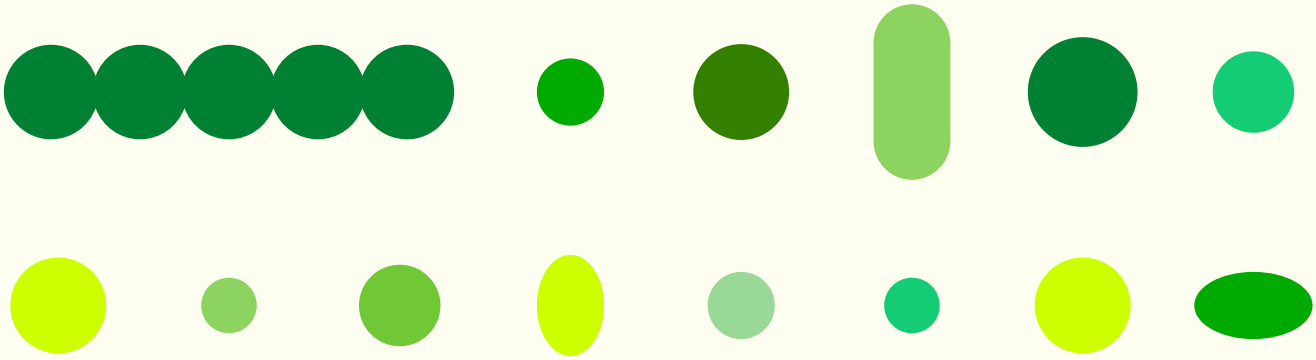
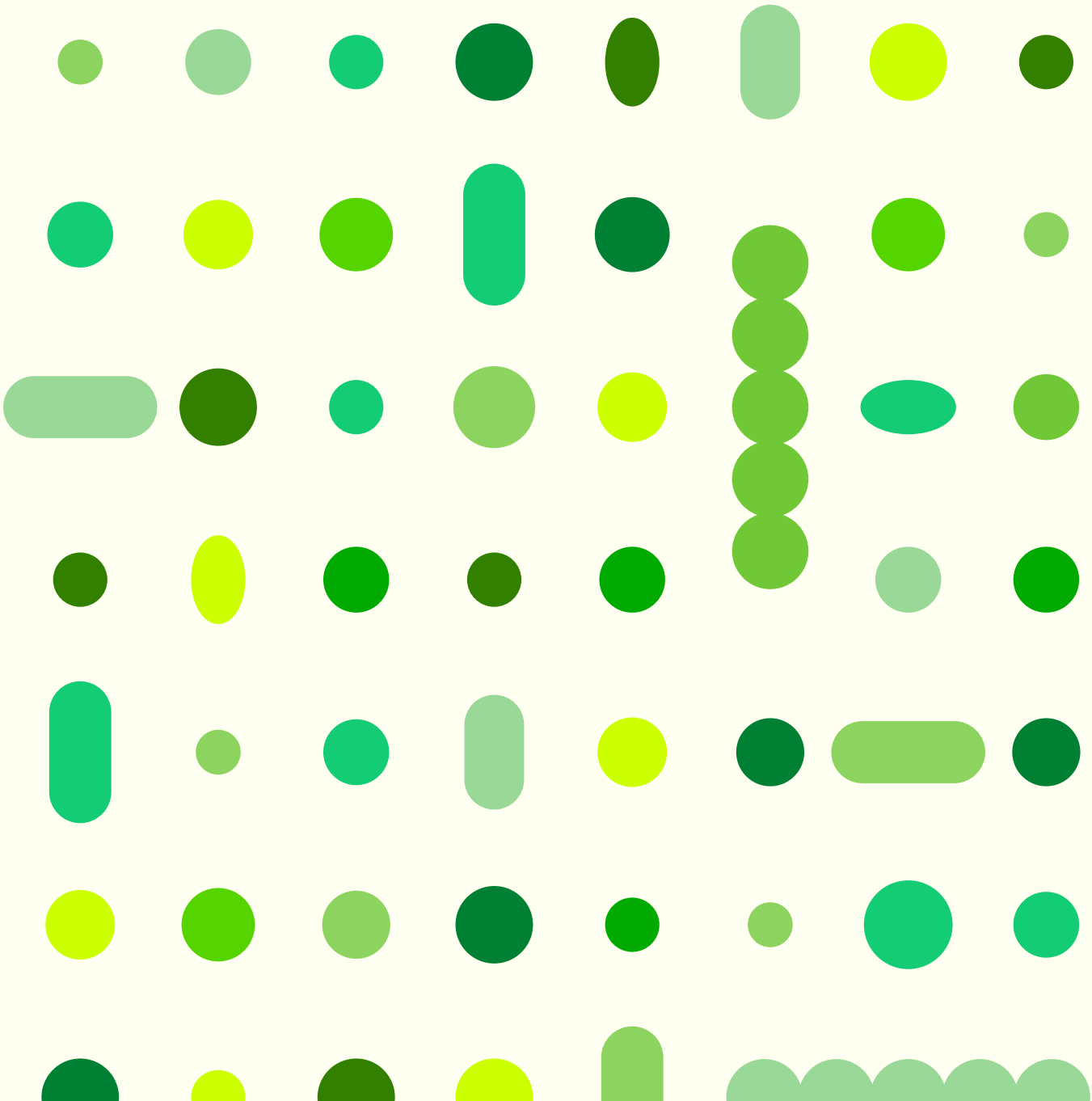


Nicolas M. Schmelling

A Matter of Time

Diversity Analyses of The Cyanobacterial Circadian Clock



A Matter of Time

Diversity Analyses of The Cyanobacterial Circadian Clock

Inaugural Dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University Düsseldorf

presented by

Nicolas M. Schmelling
born in Düsseldorf

Düsseldorf, December 2018

at the Institute for Synthetic Microbiology
of the Heinrich Heine University Düsseldorf

printed with the permission of
the Faculty of Mathematics and Natural Sciences
of the Heinrich Heine University Düsseldorf

Supervisor: Jun.-Prof. Dr. Ilka M. Axmann
Institute for Synthetic Microbiology
Heinrich Heine UniversityDüsseldorf

Co-Supervisor: Prof. Dr. Matias Zurbriggen
Institute for Synthetic Biology
Heinrich Heine University Düsseldorf

Date of oral examination

KBK - Joel Jordan Schmelling.

©2018 – Nicolas M. Schmelling
This work is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this thesis, unless otherwise stated. The work by Snijder and colleagues (Chapter 4) is reproduced with the permission of the American Association for the Advancement of Science and does not apply to the above stated copyright licenses.

Zusammenfassung

Zirkadiane Uhren sind ein universell auftretendes System zur zeitlichen Regulierung großer Teile der zellulären Genexpression in einem 24 Stunden Rhythmus. Dieser Regulationsmechanismus bietet den Organismen unter zyklischen Bedingungen einen Fitnessvorteil gegenüber ihren Konkurrenten, da sie die Fähigkeit verleihen, Veränderungen zu antizipieren und sich darauf vorzubereiten. Cyanobakterien sind photosynthetische Prokaryoten, die ihre Genexpression auf zirkadiane Weise koordinieren, um sich an den Tag/Nacht Zyklus anzupassen. Ihre zirkadiane Uhr besteht dabei aus nur drei Proteinen, KaiA, KaiB und KaiC, die in ein Netzwerk von *Input-* und *Output*-Faktoren eingebettet sind. Des weiteren sind Cyanobakterien auch für ihre hohe Diversität bekannt. Daher wurde in dieser Arbeit die Verteilung und Konservierung der bekannten zirkadianen Proteine aus *Synechococcus elongatus* PCC 7942 analysiert mit dem Ergebnis, dass drei verschiedene Zeitsysteme unter Cyanobakterien identifiziert werden konnten: (i) ein *Synechococcus* ähnliches System, welches ein komplettes Set der Faktoren in einfacher Form verfügt, (ii) ein *Prochlorococcus* ähnliches reduziertes System und ein (iii) *Synechocystis* ähnliches System, was über mehrere Kopien von Faktoren, inklusive KaiB und KaiC, verfügt. Darüber hinaus zeigen Untersuchungen, die ein gemeinsames Auftreten von Proteinen analysieren, eine Kerngruppe von Proteinen für dieses Zeitsystem, inklusive KaiB, KaiC, LdpA, SasA, RpaA und RpaB, sowie eine Erweiterungsgruppe von Proteinen, inklusive KaiA, CikA und LabA, die für die zirkadiane Regulation wichtig sein könnten. Die Ergebnisse aus dem umfassend vergleichenden Genomik-Ansatz unterstützen frühere Studien und erweitern das Bild zirkadianer Regulation innerhalb der Klasse der Cyanobakterien. Zudem zeigen die Ergebnisse die weit verbreitete Verteilung einiger Faktoren, unter anderem in Bakterien und Archaeen. Im letzten Teil der Arbeit werden die Fortschritte des molekularen Verständnisses über die zirkadiane Uhr aufgearbeitet, die erst durch mathematische Modelle erzielt werden konnten. Die *in vitro* Rekonstruktion dieser zirkadianen Uhr legte dabei den Grundstein für mathematische Modellierungsansätze. Seitdem wurden molekulare Details der Funktionalität über Zeitsysteme im Allgemeinen und die cyanobakterielle zirkadiane Uhr im Besonderen aufgeklärt, inklusive verschiedener Strategien zu *Entrainment* und Robustheit. Die Begutachtung beleuchtet jedoch auch einige Aspekte der zirkadianen Uhren, die bisher noch nicht ausreichend analysiert wurden. Die Ergebnisse dieser Arbeit geben einen umfassenden Überblick über die Verteilung der cyanobakteriellen zirkadianen Uhr, indem sie eine Reihe von Proteinfaktoren identifizieren, die das Zentrum für die Uhr in Cyanobakterien bilden könnten. Zusammen mit den überprüften Fortschritten über das Verständnis dieser zirkadianen Uhr durch mathematische Modellierung bildet es die Grundlage für weitere Analysen für eine reduzierte Uhr innerhalb oder außerhalb von Cyanobakterien und ermöglicht die Analyse einer ursprünglichen Uhr.

Abstract

Circadian clocks are a universally found timing system to regulate large parts of the cell's gene expression in a 24-hour rhythm. This regulatory mechanism provides a fitness advantage to the organisms over its competitor under cyclic conditions due to its ability to anticipate and prepare for recurring changes. Cyanobacteria are photosynthetic prokaryotes that coordinate their gene expression in a circadian fashion to adapt to the Day/Night cycle. Their circadian clock consists of only three proteins at its core, KaiA, KaiB, and KaiC, which are embedded into a network of input and output factors. Furthermore, Cyanobacteria are known for their high diversity. Thus in this thesis, the distribution and conservation of the known circadian clock factors from *Synechococcus elongatus* PCC 7942 was analyzed identifying three distinct timing systems among Cyanobacteria: (i) a *Synechococcus*-like system with a complete set of factors as a single copy, (ii) a *Prochlorococcus*-like reduced set, and (iii) a *Synechocystis*-like system that harbors multiple copies of some of the factors including the core factors KaiB and KaiC. In addition, co-occurrence analyses revealed a core set of protein factors, including the core factors KaiB, KaiC, the input factor LdpA, and the output factors SasA, RpaA, and RpaB, for this timings system as well as an extending set of proteins, including the core factor KaiA, the input and output factor CikA, and the output factor LabA, which might be important for circadian regulation. These results support previous findings and extend the picture of circadian regulation systems among Cyanobacteria by using a comprehensive comparative genomics approach. Further, the results highlight the widespread distribution of some of the factors among other Bacteria as well as Archaea. In the last part of the thesis, the advances in the molecular understanding of this timing system made by mathematical models, which were facilitated by the *in vitro* reconstruction of the core of this circadian clock, are reviewed. Since then the molecular details about the functionality of timing systems in general and the cyanobacterial circadian clock, in particular, have been unraveled, including various principles, like entrainment and robustness strategies. However, reviewing those achievements also highlights some of the aspects of circadian clocks that have not yet been sufficiently analyzed. The results of this thesis give a comprehensive overview of the distribution of the cyanobacterial circadian clock factors by identifying a set of protein factors that might build the core for timing systems in cyanobacteria. Together with the reviewed advancements into the understanding of this circadian clock by mathematical modeling, it builds the basis for further analyses of smaller clock system among or outside of Cyanobacteria and enables the analysis of an ancient clock system.

Contents

Progress is made by trial and failure; the failures are generally a hundred times more numerous than the successes; yet they are usually left unchronicled.

Sir William Ramsay

1

Introduction and Scientific Context

1.1 Timing of cellular processes

Regulating the expression of genes or the activity of gene products is as old as free-living life itself. The coordination of the various processes that run within a cell is of central importance for an organism to most efficiently use its precious resources in order to occupy its niche in the best potential way. Thus it is not surprising that several transcriptional regulators can already be found in the predicted genome of the last universal common ancestor[1]. However, regulating gene expression can manifest itself in a multitude of ways with a large range of complexity. Constitutive genes are transcribed constantly within a cell, e.g. genes that are needed for maintenance of the cell. However, some genes are expressed either facultatively when needed or induced by an external or internal stimulus. Regarding genes or gene products that are regulated in a temporal fashion, you can distinguish between two large classes: (i) those genes or gene products that are regulated as a reaction of an internal or external stimulus, e.g. changes in the nutrient availability, (ii) genes and gene products that are either activated or repressed before the change of the stimulus occurs, anticipating the recurring changes in the presence of a certain stimulus, e.g. light intensity.

Both classes of regulation can produce stable oscillations of gene expression. By studying such systems, conserved regulatory motifs have been identified, e.g. negative feedback control of processes[2]. Negative feedback control takes place when the output of a system reduces or negatively affects the input of the system. Those negative feedback mechanisms tend to support stability of the output signal and reduce the effect of other perturbations on the output. The gained insights into gene regulation and identified motifs even led to the construction of synthetic regulatory circuits, which create stable oscillation of gene expression[3,4].

1.1.1 Circadian clocks

One regulatory mechanism that creates stable oscillations of gene expression by anticipating recurring changes of a stimulus (Zeitgeber) is a circadian clock (*circa* = about, *diem* = a day). They are defined by three criteria: (i) persistence of oscillations without an exogenous cycle, e.g. in constant light or darkness, (ii) temperature compensation, and (iii) the entrainment of the endogenous oscillator(s) to the exogenous cycle[5]. The ability of entrainment allows the Zeitgeber to synchronize the oscillation with the exogenous rhythm. Circadian clocks thus have in the presence of the Zeitgeber a period length of around 24 hours, the length of a day on Earth[5].

Circadian clocks are an adaptation to recurring cyclic environmental conditions. Those changing conditions are usually challenging to organisms, however, they also open a chance for adaptation and hence a fitness advantage. In fact, in Cyanobacteria, a fitness advantage of a circadian clock in cyclic environments could be shown by competition experiments between cyanobacteria harboring an intact clock system and mutants with clock deletions. In constant environments, no phenotype was visible between wild-type and mutant cells highlighting that the mutation had no effect on the viability of the cells. However, under light-dark cycles, the wild-type cells with intact circadian clocks outcompete and overgrow the mutant cells showing the impact of anticipating and preparing for the recurring change[6,7].

It seems that a selective pressure by cyclic environments favored the evolution and adaptation of circadian clocks in a multitude of organisms ranging from Cyanobacteria (model organism: *Synechococcus elongatus* PCC 7942[8]) to fungi (model organism: *Neurospora crassa*[9]), plants (model organism: *Arabidopsis thaliana*[10]), and animals (model organism: *Drosophila melanogaster*[11], hamster[12], mice[13], and humans[14]). Interestingly, the differences in the amino acid sequence of protein factors in the respective circadian system suggest a convergent evolution of this mechanism[15,16].

At first, circadian clocks were only identified in eukaryotes and it was assumed that simple prokaryotic organisms would not be able to maintain oscillations with period lengths larger than their division time[5]. However, those misconceptions were disabused after the discovery of diazotrophic cyanobacteria that pho-

tosynthesize during the day and fix nitrogen at night[17] and later the finding of circadian gene expression in *Synechococcus elongatus* PCC 7942[8].

1.2 Cyanobacteria: The first known prokaryotic oxygenic phototrophs

Cyanobacteria represent one of the oldest phyla of Bacteria on Earth with fossil records dating back over two billion years[18]. They are considered to be mainly responsible for the build-up of an oxygenic atmosphere on Earth by performing oxygenic photosynthesis and producing oxygen as a by-product of fixing carbon dioxide[18]. The ability to perform oxygenic photosynthesis is unique among prokaryotes and otherwise only found in eukaryotes such as algae and plants. This is one of the multiple reasons why Cyanobacteria are regarded as the evolutionary ancestors of today's chloroplasts in algae and plant cells[19,20]. It is believed that a unique endosymbiotic event of an early eukaryotic cell with an ancient cyanobacterium led to the evolution of the chloroplast, followed by a substantial rearrangement of the cyanobacterial genome including a large number of gene transfers to the host nucleus[19,20].

Isolated cyanobacterial strains show a high diversity with respect to their genomes, morphology, and physiology[21,22]. Their genomes, as it is commonly found in bacteria, are usually composed of a circular chromosome, which is often accompanied by additional plasmids[21]. However, large differences are reported for ploidy and genome size in Cyanobacteria where genomes can vary between 1.49 Mb and approximately 12 Mb[21,22]. While several marine picocyanobacteria are mono- to diploid, there are other Cyanobacteria that are oligo- to polyploid like *Synechococcus elongatus* PCC 7942 or *Synechocystis* sp. PCC 6803, whose chromosome copy number ranges from three to multiple hundred copies per cell[21]. Regarding the morphology, Cyanobacteria are unicellular as well as multicellular with sometimes additionally differentiated cell types like heterocysts for nitrogen fixation. Based on these morphological and physiological differences Cyanobacteria are grouped into five taxonomic subsections[22]. Section I and II comprise unicellular Cyanobacteria, like *Synechococcus elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803 or the nitrogen fixing Cyanobacteria *Cyanothece*. All of the three belong to section I Cyanobacteria. Section III, IV, and V represent filamentous Cyanobacteria. While section III Cyanobacteria cannot form nitrogen fixing heterocysts, are section IV and V Cyanobacteria are able to differentiate cells into heterocysts[23].

Due to its potential of creating a variety of high-value organic compounds from light, water, and carbon dioxide, Cyanobacteria are increasingly utilized for biotechnological production of ethanol, isobutanol, sugars[24], isoprenes, or terpenes[25–27].

1.3 Circadian clocks in cyanobacteria

Cyanobacteria evolved a sophisticated and complex mechanism to capture and harness light energy as their main energy source for their metabolism. However, since the natural availability of sunlight is changing during the course of 24 hours, it is important for Cyanobacteria to regulate their gene expression in such a way that it allows the organism to harvest as much sunlight as possible in the most efficient way. One adaptation that Cyanobacteria evolved in order to cope with this problem is a circadian clock. Correctly anticipating the change from dark to light and vice versa is thus of central importance. However, in Cyanobacteria, in contrast to eukaryotic circadian systems, photoreceptors that directly transfer information about the light availability to the core clock have not been identified yet. Instead, Cyanobacteria sense the light availability indirectly through the redox and energy state of the cell[28]. The ratio of adenosine triphosphate (ATP) to adenosine diphosphate (ADP) as well as the redox state of the plastoquinone pool are two metabolically active compounds that play a critical role in the entrainment of the circadian clock in *Synechococcus elongatus* PCC 7942[29].

In the cyanobacterial circadian model organism *Synechococcus elongatus* PCC 7942 the circadian clock consists at its core of the three proteins KaiA, KaiB, and KaiC. These three proteins combined function as a post-translational oscillator (PTO), which is synchronized by environmental conditions as described before either through the energy or redox state of the cell directly or by input factors that themselves sense these states and transfer this information to the core clock.

In turn, the state of the circadian clock is transferred to different output factors depending on the time of day, which control gene expression and other cellular processes[30] (Fig. 1.1).

1.3.1 KaiC at the center of a three-part circadian clock

KaiC is at the core of the circadian clock in *Synechococcus elongatus* PCC 7942. Each KaiC monomer consists of two large domains, referred to as the CI and CII domain[33]. KaiC hexamers, which assemble ATP dependent and resemble the active form of KaiC, are able to bind up to 12 ATP molecules and are capable of autophosphorylation/-dephosphorylation[34–36]. The state of the circadian clock is determined by the phosphorylation state of the KaiC hexamer, which undergoes in interaction with KaiA and KaiB, a complete phosphorylation and dephosphorylation cycle within 24 hours. Primarily, the binding of a KaiA dimer to the A-loop of KaiC, a C-terminal extension of the CII domain, during the day stimulates the autophosphorylation of KaiC[37,38]. KaiB antagonizes the activating effect of KaiA and initiates the dephosphorylation of KaiC at the end of the subjective day[39]. KaiB binds to the phosphorylated KaiC hexamers and, in the following, sequesters and deactivates KaiA[40].

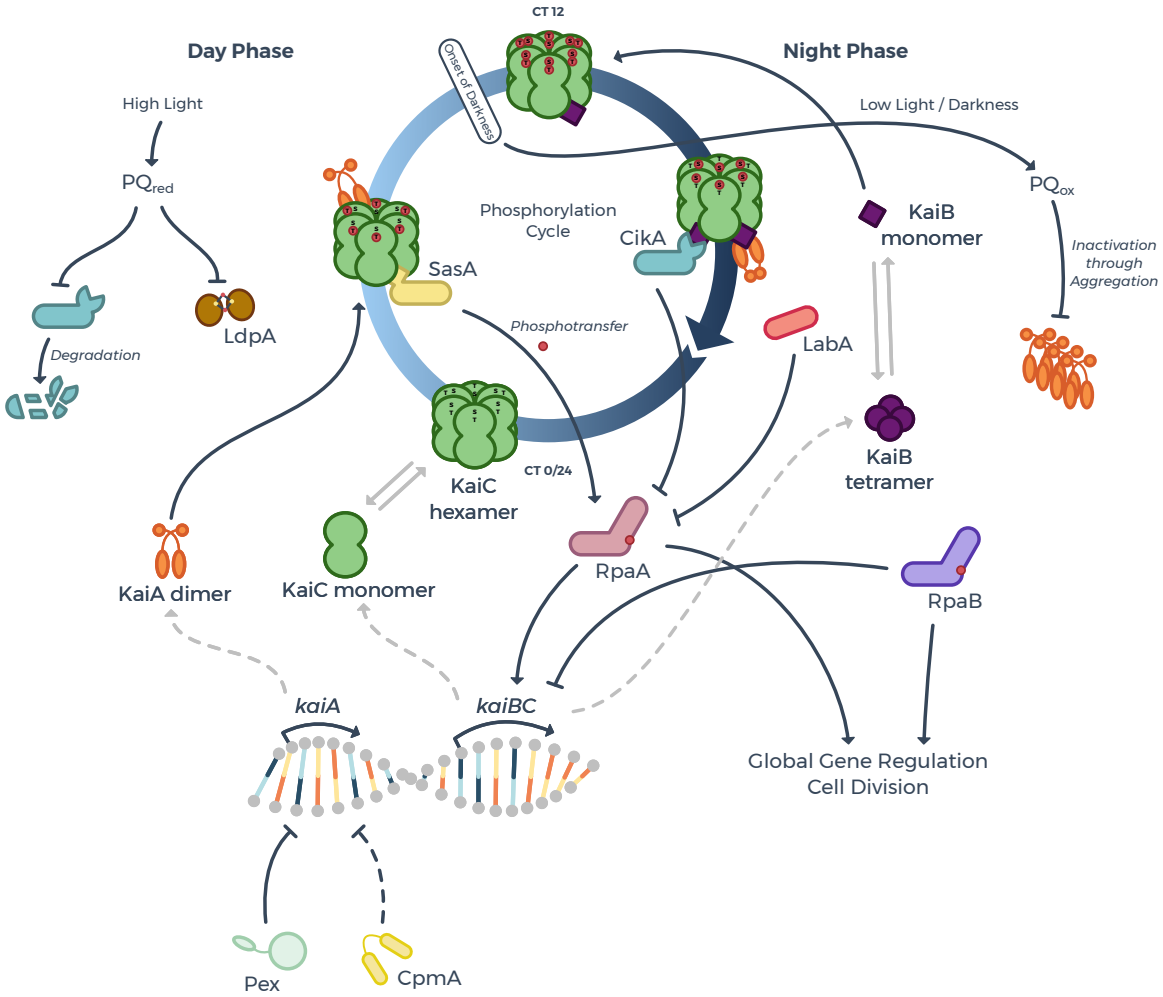


Figure 1.1: The extended network of the circadian clock in *Synechococcus elongatus* PCC 7942. At the core of the circadian clock in *Synechococcus elongatus* PCC 7942 are the three proteins KaiA, KaiB, and KaiC. The state of the clock is determined by the phosphorylation level of KaiC. KaiA stimulates the autophosphorylation of KaiC during the subjective day, whereas KaiB antagonizes the effect of KaiA by sequestering the protein and thereby initiates the dephosphorylation phase. Pex and CpmA are two transcriptional regulators of *kaiA*. The central output pathway of the clock is through RpaA, which regulates gene expression. During the subjective day, RpaA is phosphorylated by the histidine kinase SasA and dephosphorylated by CikA during the subjective night. RpaA is further inhibited by LabA. In addition, the RpaA homolog RpaB is competing for DNA binding sites. During the day, sunlight causes the reduction of the plastoquinone pool and other quinones. Reduced plastoquinones inhibit LdpA and destabilize CikA resulting in degradation of the protein. At the onset of darkness, quinones are rapidly oxidized, which cause an inactivation of KaiA through aggregation of the protein.
The figure is taken from Schmelling & Axmann 2018[31] and further extended. Graphical representations of the core proteins KaiA, KaiB, and KaiC as well as proteins from the central output pathway SasA, CikA, and RpaA are based on "Cyanobacterial Circadian Clock Output Mechanism" by The BioClock Studio (https://youtu.be/DcuKifCRx_k). The coloring of the Kai proteins is adapted to the coloring of the protein structures of the Kai proteins from Snijder and colleagues[32]. The coloring of SasA, CikA, and RpaA is adapted to those used in The Bioclock Studio video referenced before. Grey dashed lines represent transcription and translation processes. Grey solid lines represent de-/formations of homo-multimers. Dark dashed lines represent interactions of unknown mechanism. Dark solid lines represent direct interactions or effect.

As described before the core clock is able to directly sense energy or redox state of the cell. The phosphorylation of KaiC is affected by the ATP/ADP ratio directly, however, this ratio alone is not sufficient to keep the clock functional[41,42].

Even though the core factors of the cyanobacterial circadian clock are able to directly sense redox or energy states of the cell, further protein factors are needed to transfer the environmental conditions to the clock. Furthermore, protein factors are needed that read out the state of the clock and transfer this signal to control gene expression.

1.3.2 The modulation of kaiA as a circadian input

There are three studied proteins as part of the extending network of the circadian clock in *Synechococcus elongatus* PCC 7942 that either regulate the expression of *kaiA* or interact directly with this protein to transmit the environmental signal to the clock. The first input factor of the circadian clock identified in *Synechococcus elongatus* PCC 7942 was the protein period extender (Pex), named after its period-extending effect of three hours upon overexpression[43]. Pex is a transcriptional regulator, which represses the expression of *kaiA* by binding cis-element in the promoter region of *kaiA*[44]. Another protein that is thought to regulate the expression of *kaiA* is circadian phase modifier A (CpmA)[45]. However, deletion in *cpmA* also shows severe effects on growth suggesting a more general role in cellular metabolism[45].

Another factor that might modulate the activity of KaiA is the redox state of the plastoquinone pool[46,47], which is in photosynthetic organisms a function of light availability and intensity. Under high light conditions, the plastoquinone pool is reduced. Whereas darkness or even low light conditions result in an oxidation of the plastoquinone pool. The onset of darkness even causes an abrupt oxidation of quinones in Cyanobacteria. It was suggested that the N-terminal pseudo-receiver domain of KaiA is able to bind oxidized quinones, which causes KaiA proteins to aggregate and thereby silencing the stimulating effect on KaiC's phosphorylation activity[47]. Sensing of oxidized quinones by KaiA is suggested to directly signal the onset of darkness to the circadian clock[48]. However, these quinones are suggested to be some specific quinones and not from the bulk plastoquinone pool directly. These special quinones are supposed to change its redox state in interaction with other protein factors that copurify with KaiA, light-dependent period A (LdpA) and circadian input kinase A (CikA)[28,47].

LdpA is an iron-sulfur protein with two Fe_4S_4 clusters that is sensitive to the cellular redox state of the plastoquinone pool[28,49]. A reduced plastoquinone pool inactivates LdpA and sets the clock into the "short-period mode". In contrast, an oxidized plastoquinone pool activates LdpA, which then lengthens the period of the circadian clock[28]. Further, active LdpA stabilizes CikA under low light conditions or darkness[28].

KaiA is stimulating the autophosphorylation of KaiC and thus of central importance for the functionality of the circadian clock in *Synechococcus elongatus* PCC 7942. Hence, it is not surprising that KaiA is regulated at various levels. On the expression level, *kaiA* is regulated by Pex and CpmA, whereas as the KaiA protein is regulated by the redox state of quinones either directly or potentially in interaction with LdpA and CikA.

1.3.3 Antagonistic network in the readout of the clock

Reading out the state of the clock and transferring this information to downstream processes is the next step of circadian regulation in *Synechococcus elongatus* PCC 7942. The main role of the circadian clock is the orchestration of global gene expression[50,51]. However, cell division[52,53], as well as chromosome compaction[54], are suggested to be circadian regulated in *Synechococcus elongatus* PCC 7942. This process is governed by the interaction of the core clock with proteins in the output network. At the center of the output network, is the regulator of phycobilisome associated A (RpaA), which is activated and deactivated by the histidine kinase *Synechococcus* adaptive sensor A (SasA) and CikA, respectively[55,56]. In addition, low-amplitude and bright A (LabA) and regulator of phycobilisome associated B (RpaB) as well as labA-like A (LalA) and circadian rhythmicity modulator (Crm), which are not further regarded, are part of an extended output network[57–60].

During the day when KaiC phosphorylation is stimulated by KaiA, SasA interacts with KaiC through its N-terminally KaiB-like sensory domain[61]. This interaction with KaiC, in turn, stimulates the autophosphorylation activity of SasA[55,61]. Histidine-phosphorylated SasA transfers its phosphate group to the downstream partner RpaA. RpaA is an OmpR-type response regulator, which is activated by this phosphotransfer.[54,55]. Following, activated RpaA is the major regulator for circadian genome-wide gene expression in *Synechococcus elongatus* PCC 7942[30,55]. Whereas, SasA is currently the only known activator of RpaA, there are several proteins that either directly inactivate RpaA or compete with RpaA about DNA binding sites.

CikA is the main repressor of RpaA by working antagonistically to SasA[56]. CikA interacts with the KaiBC complex during the subjective night, more precisely to the KaiC-bound KaiB protein at the same site of KaiA binding[62]. This interaction enhances the dephosphorylation of RpaA through CikA[56]. It is suggested that CikA is also involved in the input pathway in *Synechococcus elongatus* PCC 7942, serving a dual role on the regulation of the circadian clock[63,64]. It has been shown that CikA is able to sense the redox state of the plastoquinone pool by binding quinones directly to its pseudo-receiver domain[64]. However, quinone binding is suggested to reduce the stability of CikA[28]. Another aspect that hints at an involvement in the entrainment of the circadian clock is the influence of CikA on the phosphorylation state of

KaiC during recovery of the clock after a dark pulse[64]. This is in agreement with previous findings, which show an inability of CikA deficient mutants to reset the phase of the clock after a five-hour dark pulse[63]. In addition, CikA deficient mutants show an amplitude reduction as well as a shortening of the circadian period by two hours[63].

Alongside the antagonistic regulatory network of SasA and CikA, RpaA is also regulated by LabA and its homologous response regulator RpaB, which is involved in circadian gene expression and control of cell growth[65,66]. This antagonistic regulation of RpaA is extended by LabA, which is believed to comprise a second negative regulation of RpaA in addition to CikA during the night[57].

In contrast, RpaB works in parallel to RpaA by regulating gene expression and competing for DNA binding sites rather than inhibiting RpaA directly. However, the dephosphorylation of RpaA and thus its deactivation is also affected by RpaB[58]. The main regulator mechanism of RpaB is, analogous to RpaA, the phosphorylation of RpaB as the protein abundance remains almost constant[67]. During the subjective night, RpaB represses gene expression, which might in some cases be supported by other factors like CikA or LabA at dusk. Whereas during the subjective day, phosphorylated RpaA is suggested to mediate dissociation of RpaB from the DNA[66]. Recently, new protein factors of the circadian network have been identified. First, the phosphate-responsive response regulator SphR, which accepts phosphate from CikA and SasA in *in vitro* assays[56]. Deletion of *sphR* alone does not interrupt circadian rhythmicity, however, in triple mutants together with *cikA* and *sasA* cell elongation phenotypes are visible[68]. The second factor is another histidine kinase CikB, which interacts with SphR and RpaA comprising an additional input mechanism in parallel to CikA. Double deficient mutants of *cikA* and *cikB* show nearly wild-type circadian rhythmicity, supporting previous findings that oxidized quinones can directly reset the clock and that the RpaA/SasA output pathway alone is sufficient to maintain circadian rhythm[48,68]. All this elucidates the complexity of the RpaA-centered output network and highlights that this system is still not yet fully understood.

1.3.4 The diversity of the circadian clock factors among cyanobacteria

All of the before described insights in the functionality of the circadian clock were studied in the model organism *Synechococcus elongatus* PCC 7942 and it is tempting to think that those insights are universally true for all Cyanobacteria. However as mentioned before, Cyanobacteria are known for their diversity, which can be observed as expected with regard to the circadian clock. Our knowledge about the diversity of circadian clock systems in Cyanobacteria is still sparse compared to the plethora of information about the system in *Synechococcus elongatus* PCC 7942.

Circadian rhythms have been identified in several cyanobacterial strains, from marine *Prochlorococcus* MED4[69]

over *Synechocystis* sp. PCC 6803[70,71] and *Synechococcus elongatus* PCC 7942[51,72] to the nitrogen-fixing cyanobacteria *Cyanothece* ATCC 51142[73,74] and *Anabaena* sp. PCC 7120[75] and the toxic cyanobacterium *Microcystis aeruginosa* PCC 7806[76]. All of these strains show oscillating gene expression, however, large differences in the percentage of genes in the genome are detectable ranging from 20 – 79%.
Beside those differences in overall transcription, variations in the protein composition are detectable. Whereas *Synechococcus elongatus* PCC 7942 has the three factors KaiA, KaiB, and KaiC at its core, *Prochlorococcus* strains only harbor two of those three proteins with KaiB and KaiC[77,78]. Interestingly, the resulting timing system in *Prochlorococcus* is not considered as a circadian clock since oscillations cease without an exogenous stimulus. Thus, this timing system is called an hourglass-like clock[77]. On the other side, three copies of the core clock factors KaiB and KaiC have been identified in *Synechocystis* sp. PCC 6803[79]. It has been shown that those KaiB and KaiC, which show the highest sequence similarity to the proteins in *Synechococcus elongatus* PCC 7942 and form a gene cluster with KaiA, also comprise the core of the clock in *Synechocystis* sp. PCC 6803 and are thus named KaiB1 and KaiC1[33]. However, it remains unknown what the function of the two other orthologs are and if they are even involved in the regulation of the circadian clock. KaiB2 and KaiC2 are only characterized in *Legionella pneumophila*, where they act in the stress response[80], while KaiB3 and KaiC3 are suggested to fine-tune the circadian clock in *Synechocystis* sp. PCC 6803[81].

1.4 Comparative genomic analyses

With the rise in high-throughput technologies and the amount of data researchers produce, it is now better than ever possible to compare organisms with each other on a genomic level. One approach that is often used in such cases is a comparative genomic analysis. Thereby, the genomic sequences of multiple organisms are compared to find sequences that are similar between these organisms and ones that differ (Fig. 1.2 A). That way it is possible to identify conserved sequences or motifs. These analyses yield information about the relation of those sequences, which are defined as analogs, homologs, orthologs, or paralogs. Analogs are sequences, which have evolved convergently and are thus unrelated, whereas, homologs are sequences that are related to each other by sharing a common ancestor. Homologs are further differentiated as orthologs or paralogs. Orthologs are sequences, which retain similar functionality in two different organisms that share one single ancestor. In contrast, paralogs are homologous sequences that arise from a gene duplication event and that do not share functional properties[82]. It is then often assumed that the conservation on a sequence level transfers into a conservation of the function of this gene or gene product between the analyzed organisms. Comparative genomic analyses are nowadays done with the help of computers in order to be able to analyze the ever-increasing numbers of genomes that are sequenced. Besides

comparing two or more genomes with each other it is also possible to search for orthologs of a gene or gene product of interest in a large number of organisms to identify those organisms that harbor this conserved gene sequence in their genome (Fig. 1.2 A). Two commonly used computer-based tools for comparing sequences with each other are the "Basic Local Alignment Search Tool" (BLAST)[83] and the CLUSTAL[84,85] (Fig. 1.2 A). Depending on the approach either of the tools is used or sometimes even both in sequence.
BLAST searches for homologous sequences to a sequence of interest in a given database. The algorithm aligns a short subsequence of a few characters, e.g. nucleotides or amino acids, to sequences in the database. Successful alignments are then extended to create an alignment that fully covers the sequence of interest. However, the results are often only partial alignments between the two sequences.
CLUSTAL, on the other side, compares two or more sequences globally, i.e. the algorithm takes all input sequences in its full-length form. The algorithm then tries to align those sequences. In order to best align all given sequences inserts CLUSTAL gaps between parts of the sequences. Further detailed information about the functionality of the two tools can be found in the respective publications.
Using a comparative genomic approach sequence orthologs can be identified between a large number of organisms. These orthologs can give first hints into a conserved functionality of a gene or protein in a related species or identify species that might harbor a certain system.

