RNase J, RNase E, RNase P, PNPase, enolase, CrhR, 6S RNA, tmRNA, the SRP, and the ribosomes. A large portion of these proteins and RNA–protein complexes co-occurs in Fractions 8–11. Most detectable RNAs occur in Fractions 8–11 as well, indicating their likely direct association with complexes composed of proteins involved in transcription, RNA degradation, RNA processing, or translation initiation. Notably, most detected RNAs fall into two groups. Group 1 comprises RNAs that sediment in high abundance in Fractions 6–7 (Cluster 10), potentially forming stable interactions with RNA chaperones. Group 2 includes RNAs that generally sediment in lower abundances in Fraction 4 (Clusters 13–14), potentially not bound to any RNA chaperone and thus are not as stable when not directly associated in a larger complex. The tRNAs are the exception, mainly sedimenting in high abundance in Fraction 5 (Cluster 8) together with the majority of tRNA synthetases and likely not bound to a larger complex, while tRNAs seem to be solely attached to the ribosomes.

Table 1 Shortlist of RNA chaperone candidates based on multiple filter criteria (shown in Figure 7)

| Locus tag | Cluster | Max Fractions | aa | Relative conservation | SVM score | Comments and annotation |
|-----------|---------|---------------|----|-----------------------|-----------|-------------------------|
| slr1424   | 10      | 6 6, 10 491   | 1  | 0.78                  | Putative SMC (structural maintenance of chromosomes) domain |
| slr1515   | 10      | 6 6 149      | 0.71| 1.97                  | glutaminate synthetase inactivating factor IF17          |
| slr1069   | 10      | 7 7 213      | 0.94| 0.99                  | CsoD1 ortholog (part of β-Carboxysome)                  |
| slr0211   | 10      | 7 4 403      | 0.99| 0.83                  | acyl-CoA N-acyltransferase domain                        |
| slr1081   | 10      | 7 4 210      | 0.75| 0.56                  | DUF820 featuring a restriction endonuclease domain       |
| ssl2064   | 10      | 7 4 75       | 0.64| 2.22                  | DUF4160                                               |
| ssl2874   | 10      | 6 6 4, 11 89 | 0.9 | 2.42                  | RemA homolog; KEGG K09777                              |
| ssl0319   | 11      | 8 3 297      | 0.57| 1.55                  | DUF3747, periplasmic protein                            |
| slr0287   | 12      | 9 5 118      | 0.9 | 1.02                  | KH domain-containing protein; KEGG K06960               |
| slr1193   | 14      | 9 3, 9, 11, 13 214 | 0.89 | 0.53                | FtsZ interacting Ftn6 homolog                          |
| slr0284   | 15      | 11 11 579   | 0.93| 0.64                  | DUF87 containing an N-terminal HAS-barrel domain; KEGG K06915 |
| ssl0639   | 15      | 12 7, 12 417 | 0.74| 0.63                  | DHH phosphoesterase domain-                             |
| sslr1238, slr0743a | 15 | 11 11 | 0.71 | 1.04 | YlxR homologue                                          |
| slr1660   | 17      | 18 18 192   | 0.91| 1.41                  | DUF3172 with N-terminal transmembrane domain            |

The locus tag is given, followed by the cluster number to which the protein was assigned, the fractions in which the abundance peaked (Max) and the fractions in which peaks occurred. The columns give the predicted number of amino acids (aa), the relative conservation and SVM scores, and finally comments and annotation. All listed proteins were assigned to the phyllogenetic group “Cyanobacteria.” For details of the taxonomic distribution and SVM scores, see Supplemental Figures 5 and 6. For a synteny analysis of selected RNA chaperone candidates, see Supplemental Figure 7.
Slr1081 (containing the domain of unknown function DUF820) featuring a restriction endonuclease domain; and Slr0284 (carrying DUF87) containing an N-terminal HAS-barrel domain similar to archaean proteins, such as the DNA double-stranded break repair helicase HerA. However, bacterial proteins containing similar domains are largely undescribed.

Another four poorly characterized proteins were Slr0211, an acyl-CoA N-acyltransferase domain containing protein; Slr0319 (DUF3747), a periplasmic protein; Slr1660 (DUF3172) containing an N-terminal transmembrane domain; and Ssl2064, a 75 amino acid DUF4160-containing protein.

Ultimately, three candidates appeared most promising. The first was Sll1424, an undescribed protein whose encoding gene is located immediately downstream of the transcriptional regulator gene ntcA in many cyanobacterial genomes (Supplemental Figure 7A). Furthermore, we noticed Ssr1238/Slr0743a, an 84 amino acid YkrR homolog with a DUF448 domain and a pronounced syntenic across cyanobacterial genomes around the ssr1238 locus, which lies between nusA, encoding the transcription termination factor NusA, and infB, encoding the translation initiation factor InfB (Supplemental Figure 7B). While the order of adjacent genes rimP-nusA-infB is conserved even in Salmonella and E. coli, a gene corresponding to ssr1238 is missing from their genomes. The relationship between ykrR with rimP-nusA-infB has also been observed in recent analyses of cyanobacterial gene linkage networks (Labella et al., 2020b).

Finally, we identified Slr0287, a K Homology (KH) domain-containing protein of 118 amino acids that shares 30% identity and 56% similarity with the corresponding protein KhpA from S. pneumonia. KhpA forms a complex with the JAG/EloR-domain protein KhpB and RNA-binding complex in S. pneumonia (Zheng et al., 2017). Indeed, we noticed that a KhpB homolog also exists in Synechocystis and is encoded by slr1472. Both proteins are widely conserved throughout the cyanobacterial phylum and even clustered together in the gradient (Supplemental Figure 5). The genes slr0287 and slr1472 were syntenic with genes encoding proteins involved in RNA metabolism and translation, such as slr0287 and rps16 (Supplemental Figure 7C) or slr1472 and rpl34, rnpA, slr1470 (encoding a membrane protein), and slr1471, encoding a YidC/Oxa1 homolog (Supplemental Figure 7D). Slr1471/Oxa1 is essential for thylakoid biogenesis and the membrane integration of the reaction center precursor protein pD1 (Spence et al., 2004; Ossenbühl et al., 2006). The synteny of rnpA, yidC/oxa1, and khpB as well as khpA and rps16 is strikingly wide, extending even to gram-positive bacteria, such as S. pneumonia (Supplemental Figure 7D).