1.5 Computational modeling of biological processes

Analyzing the presence or absence of a certain gene or gene product can help to elucidate the structure and properties of an organism's functionality. However, these kinds of computational analysis give only a static picture of reality, whereas biological processes are highly dynamic. Regulatory or metabolic processes vary during the course of a day and change with regard to the environmental conditions. Biological functionality comes from the interaction of components, like transcription factor binding to a promoter sequence on a DNA or conversion of a substrate by an enzyme. Those interactions are commonly represented as networks. These networks can describe, e.g. small regulatory interactions (Fig. 1.3 A) or even genome-scale metabolic networks (Fig. 1.3 B).
As seen before with the sophistication of technologies and the increase in the amount of data, computational tools are needed in order to sufficiently analyze them. However, for the reconstructing of the dynamic nature of biological processes, computational approaches are needed that differ from the static approaches described before. Those dynamic computational approaches fall under the concept of mathematical or computational modeling. Common to all kinds of mathematical or computational models is that they try to decode a biological process in a formal language. Representing biological processes in some

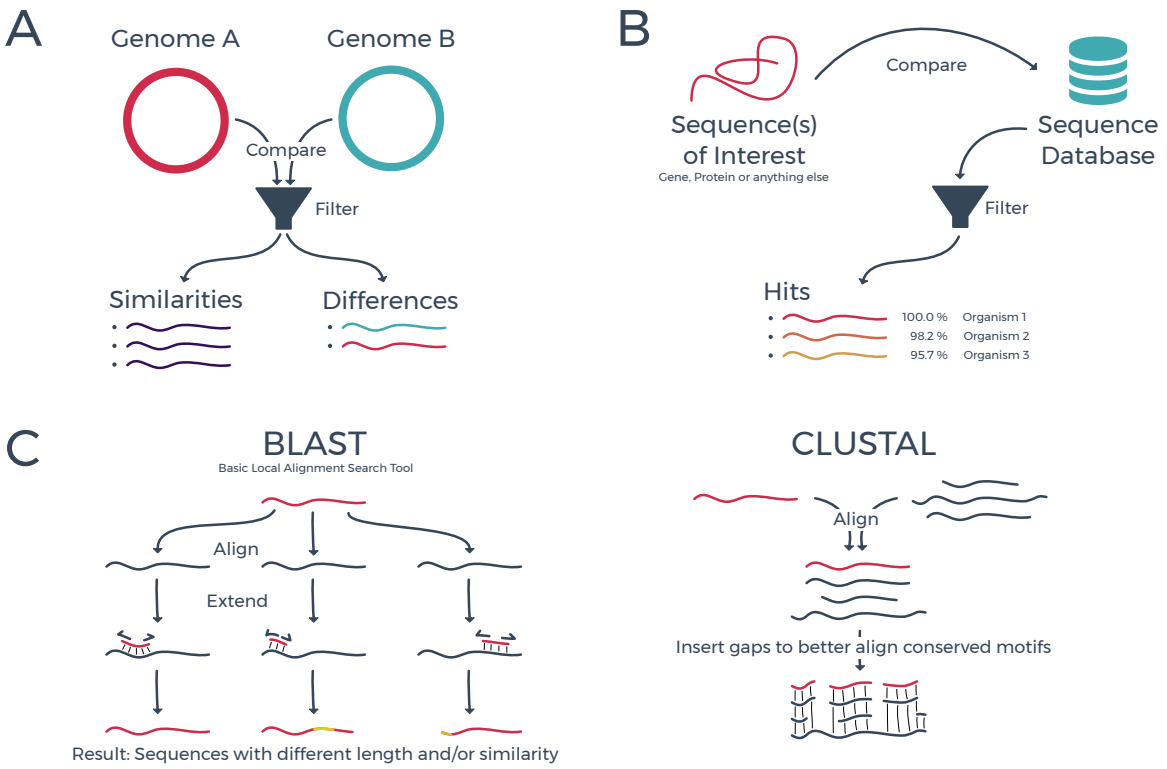


Figure 1.2: Comparative genome analyses. **A** Whole-genome comparison. Comparing all gene or protein sequences of a genome to another genome or multiple genomes yields in a set of shared sequences and sequences that differ between the input genome and the references. The results of this analysis are depending on the filter threshold for similarity between the single sequences. **B** Searching for orthologs of a given sequence or set of sequences. Instead of comparing all known sequences of two or more genomes, a single sequence or a set of sequences can be compared against a database of sequences. This database can be constructed depending on the purpose of the analysis. The results of this analysis depend on the filtering criteria and thresholds. **C** Tools for ortholog search. Two software tools that are commonly used for the before described analyses are BLAST and CLUSTAL. BLAST searches for similar sequences in a given database by aligning a short subsequence of the original sequence and then extends the alignment. Whereas CLUSTAL takes multiple sequence as an input and tries to best align those sequences by inserting gaps into the sequences.

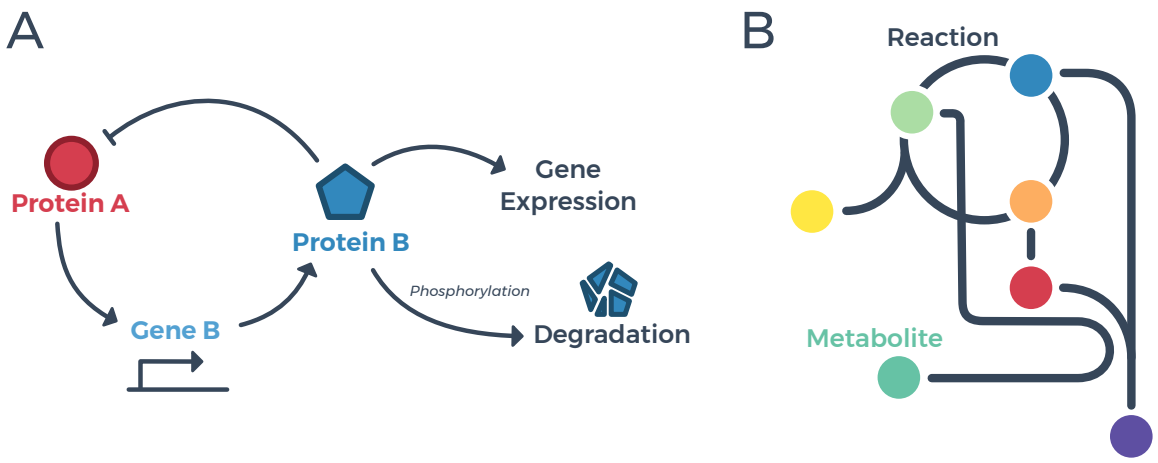


Figure 1.3: Schematic representation of mathematical models used in biological research A Schematic representation of a small regulatory interaction network. **B** Schematic representation of a genome-scale metabolic network.

kind of formal language serves two purposes according to Steuer and colleagues[86]: (i) communication of knowledge about a functional interaction, which would be impossible verbally; (ii) prediction of the behavior of a set of interactions beyond what is human intuition is able to do. For simplification in this thesis computational modeling can be divided into two classes, which are distinguished by the size of the representing network, level of detailed description, and the kind of dynamic reconstruction: large-scale models and small-scale models.

Large scale models such as a genome-scale representation of a metabolic network with all known metabolites and reactions in a cell are identified by a large number of components accompanied by a reduced level of detail about each component. This is usually due to a lack of knowledge about rate constants or other kinetic parameters. Large-scale models are represented as graph-based topological analyses or stoichiometric models. However, in those cases where detailed information is available and a higher level of detail is desired, small-scale models are more suited. Those models are most often represented by a set of ordinary differential equations (ODE).

Even though computational models help researchers to understand the behavior of a given system in more detail, those models also have their limitations as they are only an abstraction of the real system.

2

Aims of the Thesis

Comparative conversation analysis of circadian clock factors in cyanobacteria

Cyanobacteria are known for their high diversity in morphology, physiology, habitats as well as genome size and structure[21]. Furthermore, they are the best-studied prokaryotes, which regulate large parts of their genome in a circadian fashion. Thus, it is not surprising to see that there are even differences in the composition of the factors involved in the circadian clock. Previous studies with a smaller sample size already indicated that there are three different types of timing systems among Cyanobacteria: (i) the circadian clock model organisms *Synechococcus elongatus* PCC 7942, (ii) the hourglass timing system of *Prochlorococcus*, missing some of the factors known from *Synechococcus elongatus* PCC 7942, and (iii) the system from *Synechocystis* sp. PCC 6803, which harbors multiple copies of core clock factors[42,79]. Some of these studies also indicate a more widespread distribution of clock factors from Cyanobacteria in other Bacteria and even Archaea[79]. With a large increase in available genomic data, it is now possible to get a comprehensive picture of the distribution and conservation of cyanobacterial circadian clock factors. This

work aims at analyzing the diversity of circadian clock factors among Cyanobacteria as well as identifying possible KaiC-based circadian clocks in other organisms outside of the cyanobacterial phylum. With this information, conservation analyses of sequence motifs are feasible that are known to be important for the structural interaction of the core clock proteins. On the other side, it is possible to analyze the co-occurrences of these factors to further predict sets of protein factors, which have different functions within this timing system. The new insights into the distribution and co-occurrence of circadian clock factors can be used to identify novel dependencies of these factors, which can guide new experiments to unravel the molecular mechanism and interaction.

Mathematical modeling to unravel mechanisms and design principles of the cyanobacterial circadian clock

It is just over a decade ago that the circadian clock of *Synechococcus elongatus* PCC 7942 was reconstructed in a test tube[87]. Since then tremendous achievements have been made by using mathematical modeling to decipher various molecular mechanisms of this timing system and circadian systems in general. In this work, we review the detailed insights into the functionality of circadian systems and the cyanobacterial circadian clock in particular gained by mathematical modeling. These insights range from the description of the molecular functions of the core clock and its interaction over to strategies for entrainment and input compensation such as shorter phosphorylation cycles of individual KaiC hexamers under low ATP conditions. Further strategies for robustness to noise and environmental changes such as the increase in genome copy number or the uncoupling of the oscillator from transcriptional processes were identified. Another part of this review is to highlight and identify those aspects of circadian clocks that are still unanswered and need further investigation. Ones fully understood insights in circadian regulation allow for manipulation of the system with varying potential applications such as the industrial production of chemicals under natural conditions or terraforming of habitable planets and moons with different day lengths.

Zeit ist das, was man an der Uhr abliest.

Albert Einstein

3

Key results of this thesis

Circadian regulations in Cyanobacteria were first discovered over three decades ago[17]. Since then the circadian clock of the model organism *Synechococcus elongatus* PCC 7942 has been studied in detail. However, there are still open questions regarding this circadian system, including the core of the circadian clock. Furthermore, the system shows, against initial beliefs, large complexity with regard to the number of factors involved and their interactions.

The structure of the core clock proteins KaiA, KaiB, and KaiC has been resolved individually, but structural information of naturally occurring complexes of these three proteins remained unclear[40,88–90]. The results of the work described in Chapter 4 clearly show the different complexes these three proteins can be found in and their stoichiometry[32]. Furthermore, the interaction sites between the three proteins and the conformational changes that occur have been described in detail throughout the full 24-hour phosphorylation cycle[32]. Similarity analyses on the sequence of the three core clock proteins among Cyanobacteria revealed a strong conservation of these sequences, especially in those regions that are important for the interaction of the proteins[32], which hints at a conserved mechanism among Cyanobacteria.

It was suggested in previous studies that Cyanobacteria show some diversity also with regard to the cir-

cadian system[42,79]. With an increase in available genomic data, we were able to study the diversity of the circadian system more comprehensively[91]. The findings covered in Chapter 5 strengthens the previous findings of the diversity of the circadian system in Cyanobacteria. Furthermore, orthologs in other bacterial species and Archaea could be identified by these analyses, extending the previous knowledge about the distribution of those proteins. Phosphorylation studies of KaiC orthologs from Archaea also demonstrate conserved functionality of KaiC across kingdoms[91]. In addition, the co-occurrence analyses of protein factors in Cyanobacteria highlights two sets of proteins that are suggested: (i) to enable timing in cyanobacteria, which includes the core clock factors KaiB, KaiC, the input factors LdpA, IrcA, and the output factors SasA, RpaA, RpaB, and CpmA, and (ii) further to extend this ability to also circadian regulation, comprised of the core factor KaiA, the input factors CdpA, PrkE, CikA, as well as the output factors LabA, and LalA[91].

Besides the detailed insights into the functionality of the circadian clock in Cyanobacteria, generated through laboratory experiments, a tremendous understanding of this clock system, in particular, and circadian clocks, in general, have been achieved by mathematical modeling. In Chapter 6 we review a decade of theoretical studies and their remarkable findings[31]. Mathematical models have helped to understand the molecular mechanisms of the cyanobacterial circadian clock, including the 24-hour phosphorylation cycle with its interaction of KaiA, KaiB, and KaiC[92]. Insights gained through mathematical modeling revealed strategies to filter the noise of the external stimulus and the limitations of this system regarding noise[93–95]. Furthermore, some models describe strategies for entrainment and stability against changing nucleotide concentrations as well as ones for uncoupling of circadian clocks from the cell cycle[96–98].

4

Structures of the cyanobacterial circadian oscillator frozen in a fully assembled state

Joost Snijder, Jan M. Schuller, Anika Wiegard, Philip Lössl, Nicolas Schmelling, Ilka M. Axmann, Jürgen M. Plitzko, Friedrich Förster, Albert J. R. Heck (2017) Structures of the cyanobacterial circadian oscillator frozen in a fully assembled state, *Science*, 355 (6330), 1181–1184, doi: *10.1126/science.aag3218*

Author Contributions

JS, JMS, JMP, FF, and AJRH designed the structural experiments. JS, JMS, JMP, and FF performed the cryo-EM analyses. JS and PL performed the mass spectrometry analyses. AW designed the protein expression experiments and cloned, expressed, and purified recombinant proteins. NS designed and carried out bioinformatic sequence analyses. AJRH, FF, JMP, and IMA supervised the study. JS and JMS prepared the manuscript together with AW, PL, JMP, FF, and AJRH. All authors read and approved the final manuscript.

Pro rata

Bioinformatic study design: 100 %

Bioinformatic data analyses: 100 %

Summary of Publication

In this chapter, the structural interaction of the three core clock proteins is elucidated. Previously, structural information about the complex of KaiA, KaiB, and KaiC have been inconclusive and only been studied with modified or truncated Kai proteins. Thus it was unclear how the core complex forms in nature. In a collaborative work together with Joost Snijder, Jan M. Schuller, Anika Wiegard, Philip Lössl, Ilka M. Axmann, Jürgen M. Plitzko, Friedrich Förster, and Albert J. R. Heck we used mass spectrometry and cryo-electron microscopy to unravel the complex formation and interaction. The Kai proteins were modified if needed and purified by Anika Wiegard and further analyzed by Joost Snijder, Jan M. Schuller, and Philip Lössl in mass spectrometry and cryo-electron microscopy. We could identify the temporal formation of the different Kai complexes and further clearly elucidate the interaction sites between the three Kai proteins, which were previously under debate. It was of interest to us how conserved the interaction sites are in cyanobacteria. In the following, I performed similarity analyses of the three proteins in Cyanobacteria and mapped the conservation of each amino acid onto the original sequence of the Kai proteins in *Synechococcus elongatus* PCC 7942 and thereby identifying a high conservation of the sites involved in the interaction between the three proteins. The results of this work were published in *Science* and can be found on doi: *10.1126/science.aag3218*.

4.1 Abstract

Cyanobacteria have a robust circadian oscillator, known as the Kai system. Reconstituted from the purified protein components KaiC, KaiB, and KaiA, it can tick autonomously in the presence of ATP. The KaiC hexamers enter a natural 24-hour reaction cycle of autophosphorylation and assembly with KaiB and KaiA in numerous diverse forms. We describe the preparation of stoichiometrically well-defined assemblies of KaiCB and KaiCBA, as monitored by native mass spectrometry, allowing for a structural characterization by single-particle cryo–electron microscopy and mass spectrometry. Our data reveal details of the interactions between the Kai proteins and provide a structural basis to understand periodic assembly of the protein oscillator.

4.2 Introduction

Many organisms, from Cyanobacteria to animals, have adapted to Earth's day-night cycle with the evolution of an endogenous biological clock. These clocks enable circadian rhythms of gene expression and metabolism with a period close to 24 hours. Many circadian rhythms rely on complex networks of transcription-translation feedback, but simpler posttranslational oscillations have also been described in both Cyanobacteria and human red blood cells[99]. The circadian oscillator of Cyanobacteria is composed of three components: the proteins KaiC, KaiB, and KaiA[100]. This posttranslational oscillator is robust enough to allow reconstitution simply through incubation of purified recombinant KaiC, KaiB, and KaiA in the presence of ATP[87]. The in vitro oscillator can maintain a stable rhythm for weeks[101,102], allowing for its detailed study. The proteins of the Kai system collectively generate a circadian rhythm based on assembly dynamics associated with KaiC autophosphorylation and dephosphorylation[103,104]. KaiC forms a homohexamer consisting of two stacked rings of domains CI and CII, which have ATPase and kinase activity, respectively[105]. During the subjective day, the kinase activity of KaiC is stimulated by the binding of KaiA to the intrinsically disordered C-terminal regions of KaiC, resulting in sequential autophosphorylation at Thr^{432} and Ser^{431} of KaiC[38]. During the subjective night, KaiB interacts with phosphorylated KaiC, forming the KaiCB complex[105]. Binding of KaiB to KaiC changes the activities of SasA and CikA, which are key signaling proteins of clock output pathways that modulate transcription[30]. Moreover, the KaiCB complex exposes an additional KaiA-binding site, sequestering KaiA and thus preventing its productive association with KaiC[106]. The sequestration of KaiA allows KaiC to readopt its default autodephosphorylation activity, thereby slowly resetting the protein clock to an unphosphorylated state[101].

Atomic-level structures of the individual Kai proteins are available[88–90], but structural information on the KaiCB and KaiCBA complexes is still ambiguous[40,107–109]. KaiB forms monomers, dimers, and tetramers

in solution, with six KaiB monomers binding cooperatively to one KaiC hexamer[110]. It has been unclear whether KaiB binds to the CI or CII domain of KaiC[106,107,110–112]. Nuclear magnetic resonance (NMR) spectroscopy studies of engineered and truncated Kai proteins suggested that KaiB binds the KaiC-CI domain and only one subunit of a KaiA dimer[106], but it is unclear whether the wild-type, full-length proteins arrange similarly in the KaiCBA complex. Here we use mass spectrometry (MS) and cryo-electron microscopy (cryo-EM) to study the assembly and structures of the full-length clock components to provide a structural basis for the assembly dynamics of the *in vitro* circadian oscillator.

4.3 Materials and Methods

4.3.1 Cloning, Expression and Purification of Kai proteins

Plasmids pGEX-*kaiA*, pGEX-*kaiB* and pGEX-*kaiC* for expression of GST-KaiA, GST-KaiB and GST-KaiC were kindly provided by T. Kondo, Nagoya University, Nagoya, Japan. To express GST-KaiB-K42A, GST-KaiC-A108L and GST-KaiC-A108E mutations were introduced into pGEX-*kaiB* or pGEX-*kaiC* using the primers listed in Supplementary Table A.4 and the Q5® Site-Directed Mutagenesis Kit (New England Biolabs). Protein expression and purification from *Escherichia coli* BL21 (DE3) cells was carried out as previously described[33,110]. After cell lysis, GST-fused proteins were bound to Protino Glutathione Agarose 4B (Macherey and Nagel) and cleaved of their GST-tag by incubation with PreScission Protease (GE Healthcare). Homogeneity of the recombinant proteins was controlled by separating them via SDS-PAGE. If it was not sufficient, proteins were applied to a MonoQ-column (GE healthcare) and separated. After dialysis in Reaction buffer (20 mM Tris·HCl, pH 8, 150 mM NaCl, 0.5 mM EDTA, 5 mM $MgCl_2$, and 1 mM ATP) protein concentration was determined using infrared spectroscopy (Direct Detect, Spectrometer, Merck Milipore) or the Bradford method.

4.3.2 Sequence conservation analysis

The coding sequences of the genome of *Synechococcus elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803 were obtained from the NCBI FTP server (May 2016). Further, the sequences of KaiA from *Synechococcus elongatus* PCC 7942 and KaiB1 and KaiC1 from *Synechocystis* sp. PCC 6803 were retrieved from the downloaded coding sequences and compared to the entries in the Cyanobase Database[113] to assure correctness. The KaiB and KaiC homologs from *Synechocystis* sp. PCC 6803, rather than the ones from *Synechococcus elongatus* PCC 7942, were selected to exclude any hits to other naturally occurring homologs of these proteins, which do not resemble the core clock of cyanobacteria. For the construction of a database all entries in the genbank protein database[114] that were labeled as "Complete Genome" or "Chromosome"

(May 2016) were downloaded. The amino acid sequences of KaiA, KaiB1, KaiC1 were set as queries for a homology search using the standalone version of BLASTP 2.2.30+[83] and compared against the constructed database (May 2016; Settings: e-value cut-off: 10^{-5}, max. number of alignments: 10,000, wordsize: 3, matrix: BLOSUM62). In a reciprocal BLAST analysis the generated matches were used as queries to eliminate false positive results. Only the reciprocal best hit was recorded for further analyses. This time the coding sequences of *Synechococcus elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803 were used as databases (Settings: e-value cut-off: 10, max. number of alignments: 1, wordsize: 3, matrix: BLOSUM62). Afterwards all matches were filtered regarding their alignment partner. Only those protein records that aligned to the KaiA, KaiB1, or KaiC1, respectively, were kept and considered as a validated hit. Multiple sequence alignments of KaiA, KaiB1, KaiC1 homologs for only those cyanobacteria, which showed homologs to all three protein, were constructed with the standalone version of CLUSTAL Omega 1.2.1-1[85] (May 2016; Settings: iterations: 20, max. guidetree iterations: 1 dealign: yes). Afterwards the alignments were edited with Jalview[115] such that all gaps in the sequences of the respective *Synechococcus elongatus* PCC 7942 homolog were removed. Edited multialignments were used to create WebLogos[116].

4.3.3 Complex Assembly

Kai proteins were incubated for assembly in Reaction Buffer, For the 24 hour time course experiments KaiC, KaiB and KaiA were incubated at a final concentration of 3 μM, 3 μM and 1.5 μM, respectively, at 30 °C or 4 °C, as indicated. To achieve complete formation of $KaiC_6B_6$, KaiC and KaiB were incubated for one week at 4 °C, at final concentrations of 3 μM and 9 μM, respectively. To achieve complete formation of $KaiC_6B_6A_{12}$, KaiC, KaiB and KaiA were incubated for one week at 4 °C, at final concentrations of 3 μM, 9 μM and 9 μM, respectively. The same conditions were used to test the effect of single point mutations (KaiC-A108L, KaiC-A108E and KaiB-K42A) on KaiCBA complex assembly.

4.3.4 Native Mass Spectrometry

Kai samples in Reaction Buffer were transferred to ice-cold MS-compatible buffer containing 75 mM ammonium acetate (pH 6.8) using Vivaspin 500 centrifugal filter units with a 5 kDa molecular weight cutoff. The centrifugation steps were performed at 4 °C and samples were kept on ice until analyzed. Aliquots of 1-2 μL were loaded in gold-coated borosilicate capillaries prepared in-house for nanoelectrospray ionization. The samples were analyzed on a Waters QTOF2 instrument, modified for native MS of large protein complexes. The instrument was operated at 10 mbar source pressure, 1,300–1,500 V capillary voltage, 160 V cone voltage, 50 V collision energy with $1–2\times10^{-2}$ mbar pressure in the collision cell using xenon as collision gas. In addition, the 24 hour time course experiments were also analyzed on a Micromass LCT1 mass

spectrometer, modified for high mass analysis. The instrument was operated at 9 mbar source pressure, 1,300–1,500 V capillary voltage, 160 V cone voltage and the pressure in the first hexapole was kept at 4×10^{-2} mbar using argon as a buffer gas. Peaks were manually identified in the mass-to-charge ratio spectra and matched by charge state for a given assembly. From matched peaks, the exact charge state can be calculated for each peak which is subsequently used to calculate the mass by multiplication with the mass-to-charge ratio (see Fig A.2). The masses of Kai assemblies reported in Supplementary Table A.1 correspond to the average over all charge states identified for that complex. The assignments for Kai complexes listed here have additionally been verified by tandem MS experiments.

4.3.5 Hydrogen/deuterium exchange MS

Free KaiC, free KaiB and free KaiA were incubated for one week in Reaction Buffer at 4 °C at a concentration of 30 μM. KaiB-bound KaiC was incubated for one week in Reaction Buffer at 4 °C at a concentration of 30 μM in the presence of 90 μM KaiB. KaiC-bound KaiB was incubated for one week in Reaction Buffer at 4 °C at a concentration of 30 μM in the presence of 90 μM KaiC. All components in the KaiCBA complex were analyzed from KaiC, KaiB and KaiA incubated for one week in Reaction Buffer at 4 °C at a concentration of 30 μM, 90 μM and 90 μM, respectively. The HDX reaction was started by diluting the above stock solutions 20-fold to a final volume of 40 μL in D_2O. The reaction was carried out on ice for 1, 10, and 60 min before quenching by 2:1 dilution into ice-cold 6 M guanidine HCl, 300 mM Tris (2-carboxyethyl) phosphine, with pH adjusted to give a final pH of 2.5. Immediately after quenching, the sample was injected into a Waters HDX/nanoAcquity system for digestion on an online pepsin column (25 °C, at a flow rate of 50 μL min^{-1}) followed by separation on a 10-min RP-UPLC gradient at 0 °C and MS on a Waters Xevo QToF G2. For peptide identification, samples of KaiC and KaiB were prepared under identical conditions in H_2O and analyzed using MS^e data acquisition. Data for peptide identification was processed with ProteinLynx Global Server 2.5 software. Deuterium uptake was calculated compared with the control samples in H_2O using Waters DynamX 1.0.0 software. The experiments were performed in triplicate.

4.3.6 Cross-linking MS

The individual Kai proteins were transferred from Reaction buffer to 25 mM HEPES pH 8, 150 mM NaCl using Vivaspin 500 centrifugal filter units (Sartorius) with a 5 kDa (KaiB) or 10 kDa (KaiA and KaiC) molecular weight cutoff. Subsequently, the $KaiC_6B_6A_{12}$ complex was assembled as described above. A total amount of 300 μg $KaiC_6B_6A_{12}$ was cross-linked at a protein complex concentration of approximately

1 mg/ml using 1 mM bis(sulfosuccinimdyl)suberate (BS3, Thermo Fisher Scientific). The cross-linking reaction proceeded at 4 °C and was quenched after 4 hours by adding Tris-HCl (pH 7.6) in 20-fold molar excess to the BS3 concentration. The cross-linked proteins were denatured in 8 M urea, reduced with 4 mM DTT for 30 min at 56 °C, alkylated with 8 mM iodoacetamide for 30 min in the dark, again supplemented with 4 mM DTT and finally digested overnight at 37 °C using trypsin at a 1:100 (*w*/*w*) enzyme:substrate ratio. The resulting peptide mixture was desalted using Sep-Pak C18 cartridges (Waters), dried under vacuum, dissolved in 10 % (*v*/*v*) formic acid and fractionated by strong cation exchange (SCX) chromatography as described previously[117,118]. The elution fractions that, according to the SCX chromatogram, contained predominantly higher charged peptides ($z \geq 3$) were analyzed by reversed-phase liquid chromatography/tandem-mass spectrometry (LC/MS-MS) using a Proxeon EASY-nLC 1000/Orbitrap Elite system (both Thermo Fisher Scientific) and in-house packed C18 columns (Poroshell 120 EC-C18, 2.7 µm (Agilent Technologies)). Peptides were loaded using 100 % solvent A (0.1 % (*v*/*v*) formic acid in water) and separated over 60 min at 350 nL/min flow rate by increasing the solvent B concentration (0.1 % (*v*/*v*) formic acid in acetonitrile). Precursor and fragment ion mass spectra were acquired in the Orbitrap mass analyzer, respectively operating with 60,000 and 15,000 mass resolution (at m/z 200). The 5 most abundant precursor ions carrying at least 3 positive charges were fragmented by subjecting them to sequential collision-induced dissociation (CID) and electron-transfer dissociation (ETD). Subsequently, all acquired fragment ion spectra were deconvoluted to charge state 1 using the MS2 spectrum processor add-on node in Proteome Discoverer v1.4 (Thermo Fisher Scientific) and analyzed with the XlinkX crosslink search engine[118]. The XlinkX search was performed in enumeration mode with 10 ppm precursor mass tolerance and 20 ppm fragment mass tolerance, using a database that was derived from the three Kai protein sequences and nine shuffled decoy sequences. This target-decoy crosslink database was generated using the following settings: peptide length = 3–40 amino acids, allowed missed cleavages = 3, enzyme = trypsin, linked residue = Lys, fixed modification = Cys carbamidomethylation, variable modifications = ≤2 Met oxidations. Based on the individual confidence scores (N scores) of the matched target- and decoy-peptides, an N score-dependent false-discovery rate (FDR) was calculated. A crosslink was accepted when the N scores for both crosslinked peptides were $\leq 6e^{-10}$, which corresponds to $\approx$ 1 % FDR. As an additional criterion, accepted crosslinks had to be confirmed by at least 2 independent fragment ion mass spectra.

4.3.7 Cryo-electron microscopic analysis of the KaiCB and KaiCBA complex

We determined the structure of KaiCB and KaiCBA complexes, prepared as described above, by single particle cryo-electron microscopy (cryoEM). For both complexes, about 4 µL of protein was applied to a lacey carbon copper grid, incubated for 30 s, manually blotted with filter paper and washed two times

with 4 µL buffer before final blotting and vitrification in liquid ethane cooled by liquid nitrogen. The grids were imaged using a Titan Krios transmission electron microscope (FEI Company) operated at 300 kV. All images were recorded on a post-GIF K2 Summit camera (Gatan Inc,) operated in super-resolution counting mode with a defocus range between 1 and 3.5 µm, using a dose rate of 3 $e^-/Å^2$ at the specimen plane and a total exposure time of 15.2 s. Every 0.4 s an intermediate frame was recorded resulting in a total of 38 frames with an accumulated dose of 45 $e^-/Å^2$. Images were collected at 105,000 x magnification (1.35 Å/pixel). TOM² automated acquisition software[119] was used to collect a total of 1,694 micrographs for the KaiCB dataset and 3,181 micrographs for the KaiCBA dataset.

The movie stacks were aligned using an in-house implementation of the algorithm described in[120]. CTF parameter determination was carried out on the whole micrograph level using ctffind4[121]. All image processing was performed in RELION[122]. Particles were picked manually in a subset of micrographs using e2boxer[123] to generate a library of 2D class averages for subsequent automatic particle picking[124]. A total of 213,329 particles for KaiCB and for 77,079 particles for KaiCBA were selected and subjected to iterative rounds of 2D alignment and classification. The results of the 2D analysis were used to remove damaged, aggregated, or falsely selected particles from the dataset for subsequent 3D analysis. As an initial model for 3D Classification a cylinder with the approximate dimensions of the 2D class averages was chosen (12 nm diameter and height of 15 nm) to avoid starting model bias. To assess conformational heterogeneity 3D classification was carried out. The classes with the best-resolved features were combined, which resulted in 5,137 particles and 32,498 particles for KaiCB and KaiCBA, respectively. From the remaining particles resolutions of 11.5 Å for the KaiCB and 7.2 Å for the KaiCBA complex could be obtained. The map was subjected to rotational correlation analysis[125], yielding pronounced peaks at multiples of 60° indicating six-fold symmetry. Applying 6-fold symmetry improved the resolution to 7.8 Å for the KaiCB and to 5.1 Å for the KaiCBA complex.

The aligned particles were used for the "particle polishing" method within RELION[126]. The particle motion trajectories were estimated and fitted using a running average window of 3 frames. Individual frame-reconstructions were generated and per-frame B-factors were estimated. Based on the quality of the individual reconstructions and their respective B-factors, frames 2-24 were kept and weighted accordingly in the sums. For the KaiCB and KaiCBA complexes these polished particles further improved the resolution to a global resolution of 7 and 4.7 Å, respectively. The final maps were sharpened with a B-factor of -205 $Å^2$ and -245 $Å^2$ for the KaiCB and KaiCBA complexes, respectively. Local resolutions of the complexes were computed using Bsoft[127]. To prevent over-interpretation in the model building, the map was filtered accordingly in the analyzed areas.

4.3.8 Modeling

An atomic model was built for the KaiCBA complex. The crystal structures of KaiC in different functional states unambiguously fitted one handedness of the density. Our density of the CI ATPase of KaiC fits best to the crystal structure of the KaiC CI domain in the presence of ADP (PDB 4TLA ChainC). For the CII Kinase domain the crystal structure of the full length KaiC (PDB 3dvl) was used as a model. The KaiA C-terminal domain was taken from the full-length KaiA structure (PDB 1R8J). We generated an atomic model of KaiB in the thioredoxin-like fold using SasA (PDB 1T4Y) as a template in MODELLER[128,129]. All structures were pruned of flexible tails and initially rigid body fitted using UCSF Chimera[130]. To correct for clashes the model was subjected to geometry minimization using the PHENIX software package[131], as previously described[132]. After minimization the cross-resolution between the model and the reconstruction was calculated to be 4.9 Å. All the 3D structures were displayed using UCSF Chimera[130].

4.3.9 Data deposition:

The 4.7-Å cryo-EM KaiCBA structure and the 7-Å cryo-EM KaiCB structure are available from the EMDB with accession codes EMD-XXXX and EMD-XXX, respectively. The rigid body fitted KaiCBA model generated in this study has been deposited in the PDB with the accession code XXXX.

4.4 Results and Discussion

The standard *in vitro* Kai oscillator consists of a 2:2:1 molar ratio of KaiC:KaiB:KaiA in the presence of excess MgATP, incubated at 30 °C[87]. We tracked the phosphorylation-dependent assembly of the Kai proteins under these conditions and used native MS to determine the masses and stoichiometries of the formed noncovalent assemblies[133]. For the *in vitro* Kai oscillator, we simultaneously detected multiple co-occurring Kai-protein complexes, revealing more than 10 different Kai protein–assembly stoichiometries over the course of 24 hours (Fig. 4.1A and Table A.1). The KaiC starting material had low amounts of phosphorylation. Upon initial mixing, most KaiC therefore existed as a free hexamer, whereas a small fraction formed a complex with one or two KaiA dimers (Fig. A.1A). These KaiCA complexes have autophosphorylation activity, which led to cooperative formation of phosphorylated $KaiC_6B_6$ complexes through a $KaiC_6B_1$ intermediate[110]. In our samples, formation of KaiCA and KaiCB complexes peaked at 4 to 8 hours incubation time. The formation of higher-order KaiCBA complexes followed the formation of KaiCB complexes, with a maximum at 8 to 12 hours of incubation followed by a steady decline toward 24 hours. We observed $KaiC_6B_6$ with between one and six KaiA dimers bound. Detailed assignments of peaks and repeated measurements are shown in Fig. 4.1B and Figs. A.1 to A.3. During the dephosphory-

lation phase (16 to 24 hours), as KaiCBA complexes disassemble, we detected $KaiA_2B_1$ complexes in the lower mass region of the spectra (Fig. A.1B). Thus, the disassembly pathway of the KaiCBA complexes appears not to be simply the reverse of the assembly pathway but rather a distinct route.

We attempted to freeze Kai-protein assembly in specific states, producing particles amenable to more detailed structural characterization. Whereas at 30 °C the default activity of KaiC is autodephosphorylation, autophosphorylation is favored at 4 °C[104,109]. Therefore, we tested how a lower incubation temperature affected assembly of the complete *in vitro* oscillator. At 4 °C, KaiCBA complex formation was slower than at 30 °C. However, KaiCBA abundance steadily increased, and, even after 24 hours, it did not peak (Fig. 4.1A). This indicated a possible route for preparation of KaiCBA complexes with full occupancy of the KaiA-binding site. Therefore, we incubated KaiC, KaiB, and KaiA at a 1:3:3 molar ratio at 4 °C for one week in the presence of MgATP. We observed near-complete occupancy of the KaiA-binding site, as seen from the predominant formation of $KaiC_6B_6A_{12}$ assemblies (Fig. 4.1C). The measured mass of this complex was 823.3 ± 0.5 (standard deviation) kDa, compared to a theoretical mass of 821.3 kDa for $KaiC_6B_6A_{12}$ (Table A.1). Similarly, prolonged incubation of KaiC with KaiB at 4 °C resulted in the efficient formation of $KaiC_6B_6$ complexes (measured: 426.9 ± 0.1 kDa; theoretical: 426.4 kDa; Table A.1). Further experiments revealed that formation of the KaiCB complex is the limiting step for the complete assembly of KaiCBA (Fig. A.1C). Using the protocol described above, we obtained near-homogeneous $KaiC_6B_6A_{12}$ and $KaiC_6B_6$ assemblies, which we further structurally characterized by single-particle cryo-EM (Fig. 4.2). Preferred orientations of particles in ice limited the overall resolution of the KaiCB reconstruction to 7 Å, whereas the KaiCBA map was resolved to 4.7 Å (Figs. A.4 and A.5). Superposition of the KaiCB and KaiCBA maps indicated that the KaiCB subcomplex remained essentially invariant in the KaiCBA complex (Fig. A.4). Both maps clearly show that KaiB binds to the KaiC-CI domain, resolving a controversy in the field[106,110–112,133]. This architecture was further confirmed by cross-linking MS experiments (Fig. A.6 and Table A.2). For a molecular interpretation of the cryo-EM maps, we fitted available atomic models of their constituents (Fig. 4.2C).

The KaiCB structure is composed of three stacked rings (Fig. 4.2). Fitting of the homohexameric KaiC crystal structure[88] showed that the bottom two rings correspond to KaiC, and the upper ring is accordingly assigned to KaiB. A comparison of the map to the various nucleotide bound states of the KaiC-CI domain indicated that this domain is in an ADP–bound state (Fig. A.7)[134]. The nucleotide binding sites at the CI domain showed an unaccounted for density, which was hence assigned to a bound nucleotide (Fig. 4.2D). The KaiB subunits are arranged in a six-fold symmetrical ring, stacked on the lids of the small KaiC-CI–ATPase subdomains (Fig. A.8). Isolated KaiB of the cyanobacterium *Synechococcus elongatus* PCC 7942 exists in two different folds[135]. One fold, seen only in KaiB to date, has been observed in protein crys-

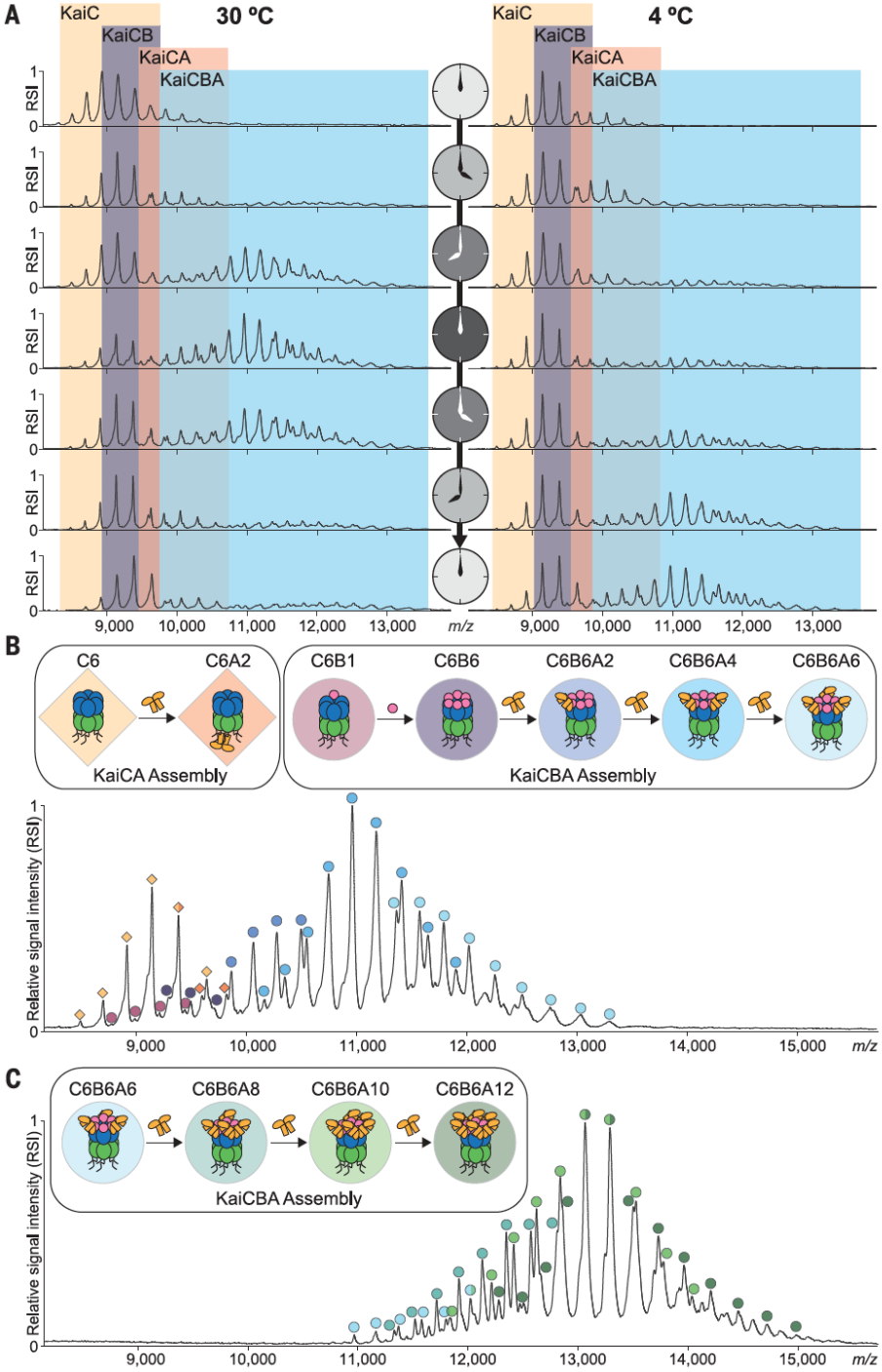


Figure 4.1: Monitoring KaiCBA assembly dynamics by native MS. (**A**) Native mass spectra of the in vitro oscillator at 30 ° or 4 °C, as indicated. The relative signal intensity (RSI) is plotted against the mass-to-charge ratio (*m*/*z*). Areas of the spectra corresponding to KaiC, KaiCA, KaiCB, and KaiCBA are indicated. (**B** and **C**) Enlarged mass spectra with full peak annotation. The identified Kai complexes are schematically represented above the spectra (KaiC-CI, green; KaiC-CII, blue; KaiA, yellow; KaiB, pink). The complexes are highlighted with differently colored circles and diamonds that match the symbols used to label the mass spectrum. A detailed explanation of the peak assignment is provided in Fig. A.2. An overview of all mass assignments is given in Table A.1. (**B**) Mass spectra of oscillator at 30 °C after 12 hours of incubation. (**C**) Mixture of KaiCBA containing excess KaiA and KaiB incubated for 1 week at 4 °C. These Kai complexes have near-complete occupancy of the KaiA-binding site.

tals[90,107,136,137]. NMR spectroscopy experiments suggested that KaiB switches from the fold observed in crystal structures to a thioredoxin-like fold upon binding to KaiC[135]. The cryo-EM map of KaiCBA confirmed that KaiC-bound KaiB adopts the thioredoxin-like fold (Fig. A.9). The observed KaiC-KaiB interface is further supported by hydrogen-deuterium exchange (HDX)–MS experiments (Fig. A.8, Table A.3, and data files S1 to S3). The KaiCBA model predicted that KaiC-Ala[108] is an essential part of the KaiC-KaiB interface. Indeed, by native MS we observed loss of binding upon mutation of Ala[108] (Fig. A.10). The position of individual KaiB subunits in the KaiCBA model also suggests possible KaiB-KaiB contacts that could promote cooperativity of KaiB binding to KaiC (Fig. A.11). The KaiA protein from *Synechococcus elongatus* PCC 7942 is composed of an N-terminal pseudoreceiver (PsR) domain and a C-terminal α-helical domain that takes part in homodimerization[89]. KaiA dimerization is consistent with the $KaiC_6B_6A_{12}$ stoichiometry determined for the fully assembled complex. Fitting of the KaiA dimer structure into the KaiCBA map (Fig. 4.2) yields excellent colocalization of secondary-structure elements in the map and the model for the C-terminal dimerization domain and is further supported by HDX-MS experiments (Fig. A.12). KaiB binds to KaiA most prominently with its β_2 strand, which is present in both KaiB folds and comprises the evolutionarily most-conserved residues of the protein (Fig. A.13). The KaiCBA model predicts that KaiB-Lys[42] is important for the KaiB-KaiA interaction, which was confirmed by loss of KaiA binding upon mutation of KaiB at this site, as observed in native MS experiments (Fig. A.10). The presence of KaiB-Lys[42] is also important for the *in vivo* clock in *Synechococcus elongatus* PCC 7942 and *Thermosynechococcus*, as *kaiBC* and *psbA* promoter activities become arrhythmic upon mutation of this site[136].

For every KaiA dimer in the KaiCBA model, only one monomer is in contact with KaiB. The HDX-MS data also showed signs of asymmetric binding, confirming that the two KaiA protomers in the dimer are distinct in the KaiCBA complex (Fig. A.14). The density assigned to KaiA in the KaiCBA map does not cover most of the N-terminal PsR domain, indicative of the domain's structural flexibility. Positioning of the PsR domains according to the fitted domain-swapped crystal structure also results in extensive clashes with KaiB (Fig. A.15). HDX-MS experiments do not indicate that the PsR domain unfolds or becomes disordered (Table A.3 and data file S1). We therefore suspect that the PsR domain is still folded, but attached with a flexible linker, which would explain the lack of density in the cryo-EM map of the KaiCBA complex. We did observe a small, unassigned KaiA-density segment near the cleft between the homodimeric C-terminal domains. We tentatively assigned this segment to residues 147 to 172, which form the small cross–β sheet and the N-terminal α5 helix in the KaiA crystal structure (Fig. A.15). Binding of KaiB to the linker region of KaiA therefore appears to dissociate the two strands in the dimer, resulting in a large displacement of the PsR domain. The α5 helix of KaiA likely occludes the site to which the flexible C-termini

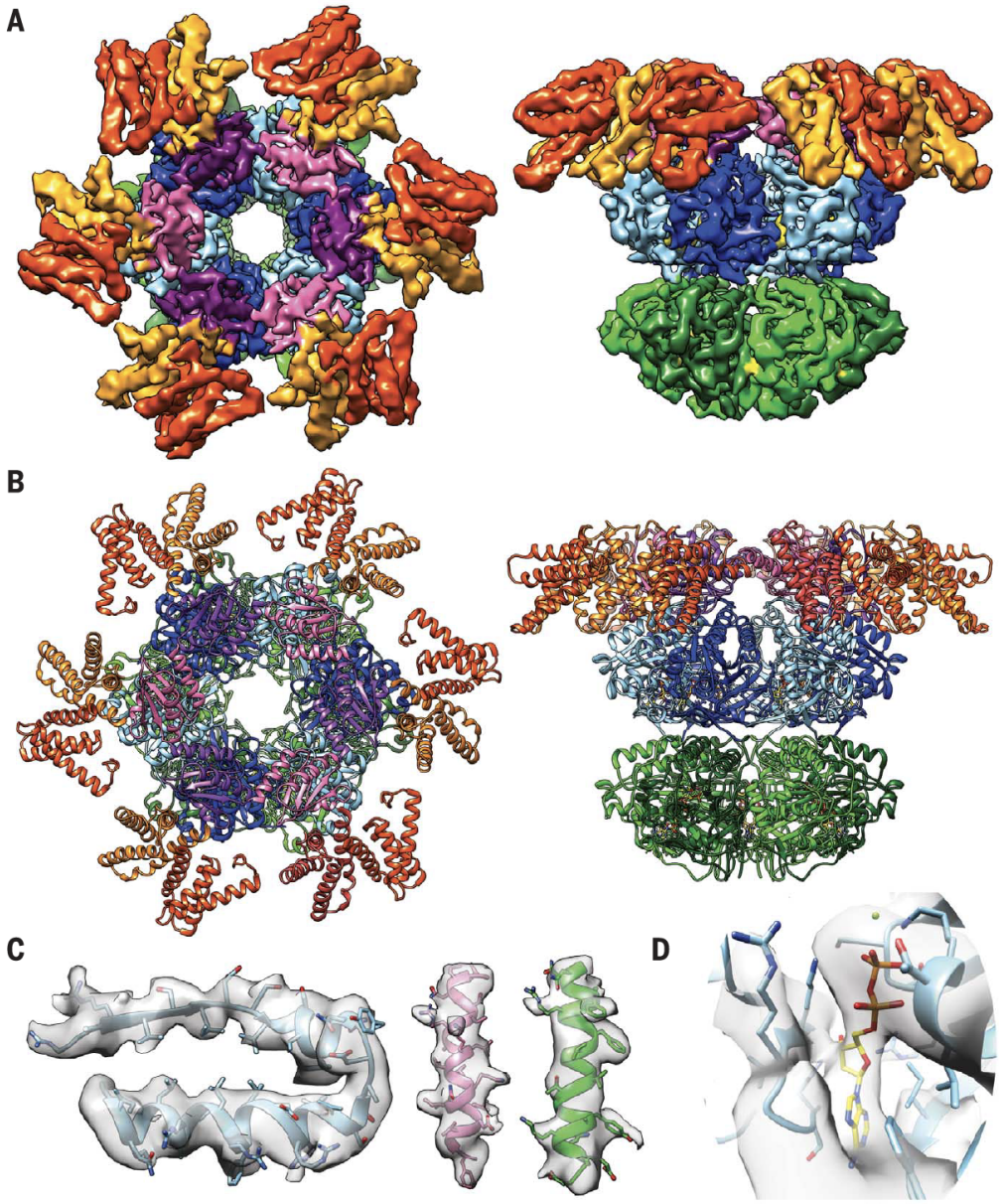


Figure 4.2: Cryo-EM map and pseudoatomic model of the KaiCBA complex. (**A**) Top and side view of the three-dimensional (3D) reconstruction of the KaiCBA complex. The CII and CI domains of KaiC are colored in dark green and green and in blue and light blue, respectively. The segmented density corresponding to KaiB is colored alternating in pink and purple, and the individual KaiA homodimers are colored orange and orange-red. (**B**) Top and side view of the model of the KaiCBA complex. Colors are the same as in (**A**). (**C**) Selected examples of the quality of the map. (**D**) Density in the nucleotide binding pocket of the KaiC-CI domain superimposed with the nucleotide bound in the KaiC crystal structure bound to ADP (Protein Data Bank 4TLA chain C).

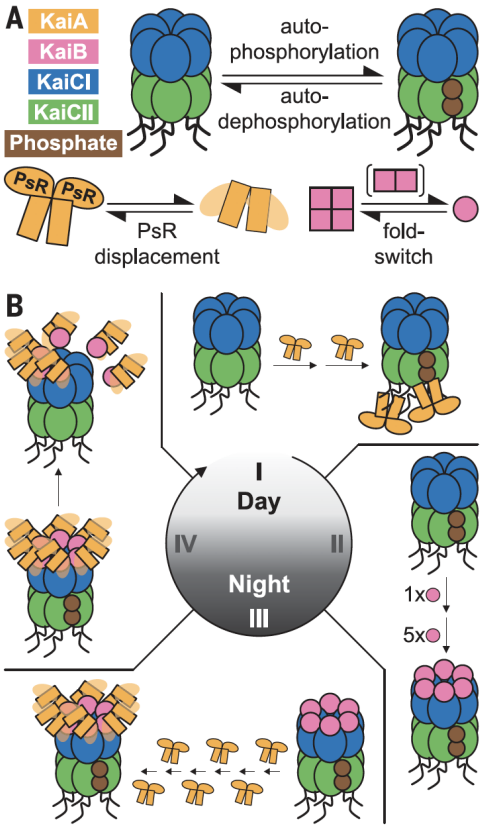


Figure 4.3: The structural basis of periodic assembly in the cyanobacterial circadian clock. (**A**) Structural transitions of the individual Kai proteins during the circadian cycle. (**B**) Molecular changes in the KaiCBA oscillator. Stepwise binding of two KaiA dimers triggers KaiC autophosphorylation at Thr^{432} and Ser^{431} (I). These phosphorylation events enable cooperative binding of fold-switched KaiB monomers to the KaiC-CI domain, forming the KaiCB complex (II). KaiCB provides a scaffold for the successive sequestration of KaiA in ternary KaiCBA assemblies, concurring with a rearrangement of the KaiA PsR domains (III). KaiA sequestration promotes KaiC autodephosphorylation, resulting in the regeneration of free KaiC through release of KaiBA subcomplexes (IV).

of KaiC bind[138]. On the basis of these structures and our native MS data, we propose a detailed model for the cyclic phosphorylation-dependent assembly of Kai components in the *in vitro* oscillator (Fig. 4.3). Upon mixing the protein components of the *in vitro* oscillator, unphosphorylated KaiC hexamers bind one or two copies of a KaiA dimer on the C terminus of the KaiC-CII domain[38]. Binding of the second KaiA dimer stimulates autophosphorylation of KaiC, first at Thr^{432} and then at Ser^{431}[104]. Serine phosphorylation triggers binding of KaiB in a fold-switched state. Six copies of KaiB bind cooperatively[110] to form phosphorylated $KaiC_6B_6$ complexes. The bound KaiB subunits present alternative binding sites for KaiA, away from a phosphorylation-stimulating interaction with the KaiC-CII domain. The KaiA dimer binds asymmetrically through its linker region to KaiB, resulting in a wide displacement of the PsR domain. As the pool of free KaiA dimers is depleted, KaiC switches back to autodephosphorylation activity. Complete dephosphorylation of KaiC results in dissociation of the KaiCBA complex by loss of $KaiA_2B_1$ subcomplexes, thereby completing one cycle of the oscillator. In cyanobacterial cells, KaiC and KaiB are

produced from the same operon and in 10- to 100-fold excess to KaiA[39]. The high excess of KaiCB over free KaiA could promote efficient sequestration of KaiA *in vivo*. The model presented here can thus serve as a framework to better understand the circadian clock in cyanobacterial cells.

5

Minimal tool set for a prokaryotic circadian clock

Nicolas M. Schmelling, Robert Lehmann, Paushali Chaudhury, Christian Beck, Sonja-Verena Albers, Ilka M. Axmann, and Anika Wiegard (2017) Minimal tool set for a prokaryotic circadian clock, *BMC Evolutionary Biology*, 17:169, doi: *10.1186/s12862-017-0999-7*

Author Contributions

IMA and AW supervised the study. NMS, CB, and RL designed and NMS and RL carried out bioinformatic data analyses. AW designed and performed sequence analysis and *in vitro* experiments. AW and PC purified recombinant proteins. NMS, RL, CB, AW, and IMA prepared the manuscript. PC and SVA contributed to the interpretation of the data and provided intellectual input. All authors read and approved the final manuscript.

Pro rata

Bioinformatic study design: 40%

Bioinformatic data analyses: 50%

Manuscript preparation: 40%

Summary of Publication

This chapter contains the collaborative work with Robert Lehmann, Paushali Chaudhury, Christian Beck, Sonja-Verena Albers, Ilka M. Axmann, and Anika Wiegard. Cyanobacteria are known for their large diversity and previously different timing systems have been identified in Cyanobacteria with shared protein factors [21,42,79]. This motivated us to analyze the distribution of protein factors of the circadian clock system in *Synechococcus elongatus* PCC 7942 as well as some factors of *Synechocystis* sp. PCC 6803 in an in-depth search for orthologs in other prokaryotes as well as eukaryotes. We further compared microarray data of several cyanobacterial strains in order to identify common gene expression patterns. Last, we biochemically characterized the kinase activity of some of the KaiC orthologs in Cyanobacteria and Archaea. The microarray analyses were carried out by Robert Lehmann and the biochemical characterization of KaiC homologs was carried by Anika Wiegard. I could find in a comparative genomics approach that protein factors involved in the circadian clock of Cyanobacteria are present in a large variety of prokaryotic organisms, however, no eukaryotic genome could be identified to contain KaiC in combination with other factors of the circadian clock, which was a central criterion in our analysis. I could further identify three classes of the circadian clock system in cyanobacteria, which were identified by co-occurrence analyses of the cyanobacterial circadian clock factors.

The results of this work, a comparative genomic analysis as well as co-occurrence analyses of cyanobacterial circadian clock factors, were published in *BMC Evolutionary Biology* and can be found on doi: *10.1186/s12862-017-0999-7*.

5.1 Abstract

Circadian clocks are found in organisms of almost all domains including photosynthetic Cyanobacteria, whereby large diversity exists within the protein components involved. In the model cyanobacterium *Synechococcus elongatus* PCC 7942 circadian rhythms are driven by a unique KaiABC protein clock, which is embedded in a network of input and output factors. Homologous proteins to the KaiABC clock have been observed in Bacteria and Archaea, where evidence for circadian behavior in these domains is accumulating. However, interaction and function of non-cyanobacterial Kai-proteins as well as homologous input and output components remain mainly unclear.

Using a universal BLAST analyses, we identified putative KaiC-based timing systems in organisms outside as well as variations within Cyanobacteria. A systematic analyses of publicly available microarray data elucidated interesting variations in circadian gene expression between different cyanobacterial strains, which might be correlated to the diversity of genome encoded clock components. Based on statistical analyses of co-occurrences of the clock components homologous to *Synechococcus elongatus* PCC 7942, we propose putative networks of reduced and fully functional clock systems. Further, we studied KaiC sequence conservation to determine functionally important regions of diverged KaiC homologs. Biochemical characterization of exemplary cyanobacterial KaiC proteins as well as homologs from two thermophilic Archaea demonstrated that kinase activity is always present. However, a KaiA-mediated phosphorylation is only detectable in KaiC1 orthologs.

Our analysis of 11,264 genomes clearly demonstrates that components of the *Synechococcus elongatus* PCC 7942 circadian clock are present in Bacteria and Archaea. However, all components are less abundant in other organisms than Cyanobacteria and KaiA, Pex, LdpA, and CdpA are only present in the latter. Thus, only reduced KaiBC-based or even simpler, solely KaiC-based timing systems might exist outside of the cyanobacterial phylum, which might be capable of driving diurnal oscillations.

5.2 Background

Life on Earth is under the influence of changing environmental conditions, which not only pose a challenge to organisms, but also present a chance of adaptation and therefore a possible fitness advantage over competitors[6,7]. Using inner timing systems organisms can coordinate their physiology and behavior according to the daily recurring changes. Simple timing systems work in an hour glass like fashion and need to be reset every day by environmental stimuli, whereas true circadian clocks generate self-sustained and temperature-compensated 24-hour rhythms of biological activities[139,140].

Circadian clocks are found in many eukaryotes such as algae, plants and mammals[16]. Even though circadian

clocks seem like a conserved trait in evolution, differences in the protein components, involved in circadian timing, suggest a convergent evolution of timing mechanisms[16]. For many years it has been believed that something as complex as a circadian clock could not have been evolved in unicellular organisms like prokaryotes[16,141,142]. However, the existence of temperature compensated 24-hour rhythms of cell division in *Synechococcus* sp. WH 7803 and circadian nitrogen fixation in *Cyanothece* sp. PCC 8801 proved otherwise[143–147]. The molecular basis of the cyanobacterial circadian clock was intensively investigated in *Synechococcus elongatus* PCC 7942, where the core clockwork resembles a posttranslational oscillator[87,148,149]. In contrast, eukaryotic circadian rhythms are believed to be mainly based on transcriptional-translational feedback loops. However, findings on post-translational systems are accumulating[99,150,151] and might exist also in Archaea[150].

Light is assumed to be the driving stimulus in circadian clock entrainment[152]. In *Synechococcus elongatus* PCC 7942, contrary to eukaryotic circadian clock systems, a photoreceptor in the input pathway of the clock could not be detected thus far. Instead, *Synechococcus elongatus* PCC 7942 cells sense light indirectly through the redox and energy state of the cell[28]. Here, two metabolic components are considered to play a major role[29]: The ATP/ADP ratio and the redox state of the plastoquinone pool[48,153]. The core of the circadian clock in *Synechococcus elongatus* PCC 7942 consists of three proteins KaiA, KaiB and KaiC. KaiC monomers are composed of two domains, which assemble into two hexameric rings[34,154]. The C-terminal ring is capable of autophosphorylation and –dephosphorylation[35,36]. KaiC phosphorylation is stimulated by the interaction with KaiA[37,38], and additionally, affected by the ATP/ADP ratio of the cell[41]. KaiB inhibits the activating effect of KaiA and initializes dephosphorylation[39]. Altogether, KaiC hexamers phosphorylate and dephosphorylate rhythmically during the course of a day. Two very recent studies provide a structural basis for the dynamic assembly of clock proteins by using crystallography and NMR[62], and mass spectrometry and cryo-electron microscopy of the native proteins[32]. The binding of oxidized quinones to KaiA has been suggested to stop the clock directly by causing KaiA aggregation[46,47]. The KaiABC core clock is embedded into a network of input and output pathways. The input factors that interact with the core clock are Pex, LdpA, PrkE, NhtA, IrcA, CdpA[28,43,44,63,64,155]. Output factors are SasA, LabA, LalA, CpmA, Crm, RpaA, and RpaB, as well as CikA, which is functioning both in input and output pathway of the circadian clock[30,45,54–61,65–67].

Sequence analysis indicated that at least three different types of timing systems are present in Cyanobacteria, (i) a KaiABC-based system as in *Synechococcus elongatus* PCC 7942, (ii) a reduced system with a KaiBC core and a reduced set of input/output factors as in *Prochlorococcus* and (iii) a reduced KaiABC system as in *Synechococcus* sp. WH 8102, which despite including all three *kai* genes, has the same input/output factors as the reduced KaiBC system[156]. Furthermore, multiple *kai* genes can exist in an organism[42,79]. In *Syne-*

chocystis sp. PCC 6803 for example, three copies of both *kaiB* and *kaiC* are found. KaiA, KaiB1 and KaiC1, most similar to the *kai* genes of *Synechococcus elongatus* PCC 7942 [33,81], are thought to resemble the core clock. KaiB3 and KaiC3 are thought to function as fine tuning factors for the circadian clock in *Synechocystis* sp. PCC 6803, whereas no circadian function has been found for KaiB2 and KaiC2 [81]. However, recently *kaiC2*-dependent adaptive growth and diurnal rhythms of nitrogen fixation were observed in *Rhodopseudomonas palustris* [157]. Homologs of *kaiB* and *kaiC* genes also exist in other Bacteria and even Archaea, where a shortend KaiC most often resembles only one domain [79]. An archaeal one-domain KaiC homolog was shown to form a hexameric ring, similar to the two duplicated domains of *Synechococcus elongatus* PCC 7942 KaiC [158]. In *Haloferax volcanii* the transcripts of four *kaiC* homologs display diurnal accumulation profiles [159]. However, the function of non-cyanobacterial Kai-proteins is mainly unclear, so far. Although some of the input- and output factors were also found in prokaryotes other than Cyanobacteria [45,156,160,161], it is unknown whether the Kai homologs outside the cyanobacterial phylum or the additional cyanobacterial Kai homologs interact with (other) in- and output factors.

In this study, we performed BLAST analyses to first identify possible KaiC-based timing systems in organisms outside of Cyanobacteria and second to explore variations in circadian clocks within Cyanobacteria. Further, we examined variations in circadian gene expression between different cyanobacterial strains using microarray data. Together, this aims at decoding the correlation between Kai proteins and additional clock components. Based on the co-occurrence of clock components known from *Synechococcus elongatus* PCC 7942 we propose putative networks of reduced and fully functional clock systems. Further, we used the sequence information of KaiC and its homologs in Cyanobacteria to determine the similarities at important sites of the protein. We chose cyanobacterial KaiC proteins as well as homologs from two thermophilic Archaea and demonstrated that kinase activity is always present. However, a KaiA-mediated phosphorylation is only detectable in cyanobacterial KaiC1 homologs.

5.3 Methods

5.3.1 Programming languages

The programming languages Python (version 3.5.1) and R (version 3.2.3) were used in this work. The processing and analysis of the microarray time series datasets was performed using R. Regarding the distribution analysis, the Biopython project ([162]; version 1.66) was used to download from GenBank as well as to work with FASTA files. Besides Biopython, the Python packages: IPython ([163] version 4.1.1) as an interactive Python environment with the IPython notebook; numpy and scipy ([164]; version 1.10.4, version 0.17.0) for numerical operations; matplotlib ([165]; version 1.5.1) for data visualization; and pandas ([166]; ver-

sion 0.17.1) for data analyses were used. The Python code necessary to reproduce the BLAST analyses is available on GitHub (http://doi.org/10.5281/zenodo.229910)[91].

5.3.2 Reciprocal BLAST and NCBI

The coding sequences of all entries in the genbank protein database[114], which were labeled as "Complete Genome" or "Chromosome", were downloaded from the NCBI FTP server (version May 2016). These sequences, including the coding sequences of *Synechococcus elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803, were used to construct a custom protein database for the homology search. Further, protein sequences of the 23 clock related proteins (Table B.1), from *Synechococcus elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803, respectively, were checked against the entries in the Cyanobase Database to ensure correctness[113] (version May 2016). These 23 protein sequences were used as queries for a search of homologs within the custom protein database, applying the standalone version of BLASTP 2.2.30+[83] (May 2016, dx.doi.org/10.17504/protocols.io.grnbv5e) with standard parameter (wordsize: 3, substitution matrix: BLOSUM62). The 10,000 best hits with an e-value of 10^{-5} or lower were filtered for further analyses. The first BLAST run returned circa 65,000 hits for all 23 cyanobacterial proteins combined.
These hits were used as queries for a second reverse BLASTP run, searching for homologs in *Synechococcus elongatus* PCC 7942 or *Synechocystis* sp. PCC 6803 genomes using the same parameters as above with an altered e-value of 10. Only hits with the original query protein as best reversal hit were accepted for further analyses, thus minimizing false positive results.
Raw and processed data is available on figshare (https://dx.doi.org/10.6084/m9.figshare.3823902.v3, https://dx.doi.org/10.6084/m9.figshare.3823899.v3)[167,168].

5.3.3 Testing of co-occurence

Co-occurrence of circadian clock proteins was examined by using the right-sided Fisher's exact test[169]. For each of the 94 cyanobacterial strains, all identified homologous clock genes were gathered into one set. The phylogenetic distribution of Cyanobacteria in the NCBI genbank database is very imbalanced. Some genera (e.g. *Prochlorococcus* and *Synechococcus*) are covered better than others. To avoid selection bias, we removed sets with identical combinations of genes, resulting in 69 unique clock systems. Null hypothesis of Fisher's exact test is a pairwise independent distribution of the proteins across all clock systems. P-values were corrected for multiple testing after Benjamini-Hochberg[170] with an excepted false discovery rate of 10^{-2}. We denote that due to the nature of statistical testing, proteins appearing in almost all clock systems are always virtually independent to others. All proteins were clustered according to their corrected p-values.

5.3.4 Microarray analysis

The diurnal expression program of six cyanobacterial strains was probed using microarray time series datasets (Table B.3). Unfortunately, the data for the two reported *Synechocystis* sp. PCC 6803 experiments by Labiosa and colleagues[171] and Kucho and coworkers[172] could not be obtained. The study by Toepel and coworkers[173] had to be discarded due to the employed ultradian light cycles (6:6 LD cycles). The *Synechocystis* sp. PCC 6803 datasets, the *Synechococcus elongatus* PCC 7942 dataset of Ito and colleagues, and the *Anabaena* sp. PCC 7120 dataset were l2m transformed, while the *Cyanothece* ATCC 51142 datasets are only available after transformation. The two biological replicates of the *Anabaena* sp. PCC 7120 dataset were concatenated for the following analyses[75], similar to the *Synechocystis* sp. PCC 6803 dataset[70]. Expression profiles were smoothed using a Savitzky-Golay lowpass filter, as proposed by Yang and colleagues[174], in order to remove pseudo peaks prior to the detection of periodic genes. Diurnally oscillating expression profiles were detected using harmonic regression analysis[175]. The derived p-values for each gene is based on the assumption of a linear background profile as compared to the sinoidal foreground model. After multiple hypothesis testing correction according to Benjamini-Hochberg[170], all datasets yielded significantly oscillating genes ($q \leq 0.05$) except for *Microcystis aeruginosa* PCC 7806, with 7 samples the shortest dataset.

5.3.5 Multiple sequence alignment

Multiple sequence alignments were constructed by the standalone version of CLUSTAL Omega 1.2.1-1[85] using 20 iterations, while only one iteration was used to construct the guide tree (May 2016, dx.doi.org/10.17504/protocols.io.gscbwaw). The sequences for the alignments were obtained from the processed data generated, as described in the method section 5.3.2. Afterwards the alignments were adjusted to *Synechococcus elongatus* PCC 7942 sequence with Jalview[115] and edited multialignments were used to create WebLogos[116].

5.3.6 Cloning, heterologous expression and purification of Kai proteins

To express KaiC1 proteins from *Synechocystis* sp. PCC 6714, *Nostoc punctiforme* ATCC 29133, *Cyanothece* sp. PCC 7424 as well as KaiC3 from *Cyanothece* sp. PCC 7424, *Microcystis aeruginosa* PCC 7806, *Pycrococcus horikoshii* OT3 PH0833, and *Thermococcus litoralis* DSM5473, the respective *kaiC* genes were amplified by PCR from genomic wildtype DNA. The ORFs and primers are listed in Table B.4. Amplified sequences were ligated into BamHI and NotI restriction sites of the plasmid pGEX-6P-1 (GE Healthcare). For *P.horikoshii* and *T.litoralis* KaiC3 the amplified PCR products were ligated into pETDuet1 Vector us-

ing BamHI/HindIII and PstI/HindIII restriction enzymes in MCS1, respectively. *Escherichia coli* DH5α or BL21 (DE3) cells were transformed with the resulting plasmids (pGEX-kaiC1-*Sy*6714, pGEX-kaiC1-*Npun*29133, pGEX-kaiC1-*Cy*7424, pGEX-kaiC3-*Cy*7424, pGEX-kaiC3-*Mic*7806, pSVA3151, pSVA3152). For expression of GST-fused KaiA-7942 and KaiC-7942 pGEX derivatives, kindly provided by T. Kondo (Nagoya University, Japan), were used. Expression of GST-Kai proteins occurred at 37 °C and 200 rpm in Terrific broth medium containing 100 μg ampicillin ml^{-1}. GST-KaiA-7942 expression was induced with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) and carried out overnight, whereas GST-KaiC homologs were expressed for 72 hours without induction. Cells were harvested and lysed in ice-cold extraction buffer (50 mM Tris/HCl (pH 8), 150 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 5 mM $MgCl_2$ and 1 mM ATP). Recombinant GST-Kai proteins were affinity purified using Protino Gluthatione Agarose 4B (Macherey–Nagel) as described in Wiegard and colleagues[33]. During the procedure the GST-tag was removed with PreScission protease in cleavage buffer (50 mM Tris/HCl (pH 8), 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 5 mM $MgCl_2$ and 0.5 mM ATP). Homogeneity of the recombinant proteins was controlled by separating them via SDS-PAGE. If it was not sufficient, proteins were further purified by anion-exchange chromatography using a MonoQ 5/50 GL or ResourceQ column (GE Healthcare). After dialysis in reaction buffer (20 mM Tris/HCl (pH 8), 150 mM NaCl, 0.5 mM EDTA, 5 mM $MgCl_2$, and 1 mM ATP) protein concentration was determined using infrared spectroscopy (Direct detect, Merck Millipore). Proteins were stored at -20 °C. For His-tagged KaiC3 homologs, *Escherichia coli* BL21 (DE3) RIL cells were transformed with pSVA3151 (*P.horikoshii*) and pSVA3152 (*T.litoralis*), and grown as preculture overnight at 37 °C in LB medium containing ampicillin (50 μg ml^{-1}) and chloramphenicol (34 μg ml^{-1}). Fresh medium containing antibiotic was inoculated with 0.1% preculture and grown at 37 °C to an OD_{600} of 0.7. After induction with 0.3 mM of IPTG, growth was continued for 16 hours at 16 °C. Cells were collected by centrifugation, frozen in liquid nitrogen, and, after storage at -80 °C, resuspended in 50 ml lysis buffer (50 mM HEPES-NaOH, pH 7.2, 150 mM NaCl) containing complete EDTA-free protease inhibitor cocktail (Roche) together with DNase I and lysed by sonication. Cell debris were removed by centrifugation at 4 °C for 30 min at 20,000 x g. Further Ni-NTA (Sigma Aldrich, Seelze, Germany) based purification was performed using columns equilibrated in purification buffer (50 mM HEPES-NaOH, pH 7.2, 150 mM NaCl). For the removal of unspecifically bound protein columns were washed with 15 column volumes of equilibration buffer including 10 mM imidazole. *T.litoralis* KaiC3 was eluted in the same buffer with 150 mM imidazole containing equilibration buffer, whereas for *P.horikoshii* KaiC3 the elution was carried out in 20 mM MES pH 6.2, 150 mM NaCl, 150 mM imidazole as this protein is stable in low pH buffer. Further purification of *T.litoralis* KaiC3 was achieved by size exclusion chromatography using Superdex 200 10/300 GL. *P.horikoshii* KaiC3 was incubated at 50 °C for 20 min and centrifuged for

15 min at 10.000 x g. Subsequently, the supernatant was dialyzed overnight against 20 mM MES pH 6.2, 150 mM NaCl buffer. As a quality control, proteins were separated via SDS-PAGE and pure proteins were frozen in liquid nitrogen and kept at -80 °C.

5.3.7 *In vitro* phosphorylation assays

To investigate KaiA dependent phosphate uptake 12 μg of KaiC-7942, KaiC1-*Sy*6714, KaiC1-*N*29133, KaiC1-*Cy*7424, KaiC3-*Cy*7424, KaiC3-*Mic*7806, KaiC3-*T.lit* or KaiC3-*P.hor* were mixed with 10 μCi γ-P^{32}-ATP in 60 μl Tris reaction buffer (20 mM Tris/HCl (pH 8), 150 mM NaCl, 0.5 mM EDTA, 5 mM $MgCl_2$, 1 mM ATP) in the presence or absence of 6 μg KaiA-7942. 10 μl aliquots were taken after 0, 0.75, 1.5, 3 and 22 hours of incubation at 30 °C and reaction was stopped by adding SDS-sample buffer. Proteins were separated in high-resolution polyacrylamide gels (10% T, 0.67% C) by SDS-PAGE (modified from[176]), stained with Coomassie brilliant blue and subjected to autoradiography. Signals were analyzed using a Fujifilm FLA-3000 (FUJIFILM). To analyze *in vitro* phosphorylation of KaiC3-*T.lit* and KaiC3-*P.hor* at higher temperatures, 10 μg of the recombinant proteins were incubated with 10 μCi γ-P^{32}-ATP in 50 μl HEPES reaction buffer (50 mM HEPES (pH 7.2), 150 mM NaCl, 5 mM $MgCl_2$, 1 mM ATP) or 50 μl MES reaction buffer (50 mM MES (pH 6), 150 mM NaCl, 5 mM $MgCl_2$, 1 mM ATP), respectively, at 75 °C. After 0, 5, 10 and 15 minutes 10 μl aliquots were taken and analyzed by SDS-PAGE and autoradiograpy as described above. Comprehensive protocols are available on protocols.io (dx.doi.org/10.17504/protocols.io.g3gbyjw, doi.org/10.17504/protocols.io.gysbxwe).