Data accessibility and visualizations
To facilitate maximum usability of this resource, we have made the entire dataset available at https://sunshine.biologie.uni-freiburg.de/GradSeqExplorer/, providing multiple tools to focus on different fractions of the proteome or transcriptome, from groups of functionally related proteins and transcripts to the level of the individual protein or sRNA. Multiple filter options, a customizable graphical output, as well as a tabular output options, provide extensive background information to the findings presented in this analysis.

Discussion
Global analysis of RNA–protein complexes in a photosynthetic cyanobacterium
Based on their interaction with different types of RBPs, five major classes of ncRNAs can be differentiated that are common to many bacteria (Hör et al., 2018). These include the 6S RNA as a global modulator of RNA polymerase (RNAP) specificity and activity (Wassarman and Storz, 2000; Barrick et al., 2005), Hfq-binding sRNAs (Zhang et al., 1998; Møller et al., 2002; Moll et al., 2003), ProQ-binding sRNAs (Smirnov et al., 2016), CsrB-like sRNAs that antagonize the translational repressor CsrA (Liu et al., 1997), and CRISPR RNAs (crRNAs), which form ribonucleoprotein particles together with the CRISPR-associated (Cas) proteins for defense (Brouns et al., 2008; Hale et al., 2009). Transacting sRNAs and their effects on gene expression have been well studied in E. coli and S. enterica (Hör et al., 2020c) but much less so in photosynthetic cyanobacteria. While 6S RNA and crRNAs exist in cyanobacteria, the association of certain sRNAs to certain RBPs has remained unresolved.

Our Grad-seq analysis of Synechocystis provides an in-depth global analysis of major stable RNA–protein complexes in a photosynthetic cyanobacterium. We anticipate that our data will not only provide crucial knowledge for this particular model cyanobacterium but also constitute a valuable resource for other cyanobacteria, including the models Anabaena (Nostoc) sp. PCC 7120 and Synechococcus elongatus.

Synechocystis is a cyanobacterial model that deviates substantially from other bacteria investigated in previous Grad-seq studies (Smirnov et al., 2016; Hör et al., 2020b). One aspect is the magnitude of internal membrane systems and membrane-bound protein complexes. We showed that even macromolecular membrane-bound protein complexes remained largely intact during the extraction and gradient process, and their in-gradient localization matched the pigment composition of their corresponding fraction in a meaningful way. Hence, the analysis provides a spatially resolved proteome dataset of high quality for the macromolecular complexes of a photosynthetic bacterium. In addition, the occurrence of subunits specific for certain subcomplexes or specialized functions provides information about their in vivo state and illustrates the dynamics in complex formation. Examples for such acclimation processes include the incorporation of PsaK2 into PSI as the high light-adapted form instead of the constitutive form, which mainly incorporates PsaK1. In our data, the occurrence of PsaK2 was restricted to disassembled and monomeric PSI, while PsaK1 occurred mainly in fully assembled monomeric and trimeric PSI, capturing the in vivo state of Synechocystis.
PSI and a multitude of other complexes under standard growth conditions.

Our data not only provide information about known protein–protein interactions but also provide clues about currently unknown interactions, such as the given examples of the PSI assembly factors Slr0933, Slr1768, and Slr2013, the PSI assembly factor VIPP1 homolog, the Pr-interacting protein PipX or the membrane protein PamA with unexplained occurrences in the heavy fractions and outside the range of their known interaction partners, which illustrate the potential value of this dataset.

Protein candidates for interaction with sRNAs
To our surprise, the gradient profiles of most RNAs diverged vastly from those of most detected proteins, locating generally in the heavier fractions and with largely similar sedimentation patterns, independently from transcript lengths. This fact is a strong indication of association with one or more major RNA chaperones or to large intact protein complexes involved in transcription or translation, such as RNAP, ribosomes, and associated factors.

The colocalization of many known intact RNA–protein complexes in Fractions 8–11 (Clusters 11–15) involved in transcription, RNA metabolism, and translation initiation shows the stability of such complexes, while the co-occurrence of a large number of RNAs illustrates that those complexes were still capturing the in vivo state of Synechocystis regulatory processes by binding RNAs. Furthermore, the fact that these fractions exhibited poor protein complexity but still harbored most of the well-studied factors involved in protein biosynthesis makes them interesting for the identification of novel candidates functionally linked to posttranscriptional regulation.

Another surprising observation was that the majority of detected transcripts were largely not associated with the previously mentioned complexes but mainly sedimented in Fractions 6–7 (Cluster 10), indicating the involvement of one or more RNA chaperones stably binding their target RNAs. The data permit the identification of candidate RBPs involved in the formation of major RNA–protein complexes in this group of bacteria. Although Synechocystis expresses multiple sRNAs of crucial relevance for oxygenic photosynthesis and other important physiological processes (Georg et al., 2014, 2017; Klähn et al., 2015; de Porcellinis et al., 2016), homologs of CsrA, ProQ or Hfq do not exist in cyanobacteria or do not bind RNA (Schuergers et al., 2014). A prediction from our Grad-seq clustering analysis is that cyanobacteria may not utilize a general RBP analogous to Hfq or ProQ, as characterized in enterobacteria, where these two RBPs are associated with more than 80% of all sRNAs (Holmqvist and Vogel, 2018). Instead, cyanobacterial sRNAs may frequently function independently of a common RNA chaperone, or they may rely more upon specialized RBPs. In this respect, our data echo the conclusions recently drawn for Gram-positive bacteria based on the Grad-seq analysis of the human pathogen S. pneumonia (Hör et al., 2020b). We filtered out a short list of previously uncharacterized protein candidates for the interaction with sRNAs. Among them, we discovered a cyanobacterial homolog of the KhpA/B complex (Zheng et al., 2017). The nearly congruent peaks of the homologs Slr0287 and Slr1472 in gradient Fractions 8–10 indicate that these proteins are part of a higher molecular weight complex, in fact much larger than the complex reported for S. pneumonia, which sedimented in Fractions 2 and 3.