5.4 Results and Discussion

5.4.1 Most complete set of circadian clock orthologs in Cyanobacteria

Growing evidence hints at circadian rhythms in many Cyanobacteria and first sequence analyses revealed that the core genes, known from *Synechococcus elongatus* PCC 7942, are conserved in almost all cyanobacterial species[33,78,79,140]. Even though daily rhythms seem to be conserved in Cyanobacteria, composition and quantity of corresponding genes on the genome level show high variability[42]. Dvornyk and colleagues first attempted to describe the variety of cyanobacterial core circadian clock systems in 2003[79]. Since then, the amount and depth of sequencing data increased manifold, which allowed us to perform a detailed analysis of the KaiC-based circadian clock including the input and output pathways. The circadian clock proteins of *Synechococcus elongatus* PCC 7942 (Table B.1) and the three diverged KaiB and KaiC homologs from *Synechocystis* sp. PCC 6803 (Table B.1) were the basis for our analysis. Their protein sequences served as basis for a reciprocal best hit BLAST analysis. Organisms with at least one homolog to KaiC were retained

for further analysis. This stringent filter was essential since KaiC represents the core of the circadian clock in *Synechococcus elongatus* PCC 7942. These organisms were grouped by their corresponding genus for a first overview of the homolog distribution (Fig. 5.1A). We found homologs in Cyanobacteria, Proteobacteria, Archaea, as well as other Bacteria such as *Chloroflexi*. This finding is in good agreement with previous studies[33,79,177]. However, our comprehensive study identified a plethora of new bacterial and archaeal genera harboring homologs to the circadian clock genes (Fig. 5.1A). Nevertheless, Cyanobacteria represent the phylum with the highest degree of sequence similarity and integrity of the system followed by Bacteria, mostly Proteobacteria. In Archaea, homologs to only a fraction of core genes could be identified (Fig. 5.1A).

5.4.2 Core circadian clock factors KaiB and KaiC beyond the phylum of Cyanobacteria

Four out of seventeen studied factors are exclusively found in Cyanobacteria (Fig. 5.1A). One of these factors is KaiA, as previously reported in studies with a smaller sample size[79,177]. However, even some Cyanobacteria, like *Candidatus Atelocyanobacterium thalassa* (previously named Cyanobacterium UCYN-A), and all representatives from the genus *Prochlorococcus* lack *kaiA* (Fig. 5.1A)[42]. Interestingly, multiple copies of *kaiA* in a single cyanobacterial genome could not be identified. This is of special interest, because we could observe strong sequence length variations for KaiA (Fig. 5.5A, C) and multiple copies for the other core proteins KaiB and KaiC have been reported (Fig. 5.1B)[33,42,79]. With the *kaiA*-lacking *Candidatus Atelocyanobacterium thalassa* and a *kaiA*-containing cyanobacterial endosymbiont[184] only two out of 94 studied Cyanobacteria do not contain a *kaiB*, whereas a third of the cyanobacterial genera contain multiple copies of *kaiB* and *kaiC* (Fig 5.1B, C). There are a few exceptional cyanobacteria like *Gloeobacter violaceus*[185], which are even lacking the *kaiC* gene and are thus not detected in this analysis due to the previously described filtering criteria. However, all Cyanobacteria having homologs to *kaiB* and *kaiC* are coding for at least one pair of proteins most similar to KaiB1 and KaiC1 from *Synechocystis* sp. PCC 6803 (Fig. 5.1B, C), which is consistent with previous studies[33,79]. Only *Cyanothece* sp. PCC 7822 has homologs similar to all three *kaiB* and *kaiC* copies from *Synechocystis* sp. PCC 6803 (Fig. 5.1C). The majority (67.68%) of the bacterial genera, outside of the cyanobacterial phylum, encode KaiC3-like proteins. Approximately half of them also contain additional KaiC homologs, most similar to KaiC1, KaiC2, or sometimes even both (Fig. 5.1B). This observation does not hold true for KaiB homologs, here KaiB2 is the major KaiB homolog (75.86%). One group of organisms, all belonging to the phylum of *Bacteroidetes*, stands out in this analysis, because they encode only proteins similar to KaiB2/KaiC2 (Fig. 5.1B). Archaeal genera show mainly homologs with highest identity to KaiC1 or KaiC3. Furthermore, almost all of the Archaea have multiple

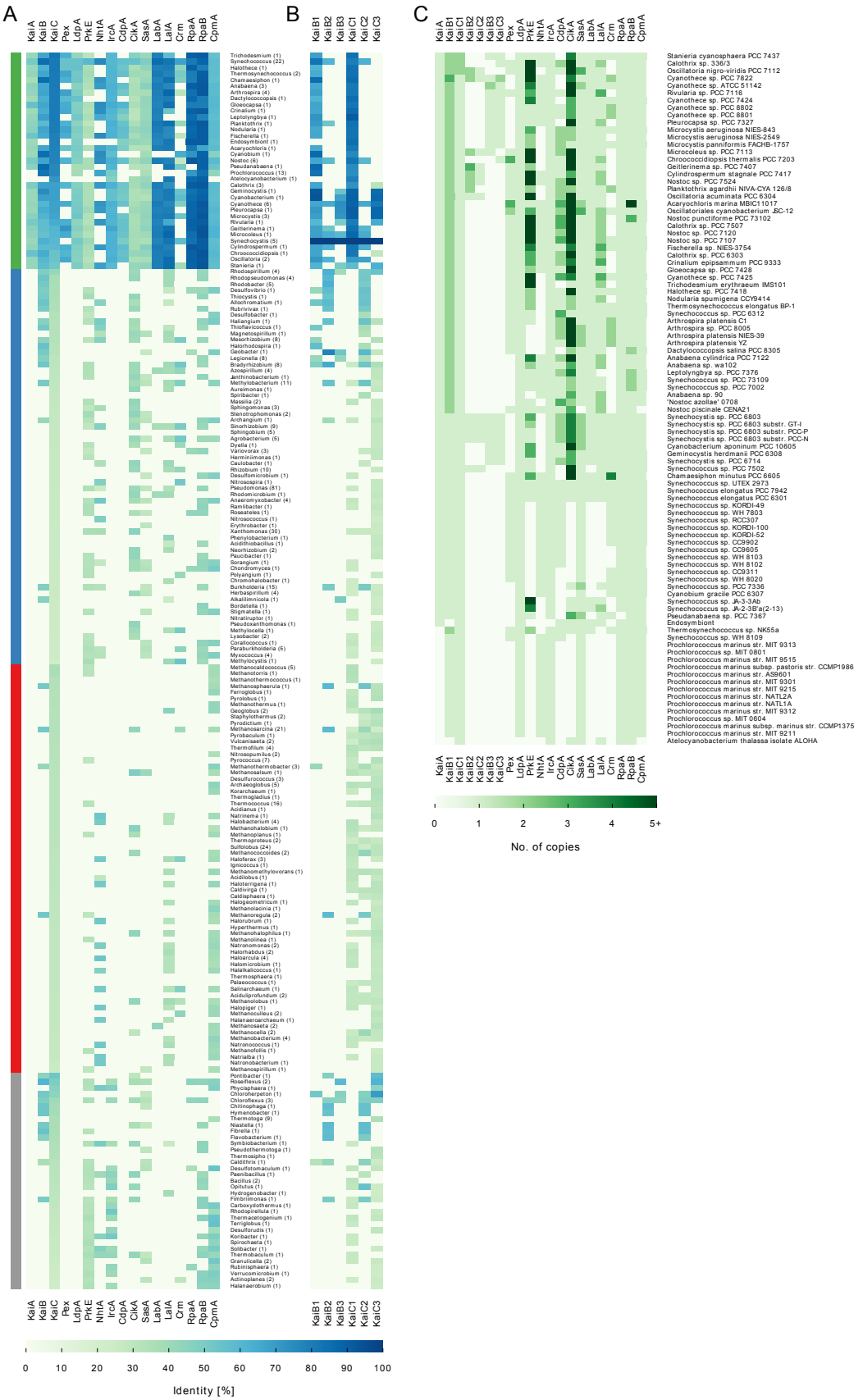


Figure 5.1: (Continued on the following page.)

Figure 5.1: Distribution of circadian clock proteins. (**A**,**B**) Shown are the mean sequence similarites of each protein for each genus that contains a KaiC homolog from *Synechococcus elongatus* PCC 7942. The genera are sorted by their group and KaiC similarity. The number in parenthesis represents the number of individual genomes per genus. (**A**) shows the mean similarities for the circadian clock proteins of *Synechococcus elongatus* PCC 7942. (**B**) shows the mean similarities for the circadian clock proteins of *Synechocystis* sp. PCC 6803. The four taxonomic main groups are highlighted: Cyanobacteria (Green), Proteobacteria (Blue), Archaea (Red), Other (Grey). (**C**) shows the number of homologs for each protein in each cyanobacterium. Copy numbers higher than five were condensed. Cyanobacterial strains with true circadian clocks[72,75,143,172,173,178–183] are highlighted with a yellow hexagon and the strain with hourglass like timing mechanism[77] with a green triangle. We could not find references for all strains, however, we identified strains with circadian rhythms, which are not listed here as their genomes are still not fully assembled.

copies of the KaiC protein. Whereas only four have homologs to KaiB, which is either similar to KaiB2 or in thermophilic *Methanothermobacter* similar to KaiB1 (Fig. 5.1B).

5.4.3 Circadian clock factors solely involved in cyanobacterial input pathways

Multiple clock input factors are present in all four taxonomic groups namely PrkE, NhtA, and CikA. The latter also acts in the output of the circadian clock (Table B.2, Fig. 5.1A)[56,63,64]. The prevalence of CikA is in good agreement with its tremendous effect on resetting the circadian clock in *Synechococcus elongatus* PCC 7942[63], where its interaction with the Kai complex is mediated by additional proteins[64,155]. CikA destabilizes in the presence of oxidized quinones and thereby integrates information about the cellular redox state into the oscillator[64]. Naturally, CikA-lacking *Prochlorococcus* are not able to reset their timing mechanism in continuous light[77], pointing at the importance of this factor. However, single mutations in *sasA* can restore circadian properties in *Synechococcus elongatus* PCC 7942 *cikA* mutants and it has been suggested that a simpler network with modified interactions of the other clock proteins can exist[186].

The input factors Pex, LdpA, and CdpA are found to be unique for Cyanobacteria (Fig. 5.1A), confirming a previous analysis of *ldpA*[187]. Hence, with Pex, a transcriptional repressor of *kaiA*[43], and LdpA, a redox-sensing protein[28] two factors, which sense the cellular metabolic state of the circadian clock in *Synechococcus elongatus* PCC 7942, are missing outside of the cyanobacterial phylum. LdpA is also the only input factor present in the reduced but functional timing system of *Prochlorococcus*[42,78]. This might indicate the necessity of this factor for the entrainment of the clock[160]. Nevertheless the possibility of a functional clock without Pex and LdpA remains, since LdpA and Pex mutants in *Synechococcus elongatus* PCC 7942 are only altered in their period length[44,49] and Pex is also missing in the KaiA-lacking *Prochlorococcus* and the KaiA-containing *Synechocystis* sp. PCC 6803 (Fig. 5.1A). In *Synechococcus elongatus* PCC 7942 the third unique cyanobacterial input factor, CdpA, influences phase resetting and acts in parallel to CikA[155]. Since CdpA seems to be essential in *Synechococcus elongatus* PCC 7942[155], it likely has a prior role

in processes other than phase resetting and is hence dispensable for input pathways in other organisms than Cyanobacteria.

Altogether, the absence of three important input factors outside of the cyanobacterial phylum suggests that other entrainment systems might be used for putative timing systems. In *Rhodobacter sphaeroides*, which displays circadian gene expression rhythms, a histidine kinase is encoded in an operon with *kaiBC* and was suggested as a candidate for transducing the redox signal to KaiBC[188]. Further, the direct entrainment by the ATP/ADP ratio [41,189] might be the primary mechanism to synchronize the circadian clock with metabolism and the environment.

5.4.4 Central output factors are missing in Archaea and non-cyanobacterial genera

The output pathway of the circadian clock in *Synechococcus elongatus* PCC 7942 involves eight proteins (see Fig. 5.7). RpaA serves as a key regulator[30]. Its activity is indirectly modulated depending on the phosphorylation state and the ATPase activity of KaiC[53,56]. SasA (antagonistically to CikA) connects the core clock to RpaA, which in turn regulates global gene expression, including the *kaiBC* promoter[30,55,56,190]. LabA, Crm and RpaB are also known to affect RpaA[57,58,60,66,190]. CpmA modulates *kaiA* expression by an unknown mechanism[45]. In contrast to the unique cyanobacterial input factors, we found none of the eight output proteins exclusively in Cyanobacteria. Homologs of five factors (SasA, CikA, LalA, Crm, CpmA) are present in all four investigated taxonomic groups (Table B.2). CpmA is a member of a superfamily essential for purine biosynthesis and thus likely to have orthologs in other organisms[45]. However, RpaA and RpaB are not present in Archaea and SasA is only found in the methanogenic genera *Methanospirillum*, and *Methanosalsum*. Hence, the entire central output pathway is missing in Archaea (Fig. 5.1A). In addition, orthologs for RpaA are found in only nine non-cyanobacterial genera, questioning whether another transcription factor might read out the putative core timer in other Bacteria (Fig. 5.1A). A previous BLAST search by Dvornyk and colleagues revealed that SasA homologs in non-cyanobacterial prokaryotes lack the KaiB-like domain[161]. This finding is confirmed in our analysis, indicating that stimulations of SasA homologs by KaiC outside of Cyanobacteria are very unlikely, because interaction occurs via this KaiB-like domain, which adopts a thioredoxin-like fold[61,191]. Altogether, our analysis reveals that possible circadian clocks of Bacteria and Archaea must use an output pathway that is different from the one described in *Synechococcus elongatus* PCC 7942.

5.4.5 Co-occurrence analysis hints at the core module for circadian timing

The previous analysis revealed substantial differences in the composition of the clock components between Cyanobacteria, other Bacteria, and Archaea. Even within Cyanobacteria there is a huge variety in the composition of the potential circadian clocks (Fig. 5.1C). Cyanobacteria have either a severely reduced timing systems, such as the one in *Prochlorococcus*, a standard system as seen in *Synechococcus elongatus* PCC 7942, or an inflated system as found in *Synechocystis* sp. PCC 6803 (Fig. 5.1C). This trichotomy of systems raises questions about essentiality and pairwise co-occurrence of circadian clock proteins. These questions were answered in a series of right-sided Fisher's exact tests. To avoid systematic biases due to an overrepresentation of closely related strains[192], we extracted 69 unique combinations of the 21 circadian clock factors as described in 5.3.

Within these 69 unique systems two factors are always present: (i) KaiC, because we selected for organisms containing at least one KaiC-like protein and (ii) RpaB, which is associated with cell size and circadian gene expression[65,66]. RpaB competes for promoter binding sites with RpaA and its phosphorylated state is thought to inhibit the phosphorylation of RpaA[58]. Other factors present in the majority ($\geq$ 90%) of the observed unique clock systems are KaiA, KaiB, LdpA, IrcA, CikA, SasA, RpaA and CpmA. Because of their abundance, most of these factors show no pairwise co-occurence. For example RpaA and RpaB are found in 68 and all 69 clock systems, respectively. Thus their joint presence comes to no surprise. Instead, the finding confirms essential roles in global transcription regulation of Cyanobacteria[193]. With Fisher's exact test we seek to identify gene pairs rather unexpectedly co-occuring in a smaller subset of organisms. Such findings can indicate a common function in the circadian clock system. Only KaiA and CikA, out of the most abundant factors, show significant co-occurence with other factors.

Within the input pathway we detected three siginficantly co-occuring pairs (Fig. 5.2): (i) CikA and its interaction partner PrkE[155] (ii) PrkE and CdpA and (iii) the *kaiA* repressor Pex and CdpA. The first two results are in good agreement with a previous study[155]. Within the output pathway, there is a significant co-occurrence between LabA and its ortholog LalA. Interestingly, we identified several significant co-occurrences between factors of the input and the output pathway. CikA, which functions in the input and output of the clock, co-occurs significantly with LabA, and LalA (Fig. 5.2). This fits well in the overall picture as CikA and LabA are thought to regulate the activity of RpaA[56,57]. Additionally, PrkE shows also significant occurrences with both LabA and LalA (Fig. 5.2). Furthermore, CdpA was found to co-occur significantly with LabA. (Fig. 5.2) This is of special interest since PrkE and CdpA are only known as interaction partners of CikA, and both are involved in phase resetting, and cell division, respectively[155]. This result, however, hints at potential increased involvement of PrkE and CdpA in the RpaA regulation and

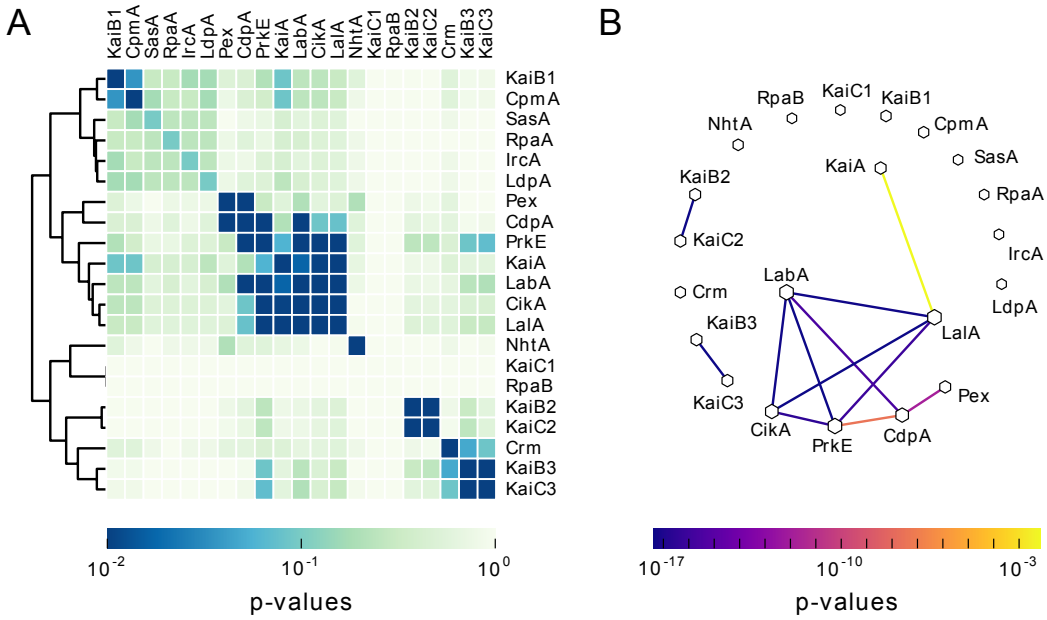


Figure 5.2: Co-occurrence of circadian clock proteins in cyanobacteria. (**A**) The p-values, calculated by pair-wise Fisher's exact tests, are visualized in a heatmap. Only p-values ≤ 0.01 are considered as significant. Proteins are sorted by a hierarchical agglomerative clustering algorithm. (**B**) Network of significant co-occurring circadian clock factors in cyanobacteria, calculated in regard to the results of the pair-wise Fisher's exact test. The line color corresponds to the level of significance. Missing links are those that had a higher p-value than 0.01. Node size is proportional to the degree of that node.

supports the view of an integrated network with overlapping interactions of input and output factors[194]. Notably, no co-occurrence of NhtA and LdpA was detected, although it was suggested that NhtA might be involved in assembly of the iron-sulfur cofactor of LdpA[155]. Significant co-occurrences with core factors could only be observed between KaiA and LalA (Fig. 5.2). However, KaiA shows a strong, but not significant, co-occurrence with CikA ($p = 0.0105$). Lastly, we also found significant co-occurrences between KaiB2 and KaiC2 as well as KaiB3 and KaiC3 (Fig. 5.2). This indicates two distinct function of the two pairs. In this context it is worth mentioning that *kaiB2*-containing archaeal genomes always encode a KaiC2 homolog.

In summary, we identified a conserved set of factors (Fig. 5.2), both in input and output that show significant co-occurrences. This set, composed of KaiA, PrkE, CdpA, CikA, LabA, and LalA, is found in Cyanobacteria with a true circadian clock such as *Synechococcus elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803[172] but is missing in cyanobacterial strains with reduced timing mechanisms such as *Prochlorococcus*. Additionally, PrkE, CikA, LabA, and LalA are also missing in most marine *Synechococcus* species. This finding hints at the importance of these factors for the functionality of a circadian clock. On the other hand, NhtA and Crm seem to play only a minor or extending role in clock regulation, because they are neither always present nor show significant co-occurrence with other factors.

5.4.6 A systematic analysis of circadian expression in Cyanobacteria

Genome-wide time-resolved expression measurements in a range of cyanobacterial strains have repeatedly indicated substantial fractions of genes with circadian regulation patterns[50,72,195]. Considering that all Cyanobacteria share the challenge of a photoautotrophic lifestyle, which requires major changes in the metabolism between day and night, one might expect a common transcriptional regulatory pattern. Thus, we compared a total of nine published microarray time-series datasets of different cyanobacterial strains under constant light or diurnal light conditions (for details see Table B.3), which were available and applicable for this analysis. Not all of the chosen microarray experiments were conducted under constant light conditions, which leads to a combination of circadian-clock regulated and light-induced genes. We therefore refer to genes with oscillating expression as diurnally regulated instead of circadian. For allowing a direct comparison, we reprocessed the raw-data and subjected the resulting expression time series to a harmonic regression oscillation detection. This method assumes a sinusoidal shape of circadian expression profiles and uses linear expression profiles as background, yielding estimates of the peak phase and amplitude of each gene.

In a first step we compared biological replicate datasets to establish the reproducibility of strain-specific circadian expression programs. Similarity between two circadian expression programs was established using the circular correlation coefficient ϱccc as described by Jammalamadaka and SenGupta[196] applied to estimated peak expression phases. The following analyses were limited to genes with oscillating expression profiles in both compared datasets since only in these cases phase and amplitude estimates are meaningful descriptors. Direct comparison of the oscillation phases and amplitudes indicates good reproducibility between two respective measurements in form of statistically significant elevated correlation of the diurnal expression patterns in *Synechococcus elongatus* PCC 7942 (ϱccc = 0.61, $p \ll 0.01$), *Synechocystis* sp. PCC 6803 (ϱccc = 0.31, $p \ll 0.01$), and *Cyanothece* sp. ATCC 51142 (ϱccc = -0.51, $p \ll 0.01$) (Figure 5.3 top row). While the *Synechocystis* sp. PCC 6803 datasets show significant similarity, the correlation is diminished by the distinct concentration of expression phases during the day in the beck14[71] dataset compared to the leh13[70] measurements. Interestingly, both *Cyanothece* sp. ATCC 51142 datasets exhibit a good agreement of peak expression phases with large early day and early night clusters, but the large negative correlation emphasizes the presence of a significant number of anti-phasic gene pairs. The corresponding oscillation amplitude values exhibit high statistically significant correlations (ϱccc > 0.77) for all three datasets (Fig. 5.3 bottom row).

This observation motivated the second step of the analysis, the comparison of expression patterns between different cyanobacterial strains. To facilitate this comparison, the prediction of homologous genes in the

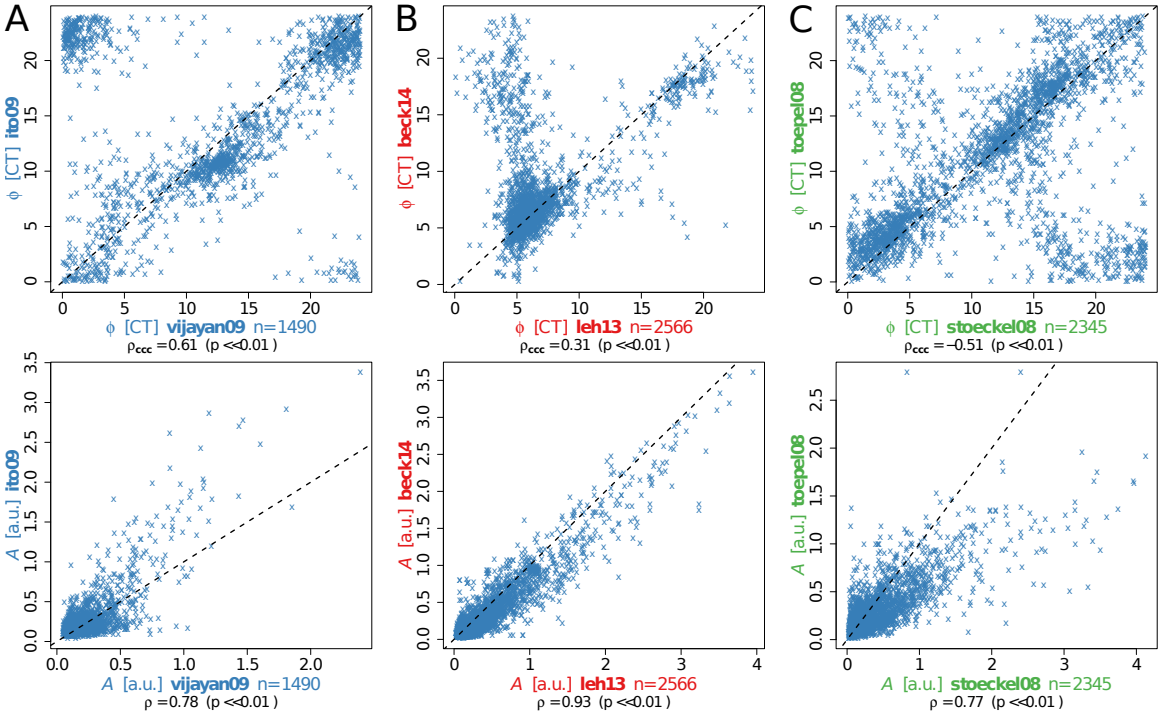


Figure 5.3: Phase and amplitude reproducibility of diurnal genes within cyanobacterial strains. Comparison of expression phase ϕ (top, [CT]) and amplitude A (bottom, [a.u.]) of diurnal genes shared between independent datasets of the same cyanobacterial strain. (**A**) *Synechococcus elongatus* PCC 7942 datasets of Vijayan and colleagues[51] (x-axis) and Ito and coworkers[72] (y-axis), (**B**) *Synechocystis* sp. PCC 6803 in the datasets of Lehmann and coworkers[70] (x-axis) and Beck and colleagues[71] (y-axis), and (**C**) *Cyanothece* in Stöckel and colleagues[73] (x-axis) and Toepel and colleagues[74] (y-axis). The number of genes found to oscillate significantly and the corresponding Pearson correlation coefficient ϱ between ϕ and A is provided below each panel, followed by the respective p-value for ϱ differing from 0. For ϕ, the circular correlation coefficient ϱccc is also provided. Axis labels are shown in the strain-specific color.

cyanobacterial clade by Beck and colleagues[71] was used as starting point. We focused first on the set of genes with oscillating expression patterns in all datasets. A summary of the cellular functions represented in the core diurnal genome is shown in Figure 5.4A (Table B.5), listing the frequencies of gene function category annotations of Cyanobase[113] for the respective *Synechocystis* sp. PCC 6803 genes. Three categories can be seen to dominate, i.e. Translation, Photosynthesis and Respiration, and Hypothetical. Most strikingly, 18 out of 64 genes annotated by Cyanobase[113] as "ribosomal protein" genes in *Synechocystis* sp. PCC 6803 fall into the core diurnal set, furthermore seven out of 27 genes annotated with "photosystem II". The remaining diurnally expressed genes are found interspersed across the metabolic network. While genes coding for parts of photosystem II, the RNA polymerase, or the ribosomal proteins can be expected to serve important roles in the adaptation to photic and aphotic phases, this analysis ascribes similar importance to other metabolic processes e.g. in the repair of UV-damaged Photosystem II centers (*slr1390*)[197], in the phosphate transport system (*pstB2*), pyrimidine and arginine biosynthesis (*sll1498*), or the glycolysis/gluconeogenesis via the fructose-bisphosphate aldolase (*sll0018*). Marker gene candidates for a working clock can be derived from the core diurnal genome. Importantly, less confidence can be placed in genes, which exhibit peak expression phases shortly past dawn, since these can either be clock regulated or simply induced by light. Two high confidence candidates are the light-independent protochlorophyllide reductase subunit ChlB and the fructose-bisphosphate aldolase *fbp*. Indeed, *fbp* also shows circadian expression patterns in *Chlamydomonas reinhardtii*[198] and *Arabidopsis thaliana* where its late-night peaks may reflect the great importance of these aldolases in higher plants for the mobilization of plastidic starch[199]. Particularly in higher plants, the mobilization of starch, the conversion into sucrose, and its transport to other parts of the plant occur mainly at night.

The group of 15 "hypothetical protein" genes in the core diurnal genome constitutes an excellent candidate set for novel clock-driven genes in strains with a working core clock. Interestingly, several of these genes are implicated with cell division, such as the YlmG-related hypothetical gene *ssl0353*, which is required for proper distribution of nucleoids in Cyanobacteria and chloroplasts[200]. Similarly, the hypothetical protein *slr1577* is suggested to function in the separation of chromosomes during cell division (Uniprot entry P74610). For the gene *slr1847* (Uniprot entry P73057) a DNA binding capability is suggested, which could therefore regulate expression, aid nucleoid organization, or protect the DNA.

The core clock genes *kaiA*, *kaiB*, and *kaiC* are a good starting point for a detailed comparative expression analysis. Only the *kaiB1* is significantly oscillating in all considered datasets. Interestingly, the *kaiA* gene features only very low amplitude expression oscillations and is arrhythmic in the vijayan09 *Synechococcus elongatus* PCC 7942 dataset. The expression phases vary from dawn (*Microcystis aeruginosa* PCC 7806) to morning (*Synechocystis* sp. PCC 6803), over midday (stoeckel08 dataset of *Cyanothece* sp. ATCC 51142),

and dusk (it009 dataset of *Synechococcus elongatus* PCC 7942), into night (*Anabaena* sp. PCC 7120). The observed expression phases of *kaiB1* are comparable to those of *kaiA*, but with significantly larger amplitude in *Synechococcus elongatus* PCC 7942 datasets. The phase of the *kaiB1* homolog in *Prochlorococcus marinus* MED4 peaks before dawn, comparable to *Anabaena* sp. PCC 7120. The *kaiC1* expression phases and amplitudes match those of *kaiB1*, with the notable exception of *Cyanothece* sp. ATCC 51142 for which antiphasic late-night peaks are observed. In *Prochlorococcus marinus* MED4, *kaiC1* peaks during the early night in contrast to the late night phase of *kaiB1*.

Many aspects agree well with previous knowledge. In *Synechococcus elongatus* PCC 7942, the core clock genes *kaiB* and *kaiC* are arranged in the *kaiBC* operon resulting in similar expression patterns[201–203], while *Cyanothece* sp. ATCC 51142 features the *kaiAB1C1* operon[204]. Interestingly, *Cyanothece* sp. ATCC 51142 features consistent anti-phasic expression of *kaiB1* and *kaiC1* whereas the remaining strains show co-expression, hinting at *Cyanothece*-specific post-transcriptional regulation of *kaiB1* or *kaiC1*. Oscillations in the *kaiA* gene expression, as reported by Ishiura and colleagues, feature small expression amplitudes compared to *kaiB* and *kaiC*[100]. In fact, *kaiA* consistently falls below the threshold of 2-fold expression change for the classification as circadian oscillator, which is commonly employed in microarray studies.

In the second step we generalized the detailed analysis of expression phases as presented for the core diurnal genome. We applied the circular correlation measure to all possible combinations of expression datasets. The resulting distribution reveals a clear separation between pairs of biological replicate datasets, featuring large numbers of shared oscillating genes and more extreme correlations (Fig. 5.4B red), and pairs of different cyanobacterial strains with fewer shared oscillating genes and much less extreme correlation coefficients (Fig. 5.4B blue). The only exception to this separation is the *Cyanothece* sp. ATCC 51142 dataset stoeckel08 (Fig. 5.4B green), which shares many oscillating genes with both *Synechocystis* sp. PCC 6803 datasets (leh13, beck14). The corresponding correlation coefficients are, however, similarly small compared to other inter-strain pairs. The full set of pairwise phase comparisons, which underlay this analysis, are shown in Figure B.1. This result indicates that the diurnal peak expression phase is not preserved amongst homologous genes in the cyanobacterial clade but might instead be tuned according to the metabolic gene outfit and the environmental needs of the respective strain.

5.4.7 Core circadian clock proteins, KaiA, KaiB, KaiC, vary in number and length

The previously observed diversity of circadian clocks within Cyanobacteria, and between other Bacteria and Archaea prompted further sequence analyses of the core clock proteins KaiA, KaiB and KaiC. Length comparisons gave rise to some new features of variations between the core factors of the circadian clock (Fig. 5.5). As described in the preceding, our BLAST analysis detected KaiA exclusively in Cyanobacte-

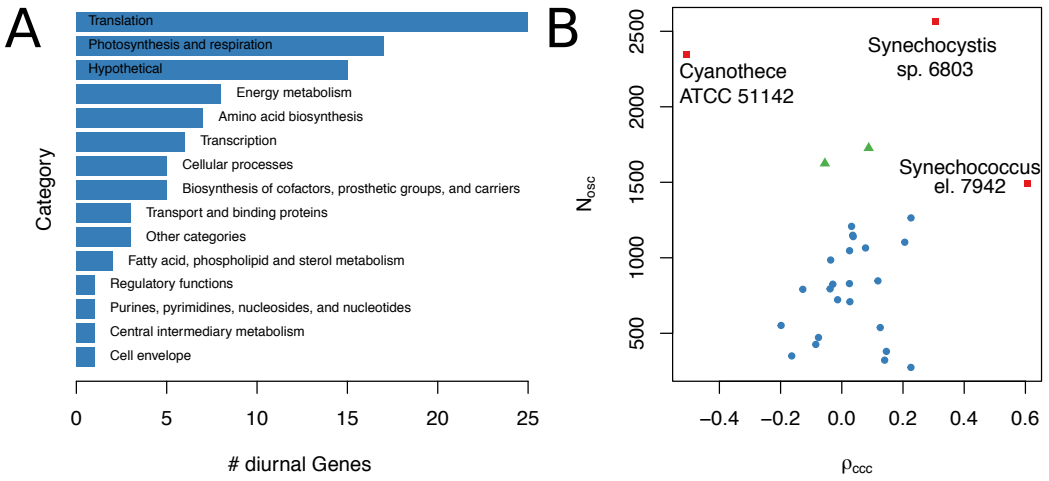


Figure 5.4: The functional composition of the core diurnal transcriptome and expression phase similarity between all datasets. (**A**) The gene functional categories of the core diurnal gene set using the annotations provided by Cyanobase for *Synechocystis* sp. PCC 6803. (**B**) The comparison is carried out between all dataset combinations. Biological replicate experiments available for three cyanobacterial strains show a large number of shared oscillating genes as well as large correlation coefficients (red). The comparison of the stoeckel08 dataset (*Cyanothece*) with both *Synechocystis* datasets is marked in green. The remaining comparisons between various strains are shown in blue.

ria. Interestingly, we could distinguish three subtypes of KaiA. While the sequence length of most KaiA is around 300 amino acids (AA) (*Synechococcus elongatus* PCC 7942: 284 AA) some stains have shortend homologs with a length of roughly 200 and 100 AA, respectively (Fig. 5.5A, C). Truncated KaiA proteins are almost exclusively found in members of the order *Nostocales*. Similar results were reported previously by Dvornyk and colleagues, who also observed a higher degree of polymorphism for the *kaiA* gene in comparison to *kaiB* and *kaiC*[79,156]. Multiple alignments of the KaiA proteins (Fig. B.2) verified that the truncated KaiA proteins have a shortened N-terminal sequence, which functions in the complete protein as the amplitude amplifier[205]. However, all of these KaiA orthologs contain the C-terminal part important for clock oscillation[205].

Distribution of the KaiB protein length reveals two distinct groups. The KaiB homologs are either as long as the one from *Synechococcus elongatus* PCC 7942 (102 AA) or about 250 AA in length (Fig. 5.5B, C). *Microcoleus* sp. PCC 7113 even has a KaiB with the length of 381 AA. KaiB homologs with query length are present in all four groups (Fig. 5.5B). Elongated KaiB proteins are mainly present in Cyanobacteria in the subclass *Oscillatoriophycideae* and the order *Nostocales*, specifying findings of Dvornyk[79] (Fig. 5.5C). BLAST analyses using the KaiB homologs from *Synechocystis* sp. PCC 6803 revealed that elongated variants are most similar to KaiB1. Elongation via concatenation of two KaiB was ruled out by visually inspecting alignments with two artificially concatenated KaiB1. Instead the elongated KaiB1 have a ∼150 AA N-terminal extension. BLAST searches of the N-terminal region showed no homologous sequences in other organisms than Cyanobacteria, and no putative conserved domains could be identified. However,

the N-terminal part is highly conserved within the Cyanobacteria having this KaiB variant. Interestingly, those *Nostocales* with an elongated KaiB1, also show a truncated KaiA.

The KaiC protein from *Synechococcus elongatus* PCC 7942 is 519 AA in length and is build up by two domains, the CI and CII domain, which have a high similarity and are connected by a linker-domain[103,206]. The C-terminal CII domain of KaiC comprises the interaction sites with KaiA as well as the specific phosphorylation sites[38,207–209]. KaiC homologs were detected with lengths varying between 101 AA and 741 AA (Fig. 5.5A, B). There are KaiC homologs in Archaea representing the whole observed length spectrum of KaiC, whereas bacterial KaiCs are almost always about 500 AA in length (Fig. 5.5A, B). Furthermore, in Bacteria and Archaea KaiB (and KaiA in Cyanobacteria) is only found when a "full-length" KaiC is present. In these bacterial organisms the length of KaiC is almost constant, regardless of the length of KaiA and KaiB homologs (Fig. 5.5A, B). However, in KaiB-possessing Archaea additional shorter KaiC homologs are found (Fig. 5.5B).