Limitations of the study
Our dataset is of high complexity, as indicated by the large numbers of different proteins and transcripts detected in the three replicates for each fraction (Supplemental Data Set 1). This complexity does not simply reflect the true biological complexity but also in part results from technical variations that were taken into account by the clustering approach selected here. We observed that some subcomplexes likely became unstable during manipulation and dissociated from their respective main complex, such as the oxygen-evolving complex of PSI. Similarly, PSI appears to have been almost completely monomerized (Figure 5), CP43 (Psbc) peaked in Fractions 4 and 5, whereas the other PSI core complex subunits were found in Fractions 6 and 7 (Figure 5). Similarly, NDH-I subunits (Supplemental Figure 4) were distributed through Fractions 3–6 of the gradient with different abundances, also indicating partial complex disruption. Hence, the possibility that some complexes are fully stable only when membrane-bound cannot be excluded and cannot be avoided during the solubilization of the membranes. Moreover, the high number of detected proteins, together with a relatively low number of fractions limited the resolution, especially in the lower Fractions 1–6 with high protein complexity. By contrast, the resolution appeared better in the higher Fractions 7–18, which exhibited lower protein complexity and where most transcripts and large protein complexes were detected, making these fractions suitable for the detection of RNA–protein complexes. Together with the chosen clustering approach to divide the content of the dataset into groups of similarly sedimenting proteins and transcripts and extensive bioinformatic analysis, our combined approach has aided the identification of previously uncharacterized protein candidates likely involved in the metabolism of RNA and post-transcriptional regulation.

Materials and methods

Cell cultivation, cell lysis, gradient preparation, and fractionation
Triplicates of 400 mL of Synechocystis liquid cultures were cultivated in BG11 medium under constant shaking in standard growth conditions (30°C, 50 μE) until reaching exponential phase (OD750 = 0.8). The cultures were harvested by centrifugation at 4,000g for 15 min at RT. After centrifugation, the cell pellets were resuspended in ice-cold lysis buffer (20 mM Tris-HCl pH 7.5, 150 mM KCl, 1 mM MgCl2, 1 mM DTT, 0.2% [v/v] Triton X-100, Ribolock RNase Inhibitor, DNase I, cOmplete Protease Inhibitor cocktail) and
mechanically lysed with glass beads at 4°C (Precellys). From this point, the samples were kept at 4°C. Unbroken cells were removed by centrifugation for 5 min at 4,000g, and the supernatant was transferred to a fresh tube. Membrane proteins were solubilized by incubation with shaking for 60 min in the dark with DDM in a 3:1 ratio according to the protein content (1% [w/v], 1 g/100 mL initial culture). RNase inhibitor was added to minimize the risk of RNA degradation during the 1 h solubilization step; visual inspection by Fragment Analyzer (Agilent) of RNA preparations after fractionation showed no sign of substantial RNA degradation. Thylakoid membranes were removed by centrifugation at 30,000g for 30 min, and the supernatant was applied on top of linear sucrose gradients. We modified the original Grad-seq protocol by using DDM instead of Triton X100. DDM is a mild detergent that allows the inclusion of membrane proteins as it is an efficient membrane solubilizer that is frequently used for the investigation of photosynthetic membrane complexes, both in cyanobacteria and in plants (Eshaghi et al., 2000; Ma et al., 2006; Barera et al., 2012; Golub et al., 2020). Moreover, based on the results of pilot experiments, we generated density gradients with sucrose instead of glycerol. The 10%–40% sucrose gradients were made using a gradient mixer with two solutions of lysis buffer, supplemented with either 10% or 40% sucrose (w/v).

Ultracentrifugation of the lysates was performed in a swinging-bucket rotor (Beckman SW40 Ti) for 16 h at 285,000g. Fractions of equal volumes (600 μL) were collected manually, and the fractions were split into samples of equal volumes for MS (MS analysis, 100 μL) and library preparation for transcriptome deep sequencing (RNA-Seq, 360 μL). The protein complexity of individual fractions was determined in pre-experiments that revealed the low complexity of Fractions 14–17. Therefore, Fractions 14 and 15 (F14–15), as well as Fractions 16 and 17 (F16–17) were pooled prior to downstream sample preparations out of economic considerations, resulting in a total of 16 samples per replicate for transcriptome and proteome analyses. Fraction 18 represents the gradient pellet.

Spectroscopy
Absorption spectra were measured using a Specord® 250 Plus (Analytik Jena) spectrophotometer at room temperature and were normalized to equal fraction volumes.

RNA sample preparation, generation of cDNA libraries, and sequencing
The 360 μL fraction samples were spiked with a barcoded 197 bp-long spike-in RNA mix ranging from 1 to 100 fmol (sum 200 fmol) and subjected to proteinase K treatment (1% SDS, 10 mM Tris-HCl pH 7.5, proteinase K) for 30 min at 50°C, followed by RNA extraction with PGTX solution (Pinto et al., 2009). After the first phase separation step, the aqueous phase was transferred to a fresh microcentrifuge tube and mixed 1:1 (v/v) with 100% EtOH, and RNA was further extracted using the RNA clean and concentrator kit (Zymo Research) according to the manufacturer’s instructions. The spike-in RNA mix was created by in vitro transcription (Mega-shortscript, Thermo Fisher) of PCR-amplified fragments from the PhiX174 bacteriophage genome (Supplemental Table 2) and purified from a 10% urea-polyacrylamide gel with the RNA PAGE Recovery kit (Zymo Research). Determination of RNA concentration was performed using the Qubit RNA HS Assay kit (Thermo Fisher). For the control, equal volumes of purified RNA were separated on a 10% denaturing urea-polyacrylamide gel.

Library preparation and paired-end sequencing (2 × 75 bp read length) were performed by Vertis Biotechnologie AG. In short, RNA samples were fragmented using ultrasound (4 pulses of 30 s each at 4°C), followed by 3’ end adapter ligation and first-strand cDNA synthesis using M-MLV reverse transcriptase. The 5’ Illumina TrueSeq sequencing adapter was ligated to the 3’ end of the purified first-strand cDNAs, and the resulting cDNAs were PCR amplified (11 cycles) and purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics). For Illumina NextSeq sequencing, the samples were pooled, and the cDNA pool was eluted in the size range of 160–500 bp from a preparative agarose gel. The primers used for PCR amplification were designed for TruSeq sequencing according to the instructions of Illumina (Supplemental Table 2).

Read mapping and normalization
The analysis of the raw reads from the RNA-sequencing was performed using the galaxy web platform. The workflow can be accessed and reproduced at the following link: https://usegalaxy.eu/u/mr559/w/grad-seq-pipeline. In short, read quality was checked at the beginning and after each step with FastQC. Adapter trimming was performed using Cutadapt (Martin, 2011), and the trimmed reads were mapped to the Synechocystis genome using segemehl (Otto et al., 2014) with default parameters and were assigned to annotated regions using FeatureCounts. Prior to normalization, all features with a maximum read count ≤5 across all fractions were filtered prior to downstream analysis. Fraction-wise normalization was performed against the spike-in RNA mix, which was added in equal amounts to each fraction prior RNA sample preparation. The relative RNA amounts per transcript refer to the spike-in normalized reads across the gradient. Reproducibility between replicates was checked by comparing the distribution of the Spearman correlation coefficient (calculated with the cor() function from the R stats package) from the relative distribution of each RNA within a replicate with the same RNAs in the other replicates (Supplemental Figure 1A).

Sample preparation for MS-based proteome measurements
As an initial step, proteins were purified by acetone/methanol precipitation. Therefore, 100 μL of each of the initial 18 fractions was mixed with an ice-cold mixture of 800 μL acetone and 100 μL methanol. After incubation overnight at −20°C, precipitated proteins were collected by centrifugation for 16 h at 1,000g at 4°C. Subsequently, protein pellets
were washed twice with 1 mL cold 80% (v/v) acetone (aq.). During the washing procedure, precipitated sucrose was removed almost completely, even in fractions with high sucrose content.

Dried protein pellets were then redissolved in 50 μL denaturation buffer (6 M urea, 2 M thiourea in 100 mM Tris-HCl pH 7.5). Intradichl disulfide bonds were reduced with dithiothreitol and the resulting thiol groups were alkylated with iodoacetamide, as described previously (Spät et al., 2015). Subsequently, proteins were predigested with endoprotease Lys-C for 3 h, further diluted with 200 μL 20 mM ammonium bicarbonate buffer, pH 8.0, and digested with trypsin overnight, applying a protease:protein ratio of 1:100 for both enzymes. The resulting peptide solutions were acidified to pH 2.5 with 10% (v/v) trifluoroacetic acid (aq.) and purified following the stage tip protocol (Rappsilber et al., 2007).

For MS-based proteomics measurements, a constant sample volume was analyzed for each fraction of each replicate to account for the different protein amounts between the fractions. Consequently, the measured protein yields ranged between a maximum of 1,000 ng and <100 ng in fractions with the lowest protein concentrations (Supplemental Table 3). The relative protein amounts refer to the relative intensity-based absolute quantification (iBAQ) values from equal fraction volumes across the gradient (equal volume). As a control, equal volumes of the initial gradient fractions were separated in 15% SDS-polyacrylamide gels and visualized by Coomassie staining.

**MS measurements**

For nanoLC-MS/MS-based proteome measurements, purified samples were loaded onto an in-house C18 nanoHPLC column (20 cm PicoTip fused silica emitter with 75 μm inner diameter (New Objective, Woburn, USA), packed with 1.9 μm ReproSil-Pur C18-AQ resin (Dr. Maisch, Ammerbuch, Germany)) on an EASY-nLC 1200 system (Thermo Scientific, Bremen, Germany). Peptides were separated using a 49 min segmented linear gradient (Supplemental Table 4), and eluting peptides were directly ionized through an on-line coupled electrospray ionization (ESI) source and analyzed on a Q Exactive HF mass spectrometer (Thermo Scientific, Bremen, Germany), operated in the positive-ion mode and data-dependent acquisition. MS full scans were acquired with a mass-to-charge (m/z) range of 300–1,650 at a resolution of 60,000. The 12 most intense multiple charged ions were selected for fragmentation by higher-energy collisional dissociation (HCD) and recorded in MS/MS scans with a resolution of 30,000. Dynamic exclusion of sequenced precursor ions was set to 30 s. Wash runs (Supplemental Table 4) were performed after each of the three samples.