Moreover, the length distribution of KaiC revealed a substantial amount of KaiC homologs with a length of circa 250 AA, which is approximately the length of one KaiC domain. This KaiC variant is mainly found in Archaea, but also in a few Bacteria. In these bacterial species no KaiB homolog could be identified. Shorter KaiC homologs do not contain the important phosphorylation sites for maintaining the oscillator function. Therefore, they might not restore the full functionality of the *Synechococcus elongatus* PCC 7942 KaiC, but can rather answer questions about the evolution of KaiC[79]. Regarding the evolution of KaiC, two valid hypotheses exist, both of which state that KaiC arose from a shorter ancestral *recA* gene followed by a gene duplication and fusion. However, on the one hand Leipe and colleagues[210] hypothesize that an ancestral single-domain KaiC originated in Bacteria, was transferred into Archaea, where its two-domain version originated, and a second lateral transfer event introduced the double domain KaiC into cyanobacteria. On the other hand Dvornyk and coworkers[79] argue in a follow up study that KaiC has to be of cyanobacterial origin. Given the amount of new genomic data further studies would help to unravel the evolutionary history of KaiC.

5.4.8 Conserved motifs and activities in the cyanobacterial KaiC subgroups

For KaiC2 homologs outside of the cyanobacterial phylum an involvement in stress response (*Legionella pneumophila*[80]) and adaptive growth under rhythmic conditions (*Rhodopseudomonas palustris*[157]) has been demonstrated. Both proteins display autophosphorylation and KaiC2 from *Rhodopseudomonas palustris* shows elevated ATPase activity[157]. Nevertheless, the function of cyanobacterial KaiC2 and KaiC3 homologs remains unclear. We already demonstrated that KaiC2 and KaiC3 from *Synechocystis* sp. PCC 6803 displays kinase activity, which is independent of KaiA, whereas KaiC1 behaved like its *Synechococ-*

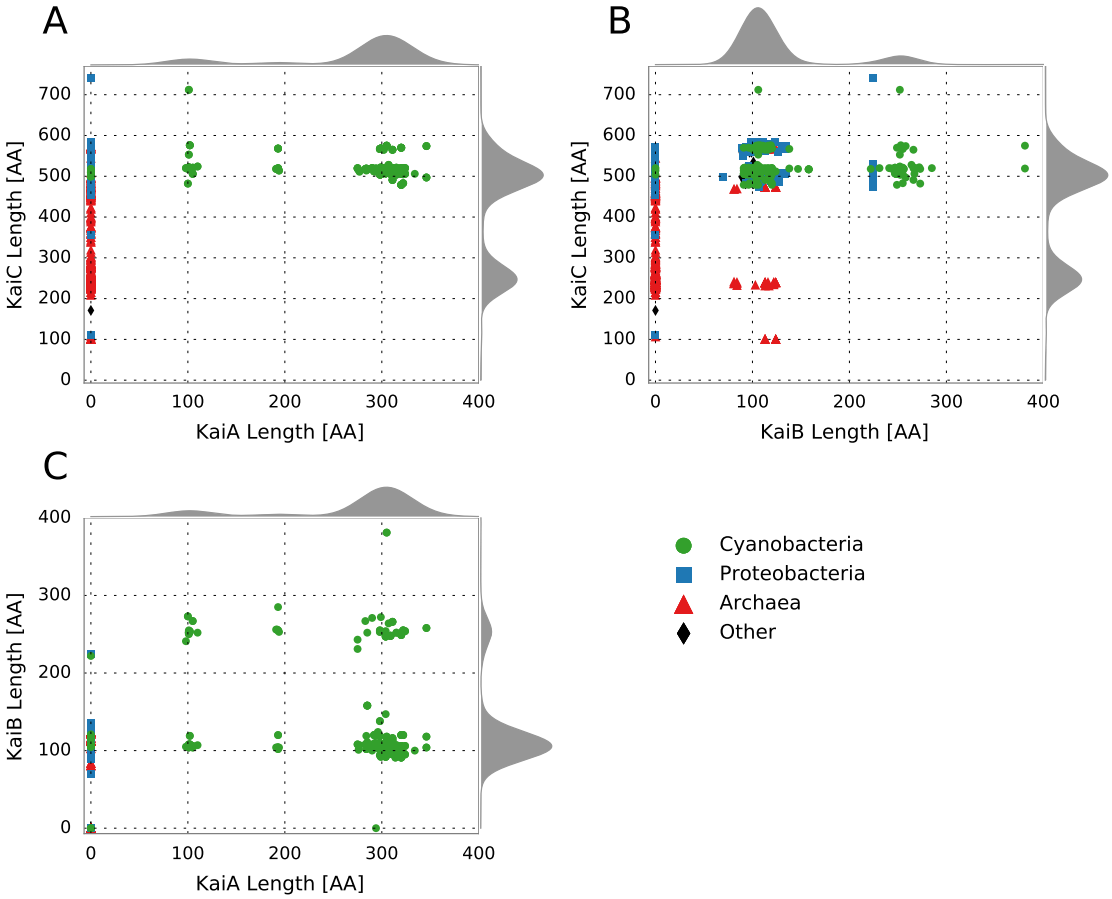


Figure 5.5: Protein length distribution of the circadian clock factors from *Synechococcus elongatus* PCC 7942 and its homologs. The four taxonomic main groups are highlighted: Cyanobacteria (Green), Proteobacteria (Blue), Archaea (Red), Other (Grey). The curves outside of the plot represent the cumulative density distribution of the respective protein. (**A**) KaiC length distribution in dependency of the KaiA length. (**B**) KaiC length distribution in dependency of the KaiB length. (**C**) KaiB length distribution in dependency of the KaiA length.

cus elongatus PCC 7942 ortholog[33]. Those activities could also be predicted from the C-terminal amino acid sequences[33]. To test whether general features of the three KaiC subgroups can be predicted, multiple alignments of the cyanobacterial KaiC1, KaiC2 and KaiC3 sequences were constructed. A WebLogo analysis revealed that relevant motifs for phosphorylation and dephosphorylation in the CII domain are highly conserved. The ATP-binding Walker Motif A (P-loop in Fig. 5.6A, GXXXXGKT,[88,100,211]) is present in all three KaiC subgroups. Strikingly, the respective sequence of KaiC-7942 (GATGTGKT) shows almost no modifications in KaiC1 and KaiC3 proteins. Furthermore, catalytic glutamates (EE in Fig. 5.6A,[212,213]), the R-finger contacting the γ-phosphate of ATP[214], and the truncated Walker motif B (Walker B in Fig. 5.6A,[35,100,211]), were found in all cyanobacterial KaiC subgroups. Notably, the arginine residue of the *Synechococcus elongatus* PCC 7942 Walker B motif is not conserved in KaiC2 homologs. Serine and subsequent threonine are the dominant phosphorylation sites in KaiC1 and KaiC3 proteins, like S^{431} and T^{432} in KaiC-7942[207,208], whereas KaiC2 homologs display two serine residues. In some KaiC3 homologs a tyrosin is present as second phosphorylation site. T^{426}, which is important for dephosphorylation of KaiC-7942[207,212,215,216], is also highly conserved. Therefore, phosphorylation and dephosphorylation via autokinase[35], ATP synthase and ATPase activity[36,212], respectively, are very likely for all cyanobacterial KaiC homologs. The same holds true for the N-terminal ATPase activity: We observed high conservation of the Walker motif A (P-loop in Fig. 5.6A), the catalytic glutamate residues (EE in Fig. 5.6A) and the R-finger in the CI domains of all cyanobacterial KaiC subgroups. The presence of the R-linker in CI domains of KaiC1 and KaiC3 homologs indicate a structural coupling of the N-terminal CI and the C-terminal CII-domain as it was demonstrated for *Thermosynechococcus* KaiC[214], whereas KaiC2 homologs lack the R-linker in CI.

Crystal structures are available for KaiB1 homologs and KaiB2 from *Legionella pneumophila*[80,107,111,136,137,217]. Free KaiB1 crystallizes in a unique fold, but can adopt a fold switched state, which is a prerequisite for KaiC binding and is similar to the thioredoxin like fold observed in KaiB2 crystals[80,135]. Residues K^{57}, G^{88} and D^{90}, which were suggested to regulate fold switching of KaiB-7942[135], are the most prominent residues at the mentioned positions in KaiB1 proteins, but not highly conserved (Fig. B.3). Hence, the question arises whether fold switching is conserved among KaiB1 proteins. Interestingly, G^{88} is highly conserved in KaiB2 proteins and hence free KaiB2 proteins may not necessarily favor the fold changed state. In KaiB3 all three residues show high conservation, which implies adoption of the ground state by the free protein and fold switching for putative KaiC binding.

KaiC homologs from the genus *Prochlorococcus* were classified as KaiC1 orthologs in our BLAST analysis (Fig. 5.1B). However, *Prochlorococcus* strains do not contain KaiA, and KaiC from *Prochlorococcus marinus* MED4 was demonstrated to phosphorylate independently of KaiA due to a modified A-loop sequence[78].

Therefore, we compared WebLogos of the A-loop sequence[37] for KaiC1 orthologs from Cyanobacteria with and without KaiA (Fig. 5.6A). The most obvious difference to KaiC1 proteins from cyanobacterial strains with KaiA is the presence of neutral glutamine in the second position, instead of a positively charged arginine. In KaiC2 and KaiC3 Weblogos motifs are even less conserved as already described for the *Synechocystis* sp. PCC 6803 representatives[33]. Interestingly, the KaiC3 WebLogo motif does not display any charged residue anymore. The absence of A-loop residues that are important to keep it in the buried state[37] indicate that the KaiA-independent phosphorylation is characteristic for KaiC2 and KaiC3 homologs. This is supported by the low conservation of the 438-444-loop and/or the 422-loop, which are part of the interaction network that mediates inhibition of phosphorylation by the buried A-loops in *Synechococcus elongatus* PCC 7942[37,103]. The high conservation of these loops in KaiA-lacking strains remains enigmatic. KaiA from *Thermosynechoccus* BP-1 interacts with residues of the A-loop and further C-terminal amino acids of *Thermosynechococcus* BP-1 KaiC[209,218]. Direct comparison of KaiC1 proteins from strains with and without KaiA reveals that the lack of KaiA goes along with dramatically decreased conservation of the KaiA interacting residues (Fig. 5.6A). Due to the high variation of those residues in KaiC2 proteins, we predict that they do not interact with KaiA. However, several KaiA interacting amino acids are present or conservatively substituted in KaiC3 proteins. Hence, complex formation of KaiC3 and KaiA might be possible, but stimulation of kinase activity by KaiA is unlikely.

To test this hypothesis, phosphate uptake as an exemplary KaiC activity was analyzed for representative cyanobacterial KaiC1 and KaiC3 proteins by incubation with γ-P^{32}-ATP at 30 °C in the presence and absence of *Synechococcus elongatus* PCC 7942 KaiA (KaiA-7942). The well-studied KaiC from *Synechococcus elongatus* PCC 7942 (KaiC-7942) served as control. As demonstrated in Figure 5.6B and 5.6C all recombinant KaiC proteins incorporated phosphate over time. The intrinsic kinase activity of KaiC1 homologs from *Nostoc punctiforme* ATCC 29413 (KaiC1-N29413), *Synechocystis* sp. PCC 6714 (KaiC1-Sy6714), and *Cyanothece* sp. PCC 7424 (KaiC1-Cy7424) was stimulated by KaiA-7942, similar to KaiC-7942 (Fig. 5.6B). As expected, KaiA had no effect on autophosphorylation of KaiC3 from *Cyanothece* sp. PCC 7424 (KaiC3-Cy7424) and *Microcystis aeruginosa* PCC 7806 (KaiC3-Mic7806, Fig. 5.6C). To extend the analysis to non-cyanobacterial proteins, KaiC3 from the hyperthermophilic Archaea *Thermococcus litoralis* (KaiC3-T.lit) and *Pyrococcus horikoshii* (KaiC3-P.hor), which show optimal growth at 85 °C and 98 °C[219,220], were analyzed in a similar way. Again both recombinant KaiC3 proteins displayed phosphorylation at 30 °C, which was independent of KaiA-7942 (Fig. 5.6C). Incubation at 75 °C indicated that the two archaeal KaiC3 homologs display kinase activity also at high growth temperatures (Fig. 5.6D). Hence kinase activity of KaiC proteins seems to be well-conserved, independent of the growth conditions of the strains, they are originating from.

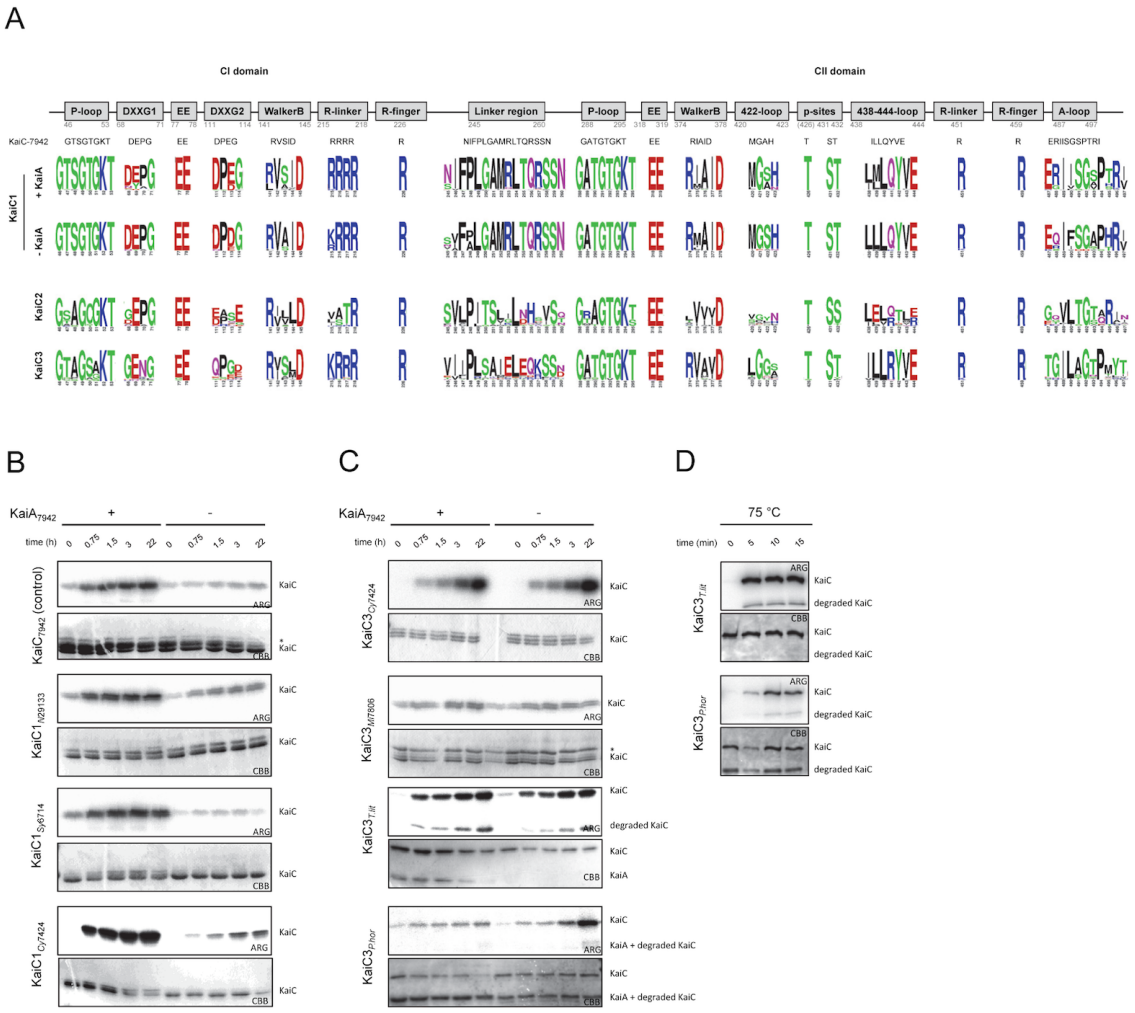


Figure 5.6: Activity of diverged KaiC homologs. (**A**) Conservation of important motifs in cyanobacterial KaiC1, KaiC2 and KaiC3 homologs based on WebLogo analyses. Motifs in KaiC1 homologs are displayed for proteins from organisms, which encode a KaiA protein or lack KaiA, respectively. Numbers indicate the residues in KaiC-7942. Properties of the residues are displayed as follows: polar (green), neutral (purple), basic (blue), acidic (red), and hydrophobic (black). (**B**,**C**) Phosphate uptake analyses of selected KaiC1 (**B**) and KaiC3 (**C**) homologs at 30 °C in dependence of KaiA. KaiC proteins were incubated with or without KaiA-7942 in the presence of γ-P^{32}-ATP. After 0, 0.75, 1.5, 3, and 22 hours samples were separated via SDS-PAGE, stained with Coomassie (CBB), and subjected to autoradiography (ARG). The asterix indicates a contaminating protein. (**D**) Kinase activity of archaeal KaiC3 homologs at 75 °C. KaiC3 proteins from *Thermococcus litoralis* and *Pyrococcus horikoshii* were incubated with γ-P^{32}-ATP and autophosphorylation was analyzed after 0, 5, 10, and 15 minutes. Shown are the Coomassie stained proteins (CBB) and autoradiography (ARG).

5.5 Conclusion

5.5.1 A core module for circadian regulation

Our analysis of 11,264 genomes clearly demonstrates that components of the *Synechococcus elongatus* PCC 7942 circadian clock are present in various bacteria and archaea. However, the frequency of Kai-clock related proteins is highest in Cyanobacteria. In fact KaiA, Pex, LdpA, and CdpA are exclusive to organisms of this phylum. In other organisms, e.g. *Rhodobacter sphaerodies*, reduced KaiBC-based clock systems are likely able to drive circadian oscillations[188]. An even simpler system solely dependent on KaiC might enable diurnal rhythms in *Haloferax volcanii*[159], probably using the ATP/ADP ratio for clock entrainment. Predictions for KaiC activities based on sequence alignments and motif analyses were validated through biochemical experiments. We confirmed kinase activity for "full-length" KaiC proteins composed of one CI and one CII domain, even in organism without *kaiA* or *kaiB*. KaiA from *Synechococcus elongatus* PCC 7942 enhanced KaiC-phosphorylation only in strains naturally possessing a *kaiA* gene.

Our co-occurrence analysis hints to a conserved extension set for circadian regulation, which is present in Cyanobacteria with observed circadian behavior and absent in Cyanobacteria having a diurnal, hourglass-like lifestyle only (see also Fig. 5.7). A diurnal core set, which is important to enable an hourglass-like timing system that resets every day, might be composed of KaiB, KaiC, LdpA, IrcA, SasA, RpaA, RpaB, and CpmA. However, our identified circadian core set, which potentially enables a selfsustained clock, additionally consists of KaiA, the two input factors CdpA, and PrkE as well as the input and output factor CikA, and the output factors LabA, and LalA.

The systematic comparison of microarray timeseries datasets indicates that the diurnal peak expression phase is not conserved amongst homologous genes in the cyanobacterial clade. Instead, the expression phase may be tuned according to the gene outfit and varying environmental needs. The analysis yielded a set of 95 genes in the core diurnal genome, which can be considered critical for the adaptation to day and night. Particularly, the subset of non-light induced genes are prime candidates for circadian clock marker genes. This set furthermore contains several hypothetical genes, which are interesting candidates for novel clock-driven genes.

The gained insights about the diversity within the composition of the components involved in the circadian protein clock as well as the diversity on the sequence level of the core factors call for further modification and simplification of the clock. The exponential increase of molecular tools for synthetic applications in recent years sets the stage for such ambitious projects. A future goal could be the reduction of complexity by removing as many factors as possible so that an integration of a circadian clock in synthetic and industrially valuable organisms becomes feasible.

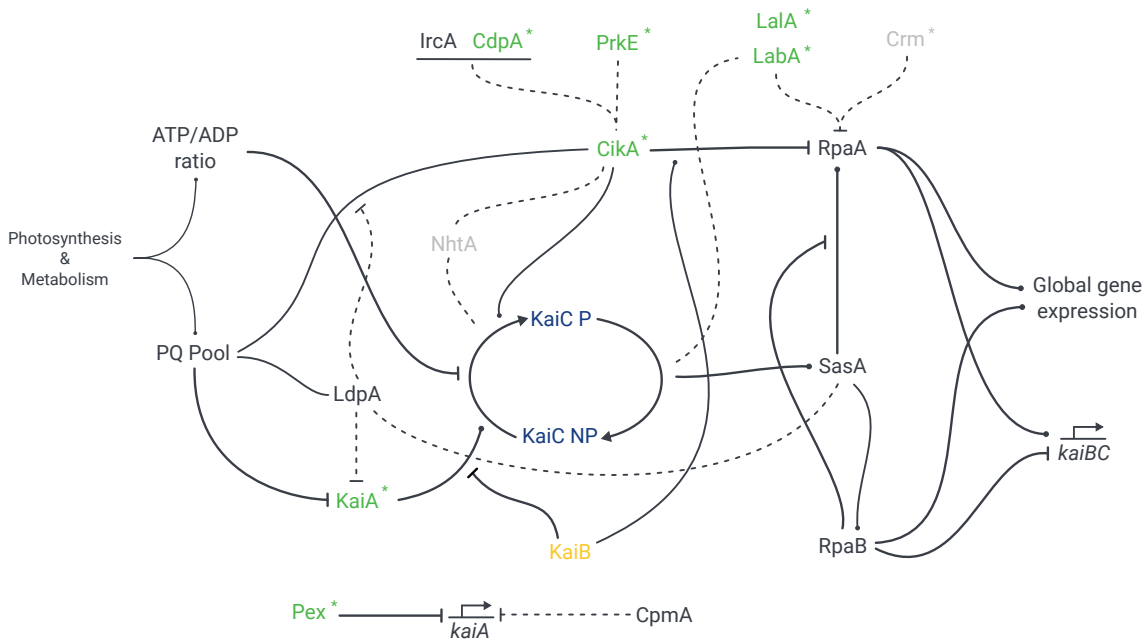


Figure 5.7: Schematic overview of the circadian clock in *Synechococcus elongatus* PCC 7942. Shown is the protein network in which the clock proteins are integrated based on [28,30,43–45,54–61,63–67,135,155,194,221]. The plastoquinone pool and ATP/ADP ratio serve as input signals to entrain the circadian clock with the environment and metabolic state of the cell. The signals are recognized by the input factors LdpA and CikA and the core clock factors KaiA and KaiC. Other input factors are Pex, and the CikA interaction partner NhtA, PrkE, IrcA, and CdpA. The stimulating effect of KaiA on KaiC is antagonized by KaiB. The output of the clock is comprised by SasA, CikA, RpaA, RpaB, LabA, LalA, Crm, and CpmA. SasA interacts with KaiC and further phosphorylates RpaA, whereas CikA acts as a phosphatase on RpaA. RpaA together with RpaB controls global gene expression as well as the expression of the *kaiBC* cluster. Solid lines represent established biochemical interactions whereas dashed lines represent inferred interactions. The seven most interconnected factors we found in our co-occurrence analysis are highlighted in green. Factors colored in light grey showed no co-occurrence in the Fisher's exact test and are present in $<$ 90% of all observed systems. Asterisks (*) indicate the factors missing in *Prochlorococcus*.

6

Computational modeling unravels the precise clockwork of cyanobacteria

Nicolas M. Schmelling and Ilka M. Axmann (2018) Computational modeling unravels the precise clockwork of cyanobacteria, *Interface Focus*, 20180038. doi: *10.1098/rsfs.2018.0038*

Author Contributions

NMS performed the literature search, designed the figures, and wrote the manuscript. IMA supervised the work, gave valuable comments on the manuscript throughout the writing process, and proofread the manuscript. Both authors gave final approval for publication.

Pro rata

Literature research: 100%

Manuscript preparation: 80%

Summary of Publication

It is just over a decade ago that the circadian clock of *Synechococcus elongatus* PCC 7942 was reconstructed in a test tube showing stable oscillations for weeks[87]. Since then tremendous achievements have been made by using mathematical modeling to decipher various molecular mechanisms of this system and timing and circadian systems in general. In this work, we review the detailed insights into the functionality of the circadian clock gained by mathematical modeling. These insights range from the description of the molecular functionality of the core clock and its interaction over the strategies for entrainment and input compensation such as shorter phosphorylation cycles of individual KaiC hexamers under low ATP conditions to strategies for robustness to noise and environmental changes such as the increase in genome copy number or the uncoupling of the oscillator from transcriptional processes. The review was published in *Interface Focus* and can be found on doi: *10.1098/rsfs.2018.0038*.

6.1 Abstract

Precisely timing the regulation of gene expression by anticipating recurring environmental changes is a fundamental part of global gene regulation. Circadian clocks are one form of this regulation, which is found in both eukaryotes and prokaryotes, providing a fitness advantage for these organisms. Whereas many different eukaryotic groups harbor circadian clocks, Cyanobacteria are the only known oxygenic phototrophic prokaryotes to regulate large parts of their genes in a circadian fashion. A decade of intensive research on the mechanisms and functionality using computational and mathematical approaches in addition to the detailed biochemical and biophysical understanding make this the best understood circadian clock. Here, we summarize the findings and insights into various parts of the cyanobacterial circadian clock made by mathematical modeling. These findings have implications for eukaryotic circadian research as well as synthetic biology harnessing the power and efficiency of global gene regulation.

6.2 Introduction

Life on Earth has evolved under the influence of changing environmental conditions. While some environmental conditions may fluctuate without any apparent regularity, a large class of environmental changes exhibits regular cycles on daily to annual timescales. The most prominent of these environmental cycles is the daily change in temperature and light availability caused by Earth's rotation itself. Circadian clocks are an adaptation to this recurring cycle.

Circadian clocks are biochemical oscillators that function as an endogenous timekeeper within organisms. Circadian clocks are characterized by three properties: (i) persistence of oscillations without an exogenous cycle, e.g. in constant light, (ii) temperature compensation, and (iii) the entrainment of the endogenous oscillator(s) to the exogenous cycle of light and darkness[5]. Without the exogenous stimulus ("Zeitgeber") the period length of the oscillator can vary between organisms. The ability of entrainment, however, allows the Zeitgeber to synchronize the oscillation with the exogenous rhythm.

While changing environmental conditions typically pose a challenge to organisms, changing conditions also present the possibility of adaptation and therefore a fitness advantage over competitors. The fact that circadian clocks can be found in a variety of organisms from mammals to plants, and fungi suggests a selective pressure by environmental light/dark cycles that favored the development of such a complex timing system. A fitness advantage of circadian clocks has been elegantly shown in competition experiments using cyanobacteria with intact circadian clocks and clock mutants[6,7]. Circadian clocks appear to be a conserved trait in evolution. However, differences in the sequences of proteins that are involved in the circadian clock suggest a convergent evolution of timing mechanisms[15,16].

The history of research on the genetic basis of circadian clocks begins in the 1970s when Konopka and Benzer first identified clock components in *Drosophila melanogaster*[11]. Afterwards, circadian clocks were also identified in mammals, including hamster, mice, and humans[12–14], as well as fungi with *Neurospora crassa*[9] and plants with *Arabidopsis thaliana*[10]. In the beginning, it was believed that circadian clocks were exclusive to eukaryotic organisms and that simpler prokaryotic organisms could simply not maintain independent circadian oscillation as they divide multiple times during a day[5]. However, in the mid-1980s, researchers identified oscillations in different diazotrophic cyanobacteria, performing photosynthesis during the day and fixing nitrogen at night[5]. At least since the remarkable findings of Kondo and his colleagues in 1993 identifying circadian gene expression in Cyanobacteria and later the central three proteins of the cyanobacterial circadian clock KaiA, KaiB, and KaiC all doubts that simple prokaryotic organisms can harbor such complex systems as circadian clocks were removed[8,100].

In this review, we will cover various computational concepts of circadian clocks. We will first introduce basic functionality and the differences between eukaryotic and cyanobacterial circadian clocks. Afterward, we will focus on the cyanobacterial circadian clock system as this is the evolutionary oldest known yet functional clock system. Furthermore, the cyanobacterial clock is special in the sense that it does not rely on transcriptional process and can stably oscillate in a test tube for weeks. In addition, over the last decades central processes and characteristics of circadian clocks have been studied experimentally and computationally, including entrainment strategies, the adaptation to noise and environmental changes, as well as robustness concepts, using the cyanobacterial system. Further aspects and information about computing by biological clocks, which are not covered in this review, can be found in a recent review by Dalchau and colleagues[222].

6.3 Basic Functionality of Circadian Clocks

Circadian clocks are a prominent way to coordinate and regulate gene expression on a global cell scale. Instead of regulating single genes or an operon, circadian clocks are able to regulate hundreds to thousands of genes. In *Synechococcus elongatus* PCC 7942, the model organism for cyanobacterial circadian clock research, up to 64% of the genome is under the control of the circadian clock[51]. Other cyanobacteria also show oscillations of large portion of their genomes ranging 20% - 79%[91] compared with eukaryotic organisms, which show only 5% - 15% of rhythmically oscillating transcripts[223].

However, not everything that oscillates is under the control of a circadian clock. As mentioned above, circadian clocks are characterized by three essential criteria[5]. In *Saccharomyces cerevisiae*, even under constant optimal growth conditions, global gene expression oscillates. So far no circadian clock has been identified

in yeast and these oscillations are a result of a rather general growth principle where changes in genome configuration are due to changes in the energy levels of the cell[224,225].

Even though there are phenomena, which cause stable oscillations in global gene expression, e.g. supercoiling of the DNA, circadian clocks are a prevalent form found in almost any group of organisms. It has been hypothesized that the cyanobacterial clock, however, is influencing the supercoiling state of the DNA thereby connecting the circadian system's output with a global gene expression system[51,223]. Circadian clocks are divided into two groups in regard to their functionality: Eukaryotic circadian clocks comprise nested transcription-translation feedback loops (TTFL), whereas cyanobacterial circadian clocks, as the only known example for prokaryotic circadian clocks, comprise a post-translational oscillator (PTO)[16]. As mentioned before, sequence analyses suggest a convergent evolution with multiple origins of circadian clocks, which is supported by the different functionality of known circadian clocks.

Even though eukaryotic circadian clocks do not share sequence similarities between the protein factors involved, they share structural similar circuits as all of them comprise nested positive and negative gene-protein feedback loops. However, based on the activation or inhibition of gene expression of clock factors, TTFL clocks are susceptible to effects of DNA replication or gene position on the chromosome as we will discuss later.

In cyanobacteria, the circadian clock comprises three proteins KaiA, KaiB, and KaiC. At the center of the cyanobacterial circadian clock is a phosphorylation cycle. In detail, Cyanobacteria measure time by the phosphorylation state of KaiC, which is the central part of the clock. KaiC auto-phosphorylates during the day stimulated by KaiA (Fig. 6.1). At dusk, KaiC is completely phosphorylated and a conformational change allows KaiB to bind to KaiC. KaiB then antagonizes the positive effect of KaiA and initiates auto-dephosphorylation by sequestering KaiA. At dawn KaiC is completely dephosphorylated, which causes KaiB to be no longer able to bind to KaiC, the KaiCBA complex falls apart and the cycle starts all over again with free KaiA binding to KaiC and stimulating the phosphorylation[32,62].

After reconstructing the phosphorylation mechanism of KaiC in a test tube[87], multiple computational models using differential equations arose in order to elucidate the functionality of the cyanobacterial circadian clock. It was proposed that shuffling and exchange of KaiC monomers between KaiC hexamer is important for stability and synchronicity of the oscillation[226–228]. Further, the remarkably robust high-amplitude phosphorylation cycles of KaiC are achieved by sequestration of free KaiA mediating synchronization[16,101,105,229,230]. This sequestration mechanism is a conserved design principle in natural systems, which can also be found in many different systems, e.g. MAPK protein kinase cascade[231,232]. Shortly after, other models resolved the importance of KaiA and KaiB in the cyclic phosphorylation of KaiC[230,233–235]. Each KaiC monomer has two independent phosphorylation sites, and it could experimentally be shown

by Rust and colleagues that there is an ordered phosphorylation of these sites, which was before neglected by other models[92]. Further, their model showed that KaiA is inactivated when the concentration of serine phosphorylated KaiC exceeds a certain threshold. Shortly after Brettschneider and colleagues built an advanced model on top of the model proposed by Rust and coworkers, which showed for the first time the ability of entrainment by temperature[101]. This mathematical model by Brettschneider and colleagues (2010) suggested that the complex of KaiCpS/T (KaiC phosphorylated only on S^{431}) and KaiB binds KaiA dimers at a newly formed binding site thereby inhibiting KaiA-mediated stimulation of KaiC phosphorylation, which was confirmed just recently by single-particle cryo-electron microscopy and mass spectrometry[32,62].

Computational models have helped to resolve open questions about the dynamics of interactions between clock factors in cyanobacteria. However, in eukaryotic circadian clocks protein factors regularly activate or suppress gene expression of their own expression or of other clock factors[236]. In cyanobacteria, a similar negative feedback mechanism was suggested by Iwasaki and colleagues in which phosphorylated KaiC inhibits the expression of the *kaiBC* mRNA[202]. Computational models could first confirm those suggestions by showing that such a negative feedback mechanism can, in fact, explain sustained oscillation in *kaiBC* expression[237]. Later, Hertel and coworkers could also identify which of the four phosphorylation states of KaiC suppresses *kaiBC* expression and that even two forms activate its own expression[238].

The circadian clock of the model organism *Synechococcus elongatus* PCC 7942 was the basis for all of the above-described models. However, Cyanobacteria are characteristic for their diversity, which can also be observed regarding the composition of circadian clock factors. With respect to this composition, three distinct groups of cyanobacteria can be identified: (i) *Synechococcus*-like clock composition harboring single copies of each of the known factors in the circadian clock of *Synechococcus elongatus* PCC 7942, (ii) a *Synechocystis*-like clock composition harboring multiple copies of some factors including KaiB and KaiC, and (iii) a *Prochlorococcus*-like clock, which is missing multiple factors of input and output pathways as well as the central factor KaiA[42,79,91]. Interestingly, the *kaiA* gene never occurs more than once in a genome and cannot be found outside Cyanobacteria in contrast to other core clock factors. Furthermore, the *Prochlorococcus* clock is called an hourglass timing system and is not considered a circadian clock as it does not persist without an exogenous cycle but shows oscillation of global gene expression under a light-dark regime[77]. It remains an open question how the multiple copies of the core clock factors in *Synechocystis* sp. PCC 6803 are integrated into the circadian clock.

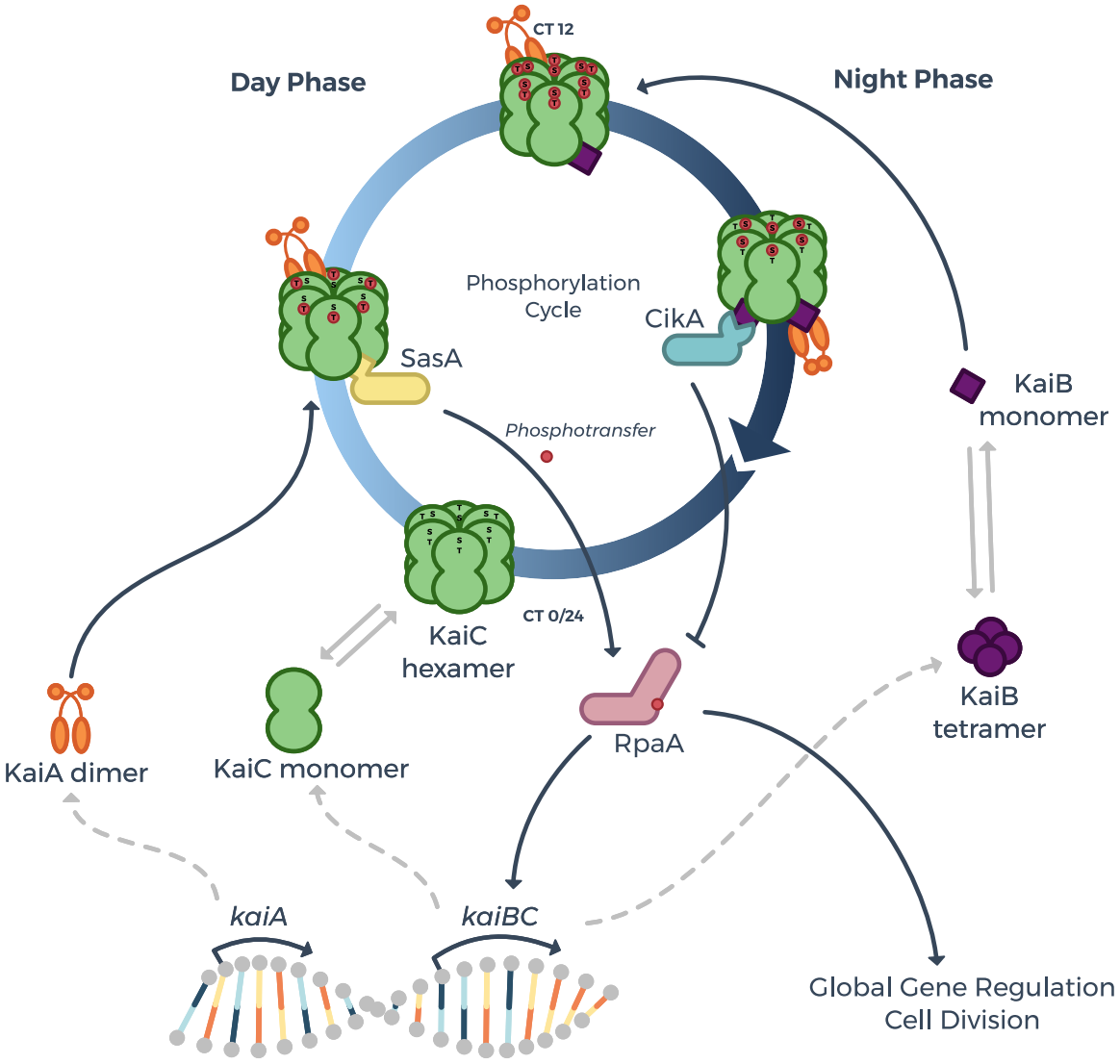


Figure 6.1: Circadian clock of *Synechococcus elongatus* PCC 7942 and its interaction network. The interactions of the core clock result in a 24 hour cycle of phosphorylation and dephosphorylation. Depending on the phosphorylation state of KaiC hexamers different proteins interact with the core clock to connect the circadian signal to global gene regulation. SasA and CikA comprise an antagonistic system, which regulates the activity of RpaA and thus its ability to bind promoters. The figure is adapted from Hertel and colleagues, Shultzaberger and coworkers as well as Schmelling and colleagues [91,194,238]. Graphical representations of the protein factors involved are based on "Cyanobacterial Circadian Clock Output Mechanism" by The BioClock Studio (https://youtu.be/DcuKifCRx_k). The coloring of the Kai proteins is adapted to the coloring of the protein structures of the Kai proteins from Snijder and coworkers [32]. Dashed lines represent transcription and translation processes. Solid lines represent physical interactions between proteins. Grey solid lines represent de-/formations of homo-multimers.