**Proteomics data processing**

Acquired raw MS data of all three replicates were processed using the MaxQuant software suite (version 1.5.2.8; Cox and Mann, 2008). Each file was defined as an individual sample. Data processing was performed with default settings, applying iBAQ, and the following search criteria: trypsin was defined as cleaving enzyme and a maximum of two missed cleavages was allowed. Carbamidomethylation of Cys residues was set as a fixed modification and Met oxidation and protein N-terminal acetylation were defined as variable modifications. MS/MS spectra were searched against a combined target-decoy database of _Synechocystis_, downloaded from Cyanobase (http://genome.microbedb.jp/cyanobase; version 06/2014), and the sequences of newly discovered small proteins (Mitschke et al., 2011; Kopf et al., 2014; Baumgartner et al., 2016), representing a total of 3,681 protein sequences plus 245 common contaminants. To enable the identification of low-abundance proteins that were not selected for MS/MS sequencing in any fraction, ancillary measurements were performed for the first replicate with a constant protein yield of 1,000 ng per fraction. These supporting measurements were included during data processing as a resource to promote overall protein identifications via the matching between runs option. Reported protein and peptide false discovery rates were each limited to 1%. The reproducibility between replicates was checked by comparing the distribution of the Spearman correlation coefficient (calculated with the cor() function from the R stats package) from the relative distribution of each protein within a replicate with the same protein in the other replicates (Supplemental Figure 1B). An additional feature to visualize constant protein amounts per fraction was implemented into the web application (constant protein amount), which proved to be useful for the detection of lowly abundant protein sub-complexes, which were within the gradient in close proximity to other highly abundant sub-complexes (e.g. PSI monomer and PSI trimer). The constant protein amount was estimated by determining the protein amount in equal fraction volumes, divided by the total protein amount per fraction. The first replicate, which was also measured with a constant protein amount of 1,000 ng per fraction, served as reference for the quality of this estimation.

**Hierarchical clustering and module detection**

Clustering and module (cluster) detection were performed using the dynamic tree cut algorithm implemented in the WGCNA R package (Langfelder and Horvath, 2008, 2012) from the standardized relative abundances (z-scores, calculated with the scale() function in R) of the normalized RNA-seq reads and iBAQ intensities. The unsupervised detection of 17 clusters was mainly dependent on the gradient sedimentation profiles of the proteins and transcripts from the input data and their variations between replicates. The parameters for the hierarchical clustering method and the dynamic tree cut algorithm were adjusted according to the standard WGCNA procedure (Zhang and Horvath, 2005). The Pearson correlation coefficients of the individual protein and transcript distribution to the assigned cluster means were calculated with the cor() function from the R stats package.
RNApred benchmarking for \textit{Synechocystis}\footnote{Schuergers et al., 2014}, thus confirming this approach. The observations of \textit{Synechocystis} failed to fulfill the set criteria by far, matching previous (Supplemental Figure 6C). Notably, the cyanobacterial Hfq as not binding RNA (Scheruergers et al., 2014), thus confirming this approach. The groups “Nucleotide binding” and “DNA binding” were distributed between the “RNA binding” and “No binding” groups with the majority of the “DNA binding” group having an SVM score below 1.07. Therefore, a conservative minimum SVM score of at least 0.5 should be sufficient to remove most proteins without nucleic acid binding properties. The median SVM score of all orthologs in the genomes give in Supplemental Table 1 was included to enhance the robustness of the approach, applying a minimum score of \( \geq 0.2 \), which is the default threshold for RBPs in the webserver. The boxplots for the SVM scores were generated using the \texttt{geom_boxplot()} function of the \texttt{ggplot2} R package. The upper and lower whiskers correspond to the 75th and 25th percentiles. The median is shown as a black line within the boxes. The outliers are shown as black dots outside of the whisker ranges. The \textit{Synechocystis} SVM scores are shown separately as red dots.

**Data availability**

The datasets produced in this study are available in the following databases:

- FastQ files from RNA sequencing in SRA (BioProject accession number PRJNA608723): \url{https://www.ncbi.nlm.nih.gov/sra/PRJNA608723}
- Galaxy workflow used for raw read processing: \url{https://usegalaxy.eu/u/mr559/w/grad-seq-pipeline}
- All code involving the bioinformatics workflow and the shiny app is available on Github (https://github.com/MatthiasRiediger/Analysis-of-a-photosynthetic-cyanobacterium-rich-in-internal-membrane-systems-via-Grad-seq.git)
- MS raw data deposited at the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Vizcaino et al, 2013) under the identifier: PXD020025.
- Data accessibility and visualization: \url{https://sunshine.biologie.uni-freiburg.de/GradSeqExplorer/}

**Supplemental data**

**Supplemental Figure 1** Correlations between replicates.

**Supplemental Figure 2** Distribution of 2,394 proteins and >4,000 different transcripts along a sucrose density gradient.

**Supplemental Figure 3** Sedimentation velocity as a function of molecular weight.

**Supplemental Figure 4** Heatmap representation of standardized relative protein abundances (z-score).

**Supplemental Figure 5** Cluster protein composition and selection process of putative cyanobacterial RNA binding proteins.

**Supplemental Figure 6** Comparison of determined SVM scores.

**Supplemental Figure 7** Conserved synteny in the cyanobacterial phylum of putative protein candidates involved in RNA metabolism.

**Supplemental Table 1** All selected genomes used for the phylogenetic occurrence analysis based on domclust of the Microbial Genome Database (MBGD).

**Supplemental Table 2** List of oligonucleotides.

**Supplemental Table 3** Total protein content per fraction measured by Bradford assay.

**Supplemental Table 4** nanoHPLC gradient used for MS-based proteome measurements.

**Supplemental Data Set 1** Overview of fraction complexity.

**Supplemental Data Set 2** Full Grad-seq dataset.

**Funding**

The study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—322977937/GRK2344, the priority program SPP2141 (grant no. HE 2544/14-1) to W.R.H. and through the “SCyCode” research group 397695561/FOR 2816 to W.R.H. and B.M.