6.4 Entrainment

One defining characteristic of a circadian clock is the ability to be entrained to an exogenous cycle. It is important that a circadian system can be synchronized to an exogenous stimulus in order to be most useful for the organism in anticipating recurring patterns. However, it is also necessary that such a system is robust against naturally occurring fluctuation of the input signal.

There are two strategies for entraining the clock to an exogenous stimulus: (i) with a direct sensing of light intensity by clock components or other components that pass the signal onto clock components which strategy is normally used by eukaryotic circadian clock systems[15] and (ii) indirect sensing of light through changes in the metabolic state of the cell, i.e. redox state or ATP/ADP ratio, which is commonly used by Cyanobacteria[41,194]. The core oscillator of the cyanobacterial circadian clock depends as above described on the phosphorylation state of KaiC. KaiC is able to autophosphorylate and -dephosphorylate, which is enhanced by the interaction with the other core factors KaiA and KaiB. Each monomer has two phosphorylation sites, which in the end results in 12 independent sites for each hexamer. The phosphorylation pattern of KaiC hexamers is highly ordered resulting in cycles of the following four states: unphosphorylated (U-KaiC), phosphorylated only on S^{431} (S-KaiC), phosphorylated only on T^{432} (T-KaiC), and phosphorylated on both S^{431} and T^{432} (ST-KaiC)[92]. It has experimentally been shown that the phase of the circadian clock is affected by the ATP/ADP ratio, which is a result of the cellular catabolic metabolism and the photosynthetic apparatus[194]. Dark phases cause a drop in the ATP/ADP ratio, which shifts the clock into the dephosphorylation phase. Rust and colleagues convincingly showed that the *in vitro* form of the cyanobacterial circadian clock reacts differently to the ATP/ADP ratio depending on the timing of the dark phase[41]. During the phosphorylation phase (subjective day) the oscillator is most susceptible to changes in the ATP/ADP ratio, whereas in the dephosphorylation phase (subjective night) the oscillator is almost insensitive. Adding the competitive inhibitor to their core clock model, they described an entrainment mechanism for the cyanobacterial circadian clock that works *in vitro* and does not rely on an additional signaling pathway, resembling the direct effect of the metabolic state of the cell on the phase of the circadian clock[41].

Such an entrainment strategy seems like a very early form as it does not rely on other protein factors or special compounds. ATP and ADP are most likely older than the KaiC or its ancestral version, meaning that the core of a clock comprised of a single protein, KaiC or its ancestor, and the cellular energy compound ATP and its derivatives could have arisen early in evolution. A recent mathematical model by Hernansaiz-Ballesteros and colleagues[239] demonstrated how a single molecule similar to KaiC can form an oscillator through ordered phosphorylation. The model was based on a two-intermediate system as an alteration of

an "Approximate Majority" algorithm and systematically removed reaction paths between the different phosphorylation states. In the end, the model predicted a system with similar ordered reaction paths as seen for KaiC[239]. Thus, KaiC could have evolved as an ATP biosensor and ancient clock and later more protein factors were added to the clock in order to make it more robust to external and intrinsic fluctuations as well as to create a more robust readout. This is in good agreement with sequencing analyses, which suggest that KaiC originated first from a RecA (bacterial DNA recombination protein) ancestor by gene duplication and subsequent fusion before KaiB and last KaiA originated[79]. Other protein factors of the circadian clock are evolutionarily younger or as old as KaiB. For example, SasA, the first downstream interaction partner of KaiC (Fig. 6.1), is thought to have originated from a fusion of a two-component histidine kinase and an ancestor of KaiB around the time when KaiB formed a cluster with KaiC[240].

However, what are the environmental conditions that led to the evolution of circadian clocks? What conditions would favor the evolution of such a complex system and provide a fitness advantage for the organism? Troein and colleagues[241] attempted to answer those questions by evolving *in silico* gene regulatory networks to best predict phases of a day/night cycle. Those gene regulatory networks best resemble the eukaryotic circadian clock systems, however, their findings can still be considered to be universally correct for other circadian clock systems. Troein and coworkers showed that only under multiple photoperiods combined with environmental noise were circadian clocks able to evolve with a high probability. In addition, real environmental variations even increase this probability. They concluded that only seasonally changing photoperiods combined with sufficient noise can lead to the evolution of circadian clocks, which is accompanied by an increase in complexity of the system[241].

It has been shown through sequence analyses that cyanobacterial clocks are evolutionarily old systems and by mathematical modeling that seasons and noise are needed to promote the evolution of such a system. Nevertheless, one aspect of the evolution of circadian clocks is still missing. All above-mentioned simulations assume a day length of roughly 24 hours, however, the length of day changed drastically over the last 2.5 billion years and thus also during the evolution of circadian clocks[242]. Based on measurements of sandstone laminae, the length of a solar day on Earth is estimated to be 17 hours[242]. Furthermore, models about the origin of the Moon 4.47 billion years ago predict a rotational speed of the Earth between 2.5 and 9 hours[243–245]. How does the length of day affect the complexity of circadian clocks? Were additional factors also needed to adjust the biochemical speed of circadian clocks in addition to increase robustness?

6.5 Adaptation of Circadian Clocks to Noise and Environmental Changes

As described before, noise and environmental changes, e.g. seasonal changes in the photoperiod, are essential for the evolution of circadian clocks [241]. The importance of noise for the evolution of circadian clocks also highlights the potential for adaptation. Circadian clocks are one of the naturally occurring adaptations to recurring noisy patterns, e.g. noisy light-dark cycles. They provide a fitness advantage to organisms by robustly buffering such noise. Further, circadian clocks are able to adjust the oscillator's period to long-term seasonal changes in the environment. As Troein and colleagues [241] showed, one strategy of adaptation is to increase the number of feedback loops and complexity of gene regulatory networks. However, the cyanobacterial circadian clock has a PTO at the center, which functions in a different manner from eukaryotic TTFLs. In addition, there are two forms of timing mechanisms in cyanobacteria: (i) a circadian clock as seen in *Synechococcus elongatus* PCC 7942, which functions similar to a limit cycle oscillator [97,246], and (ii) an hourglass timing system as seen in *Prochlorococcus*, which works more like a point attractor that stops oscillating and relaxes to a stable fixed point in the absence of an exogenous cycle [97]. Interestingly, recent findings of microbial communities with *Prochlorococcus* indicate prolonged oscillations of transcripts in extended darkness [247]. These findings suggest a more complex exogenous network in addition to the light/dark cycle to entrain the timing system in nature. The interactions in the coculture seem to cause changes of the redox state in *Prochlorococcus* cells, which are similar to the ones found under a light/dark regime and thus prolong the oscillations [247].

How do these two systems cope with noise? Before answering this question we first have to distinguish between external noise by variations in the exogenous stimulus, e.g. light availability, and internal noise due to variations in biochemical reactions. External noise such as shading by clouds can have drastic effects on the availability of light. Such fluctuations can also range in their timescale from seconds to hours, which can cause the circadian clock to be shifted back to night cycles [94]. Fluctuations of external cues have different effects on the two circadian clock systems in cyanobacteria. Whereas, the limit cycle oscillator such as the circadian clock from *Synechococcus elongatus* PCC 7942 is almost unaffected by such fluctuations [93,97], the point attractor like the hourglass timing system from *Prochlorococcus* is set in free fall towards the night state [97]. The ability to tell the accurate time increases with the size of the limit cycle and the overlap between the day and night state of the system [97]. On the other side, when considering only internal noise due to finite numbers of proteins the point attractor outperforms the limit cycle in regard to precision due to its ability to change faster between both states [97].

Limit cycle oscillators are more robust under a range of external and even internal noise levels, however, so far only noise in a 12:12 (light:dark) hour regime was analyzed. As previously described, in addition to

noise also seasonal changes are required for the evolution of circadian clocks[241]. How do these circadian clocks adapt to seasonal changes in the day length? To answer this question, Leypunskiy and colleagues analyzed the effects of different day length on the performance of the circadian clock of *Synechococcus elongatus* PCC 7942 *in vitro* and *in vivo*. Interestingly, they could show that after a transient phase the circadian clock tracks midday over a large range of day lengths ranging from 6 to 20 hours of light. This observation is also found in the *in vitro* clock in a test tube, where similar to light and dark phases, phases of high ATP or high ADP are alternating, resembling the intercellular effect of light on the levels of ATP and ADP through photosynthesis[95]. The effects of seasonality and different photoperiods on entrainment are studied theoretically in further detail by Schmal and colleagues[248].

The cyanobacterial circadian clock has the ability to perform under various conditions regarding the length of day and the level of external and internal noise, however, the molecular mechanisms of how the circadian clock is able to cope with these variations remain unknown. By comparing three models that try to explain some of the molecular mechanisms of input compensation Paijmans and coworkers[93] were able to identify a new mechanism at the individual hexamer level. Whereas the period of the circadian clock in their model[249] remains unaffected by changes in the bulk ATP fraction, the models by Rust and colleagues[92] and van Zon and coworkers[230] show decreasing or increasing period lengths by decreasing ATP amount, respectively. This high level of input compensation of the Paijmans model is achieved on the individual hexamer level rather the population level. At increasing ATP levels, the phosphorylation rate increases, which is true for all models, resulting in an increased cycle length of each hexamer through the phosphorylation cycle. Meaning that under decreasing ATP levels, input compensation is achieved through smaller cycles in the phosphorylation state space by individual hexamers and an earlier switching to the inactive state of these hexamers[93]. This mechanism provides a way to explain the robustness of this circadian clock system to external noise by changes in the light availability and a subsequent drop in the ATP levels.

Even though the circadian clock shows great robustness to noise and unexpected changes in the availability of energy through sunlight, it is still vulnerable to large drops in ATP levels at certain times of day. Looking at the phase response curves for the three previously mentioned models we see that early in the day a dark pulse has the strongest effect on the clock by delaying its phase[93]. In order to analyze the cost associated with this vulnerability, Lambert and colleagues[94] studied the effects of misalignment of the circadian clock to the environment in individual cyanobacterial cells. Early in the day cells are most vulnerable to sudden drops in ATP levels resulting in a growth arrest, which can lead to a failure to resume to growth when cells are placed into the light. At the beginning of the subjective day, the cells are rapidly growing, whereas during subjective night, especially at dusk, the cells are most resistant against starvation, which potentially corresponds to the internal glycogen storage[94]. Thus, the authors conclude that one of the

major functions of the cyanobacterial circadian clock is to balance between rapid growth and starvation resistance.

6.6 Further Strategies for Robustness of a Circadian Clock

As highlighted before, one of the most important tasks of a circadian clock is the ability to robustly measure and transfer an exogenous signal to downstream processes. We have seen that noise is essential for the evolution of the circadian clock and that the cyanobacterial clock has evolved different strategies to cope with external and internal noise in order to robustly predict the period of the exogenous cycle. Furthermore, it was shown that independent of the length of the exogenous stimulus, i.e. light, the cyanobacterial circadian clock tracks the middle of that phase in order to predict to the upcoming change.

There are, however, further strategies for robustness of circadian clocks and oscillators. Some of which are already found in the circadian clock of cyanobacteria, others are strategies that can be applied in order to make synthetic oscillators more robust.

Eukaryotic circadian clocks comprise nested TTFLs where gene products regulate the expression of other factors of the clock. One common motif in these TTFLs is the self-repression of the gene by the gene product, which is found in prokaryotic as well as eukaryotic oscillators. It has been shown that these motifs are prone to be phase-locked with the cell cycle as they are largely influenced by the effects of gene density due to genome replication events[98]. Over the course of the cell cycle, the cell volume increases steadily until cell division, however, the genome copy number doubles more or less instantaneously if there is only one genome copy per cell. In these cases, the simple negative transcriptional feedback loops are locked to the cell cycle causing them to lose their autonomous functionality as a biological timekeeper[98]. One very potent strategy to overcome this phase-locking is to increase the genome copy number normally found in cells. The circadian clock model organism *Synechococcus elongatus* PCC 7942, for example, naturally harbors four genome copies per cell. Increasing the natural copy number weakens the otherwise drastic effect of DNA replication events on the overall gene density, because normally not all genome copies are copied at the same time but rather one by one. This causes the gene density to gradually increase, which almost eliminates a phase-locking to the cell cycle[98].

This is also an important part to remember when building a synthetic oscillator. Usually, synthetic constructs are introduced to cells on plasmids, which are randomly replicated with 10-100 copies per cell, thus are not susceptible to phase-locking by the cell cycle. However, if you want to integrate your construct on the genome it is important to figure out the position on the genome as this might be essential for functionality due to the timing of replication. For two synthetic oscillators, the repressilator (Fig. 6.2 a,[3]) and the

dual-feedback oscillator (Fig. 6.2 b, [4]), genome integration was analyzed [250]. Depending on the structure of the synthetic oscillator different strategies arise regarding the position on the genome and the distance between each factor. The repressilator (Fig. 6.2 a, [3]) shows almost no locking when all three genes are close to each other and replicated at the same time. Whereas, when genes are placed further apart from each other on the genome locking to the cell cycle can be observed, increasing proportionally with the distance between genes. This effect does not disappear when noise is added to the timing of replication for each gene [250]. Interestingly, the opposite is true for the dual-feedback oscillator (Fig. 6.2 b, [4]): strongest locking can be observed when genes are placed right next to each other on the chromosome, whereas locking diminishes when the distance between the two genes increases. However, locking never disappears as for the repressilator, even though noise has similar attenuating effects on the locking to the cell cycle in the dual-feedback oscillator [250].

As mentioned above, some cyanobacteria have found a way to make their circadian clock robust against the cell cycle by harboring multiple genome copies per cell, however, they also have an additional way that helps prevent phase-locking to the cell cycle. By using a protein oscillator at the core of the circadian clock instead of nested TTFLs, Cyanobacteria uncouple their timing machinery from the cell cycle. This is already sufficient to prevent phase-locking to the cell cycle, however, increasing the genome copy number is even better [98]. However, a PTO at the center of a circadian clock alone is also susceptible to high protein decay or dilution rates. Thus, for cells dividing faster than 24 hours, which show a high protein dilution, a TTFL in addition to a PTO is required to generate robust circadian rhythms [203,251].

Most cyanobacteria, which have been shown to harbor a clock have in addition to the protein oscillator multiple genome copies, thus, their timing system is remarkably robust against the cell cycle. However, the picocyanobacteria *Prochlorococcus*, which have a reduced genome and circadian clock compared to *Synechococcus elongatus* PCC 7942, have only one genome copy per cell. In these cyanobacteria, we see a strong coupling between the circadian clock and the cell cycle as they stably divide once per 24 hours with distinctly timed cell cycle phases [77,252].

In addition to the PTO at the core of the circadian clock, Cyanobacteria use a second protein network to transfer the clock state to gene regulation, which further uncouples the clock from the cell cycle. This push-pull network with RpaA at its center is complemented by SasA, CikA, and RpaB depending on the cyanobacterial strain (Fig. 6.1, [91]). Interestingly, the precision with which circadian clocks tell time increases with the number of oscillations, i.e. genes involved in the readout of the clock [96]. Furthermore, cross-regulatory interactions of proteins in the readout of a clock, e.g. the RpaA network, can increase the transmission of temporal information [96].

We have seen that there are various strategies to uncouple the oscillator from the cell cycle and make it

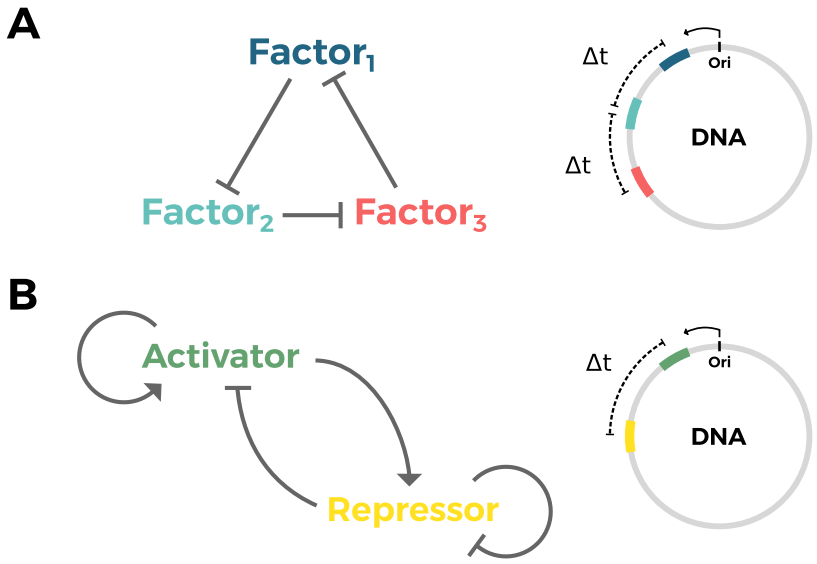


Figure 6.2: Representation of two synthetic oscillators and genome integration. (A) The network architecture of the repressilator by Elowitz & Leibler[3]: Factor1 represents a protein that represses the production of Factor2, which in turn represses the production of Factor3. Factor3 acts as a repressor for the gene expression of Factor1 again. The three factors are color coded and their location on a circular chromosome is shown on the right. The Δt represents the distance on the chromosome and thus the time difference between the replication of the genes. **(B)** The network architecture of the dual-feedback oscillator by Stricker and colleagues[4]: The activator activates its own production and enhances the production of the repressor. Whereas the repressor prevents its own production and suppresses the production of the activator. The two factors are color coded and their location on a circular chromosome is shown on the right. The Δt represents the distance on the chromosome and thus the time difference between the replication of the genes. Ori depicts the origin of replication and the arrow indicates the direction of replication. This figure is adapted from Paijmans and colleagues[250] where the effects of the position on the chromosome of these two synthetic oscillators are studied in more detail.

robust against changes in the protein and nucleotide levels. This can either be achieved by increasing the genome copy number, by placing the genes for the oscillator on plasmids with higher copy number, or by integrating one or even multiple PTOs into the circadian oscillator.

6.7 Summary

Circadian clocks are a prevalent system to coordinate gene expression of large proportions of the genome. The ability to anticipate recurring environmental changes provides a fitness advantage, which is harnessed by organisms of every group of life. However, circadian clocks are only a specialized form of timing mechanisms defined by three characteristics, i.e. entrainment, free running, temperature compensation, that produce oscillating gene expression. Regulating large parts of the genome in a timely fashion is an even more widespread phenomenon, which exhibits different forms besides circadian clocks that we did not cover in this review.

Over the last decade, different parts of the circadian clock in Cyanobacteria were characterized through mathematical modeling gaining insights into circadian regulation that go beyond Cyanobacteria allow-

ing us to understand general features of circadian clocks. The understanding of the molecular mechanisms underlying this timing system in Cyanobacteria including the ordered phosphorylation of the KaiC monomers[92], the interaction with KaiA and KaiB[230], as well as the influence on gene regulation[237] provided the basis for further studies. Through mathematical models, we have seen how this system can potentially robustly filter noise of the exogenous signal and maintain its inner rhythm[93]. Nevertheless, these models also showed us the limitations of this circadian clock when noise is too high or in an adverse moment[94]. Some of these models even though they devote themselves to the cyanobacterial circadian clock show us a universal way how to uncouple timing systems from the cell cycle making them robust against abrupt changing nucleotide concentrations[96,98]. The computer simulations by Troein and colleagues[241] helped to understand the environmental conditions that lead to the evolution of circadian clocks. However, these simulations took TTFL as the basis for their analyses and did not consider PTOs as well. Nevertheless, their results can be considered universally applicable for the evolution of circadian clocks.

In the future, it will be interesting to see if the size of the cyanobacterial circadian clock can be reduced based on the molecular understanding of interactions of protein factors and still retain full functionality. In addition, the design principles unraveled by mathematical models and the biochemical characterization of protein domains provide the basis for further protein engineering and synthetic biology approaches to building a timing system and oscillator from scratch. These newly designed timing systems can then be used to artificially regulate gene expression of engineered metabolic pathways to coordinate the production of metabolites as efficiently as possible in a natural noisy environment. Furthermore, the understanding of how biological systems robustly cope with noise can help to shape the future of other research areas beyond biology, e.g. in computing architecture and new algorithms for parallel computing.

Science is a way of thinking much more than it is a body of knowledge.

Carl Sagan

7 Discussion & Outlook

The circadian clock of *Synechococcus elongatus* PCC 7942 is one of the best-studied clock systems. The three core proteins, KaiA, KaiB, and KaiC have been thoroughly studied. Biochemical studies unraveled the 24-hour phosphorylation cycle as well as the stimulating interactions with KaiA and KaiB. The structure of all three proteins is known and just recently the structural data for the interactions during the full phosphorylation cycle has been shown[32,62]. The cyanobacterial circadian clock, however, does not only consists of these three proteins even though they build the core of the clock and are able to produce stable oscillation in a test tube with just ATP added[87]. In the natural context, the three proteins are embedded into a network of input and output factors. For the interaction factors SasA and CikA, functional data and even structural data are available clearly showing the role of these factors in the network of the circadian clock[56,62,109].

Since the protein sequences of all established clock factors in *Synechococcus elongatus* PCC 7942 are known, similarity analyses can be obtained to find orthologs to those proteins in other organisms. This has already been done in previous studies with small sample sizes for the core clock factors KaiA, KaiB, and KaiC[42,79], the input factors Pex[42], LdpA[42,187], and CikA[42] as well as the output factors SasA[42,161], LabA[42],

and RpaA[42]. For the core clock factors sequences from species outside of the cyanobacterial phylum were used[79], whereas the rest of the studies focused solely on cyanobacterial species. In order to get a comprehensive overview of the distribution and conservation of the clock factors, we decided to take advantage of the amount of available genomic data and search for orthologs in all known complete genomes[91]. Our findings confirmed those previous studies by Dvornyk and colleagues[79,161,187] and even extended our understanding of the distribution of the system by finding orthologs for multiple clock factors including the core factors KaiB and KaiC in a large number of bacterial and archaeal genera[91]. We further analyzed the conservation and diversity of the core clock factors and identified strong conservation of motifs for the interaction of KaiA and KaiB in cyanobacteria, where the β_2 sheet represents the most conserved part of the KaiB protein[32]. In addition, the autophosphorylation activity of KaiC was analyzed, which was experimentally confirmed by phosphorylation assays with purified proteins from Cyanobacteria and Archaea[91]. Those results confirm previous findings of phosphorylation studies with KaiC variants from *Synechocystis* sp. PCC 6803[33] strengthening the idea of a conserved functionality of KaiC orthologs.

Our analyses revealed large diversity in the composition of factors of this system among all identified organisms as well as within cyanobacteria. In order to get a better understanding of the diversity, we analyzed the co-occurrence of clock factors among Cyanobacteria as they show the most complete set of factors and thus most promising candidates for predicting conserved functionality across organisms. This analysis identified two sets of proteins: (i) factors, especially the core factors KaiB, KaiC, the input factor LdpA, and the output factors SasA, RpaA, and RpaB, that are found in almost all cyanobacteria, which lead us to the hypothesis that those factors comprise the core set needed for timing; (ii) factors that show strong co-occurrence since they are only found in a set of cyanobacteria. Here the core factor KaiA, the input and output factor CikA, and the output factor LabA are of special interest as they are important for circadian regulation in *Synechococcus elongatus* PCC 7942[56,57,63,202] but are lacking in *Prochlorococcus*. *Prochlorococcus* shows stable oscillation of gene expression under light-dark cycles, which disappear under continuous conditions[77,78]. Thus, we hypothesized that those factors are needed for circadian regulation. The results of the analysis give some indication for further experiments, e.g. introducing selected candidates of those factors into *Prochlorococcus* in order to potentially restore circadian regulation or to delete factors of *Synechococcus elongatus* PCC 7942 to further reduce the system. However, the study has some limitations as it only analyzed known factors and did not include all proteins from *Synechococcus elongatus* PCC 7942 to identify potentially new unknown factors involved in the circadian clock. Furthermore, the analysis needs constantly be extended whenever new proteins in the circadian clock are identified. In addition, the analysis only focuses on the presence or absence of factors rather than looking at sequence motifs that co-occur with the presence or absence of factors. However, such an advancement of the analysis requires much

more sophisticated methods to be done properly. Nevertheless, those analyses will be important in order to predict conserved functionality of the system between organisms. Depending on the environmental conditions not all of the known factors are needed to have a fully functional timing system. As mentioned before, even though the system in *Prochlorococcus* is not considered a circadian clock, as the oscillations in gene expression cease without the external stimulus, no obvious difference can be measured compared to a cyanobacterium that harbors a circadian clock like *Synechococcus elongatus* PCC 7942 under light-dark cycles. It could be shown that a deletion of *cikA* is detrimental for the functionality of the circadian clock in *Synechococcus elongatus* PCC 7942[63]. However, a single amino acid replacement from tyrosine to histidine at position 290 in the protein sequence of SasA is able to recover the phenotype[186]. Would it be sufficient to introduce this mutation into the protein sequence of SasA in *Prochlorococcus* in order to make the hourglass-like timing system a circadian clock or are other mutations in other factors needed as well? *Prochlorococcus* has, in fact, a different aromatic amino acid at this important position, however, not the newly introduced histidine[186]. Furthermore, just recently multiple single point mutations in the sequence of the output regulator RpaA have been identified increasing the growth rate of *Synechococcus elongatus* PCC 7942[253]. Detecting those naturally occurring point mutations and subsequently mapping them onto the presence or absence of other clock factors will be the next step. Results of this analysis might yield new candidates for further reduction of the system without losing the ability of robust circadian regulation. Manipulating and reducing the circadian clock to be more robust under natural conditions while either lowering the energy burden on the organisms to maintain factors of the clock or to increase productivity and growth on the organism in order to be more valuable for biotechnological production provides great potential. However, finding the minimal set of protein factors for the cyanobacterial circadian clock including potential sequence mutations in some of the factors is still a milestone to reach.

In foresight that our resources on Earth are limited and that we as a civilization are striving to colonize other planets in our solar system or even in other solar systems, there is potentially the challenge of making those planets habitable for us humans, e.g. creating an oxygenic atmosphere. One group of organisms that are among those that might enable us to colonize other planets by creating the needed conditions are Cyanobacteria with their ability to produce oxygen from carbon dioxide and water. This process is called oxygenic photosynthesis and has already created the environmental conditions we are living in today[18]. One thing that needs to be considered is the length of day on a potentially habitable planet. If the day length is far shorter or longer than the 24 hours that we have here on Earth right now, it is important to know how to adjust the period length of the timing system in order to better align the organism's clock with the day length to regulate the gene expression most effective.

There are two ways of addressing this problem either top-down by investigating the functionality of ex-

isting factors and how mutations or deletions of the factors affect the overall functionality of the system. The other way would be a bottom-up approach where a timing system would be evolved or constructed from scratch. For both cases finding strategies to manipulate the period length or robustness of the circadian clock can be analyzed by mathematical models. Mathematical models have already proven to provide invaluable insight into the mechanisms of the core clock[31]. However, so far those models mainly focused on the functionality of the core clock factors KaiA, KaiB, and KaiC or more general features of circadian clocks and feedback regulations[31]. Thus, more detailed models that include protein factors from the input or output network are needed. The insights of those models would probably mainly guide new experiments in the laboratory for a top-down strategy, but will also help to build a circadian clock from the ground up.

Building a circadian clock can either be done by assembling existing factors into a new network or evolving a new system with new factors from scratch. For the second case, one needs to understand what kind of environmental conditions the evolution of a timing system would favor. Since circadian clocks have evolved multiple times in the history of life on Earth[16], those conditions are supposed to be rather ubiquitous on Earth than a single event in Earth history. There are some first indications about these environmental conditions, i.e. noise in the external signal and seasonal changes in the length of the signal, by *in silico* evolutionary modeling by Troein and colleagues[241], however, it is yet to be shown that these conditions yield in the evolution of timing mechanisms *in vivo*.

Reconstructing the evolutionary conditions favoring circadian clocks might as well shed light on the origin of life as it would narrow down potential places for the origin of life on Earth. Circadian clocks are an adaptation to the cyclic environment on Earth as a result of Earth rotation itself. This rotation causes cycles of light or UV radiation or wet and dry phases through tides. It has been shown that both UV radiation and wet-dry cycles can produce molecules of life in primordial conditions that might have been on early Earth[254–257]. Some of these cyclic conditions even cause selective conditions[254,258]. Thus, anticipating upcoming cycles and preparing for them might have been a survival strategy early in the history of life on Earth.

Circadian clocks are a prevalent system on Earth used by organisms across the tree of life to anticipate recurring environmental changes. The circadian clock of cyanobacteria, with its model organism *Synechococcus elongatus* PCC 7942, is one of the best studied systems. The results of this thesis complete the missing piece in the understanding of the structural interaction of the core clock factors. Furthermore, the comprehensive conservation analysis yielded new insights into the distribution of the system among prokaryotes, as well as highlights the diversity and conservation among cyanobacteria. The results of this study provide the basis for further investigations regarding the co-occurrence of factors and mutations within these pro-

teins. Those studies will help to understand how to reduce the complexity of the cyanobacterial circadian clock. Last, through reviewing the discoveries achieved by mathematical modeling, this thesis identifies open questions in the understanding of the cyanobacterial system, in particular, and circadian regulation, in general, which goes as far as the evolution of circadian regulation and the origin of life on Earth.

A

Structures of the cyanobacterial circadian oscillator frozen in a fully assembled state

Joost Snijder, Jan M. Schuller, Anika Wiegard, Philip Lössl, Nicolas Schmelling, Ilka M. Axmann, Jürgen M. Plitzko, Friedrich Förster, Albert J. R. Heck (2017) Structures of the cyanobacterial circadian oscillator frozen in a fully assembled state, *Science*, 355 (6330), 1181–1184, doi: *10.1126/science.aag3218*

Table A.1: Masses of identified Kai protein and protein complex species. The molecular weights of the monomeric Kai proteins are shown in parentheses. Note that the residual masses expressed as mass deviation (%) originate from a small amount of residual buffer and solvent that remains attached to the ions, as routinely observed in native MS experiments. The table can be found online.

Table A.2: Unique crosslinks identified within the KaiCBA complex. Listed are charge state, mass and number of peptide spectrum matches (PSMs) for each crosslinked peptide pair. The crosslinked peptides (α and β) are individually characterized by their amino acid sequence (o = Met oxidation), the positions of the linked Lys residues, their molecular weight, the number of matched fragment ions (nonintegral numbers indicate a fragment ion matching to both peptides) and the N scores, as calculated by XlinkX. The table can be found online.

Table A.3: Fractional changes in Deuterium uptake of KaiA, KaiB and KaiC as free components compared to KaiCB and KaiCBA assemblies. The list includes all confidently identified peptides ordered by primary sequence on the protein level. The table can be found online.

Table A.4: Primers used for mutagenesis. Small letters indicate the inserted mutations resulting in exchange of single amino acids. The table can be found online.

Table A.5: Deuterium uptake plots for all KaiA (Data S1.), KaiB (Data S2.) and KaiC (Data S3.) peptides included in the analysis. Error bars represent standard deviations.

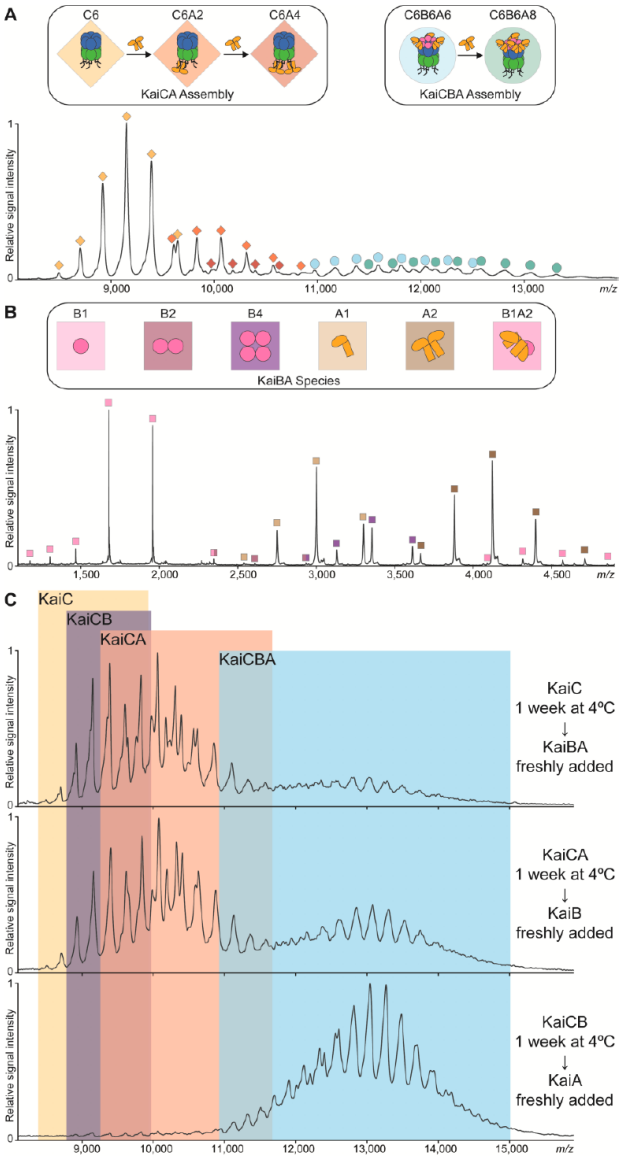


Figure A.1: Native MS of KaiCBA assembly. (**A**) Detailed view of the KaiCBA oscillator at 30 °C after 4 hours incubation. (**B**) Detailed view of the KaiCBA oscillator at 30 °C after 20 hours incubation, demonstrating the presence of $KaiB_1A_2$ subcomplexes. (**C**) Ordered assembly of KaiCBA complexes shows that formation of the KaiCB complex is limiting to formation of the KaiCBA complex.

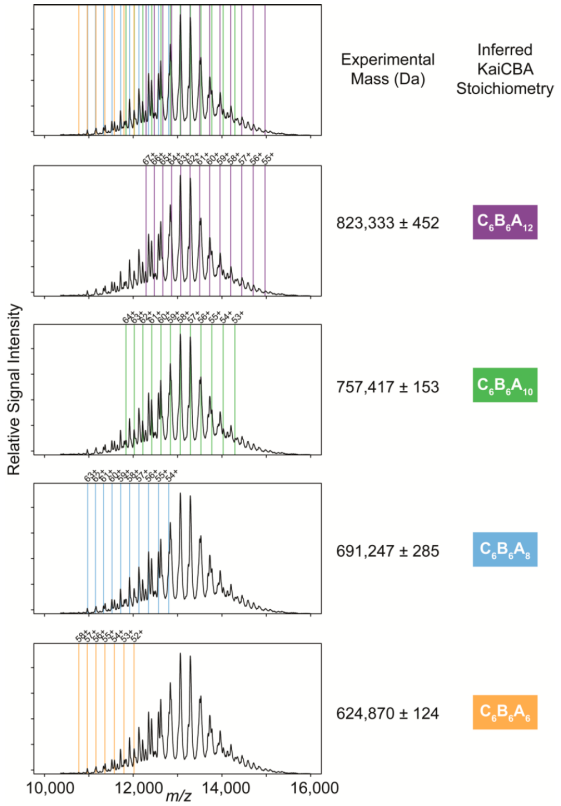


Figure A.2: Native electrospray ionization-mass spectra. Typical example displaying how species identification and molecular weight determination was performed from the native electrospray ionization-mass spectra. As a result of the electrospray ionization process, all different KaiCBA species are detected as differently charged ions. Based on these charge state distributions, the molecular weight of each species can be calculated. The theoretical mass-to-charge ratios (m/z) of the identified charge state distributions are depicted as vertical lines. As shown in the top panel, these charge state distributions can explain all major peaks of the mass spectrum, providing evidence that the listed KaiCBA complexes are the most abundant species in the sample. The displayed error range is established by taking into account all charge states of a given species, and is typically within 0.05% of the total molecular weight. All other spectra were evaluated in a similar manner.

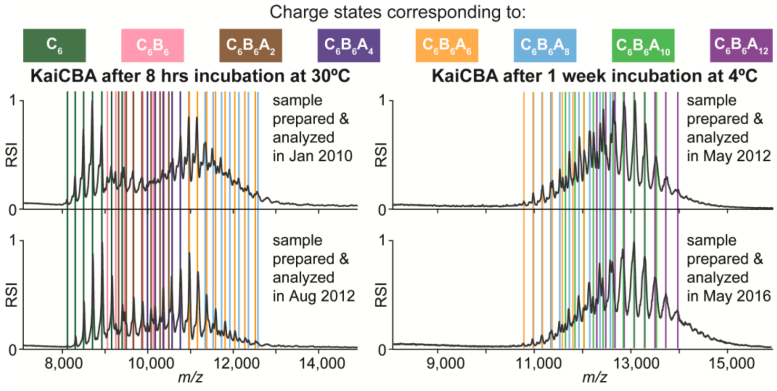


Figure A.3: Native MS measurements of the 8 hour timepoint. Repeated native MS measurements of the 8 hour timepoint from the 30 °C time course and the full KaiCBA assembly. Measurements were performed on different preps, spanning multiple years, by different experimentors. Peak assignments as explained in Fig A.2.