**Conflict of interest statement.** The authors declare that they have no conflicts of interest.

**References**

Athukoralage JS, McQuarrie S, Grüssow S, Graham S, Gloster TM, White MF (2020) Tetramerisation of the CRISPR ring nuclease Crn3/Csx3 facilitates cysolic oligoadenylate cleavage. eLife 9: e57627

Baers LL, Breckels LM, Mills LA, Gatto L, Deery MJ, Stevens TJ, Howe CJ, Lilley KS, Lea-Smith DJ (2019) Proteome mapping of a cyanobacterium reveals distinct compartment organization and cell-dispersed metabolism. Plant Physiol 181: 1721–1738

Bandrya KJ, Luisi BF (2018) RNase E and the high-fidelity orchestration of RNA metabolism. Microbiol Spectr 6: RWR-0008-2017

Barera S, Pagliano C, Pape T, Saracco G, Barber J (2012) Characterization of Psll-LHCCI supercomplexes isolated from pea thylakoid membrane by one-step treatment with \( \alpha\)- and \( \beta\)-dodecyl-D-maltoside. Philos Trans R Soc Lond B Biol Sci 367: 3389–3399

Barrick JE, Sudarsan N, Weinberg Z, Ruzzo WL, Breaker RR (2005) 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. RNA N Y N 11: 774–784

Baumgartner D, Kopf M, Klähn S, Steglich C, Hess WR (2016) Small proteins in cyanobacteria provide a paradigm for the functional analysis of the bacterial microproteome. BMC Microbiol 16: 285
Bøggild A, Overgaard M, Valentin-Hansen P, Brodersen DE (2009) Cyanobacteria contain a structural homologue of the Hfq protein with altered RNA-binding properties. FEBS J 276: 3904–3915

Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, Dickman MJ, Makarova KS, Koonin EV, van der Oost J (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321: 960–964

Cantos R, Labella JI, Espinosa J, Contreras A (2019) The nitrogen regulator PipX acts in cis to prevent operon polarity. Environ Microbiol Rep 11: 495–507

Cech TR, Steitz JA (2014) The noncoding RNA revolution-trashing old rules to forge new ones. Cell 157: 77–94

Chitnis VP, Chitnis PR (1993) Psal subunit is required for the formation of photosystem I trimers in the cyanobacterium Synechocystis sp. PCC 6803. FEBS Lett 336: 330–334

Commichau FM, Stülke J (2008) Trigger enzymes: bifunctional proteins active in metabolism and in controlling gene expression. Mol Microbiol 67: 692–702

Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantitation. Nat Biotechnol 26: 1367–1372

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

Dühring U, Ossenbühl F, Wilde A (2007) Late assembly steps and dynamics of the cyanobacterial photosystem I. J Biol Chem 282: 10915–10921

Eshagi S, Turcsányi E, Vass I, Nugent J, Andersson B, Barber J (2000) Functional characterization of the PS II-LHC II supercomplex isolated by a direct method from spinach thylakoid membranes. Photosynth Res 64: 179–187

Forcada-Nadal A, Llácer JL, Contreras A, Marco-Marin C, Rubio V (2018) The Pil-NAGK-PipX-NtcA regulatory axis of cyanobacteria: A tale of changing partners, allosteric effectors and non-covalent interactions. Front Mol Biosci 5: 91

Fujimori T, Hihara Y, Sonoi K (2005) Psak2 subunit in photosystem I is involved in state transition under high light condition in the cyanobacterium Synechocystis sp. PCC 6803. J Biol Chem 280: 22191–22197

Gao F, Zhao J, Chen L, Battchikova N, Ran Z, Aro E-M, Ogawa T, Ma E (2016) The NDH-1L-PS1 supercomplex is important for efficient cyclic electron transport in cyanobacteria. Plant Physiol 172: 1451–1464

Garcia-Dominguez M, Reyes JC, Florencio FJ (1999) Glutamine synthetase inactivation by protein-protein interaction. Proc Natl Acad Sci USA 96: 7161–7166

Garcia-Dominguez M, Reyes JC, Barquist L, Vogel J (2020) Global discovery of bacterial RNA-binding proteins by RNAse-sensitive gradient profiles reports a new FinO domain protein. RNA 26: 1448–1463

Golub M, Hussein R, Ibrahim M, Hecht M, Wieland DCF, Martel D, Machado B, Zouni A, Pieper J (2020) Solution structure of the detergent-photosystem II core complex investigated by small angle scattering techniques. J Phys Chem B 124: 8583–8592

Gunnels L, Kulkila K, Kurkela J, Wada H, Tyystjärvi E, Tyystjärvi T (2014) The omega subunit of the RNA polymerase core directs transcription efficiency in cyanobacteria. Nucleic Acids Res 42: 4606–4614

Hagemann M, Hess WR (2018) Systems and synthetic biology for the biotechnological application of cyanobacteria. Curr Opin Biotechnol 49: 94–99

Hale CR, Zhao P, Olson S, Buffett MO, Graceve BR, Wells L, Terns RM, Terns MP (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 139: 945–956

Han W, Stella S, Zhang Y, Guo T, Sulek K, Peng-Lundgren L, Montoya G, She Q (2018) A Type III-B Cmr effector complex catalyzes the synthesis of cyclic oligoadenylate second messengers by cooperative substrate binding. Nucleic Acids Res 46: 10319–10330

He Z, Mi H (2016) Functional characterization of the subunits N, H, J, and O of the NAD(P)H dehydrogenase complexes in Synechocystis sp. strain PCC 6803. Plant Physiol 171: 1320–1332

Heilmann B, Jakkola K, Georg J, Tyystjärvi T, Hess WR, Axmann IM, Dienst D (2017) 6S RNA plays a role in recovery from nitrogen depletion in Synechocystis sp. PCC 6803. BMC Microbiol 17: 229

Holmqvist E, Vogel J (2018) RNA-binding proteins in bacteria. Nat Rev Microbiol 16: 601–615

Hör J, Di Giorgio S, Cerovac M, Venturini E, Förster KU, Vogel J (2020a) Grad-seq shines light on unrecognized RNA and protein complexes in the model bacterium Escherichia coli. Nucleic Acids Res 48: 9931–9939

Hör J, Garriss G, Di Giorgio S, Hack L-M, Vanselow JT, Förster KU, Schlosser A, Henriques-Normark B, Vogel J (2020b) Grad-seq in a Gram-positive bacterium reveals exonolementic sRNA activation in competence control. EMBO J 39: e103852

Hör J, Gorski SA, Vogel J (2018) Bacterial RNA biology on a genome scale. Mol Cell 70: 785–799