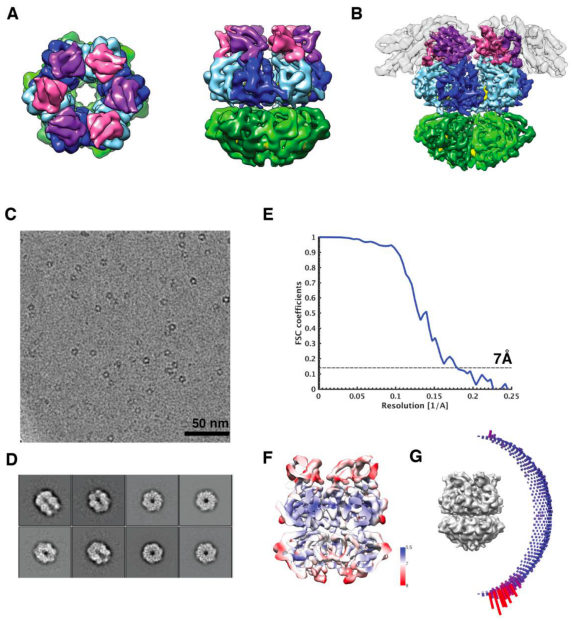


Figure A.4: Cryo-EM single particle analysis of the KaiCB complex. (**A**) Top and side view of the 3D reconstruction of the KaiCB complex. The color scheme of the cartoon representations is adapted from Figure 4.2. (**B**) Side view of the KaiCBA reconstruction with the density segments corresponding to KaiA rendered transparently, demonstrating the same overall architecture of the KaiCB part of this density. (**C**) Part of a cryo-EM micrograph of KaiCB. (**D**) Reference-free 2D class averages from the KaiCB dataset. (**E**) Fourier shell correlation (FSC) for two independently refined halves of the data. The curve for reconstruction with C_6 symmetry imposed during refinement is shown and the resolution at which FSC = 0.143. (**F**) Cut-open density of the KaiCB density colored according to its local resolution. (**G**) Orientation distribution plot of all particles that contribute to the reconstruction.

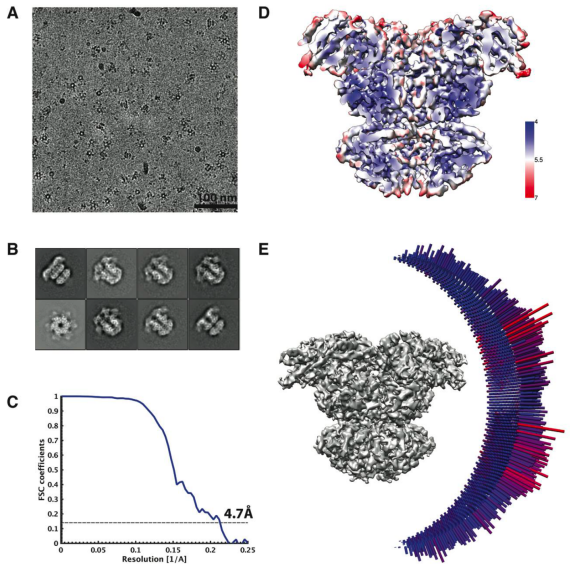


Figure A.5: Cryo-EM single particle analysis of the KaiCBA complex. (**A**) Part of a cryo-EM micrograph of KaiCBA. (**B**) Reference-free 2D class averages from the KaiCBA dataset. (**C**) Fourier shell correlation (FSC) for two independently refined halves of the data. Shown is the curve for reconstructions with C_6 symmetry imposed during refinement and the resolution at which FSC = 0.143. (**D**) Cut-open density of the KaiCBA density colored according to its local resolution. (**E**) Orientation distribution plot of all particles that contribute to the reconstruction.

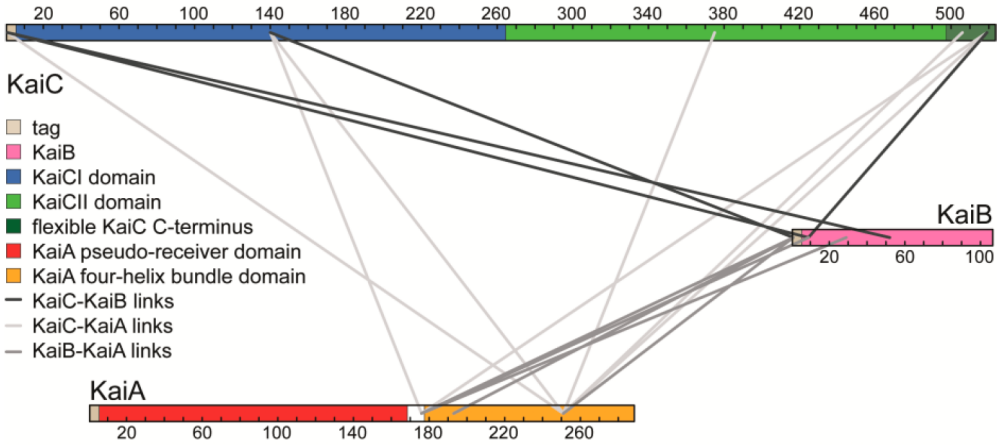


Figure A.6: Intermolecular cross-links identified in the KaiCBA complex. In total, 66 unique cross-links were identified at 1% false-discovery rate (Table A.2). In agreement with the EM density-based KaiCBA model, the cross-links indicate spatial proximity between the KaiA C-terminal α-helical domain and the N-terminal region of KaiB. KaiCBA complex formation via the KaiC-CI domain, as suggested by cryo-EM, is supported by six cross-links connecting KaiA and KaiB with KaiC-CI. KaiA and KaiB were also found to be cross-linked with the C-terminal tails of KaiC, which are highly mobile, as evidenced by the absence of assignable EM density. Hence, these crosslinks, likely arising from short-lived interactions among the KaiC tails, KaiA and KaiB, are difficult to interpret with respect to our model.

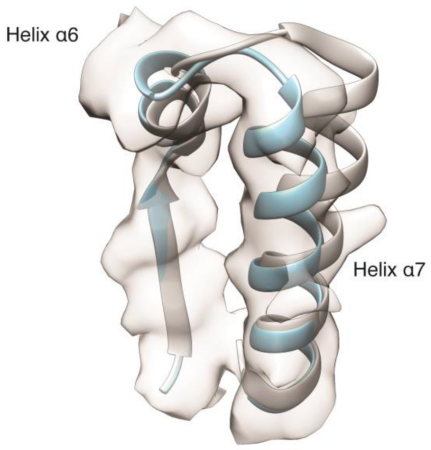


Figure A.7: KaiC adopts the posthydrolysis conformation. Overlay of the EM density with the pre- (grey, PDB 4TLA ChainA) and post-hydrolysis state (blue, PDB 4TLA ChainC) of helices 6 and 7 of KaiC-CI from Abe and colleagues[134].

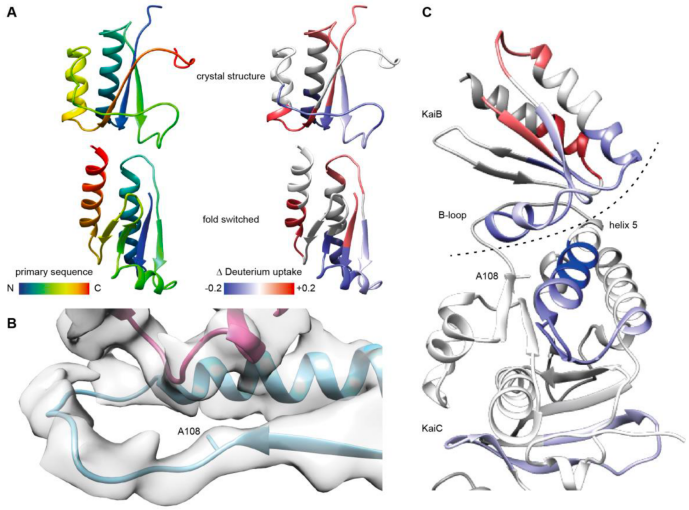


Figure A.8: Details of the KaiB-KaiC interaction. (**A**) Fractional change in deuterium uptake of free KaiB and KaiC-bound KaiB after 60 min exposure mapped on the KaiB fold observed in the crystal structure (PDB code 4KSO) or the KaiCBA map (fold switched state). (**B**-**C**) Detailed views of the KaiB-C interface in the KaiCBA model. The highly conserved KaiC Ala^{108} residue is highlighted. (**B**) EM density at the KaiB-C interface. The color scheme of the cartoon representations is adapted from Figure 4.2. (**C**) Model of the KaiB-C interface colored according to changes in deuterium uptake between the free components and the KaiCB complex after 60 minutes exposure. The labeled KaiC helix 5 shows the sharpest decrease in deuterium uptake, indicating strong protection upon KaiB binding. The color scale is the same as in (**A**).

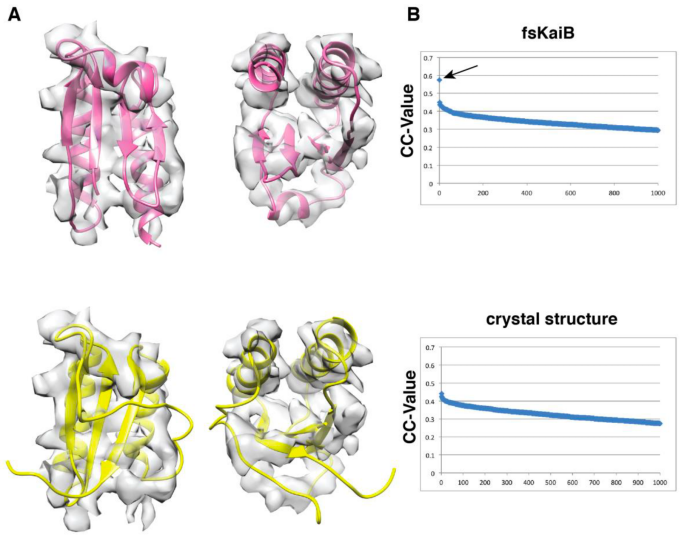


Figure A.9: Details of the KaiB fold-switch. (**A**) EM density of KaiB with the fitted structure of (i) fold-switched KaiB model (pink) and (ii) crystal structure (yellow). (**B**) Plot of the cross-correlation value between the respective model and the KaiB density. Shown are the top 1,000 positions from the global fitinmap search in UCSF Chimera. The arrow indicates that the best solution for the fold-switched model has a higher CC than the crystal structure.

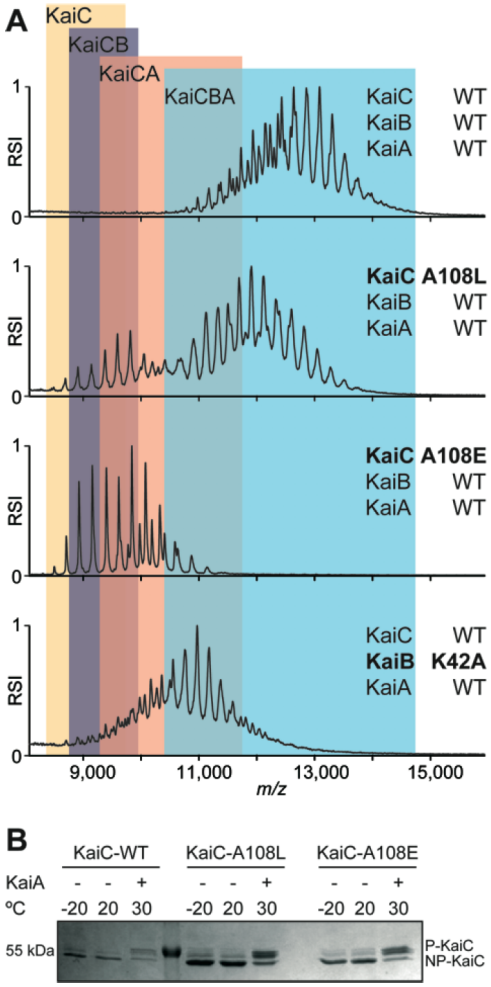


Figure A.10: Effect of Kai protein mutations on assembly of KaiCBA complexes. (**A**) Native MS of KaiCBA assembly using WT, KaiC-A108L, KaiC-A108E and KaiB-K42A. RSI = Relative signal intensity. (**B**) Verification of KaiA-stimulated KaiC phosphorylation activity in the KaiC mutants.

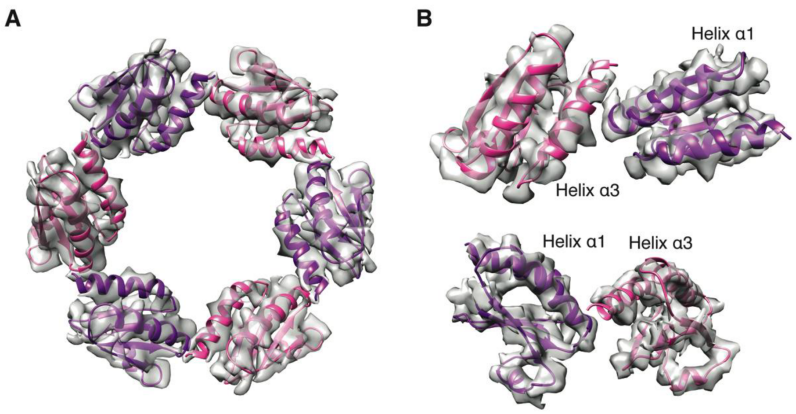


Figure A.11: Details of the KaiB-KaiB interaction. (**A**) Top view of the EM density of the KaiB ring and the fitted models. (**B**) Close-up of the KaiB-KaiB interacting region formed between the C-terminal part of helix 3 and the N-terminal part of helix 1.

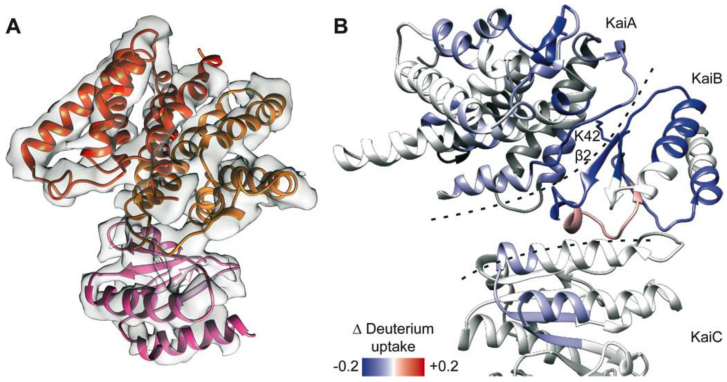


Figure A.12: Details of the KaiA-KaiCB interaction. (**A**) EM density representing Kai A and KaiB. The color scheme of the cartoon representations is adapted from Figure 4.2. (**B**) Detailed view of the KaiA-KaiCB interface in the KaiCBA model, colored according to changes in deuterium uptake between free KaiA and the KaiCBA complex, or between KaiCB and KaiCBA complexes in the case of KaiC and KaiB. The KaiA-binding β2-strand of KaiB and the functionally important KaiB Lys^{42} residue are additionally labeled.

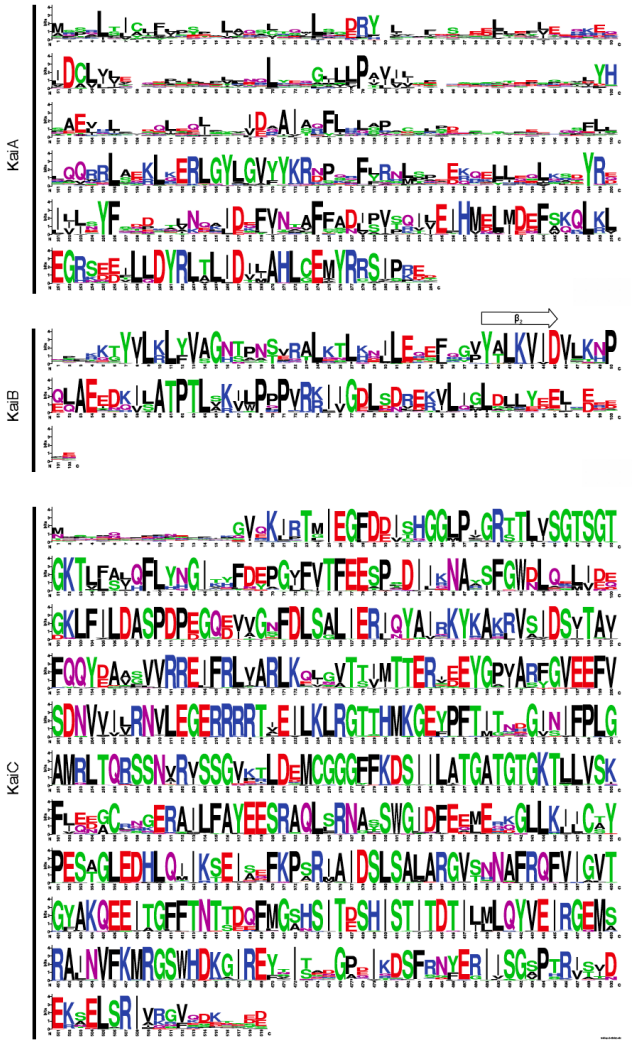


Figure A.13: Kai protein sequence conservation. Weblogos representing multiple alignments of KaiA, KaiB and KaiC from Cyanobacteria containing all three proteins. Amino acids are colored according to their chemical properties: polar (green), neutral (purple), basic (blue), acidic (red), hydrophobic (black).

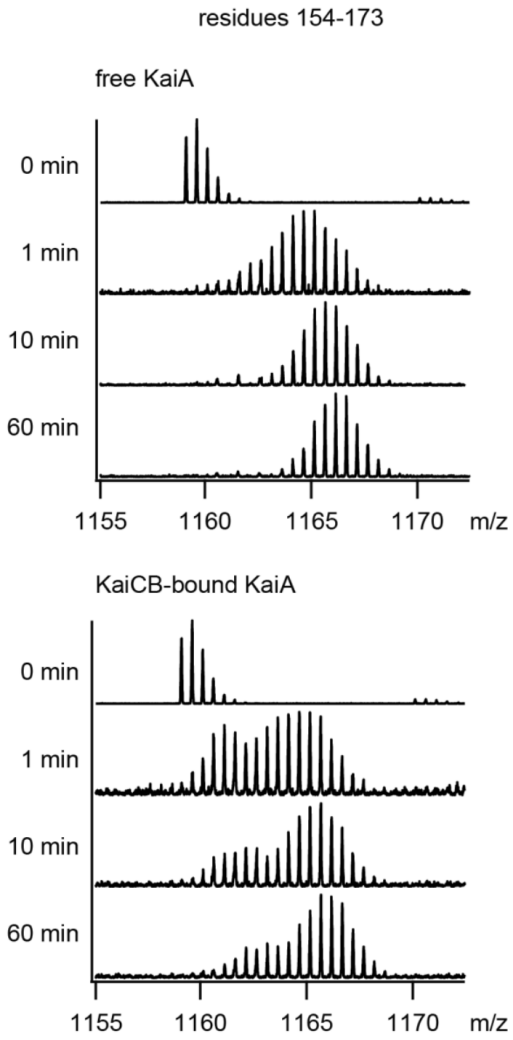


Figure A.14: Asymmetric binding of KaiA to KaiCB as detected by HDX-MS. The main KaiCB binding peptide of KaiA displays bimodal exchange in the KaiCBA complex. The relative abundance of each distribution (approximately 1:2, protected:unprotected) is consistent with asymmetric binding of the KaiA dimer to KaiCB. Note that the complexes are prepared with a 1.5-fold stoichiometric excess of KaiA; symmetric binding would have resulted in a bimodal distribution with a relative abundance of 2:1 protected:unprotected.

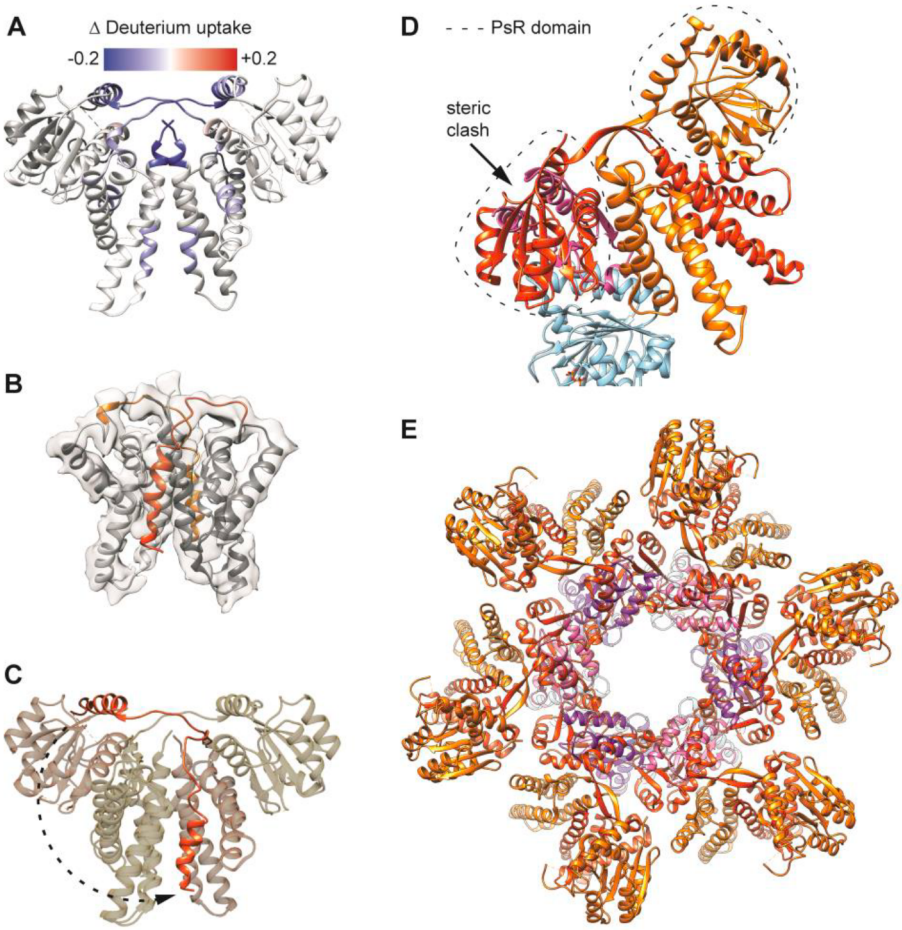


Figure A.15: KaiA sequestration is accompanied by displacement of its PsR domain. (**A**) Fractional change in deuterium uptake in free KaiA and KaiCB-bound KaiA after 60 minutes exposure mapped on the KaiA crystal structure (PDB code 1R8L). (**B**) EM density assigned to KaiA. The model of KaiA's C-terminal domain is shown in grey. Parts of the KaiA PsR domain and linker region to which an EM density could be assigned ($KaiA^{147-172}$) are shown in orange. (**C**) Alignment of the KaiA crystal structure with a KaiA dimer from the KaiCBA model. $KaiA^{147-172}$ are highlighted in orange, suggesting a wide displacement of the KaiA PsR domain via the linker region upon binding to KaiCB. (**D**) KaiA crystal structure superimposed with the KaiCB model. The non-displaced PsR-domain clashes with the KaiB protein, as indicated by an arrow. (**E**) Top-view of the KaiCB model and full-length KaiA.

B

Minimal tool set for a prokaryotic circadian clock

Nicolas M. Schmelling, Robert Lehmann, Paushali Chaudhury, Christian Beck, Sonja-Verena Albers, Ilka M. Axmann, and Anika Wiegard (2017) Minimal tool set for a prokaryotic circadian clock, *BMC Evolutionary Biology*, 17:169, doi: *10.1186/s12862-017-0999-7*

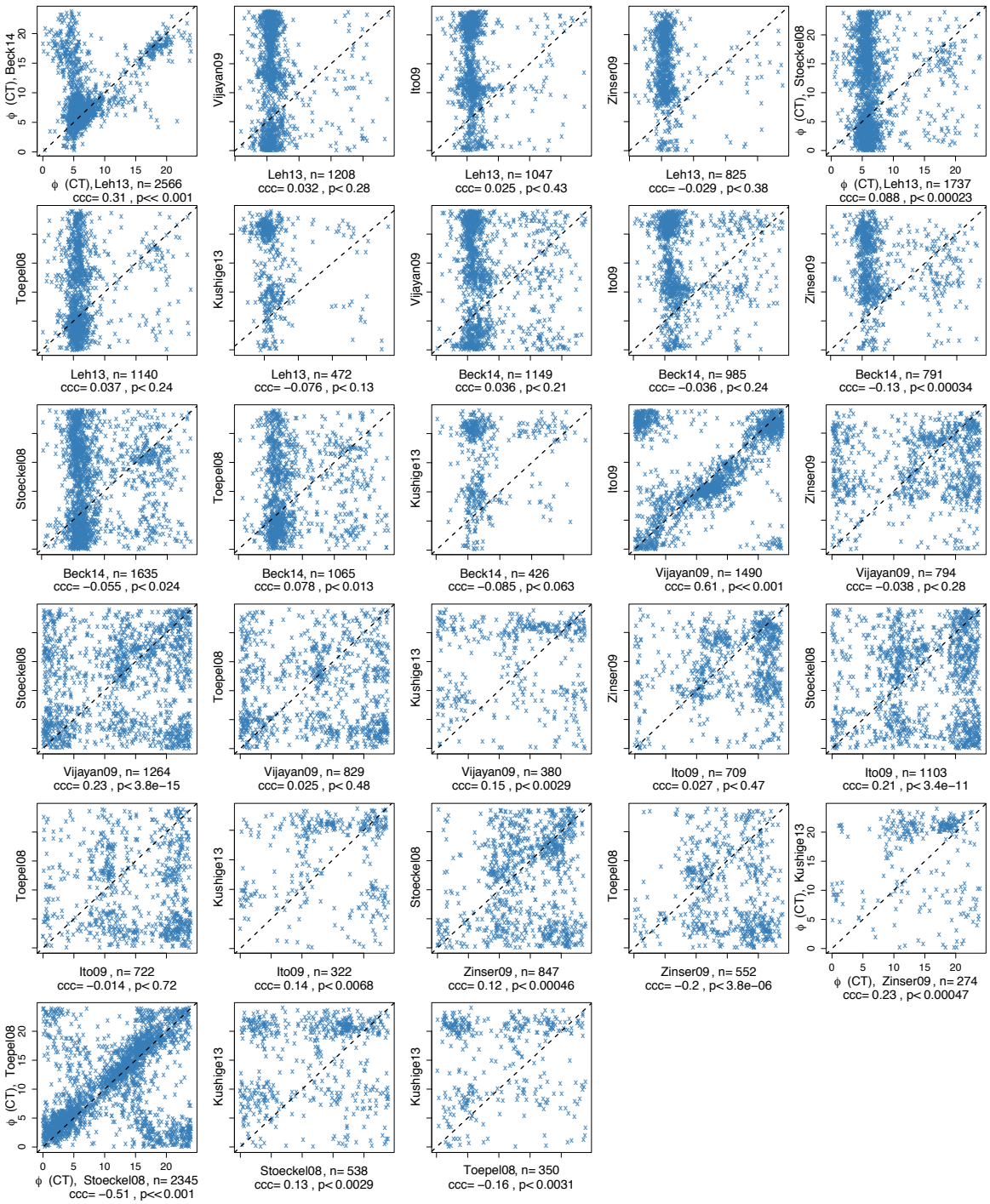


Figure B.1: Expression phase similarity across cyanobacterial strains by pairwise comparison between available datasets. All possible pairwise combinations of available circadian datasets are compared with respect to peak expression phase, considering only genes, which oscillate significantly (fdr < 0.05) in both datasets. While phases are compared directly for same-strain combinations, gene pairs across different strains are derived via homology prediction. The respective dataset is shown on each axis, the count of homologous genes significantly oscillating is provided with the x-axis label, together with the circular correlation coefficient ϱccc and the resulting p-value.

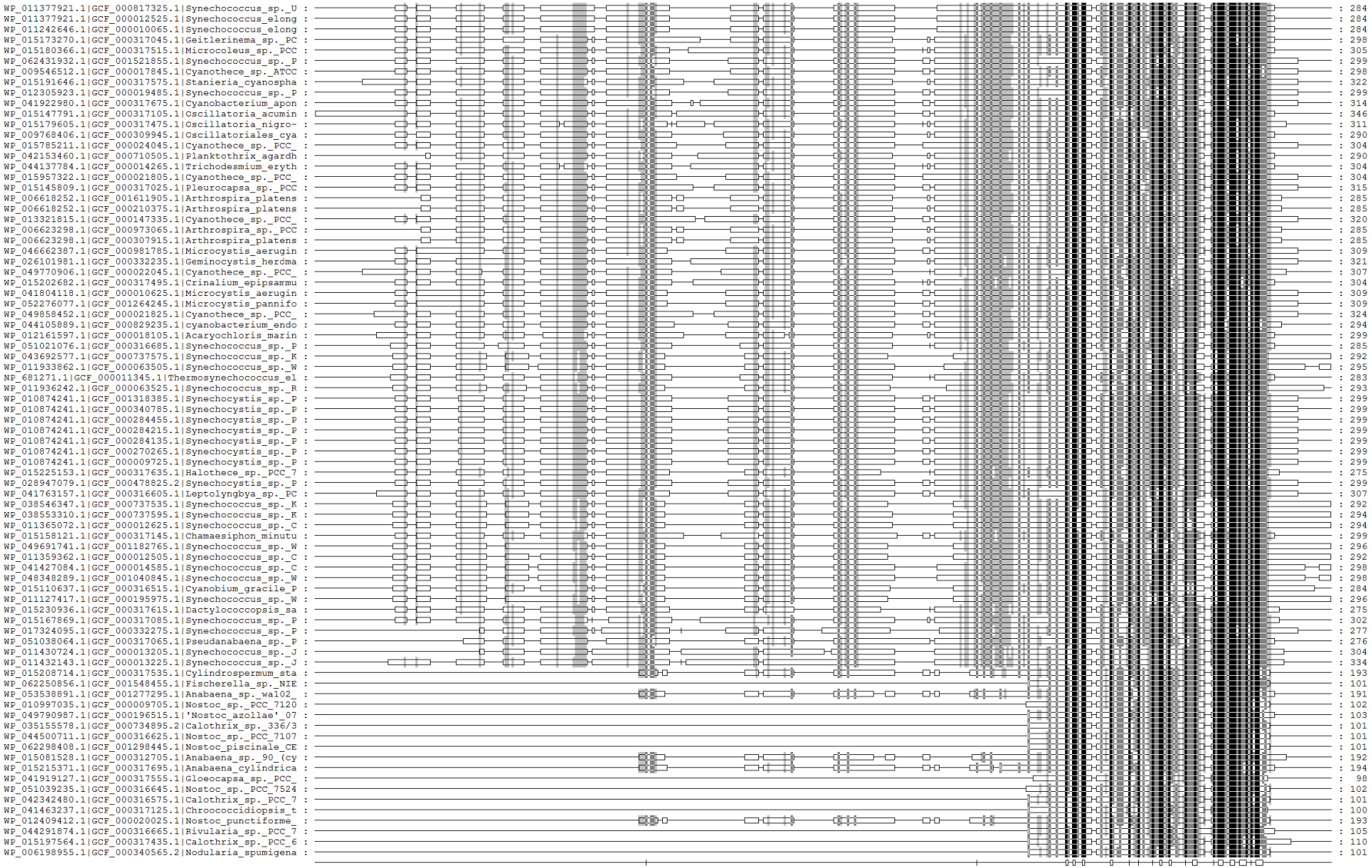


Figure B.2: Multiple sequence alignment of KaiA. Conserved regions are highlight: $\leq$100% (black), $\leq$80% (grey), $\leq$60% (light grey), $\leq$40% (white).

Table B.1: Proteins of the circadian clock used as queries for the reciprocal BLAST analysis. The association of KaiB2/B3 and KaiC2/C3 varies between the Cyanobase Database and Dvornyk and colleagues (2003)[79]

Organism Name	Function	Protein Name	Gene ID
Synechococcus elongatus PCC 7942	Core	KaiA	Synpcc7942_1218
		KaiB	Synpcc7942_1217
		KaiC	Synpcc7942_1216
	Input	Pex	Synpcc7942_0677
		LdpA	Synpcc7942_0624
		NhtA	Synpcc7942_2160
		PrkE	Synpcc7942_0600
		IrcA	Synpcc7942_2383
		CdpA	Synpcc7942_1604
	Input/Output	CikA	Synpcc7942_0644
	Output	SasA	Synpcc7942_2114
		LabA	Synpcc7942_1891
		LalA	Synpcc7942_1143
		Crm	Synpcc7942_0096
		RpaA	Synpcc7942_0095
		RpaB	Synpcc7942_1453
		CpmA	Synpcc7942_1168
Synechocystis sp. PCC 6803	Core	KaiB1	slr0757
		KaiC1	slr0758
		KaiB2	sll1596
		KaiC2	sll1595
		KaiB3	sll0486
		KaiC3	slr1942

Table B.2: Distribution of *Synechococcus* elongatus PCC 7942 based circadian clock proteins in the four main groups. The percentage indicates the coverage of the orthologs per group.

Group	Core Clock	Input Pathway	Output Pathway
Cyanobacteria	KaiA (94.29%), KaiB (94.29%), KaiC (100%)	Pex (62.86%), LdpA (97.14%), CikA (94.29%), NhtA (48.57%), PrkE (85.71%), IrcA (97.14%), CdpA (74.29%)	CikA (94.29%), SasA (100%), LabA (88.57%), LalA (91.43%), Crm (45.71%), RpaA (97.14%), RpaB (100%), CpmA (97.14%)
Proteobacteria	KaiB (28.13%), KaiC (100%)	CikA (56.25%), NhtA (26.56%), PrkE (51.56%)	CikA (56.25%), SasA (32.81%), LabA (4.69%), LalA (39.06%), Crm (20.31%), RpaA (7.81%), RpaB (48.44%), CpmA (21.86%)
Archaea	KaiB (6.06%), KaiC (100%)	CikA (12.12%), NhtA (22.73%), PrkE (13.64%)	CikA (12.12%), SasA (3.03%), LabA (1.51%), LalA (33.33%), Crm (7.58%), CpmA (63.64%)
Other	KaiB (31.43%), KaiC (100%)	CikA (25.71%), NhtA (17.14%), PrkE (71.43%), IrcA (54.29%)	CikA (25.71%), SasA (31.43%), LalA (5.71%), Crm (2.86%), RpaA (11.43%), RpaB (48.57%), CpmA (51.43%)

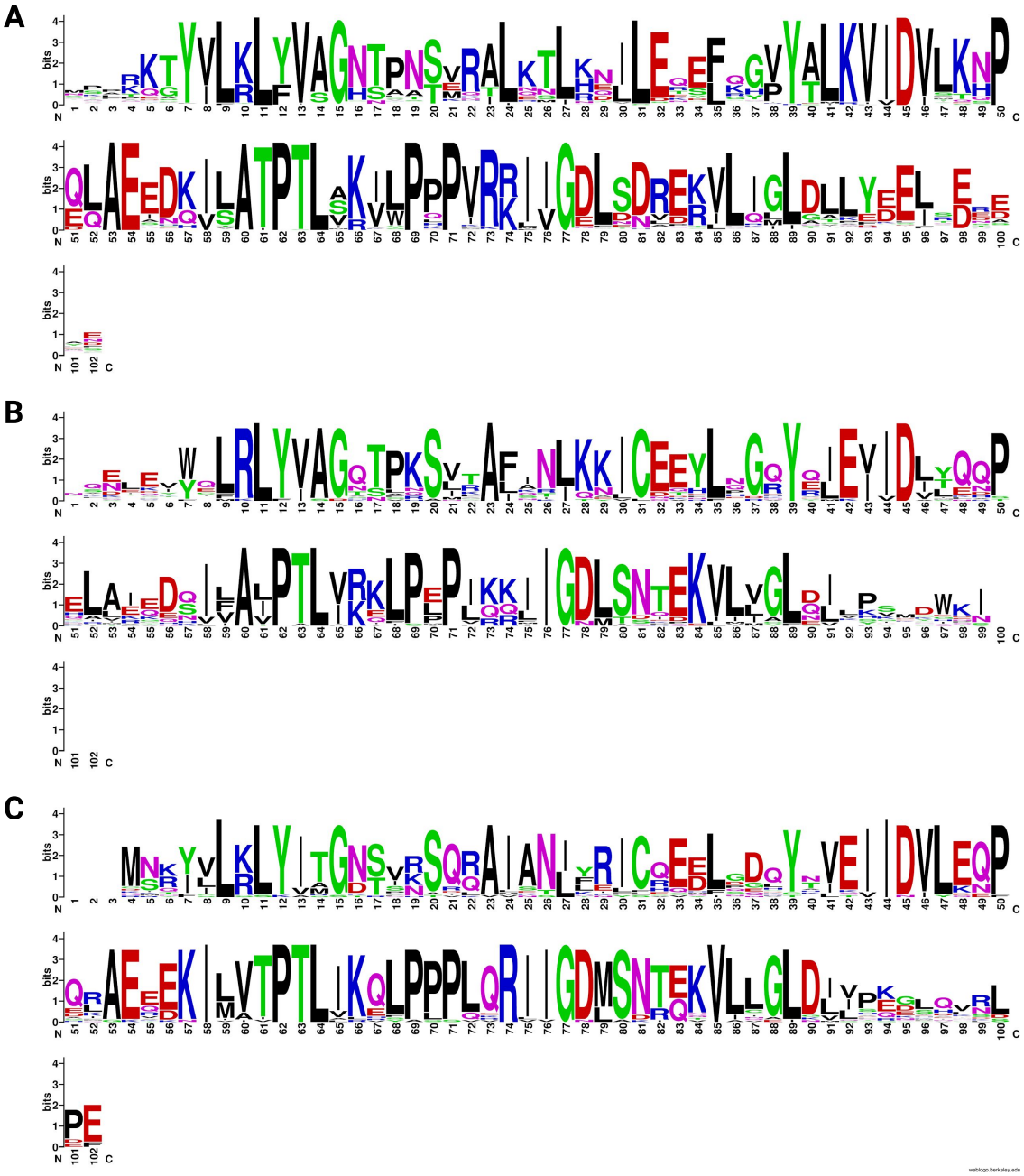


Figure B.3: WebLogos of KaiB multialignements. All multialignments are mapped onto the sequence of KaiB from *Synechococcus elongatus* PCC 7942. (**A**) KaiB1 from *Synechocysits* sp. PCC 6803 multialignment. (**B**) KaiB2 from *Synechocysits* sp. PCC 6803 multialignment. (**C**) KaiB3 from *Synechocysits* sp. PCC 6803 multialignment.