Hör J, Matera G, Vogel J, Gottesman S, Storz G (2020c) Trans-acting small RNAs and their effects on gene expression in Escherichia coli and Salmonella enterica. EcoSal Plus 9: 10.1128

Hrle A, Maier L-K, Sharma K, Ebert J, Basquin C, Urlaub H, Matchelder A, Conti E (2014) Structural analyses of the CRISPR protein Csc2 reveal the RNA-binding interface of the type I-D Cas7 family. RNA Biol 11: 1072–1082

Ishikawa Y, Schröder WP, Funk C (2005) Functional analysis of the PspB-like protein (sll1418) in Synechocystis sp. PCC 6803. Photosynth Res 84: 257–262

Iyer LM, Koonin EV, Aravind L (2004) Evolution of bacterial RNA polymerase: implications for large-scale bacterial phylogeny, domain accretion, and horizontal gene transfer. Gene 335: 73–88

Kanehisa M (2019) Toward understanding the origin and evolution of cellular organisms. Protein Sci 28: 1947–1951

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28: 27–30

Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M, Tomita M, Kurata T, Kishimoto K, Kawashima S, Ogata K (2019) New approach for understanding genome variations in KEGG. Nucleic Acids Res 47: D561–D571

Kazlauskiene M, Kostiuk G, Venclovas Č, Tamulaitis G, Siksnys V (2017) A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. Science 357: 605–609

Kiefer SN, Almendros C, Behler J, McKenzie RE, Nóbrega FL, Haagsma AC, Vink JNA, Hess WR, Brouns SJ (2018) Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation. Cell Rep 22: 3377–3383

Kinney JN, Axen SD, Kerfeld CA (2011) Comparative analysis of carboxysome shell proteins. Photosynth Res 109: 21–32

Klähn S, Bolay F, Wright PR, Attihoo RM, Brewer KL, Hagemann M, Breker RR, Hess WR (2018) A glutamine riboswitch is a key mechanism...
element for the regulation of glutamine synthetase in cyanobacteria. Nucleic Acids Res 46: 10082–10094

Klähn S, Schaal C, Georg J, Baumgartner D, Knippen G, Hagemann M, Muro-Pastor AM, Hess WR (2015) The sRNA NsIR4 is involved in nitrogen assimilation control in cyanobacteria by targeting glutamine synthetase inactivating factor IF7. Proc Natl Acad Sci USA 112: E6243–E6252

Kondo K, Geng XX, Katayama M, Ikeuchi M (2005) Distinct roles of CpcG1 and CpcG2 in phycobilisome assembly in the cyanobacterium Synechocystis sp. PCC 6803. Photosyn Res 84: 269–273

Kondo K, Mullineaux CW, Ikeuchi M (2009) Distinct roles of CpcG1-phycoylsomer and CpcG2-phycoylsomer in state transitions in a cyanobacterium Synechocystis sp. PCC 6803. Photosyn Res 99: 217–225

Kondo K, Ochiai Y, Katayama M, Ikeuchi M (2007) The membrane-associated CpcG2-phycoylsomer in Synechocystis: a new photosystem I antenna. Plant Physiol 144: 1200–1210

Kopf M, Hess WR (2015) Regulatory RNAs in photosynthetic cyanobacteria. FEMS Microbiol Rev 39: 301–315

Kopf M, Klähn S, Scholz I, Matthiessen JKF, Hess WR, Voß B (2014) Comparative analysis of the primary transcriptome of Synechocystis sp. PCC 6803. DNA Res 21: 527–539

Kopfmann S, Roesch SK, Hess WR (2016) Type II toxin-antitoxin systems in the unicellular cyanobacterium Synechocystis sp. PCC 6803. Toxins 8: 228:1–228:23

Kumar M, Gromiha MM, Raghava GPS (2011) SVM based prediction of RNA-binding proteins using binding residues and evolutionary information. J Mol Recognit 24: 303–313

Labella JL, Cantos R, Salinas P, Espinosa J, Contreras A (2020a) Distinctive features of PipX, a unique signaling protein of Cyanobacteria. Life 10: 79

Labella JL, Llop A, Contreras A (2020b) The default cyanobacterial linked genome: an interactive platform based on cyanobacterial linkage networks to assist functional genomics. FEBS Lett 594: 1661–1674

Labella JL, Obrebska A, Espinosa J, Salinas P, Forcada-Nadal A, Tremino L, Rubio V, Contreras A (2016) Expanding the cyanobacterial nitrogen regulatory network: the GntR-like regulator PlmA interacts with the PipI-PipX complex. Front Microbiol 7: 1677

Langfelder P, Horvath S (2012) Fast R functions for robust correlations and hierarchical clustering. J Stat Softw 46: i11

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9: 559

Liu H, Weiz DA, Zhang MM, Cheng M, Zhang B, Zhang H, Gerstenecker GS, Pakrasi HB, Gross ML, Blankenship RE (2005) Identification of PamA (Synechocystis PCC 6803). Wei Sheng Wu Xue Bao 53: 15397–15402

Liu MY, Gui G, Wei B, Preston JF, Oakford L, Yuksel U, Giedroc P, Zhang MM, Cheng M, Zhang B, Zhang H, Gerstenecker GS, Pakrasi HB, Gross ML, Blankenship RE (2006) The rbcS mRNAs require the 38-kDa ribosomal protein S1 but not the Shine-Dalgarno sequence: development of a cyanobacterial in vitro translation system. J Biol Chem 281: 38314–38321

Nakagawa S, Niimura Y, Miura K, Gojobori T (2010) Dynamic evolution of translation initiation mechanisms in prokaryotes. Proc Natl Acad Sci USA 107: 6382–6387

Nickelsen J, Rengstl B (2013) Photosystem II assembly: from cyanobacteria to plants. Annu Rev Plant Biol 64: 609–635

Niewoehner O, Garcia-Doval C, Rostel JT, Berk C, Schwede F, Bigler L, Hall J, Marraffini LA, Jinek M (2017) Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. Nature 548: 543–548

Ning D, Zhao W, Qian Y (2013) [A hypothetical gene pair, ssl2138sp. strain PCC 6803]. Wei Sheng Wu Xue Bao 53: 1043–1049

Ossenbühl F, Inaba-Sulpice M, Meurer J, Soll J, Eichacker LA (2003) The rise of regulatory RNA. Nat Rev Mol Biosci 3: 1432–1436

Osmarova KS, Wolf YI, Irazo J, Shmakov SA, Alkhnbashi OS, Brouns SJ, Charpentier E, Cheng D, Haft DH, Horvath P, et al. (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol 18: 67–83

Makarova KS, Anantharaman V, Grishin NV, Koonin EV, Aravind L (2014) CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. Front Genet 5: 102

Marbouty M, Saguez C, Cassier-Chauvat C, Chauvat F (2009) Characterization of the FtsZ-interacting septal proteins SepF and Ftn6 in the spherical-celled cyanobacterium Synechocystis strain PCC 6803. J Bacteriol 191: 6178–6185

Martin M (2011) Cudatope removes adapter sequences from high-throughput sequencing reads. EMBO J 17: 10

Melamed S, Adams PP, Zhang A, Zhang H, Storz G (2020) RNA-RNA interacotmes of ProQ and Hfq reveal overlapping and competing roles. Mol Cell 77: 411–425.e7

Mitschke J, Grorg J, Scholz I, Matthiessen JKF, Dienst D, Bantschell J, Voß B, Steglich C, Wilde A, Vogel J, Hess WR (2011) An experimentally anchored map of transcriptional start sites in the model cyanobacterium Synechocystis sp. PCC6803. Proc Natl Acad Sci USA 108: 2124–2129

Molina R, Stella S, Feng M, Soños N, Jauniskis V, Pozdnyakova I, López-Méndez B, She Q, Montoya G (2019) Structure of Csx1-ccOA4 complex reveals the basis of RNA decay in Type III-B CRISPR-Cas. Nat Commun 10: 4302

Moll I, Leitsch D, Steinhauser T, Bläsi U (2003) RNA chaperone activity of the Sm-like Hfq protein. EMBO Rep 4: 284–289

Moller T, Franch T, Hejrup P, Keene DR, Bächinger HP, Brennan RG, Valentin-Hansen P (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. Mol Cell 9: 23–30

Morris KV, Mattick JS (2014) The rise of regulatory RNA. Nat Rev Genet 15: 423–437

Mutsuda S, Sugiuira M (2006) Translation initiation in cyanobacteria requires the 30-kDa ribosomal protein S1 but not the Shine-Dalgarno sequence: development of a cyanobacterial in vitro translation system. J Biol Chem 281: 38314–38321

Nakagawa S, Niimura Y, Miura K, Gojobori T (2010) Dynamic evolution of translation initiation mechanisms in prokaryotes. Proc Natl Acad Sci USA 107: 6382–6387

Nickelsen J, Rengstl B (2013) Photosystem II assembly: from cyanobacteria to plants. Annu Rev Plant Biol 64: 609–635

Niewoehner O, Garcia-Doval C, Rostel JT, Berk C, Schwede F, Bigler L, Hall J, Marraffini LA, Jinek M (2017) Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. Nature 548: 543–548

Ning D, Zhao W, Qian Y (2013) [A hypothetical gene pair, ssl11092, constitutes a functional TA system on the chromosome of Synechocystis PCC 6803]. Wei Sheng Wu Xue Bao 53: 1043–1049

Osanai T, Sato S, Tabata S, Tanaka K (2005) Identification of PamA as a PII-binding membrane protein important in nitrogen-related and sugar-catabolic gene expression in Synechocystis sp. PCC 6803. J Biol Chem 280: 34684–34690

Ossenbühl F, Inaba-Sulpice M, Meurer J, Soll J, Eichacker LA (2006) The Synechocystis sp PCC 6803 Oxa1 homolog is essential for membrane integration of reaction center precursor protein pD1. Plant Cell 18: 2236–2246

Otto C, Stadler PF, Hoffmann S (2014) Lacking alignments? The next-generation sequencing mapper segemehl revisited. Bioinforma Oxf Engl 30: 1837–1843

Pinto F, Thapper A, Sontheim W, Lindblad P (2009) Analysis of current and alternative phenol based RNA extraction methodsology for cyanobacteria. BMC Mol Biol 10: 79

de Porcellinis AJ, Klähn S, Rosgaard L, Kirsch R, Gutukunst K, Georg J, Hess WR, Sakurai Y (2016) The non-coding RNA Ncr0700/PmgR1 is required for photomixotrophic growth and the regulation of glycan accumulation in the cyanobacterium Synechocystis sp. PCC 6803. Plant Cell Physiol 57: 2091–2103

Quendera AP, Seixas AF, dos Santos RF, Santos I, Silva JPN, Araújo CM, Andrade JM (2020) RNA-binding proteins driving the regulatory activity of small non-coding RNAs in bacteria. Front Mol Biosci 7:78
Zhang J, Gao F, Zhao J, Ogawa T, Wang Q, Ma W (2014a) NdhP is an exclusive subunit of large complex of NADPH dehydrogenase essential to stabilize the complex in *Synechocystis* sp. strain PCC 6803. J Biol Chem 289: 18770–18781

Zhang J-Y, Deng X-M, Li F-P, Wang L, Huang Q-Y, Zhang C-C, Chen W-L (2014b) RNase E forms a complex with polynucleotide phosphorylase in cyanobacteria via a cyanobacterial-specific nonapeptide in the noncatalytic region. RNA 20: 568–579

Zheng JJ, Perez AJ, Tsui H-CT, Massidda O, Winkler ME (2017) Absence of the KhpA and KhpB (JAG/EloR) RNA-binding proteins suppresses the requirement for PBP2b by overproduction of FtsA in *Streptococcus pneumoniae* D39. Mol Microbiol 106: 793–814