Table B.5: Diurnal core CLOGs across cyanobacterial strains excluding *Microcystis aeruginosa* PCC 7806. In every CLOG (row) at least one gene of each of the considered datasets (columns) exhibited diurnal expression. The table can be found online.

Table B.3: A collection of circadian and diurnal expression datasets in the cyanobacterial clade. Datasets with assigned abbreviation were used to determine the core oscillatory genome. Information is provided about the ability to fix nitrogen, the habitat (freshwater F, saltwater S), the total number of genes, the publication reference, the absolute and relative number of diurnally expressed genes reported in the original publications, the applied light and sampling schema, the experimental culture conditions, and the methods for microarray normalisation and oscillating gene detection. References to datasets employed in the following comparison are shown bold.

Strain	Strain Abbrev.	N_2 Fixation	Habitat	Total Genes	Ref.	Dataset Abbrev.	Diurnal Genes	Light Conditions	Culture Conditions	Normalization / Oscillation Detection
Prochlorcoccus marinus MED4	ProMED4	-	M	1766	69	zinser09	1403 (79%)	LD (14:10, T_{samp} 2h)	Pro99 Medium, stirred, 24 °C, Batch	RMA, Fourier Analysis
Synechocystis sp. PCC 6803	Syn6803	-	F	3628	70	leh13	1133 (31%)	LD (12:12, T_{samp} irreg.)	BG11 Medium, air bubbling, 30 °C, Batch	LOS, Fourier Analysis
					71	beck14	(27%)	LD (12:12, T_{samp} 2h)	BG11 Medium, air bubbling, 30 °C, Batch	LOS, Fourier Analysis
					171	-	1349 (37%)	LD (14:10, T_{samp} 2h)	BG11 Medium, stirred, 3% CO_2air bubbling, 27 °C, Turbidostat	Standorf Microarray Database standard, ANOVA and correlation with light
					172	-	237 (9%)	LL (T_{samp} 4h)	BG11 Medium, air bubbling, stirred, 30 °C, Batch with manual dilution	LOWESS, modified Cosiner
Synechococcus elongatus PCC 7942	Syc7942	-	F	2719	51	vijayan09	1748 (64%)	LL (T_{samp} 4h)	BG11 Medium, 1% air CO_2 bubbling, 30 °C, Continuous Culture	Loess and Quantile, Fourier Analysis
					72	ito09	800 (29%)	LL (T_{samp} 2h)	BG11 Medium, 30 °C, Continuous Culture	Replicate Mean Polishing, Correlation to Sine
Microcystis aeruginosa PCC 7806	Mic7806	-	F	6360	76	straub11	1344 (21%)	LD (12:12, T_{samp} irreg.)	BG11 Medium, 1% air CO_2 bubbling, 22 °C, Batch	LOWESS, significant difference to CTo
Anabaena sp. PCC 7120	Ana7120	•	F	6222	75	kushige13	78 (1.25%)	LL (T_{samp} 4h)	BG11 + N Medium, 30 °C, Contiuous Culture	Replicate Mean Polishing, Correlation to Sine
Cyanothece ATCC 51142	Cyn51142	•	M	5354	73	stoeckel08	1445 (≈ 30%)	LD (12:12, T_{samp} 4h)	ASP2 Medium, 30 °C, air bubbling, Batch	LOWESS, Correlation Network
					74	toepel08	1424 (≈ 20%)	LD (12:12)/24h, LL T_{samp} 4h)	ASP2 Medium, 30 °C, Airlift Bioreactor	LOWESS, Differential Expression
					173	-	1400 (27%)	LD (6:6, T_{samp} 2h)	ASP2 Medium, 30 °C, Airlift Bioreactor	LOWESS, Correlation Network

Table B.4: Oligonucleotides used for cloning.

Number	Primer Name	Primer Sequence	ORF
1	fw-kaiC1-Syn6714-BamHI	CTACGGATCCAACTCACCCATCGTTAACG	D082_30580
2	rev-kaiC1-Syn6714-NotI	GAAGCGGCCGCCTACTCGACGGTTTTATC	D082_30580
3	fw-kaiC1-Npun29133-BamHI	CTACGGATCCAGTCAAAACGAGCAAG	Npun_R2886
4	rev-kaiC1-Npun29133-NotI	CGAAGCGGCCGCTTAGGGTTCGGAAC	Npun_R2886
5	fw-kaiC1-Cy7424-BamHI	CATAGGATCCAATGAACCCATTCCCAACG	PCC7424_0599
6	rev-kaiC1-Cy7424-NotI	CATTGCGGCCGCTTATTCATCTAAAGTTTTATC	PCC7424_0599
7	fw-kaiC3-Cy7424-BamHI	CGAAGGATCCAATCAAGACAACGAAC	PCC7424_3006
8	rev-kaiC3-Cy7424-NotI	CTGTGCGGCCGCCTAAGACCGTTCTTCAAAC	PCC7424_3006
9	fw-kaiC3-Mic7806-BamHI	CTACGGATCCACGCAAAATAATCCCCTAG	IPF_2046
10	rev-kaiC3-Mic7806-NotI	GAAGCGGCCGCCTAACTACGATCCTCA	IPF_2046
11	fw-KaiC3-PH_RS03935 BamHI (DSM 12428)	CCCGGATCCGATGCTCTTAATTGTTGGAACTCC	Gene ID: 1443164
12	rev-KaiC3-PH_RS03935 HindIII (DSM 12428)	CGGGGAAGCTTTTACTCATAAATTTCCACCCCTC	Gene ID: 1443164
13	fw-KaiC3-OCC_RS02010 PstI (DSM 5473)	GGGCTGCAGATGAGCAGAACGGGAATTG	Gene ID: 16548747
14	rev-KaiC3-OCC_RS02010 HindIII (DSM 5473)	GCCGGGAAGCTTTTATTCATAAATTTCCACCC	Gene ID: 16548747

Abbreviations

A, Ala	Alanine
Å	Ångstrom
AA	Amino acid
ADP	Adenosine diphosphate
ARG	Autoradiography
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
°C	Degree Celsius
CBB	Coomassie Brilliant Blue
CID	Collision-induced Dissociation
CikA	circadian input kinase A
CpmA	circadian phase modifier A
Crm	circadian rhythmicity modulator
C-terminal	End of the protein sequence, marked by carboxyl (-COO^-) group
Cys	Cysteine
Da	Dalton
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
e	Electron
E	Glutamic acid
EDTA	ethylenediaminetetraacetic acid
e.g.	exempli gratia (for example)
EM	Electron Microscopy
EMDB	Electron Microscopy Data Bank
ETD	Electron-transfer Dissociation
FDR	False-Discovery Rate
FSC	Fourier Shell Correlation
FTP	File Transfer Protocol
G	Glycine
GST	Glutathione S-transferase
HDX	hydrogen-deuterium exchange
i.e.	id est (in other words)
K, Lys	Lysine
Kai	kaiten (japanese for cycle)
S-KaiC, KaiCpS/T	KaiC phosphorylated only on S^{431}
ST-KaiC	KaiC phosphorylated on both S^{431} and T^{432}
T-KaiC	KaiC phosphorylated only on T^{432}
U-KaiC	unphosphorylated KaiC

L	Leucine
LabA	low-amplitude and bright A
LalA	labA-like A
LC	Liquid Chromatography
LD	Light Dark
LdpA	light dependent period A
M	Molar
MAPK	Mitogen-activated protein kinase
Mb	Megabases
Met	Methionine
MES	2-ethanesulfonic acid
MS	Mass Spectrometry
NCBI	National Center for Biotechnology Information
NMR	Nuclear magnetic resonance
N-terminal	Beginning of the protein sequence, marked by anime ($-NH_2$) group
ODE	ordinary differential equations
PAGE	Polyacrylamide gel electrophoresis
PCC	Pasteur Culture Collection
PDB	Protein Database
Pex	period extender
PsR	Pseudoreceiver
PTO	Post-translational oscillator
PQ	Plastoquinone
RpaA	regulator of phycobilisome associated A
RpaB	regulator of phycobilisome associated B
SCX	Strong Cation Exchange
SDS	Sodium Dodecyl Sulfate
Ser	Serine
SasA	Synechococcus adaptive sensor
sp.	species
T, Thr	Threonine
Tris	tris-(hydroxymethyl)-aminomethane
TTFL	transcriptional-translational feedback loop
UV	ultraviolet
v	Volume
V	Volt
w	Weight
z	Charge

Listing of figures

List of Publications

The thesis consists of the following published research and review articles:

Chapter 4

Joost Snijder, Jan M. Schuller, Anika Wiegard, Philip Lössl, Nicolas Schmelling, Ilka M. Axmann, Jürgen M. Plitzko, Friedrich Förster, Albert J. R. Heck (2017) Structures of the cyanobacterial circadian oscillator frozen in a fully assembled state, *Science*, 355 (6330), 1181–1184, doi: *10.1126/science.aag3218*

Chapter 5

Nicolas M. Schmelling, Robert Lehmann, Paushali Chaudhury, Christian Beck, Sonja-Verena Albers, Ilka M. Axmann, and Anika Wiegard (2017) Minimal tool set for a prokaryotic circadian clock, *BMC Evolutionary Biology*, 17:169, doi: *10.1186/s12862-017-0999-7*

Chapter 6

Nicolas M. Schmelling and Ilka M. Axmann (2018) Computational modeling unravels the precise clockwork of cyanobacteria, *Interface Focus*, 20180038. doi: *10.1098/rsfs.2018.0038*

Other contributions, which are not the subject matter of the present thesis:

Articles

1. Nicolas Schmelling (2016) Selbstlernende Computerprogramme und Evolution, *Wikimedia Blog*, *https://blog.wikimedia.de/2016/12/01/selbstlernende-computerprogramme-und-evolution-ein-projektbericht-aus-dem-fellow-programm-freies-wissen/*

2. Nicolas Schmelling (2017) Data Intensive Science, *Research Ideas and Outcomes*, 3: e12032. doi: *10.3897/rio.3.e12032*

3. Nicolas M. Schmelling and Ilka M. Axmann (2018) What Time Is It? - Biological Oscillators to Robustly Anticipate Changes, In Milan Češka & David Šafránek (Eds.), *Computational Methods in Systems Biology* - 16th International Conference, CMSB 2018 Brno, Czech Republic, September 12-14, 2018 Proceedings. (pp. XIII - XIV). Cham, Switzerland: Springer International Publishing, doi: *10.1007/978-3-319-99429-1*

Poster

1. Nicolas Schmelling (2016) Open Science 4.0 - Transforming The Way We Do (Open) Science Using Digital Tools, *Figshare*, doi: *10.6084/m9.figshare.5383702.v1*

2. Nicolas Schmelling (2018) Origin of Earth and Circadian Clocks, *Figshare*, doi: *10.6084/m9.figshare.6115871.v1*

Data Sets

1. Nicolas Schmelling (2017) Minimal Tool Set for a Prokaryotic Circadian Clock Raw Data, *Figshare*, doi: *10.6084/m9.figshare.3823902.v3*

2. Nicolas Schmelling (2017) Minimal Tool Set for a Prokaryotic Circadian Clock Processed Data, *Figshare*, doi: *10.6084/m9.figshare.3823899.v3*

Protocols

1. Nicolas Schmelling (2016) Reciprocal Best Hit BLAST (Version 1), *protocols.io*, doi: *10.17504/protocols.io.grnbv5e*

2. Nicolas Schmelling (2016) Multiple Alignments and Weblogo (Version 1), *protocols.io*, doi: *10.17504/protocols.io.gscbwaw*

3. Nicolas Schmelling (2018) Reciprocal Best Hit BLAST (Version 2), *protocols.io*, doi: *10.17504/protocols.io.q3rdym6*

Software

1. Nicolas Schmelling (2017) schmelling/reciprocal_BLAST (Version v0.1), *Zenodo*, doi: *10.5281/zenodo.229910*

Acknowledgments

First and foremost, I particularly want to thank Ilka Axmann for everything she did for me over the past five years as a Master and doctoral student. I've learned so much from having the freedom to try out various things within my research project and in other things iGEM, mentoring, or teaching. I most likely would not have those opportunities in another group and I'm really thankful for this.
Many thanks go to Matias Zurbriggen. First for supporting me as my mentor during my doctorate and the great discussion about my project. Second for the great collaboration organizing the iGEM team here at the university in the last years.
I especially thank Charles Ofria for hosting me in his lab at Michigan State University and taking the time and effort to help me learn more about evolutionary modeling and apply it to my project.
A next big thank you goes the people in the open science and cyanobacterial community, especially to Titus Brown, Daniel Mietchen, and Lenny Teytelman for their support and introduction to open science.
I want to thank my colleagues Christian Beck, Robert Lehmann, and Anika Wiegard for their fantastic help in introducing me to the circadian clock of cyanobacteria and helping me with a plethora of problems that I encountered during my research project. Furthermore, I want to thank all of the lab members and colleagues I had the pleasure to work with for the great discussion on science and random things. Here I particularly want thank Stephanie Dittebrand, Kai Hußnätter, and Eric Young for endless encouragements, for discussing research projects and ideas, for starting new projects, and for proofreading blog posts and other writings.
Last but not least I want to thank all those great peoples that luckily do not work in science and give me a break from all this. Where would I be without the support of the Sharks? So thank you Robin Bröchler, Kevin Junggeburth, and Stefan Wexel. Many thanks also go to my family for their love, guidance, and support, especially to my better half Leonie-Alexa Koch.

References

[1] M. C. Weiss, F. L. Sousa, N. Mrnjavac, M. Clamp, S. Neukirchen, M. Roettger, S. Nelson-Sathi, and W. F. Martin, "The physiology and habitat of the last universal common ancestor," *Nature Microbiology*, vol. 1, 07 2016.

[2] B. C. Goodwin, "Oscillatory behavior in enzymatic control processes," *Advances in Enzyme Regulation*, vol. 3, pp. 425–437, 1965.

[3] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, 01 2000.

[4] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, "A fast, robust and tunable synthetic gene oscillator," *Nature*, vol. 456, 2008.

[5] C. H. Johnson, S. S. Golden, M. Ishiura, and T. Kondo, "Circadian clocks in prokaryotes," *Molecular Microbiology*, vol. 21, no. 1, pp. 5–11, 1996.

[6] Y. Ouyang, C. R. Andersson, T. Kondo, S. S. Golden, and C. H. Johnson, "Resonating circadian clocks enhance fitness in cyanobacteria," *PNAS*, vol. 95, no. 15, pp. 8660–4, 1998.

[7] M. A. Woelfle, Y. Ouyang, K. Phanvijhitsiri, and C. H. Johnson, "The adaptive value of circadian clocks: an experimental assessment in cyanobacteria," *Current Biology*, vol. 14, no. 16, pp. 1481–6, 2004.

[8] T. Kondo, C. A. Strayer, R. D. Kulkarni, W. Taylor, M. Ishiura, S. S. Golden, and C. H. Johnson, "Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria," *PNAS*, vol. 90, no. 12, pp. 5672–5676, 1993.

[9] J. F. Feldman and M. N. Hoyle, "Isolation of Circadian Clock Mutants of *Neurospora crassa*," *Genetics*, vol. 75, pp. 605–613, 12 1973.

[10] A. J. Millar, I. A. Carre, C. A. Strayer, N.-H. Chua, and S. A. Kay, "Circadian clock mutants in Arabidopsis identified by luciferase imaging," *Science*, vol. 267, no. 5201, pp. 1161–1163, 1995.

[11] R. J. Konopka and S. Benzer, "Clock Mutants of *Drosophila melanogaster*," *PNAS*, vol. 68, no. 9, pp. 2112–2116, 1971.

[12] M. R. Ralph and M. Menaker, "A mutation of the circadian system in golden hamsters," *Science*, vol. 241, no. 4870, pp. 1225–1227, 1988.

[13] M. H. Vitaterna, D. P. King, A.-M. Chang, J. M. Kornhauser, P. L. Lowrey, J. D. McDonald, W. F. Dove, L. H. Pinto, F. W. Turek, and J. S. Takahashi, "Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior," *Science*, vol. 264, no. 5159, pp. 719–725, 1994.

[14] D. S. Hsu, X. Zhao, S. Zhao, A. Kazantsev, R.-P. Wang, T. Todo, Y.-F. Wei, and A. Sancar, "Putative Human Blue-Light Photoreceptors hCRY1 and hCRY2 Are Flavoproteins," *Biochemistry*, vol. 35, no. 44, pp. 13871–13877, 1996.

[15] M. Rosbash, "The Implications of Multiple Circadian Clock Origins," *PLoS Biology*, vol. 7, pp. 1–5, 03 2009.

[16] C. H. Johnson, P. L. Stewart, and M. Egli, "The cyanobacterial circadian system: from biophysics to bioevolution," *Annual Review of Biophysics*, vol. 40, pp. 143–67, 2011.

[17] L. J. Stal and W. E. Krumbein, "Nitrogenase activity in the non-heterocystous cyanobacterium *Oscillatoria* sp. grown under alternating light-dark cycles," *Archives of Microbiology*, vol. 143, pp. 67–71, 10 1985.

[18] A. L. Sessions, D. M. Doughty, P. V. Welander, R. E. Summons, and D. K. Newman, "The Continuing Puzzle of the Great Oxidation Event," *Current Biology*, vol. 19, no. 14, pp. R567–R574, 2009.

[19] J. A. Raven and J. F. Allen, "Genomics and chloroplast evolution: what did cyanobacteria do for plants?," *Genome Biology*, vol. 4, 03 2003.

[20] S. B. Gould, R. F. Waller, and G. I. McFadden, "Plastid Evolution," *Annual Review of Plant Biology*, vol. 59, no. 1, pp. 491–517, 2008.

[21] W. R. Hess, "Cyanobacterial genomics for ecology and biotechnology," *Current Opinion in Microbiology*, vol. 14, no. 5, pp. 608–614, 2011.

[22] T. Dagan, M. Roettger, K. Stucken, G. Landan, R. Koch, P. Major, S. B. Gould, V. V. Goremykin, R. Rippka, N. Tandeau de Marsac, M. Gugger, P. J. Lockhart, J. F. Allen, I. Brune, I. Maus, A. Pühler, and W. F. Martin, "Genomes of Stigonematalean Cyanobacteria (Subsection V) and the Evolution of Oxygenic Photosynthesis from Prokaryotes to Plastids," *Genome Biology and Evolution*, vol. 5, no. 1, pp. 31–44, 2013.

[23] R. Rippka, J. Deruelles, J. B. Waterbury, M. Herdman, and R. Y. Stanier, "Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria," *Microbiology*, vol. 111, no. 1, pp. 1–61, 1979.

[24] D. C. Ducat, J. C. Way, and P. A. Silver, "Engineering cyanobacteria to generate high-value products," *Trends in Biotechnology*, vol. 29, no. 2, pp. 95–103, 2011.

[25] P. Lindberg, S. Park, and A. Melis, "Engineering a platform for photosynthetic isoprene production in cyanobacteria, using *Synechocystis* as the model organism," *Metabolic Engineering*, vol. 12, no. 1, pp. 70–79, 2010.

[26] E. Englund, B. Pattanaik, S. J. K. Ubhayasekera, K. Stensjö, J. Bergquist, and P. Lindberg, "Production of Squalene in *Synechocystis* sp. PCC 6803," *PLoS ONE*, vol. 9, pp. 1–8, 03 2014.

[27] A. Loeschcke, D. Dienst, V. Wewer, J. Hage-Hülsmann, M. Dietsch, S. Kranz-Finger, V. Hüren, S. Metzger, V. B. Urlacher, T. Gigolashvili, S. Kopriva, I. M. Axmann, T. Drepper, and K.-E. Jaeger, "The photosynthetic bacteria *Rhodobacter capsulatus* and *Synechocystis* sp. PCC 6803 as new hosts for cyclic plant triterpene biosynthesis," *PLoS ONE*, vol. 12, pp. 1–23, 12 2017.

[28] N. B. Ivleva, M. R. Bramlett, P. A. Lindahl, and S. S. Golden, "LdpA: a component of the circadian clock senses redox state of the cell.," *The EMBO Journal*, vol. 24, no. 6, pp. 1202–1210, 2005.

[29] G. K. Pattanayak and M. J. Rust, "The Cyanobacterial Clock and Metabolism," *Current Opinion in Microbiology*, vol. 18, pp. 90–95, 2014.

[30] J. S. Markson, J. R. Piechura, A. M. Puszynska, and E. K. O'Shea, "Circadian control of global gene expression by the cyanobacterial master regulator RpaA," *Cell*, vol. 155, no. 6, pp. 1396–1408, 2013.

[31] N. M. Schmelling and I. M. Axmann, "Computational modelling unravels the precise clockwork of cyanobacteria," *Interface Focus*, vol. 8, no. 6, 2018.

[32] J. Snijder, J. M. Schuller, A. Wiegard, P. Lössl, N. Schmelling, I. M. Axmann, J. M. Plitzko, F. Förster, and A. J. R. Heck, "Structures of the cyanobacterial circadian oscillator frozen in a fully assembled state," *Science*, vol. 355, no. 6330, pp. 1181–1184, 2017.

[33] A. Wiegard, A. K. Dörrich, H. T. Deinzer, C. Beck, A. Wilde, J. Holtzendorff, and I. M. Axmann, "Biochemical analysis of three putative KaiC clock proteins from *Synechocystis* sp. PCC 6803 suggests their functional divergence," *Microbiology*, vol. 159, no. 5, pp. 948–958, 2013.

[34] F. Hayashi, H. Suzuki, R. Iwase, T. Uzumaki, A. Miyake, J. R. Shen, K. Imada, Y. Furukawa, K. Yonekura, K. Namba, and M. Ishiura, "ATP-induced hexameric ring structure of the cyanobacterial circadian clock protein KaiC," *Genes to Cells*, vol. 8, no. 3, pp. 287–96, 2003.

[35] T. Nishiwaki, H. Iwasaki, M. Ishiura, and T. Kondo, "Nucleotide binding and autophosphorylation of the clock protein KaiC as a circadian timing process of cyanobacteria.," *PNAS*, vol. 97, no. 1, pp. 495–499, 2000.

[36] T. Nishiwaki and T. Kondo, "Circadian autodephosphorylation of cyanobacterial clock protein KaiC occurs via formation of ATP as intermediate," *Journal of Biological Chemistry*, vol. 287, no. 22, pp. 18030–18035, 2012.

[37] Y.-I. Kim, G. Dong, J. Carruthers, C. W., S. S. Golden, and A. LiWang, "The day/night switch in KaiC, a central oscillator component of the circadian clock of cyanobacteria," *PNAS*, vol. 105, no. 35, pp. 12825–30, 2008.

[38] R. Pattanayek, D. R. Williams, S. Pattanayek, Y. Xu, T. Mori, C. H. Johnson, P. L. Stewart, and M. Egli, "Analysis of KaiA-KaiC protein interactions in the cyanobacterial circadian clock using hybrid structural methods," *The EMBO Journal*, vol. 25, no. 9, pp. 2017–28, 2006.

[39] Y. Kitayama, H. Iwasaki, T. Nishiwaki, and T. Kondo, "KaiB functions as an attenuator of KaiC phosphorylation in the cyanobacterial circadian clock system," *The EMBO Journal*, vol. 22, no. 9, pp. 2127–34, 2003.

[40] M. Egli, "Intricate Protein-Protein Interactions in the Cyanobacterial Circadian Clock," *Journal of Biological Chemistry*, vol. 289, no. 31, pp. 21267–21275, 2014.

[41] M. J. Rust, S. S. Golden, and E. K. O'Shea, "Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator.," *Science*, vol. 331, no. 6014, pp. 220–3, 2011.

[42] I. M. Axmann, S. Hertel, A. Wiegard, A. K. Dörrich, and A. Wilde, "Diversity of KaiC-based timing systems in marine Cyanobacteria," *Marine Genomics*, vol. 14, pp. 3–16, 2014.

[43] S. Kutsuna, T. Kondo, and S. Aoki, "A Period-Extender Gene, *pex*, That Extends the Period of the Circadian Clock in the Cyanobacterium *Synechococcus* sp. Strain PCC 7942," *Journal of Bacteriology*, vol. 180, no. 8, pp. 2167–2174, 1998.

[44] S. Kutsuna, T. Kondo, H. Ikegami, T. Uzumaki, M. Katayama, and M. Ishiura, "The circadian clock-related gene *pex* regulates a negative *cis* element in the *kaiA* promoter region," *Journal of Bacteriology*, vol. 189, no. 21, pp. 7690–7696, 2007.

[45] M. Katayama, N. F. Tsinoremas, T. Kondo, and S. S. Golden, "*cpmA*, a gene involved in an output pathway of the cyanobacterial circadian system," *Journal of Bacteriology*, vol. 181, no. 11, pp. 3516–3524, 1999.

[46] S. B. Williams, I. Vakonakis, S. S. Golden, and A. C. LiWang, "Structure and function from the circadian clock protein KaiA of *Synechococcus elongatus*: a potential clock input mechanism," *PNAS*, vol. 99, no. 24, pp. 15357–62, 2002.

[47] T. L. Wood, J. Bridwell-Rabb, Y. I. Kim, T. Gao, Y. G. Chang, A. LiWang, D. P. Barondeau, and S. S. Golden, "The KaiA protein of the cyanobacterial circadian oscillator is modulated by a redox-active cofactor," *PNAS*, vol. 107, no. 13, pp. 5804–9, 2010.

[48] Y.-I. Kim, D. J. Vinyard, G. M. Ananyev, G. C. Dismukes, and S. S. Golden, "Oxidized quinones signal onset of darkness directly to the cyanobacterial circadian oscillator," *PNAS*, vol. 109, no. 44, pp. 17765–17769, 2012.

[49] M. Katayama, T. Kondo, J. Xiong, and S. S. Golden, "*ldpA* Encodes an Iron-Sulfur Protein Involved in Light-Dependent Modulation of the Circadian Period in the Cyanobacterium *Synechococcus elongatus* PCC 7942," *Journal of Bacteriology*, vol. 185, no. 4, pp. 1415–1422, 2003.

[50] Y. Liu, N. F. Tsinoremas, C. H. Johnson, N. V. Lebedeva, S. S. Golden, M. Ishiura, and T. Kondo, "Circadian orchestration of gene expression in cyanobacteria," *Genes and Development*, vol. 9, no. 12, pp. 1469–78, 1995.

[51] V. Vijayan, R. Zuzow, and E. K. O'Shea, "Oscillations in supercoiling drive circadian gene expression in cyanobacteria.," *PNAS*, vol. 106, no. 52, pp. 22564–8, 2009.

[52] T. Mori, B. Binder, and C. H. Johnson, "Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours.," *PNAS*, vol. 93, pp. 10183–10188, 09 1996.

[53] G. Dong, Q. Yang, Q. Wang, Y. I. Kim, T. L. Wood, K. W. Osteryoung, A. van Oudenaarden, and S. S. Golden, "Elevated ATPase activity of KaiC applies a circadian checkpoint on cell division in *Synechococcus elongatus*," *Cell*, vol. 140, no. 4, pp. 529–39, 2010.

[54] R. M. Smith and S. B. Williams, "Circadian rhythms in gene transcription imparted by chromosome compaction in the cyanobacterium *Synechococcus elongatus*," *PNAS*, vol. 103, no. 22, pp. 8564–9, 2006.

[55] N. Takai, M. Nakajima, T. Oyama, R. Kito, C. Sugita, M. Sugita, T. Kondo, and H. Iwasaki, "A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria.," *PNAS*, vol. 103, no. 32, pp. 12109–12114, 2006.

[56] A. Gutu and E. K. O'Shea, "Two Antagonistic Clock-Regulated Histidine Kinases Time the Activation of Circadian Gene Expression," *Molecular Cell*, vol. 50, no. 2, pp. 288–294, 2013.

[57] Y. Taniguchi, M. Katayama, R. Ito, N. Takai, T. Kondo, and T. Oyama, "*labA*: a novel gene required for negative feedback regulation of the cyanobacterial circadian clock protein KaiC," *Genes & Development*, vol. 21, no. 1, pp. 60–70, 2007.

[58] J. Espinosa, J. S. Boyd, R. Cantos, P. Salinas, S. S. Golden, and A. Contreras, "Cross-talk and regulatory interactions between the essential response regulator RpaB and cyanobacterial circadian clock output," *PNAS*, vol. 112, no. 7, pp. 2198–2203, 2015.

[59] Y. Taniguchi, T. Nishikawa, T. Kondo, and T. Oyama, "Overexpression of *lalA*, a paralog of *labA*, is capable of affecting both circadian gene expression and cell growth in the cyanobacterium *Synechococcus elongatus* PCC 7942," *FEBS Letters*, vol. 586, no. 6, pp. 753–759, 2012.

[60] J. S. Boyd, J. R. Bordowitz, A. C. Bree, and S. S. Golden, "An allele of the *crm* gene blocks cyanobacterial circadian rhythms.," *PNAS*, vol. 110, no. 34, pp. 13950–5, 2013.

[61] H. Iwasaki, S. B. Williams, Y. Kitayama, M. Ishiura, S. S. Golden, and T. Kondo, "A KaiC-interacting sensory histidine kinase, SasA, necessary to sustain robust circadian oscillation in cyanobacteria," *Cell*, vol. 101, no. 2, pp. 223–233, 2000.

[62] R. Tseng, N. F. Goularte, A. Chavan, J. Luu, S. E. Cohen, Y.-G. Chang, J. Heisler, S. Li, A. K. Michael, S. Tripathi, S. S. Golden, A. LiWang, and C. L. Partch, "Structural basis of the day-night transition in a bacterial circadian clock," *Science*, vol. 355, no. 6330, pp. 1174–1180, 2017.

[63] O. Schmitz, M. Katayama, S. B. Williams, T. Kondo, and S. S. Golden, "CikA, a bacteriophytochrome that resets the cyanobacterial circadian clock.," *Science*, vol. 289, no. 5480, pp. 765–768, 2000.

[64] N. B. Ivleva, T. Gao, A. C. LiWang, and S. S. Golden, "Quinone sensing by the circadian input kinase of the cyanobacterial circadian clock.," *PNAS*, vol. 103, no. 46, pp. 17468–17473, 2006.

[65] F. Moronta-Barrios, J. Espinosa, and A. Contreras, "Negative control of cell size in the cyanobacterium *Synechococcus elongatus* PCC 7942 by the essential response regulator RpaB," *FEBS Letters*, vol. 587, no. 5, pp. 504–9, 2013.

[66] M. Hanaoka, N. Takai, N. Hosokawa, M. Fujiwara, Y. Akimoto, N. Kobori, H. Iwasaki, T. Kondo, and K. Tanaka, "RpaB, another response regulator operating circadian clock-dependent transcriptional regulation in *Synechococcus elongatus* PCC 7942," *Journal of Biological Chemistry*, vol. 287, no. 31, pp. 26321–26327, 2012.

[67] F. Moronta-Barrios, J. Espinosa, and A. Contreras, "*In vivo* features of signal transduction by the essential response regulator RpaB from *Synechococcus elongatus* PCC 7942," *Microbiology*, vol. 158, no. 5, pp. 1229–37, 2012.

[68] J. S. Boyd, R. R. Cheng, M. L. Paddock, C. Sancar, F. Morcos, and S. S. Golden, "A Combined Computational and Genetic Approach Uncovers Network Interactions of the Cyanobacterial Circadian Clock," *Journal of Bacteriology*, vol. 198, no. 18, pp. 2439–2447, 2016.

[69] E. R. Zinser, D. Lindell, Z. I. Johnson, M. E. Futschik, C. Steglich, M. L. Coleman, M. A. Wright, T. Rector, R. Steen, N. McNulty, L. R. Thompson, and S. W. Chisholm, "Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *Prochlorococcus*," *PLoS One*, vol. 4, no. 4, 2009.

[70] R. Lehmann, R. Machné, J. Georg, M. Benary, I. M. Axmann, and R. Steuer, "How cyanobacteria pose new problems to old methods: challenges in microarray time series analysis.," *BMC Bioinformatics*, vol. 14, p. 133, 2013.

[71] C. Beck, S. Hertel, A. Rediger, R. Lehmann, A. Wiegard, A. Kölsch, B. Heilmann, J. Georg, W. R. Hess, and I. M. Axmann, "Daily expression pattern of protein-encoding genes and small non coding RNAs in *Synechocystis sp.* strain PCC 6803," *Applied and Environmental Microbiology*, vol. 80, no. 17, pp. 5195–5206, 2014.

[72] H. Ito, M. Mutsuda, Y. Murayama, J. Tomita, N. Hosokawa, K. Terauchi, C. Sugita, M. Sugita, T. Kondo, and H. Iwasaki, "Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in *Synechococcus elongatus*.," *PNAS*, vol. 106, no. 33, pp. 14168–73, 2009.

[73] J. Stöckel, E. A. Welsh, M. Liberton, R. Kunnvakkam, R. Aurora, and H. B. Pakrasi, "Global transcriptomic analysis of *Cyanothece* 51142 reveals robust diurnal oscillation of central metabolic processes," *PNAS*, vol. 105, no. 16, pp. 6156–6161, 2008.

[74] J. Toepel, E. Welsh, T. C. Summerfield, H. B. Pakrasi, and L. A. Sherman, "Differential transcriptional analysis of the cyanobacterium *Cyanothece* sp. strain ATCC 51142 during light-dark and continuous-light growth," *Journal of Bacteriology*, vol. 190, no. 11, pp. 3904–3913, 2008.

[75] H. Kushige, H. Kugenuma, M. Matsuoka, S. Ehira, M. Ohmori, and H. Iwasaki, "Genome-wide and heterocyst-specific circadian gene expression in the filamentous Cyanobacterium *Anabaena* sp. strain PCC 7120," *Journal of Bacteriology*, vol. 195, no. 6, pp. 1276–84, 2013.

[76] C. Straub, P. Quillardet, J. Vergalli, N. T. de Marsac, and J. F. Humbert, "A day in the life of *Microcystis aeruginosa* strain PCC 7806 as revealed by a transcriptomic analysis," *PLoS One*, vol. 6, no. 1, 2011.

[77] J. Holtzendorff, F. Partensky, D. Mella, J.-F. Lennon, W. R. Hess, and L. Garczarek, "Genome streamlining results in loss of robustness of the circadian clock in the marine cyanobacterium *Prochlorococcus marinus* PCC 9511.," *Journal of Biological Rhythms*, vol. 23, no. 3, pp. 187–199, 2008.

[78] I. M. Axmann, U. Dühring, L. Seeliger, A. Arnold, J. T. Vanselow, A. Kramer, and A. Wilde, "Biochemical evidence for a timing mechanism in *Prochlorococcus*," *Journal of Bacteriology*, vol. 191, no. 17, pp. 5342–5347, 2009.

[79] V. Dvornyk, O. Vinogradova, and E. Nevo, "Origin and evolution of circadian clock genes in prokaryotes.," *PNAS*, vol. 100, no. 5, pp. 2495–2500, 2003.

[80] M. Loza-Correa, T. Sahr, M. Rolando, C. Daniels, P. Petit, T. Skarina, L. Gomez Valero, D. Dervins-Ravault, N. Honoré, A. Savchenko, and C. Buchrieser, "The *Legionella pneumophila kai* operon is implicated in stress response and confers fitness in competitive environments," *Environmental Microbiology*, vol. 16, no. 2, pp. 359–381, 2014.

[81] S. Aoki and K. Onai, *Circadian clocks of Synechocystis sp. strain PCC 6803, Thermosynechococcus elongatus, Prochlorococcus spp., Trichodesmium spp. and other species*, book section 15, pp. 259–282. Berlin Heidelberg: Springer Berlin Heidelberg, 2009.

[82] W. M. Fitch, "Homology: a personal view on some of the problems," *Trends in Genetics*, vol. 16, no. 5, pp. 227–231, 2000.

[83] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–402, 1997.

[84] D. G. Higgins and P. M. Sharp, "CLUSTAL: a package for performing multiple sequence alignment on a microcomputer," *Gene*, vol. 73, no. 1, pp. 237–244, 1988.

[85] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins, "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega," *Molecular Systems Biology*, vol. 7, p. 539, 2011.

[86] R. Steuer, H. Knoop, and R. Machné, "Modelling cyanobacteria: from metabolism to integrative models of phototrophic growth," *Journal of Experimental Botany*, vol. 63, no. 6, pp. 2259–2274, 2012.

[87] M. Nakajima, K. Imai, H. Ito, T. Nishiwaki, Y. Murayama, H. Iwasaki, T. Oyama, and T. Kondo, "Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation *in vitro*," *Science*, vol. 308, no. 5720, pp. 414–5, 2005.

[88] R. Pattanayek, J. Wang, T. Mori, Y. Xu, C. H. Johnson, and M. Egli, "Visualizing a circadian clock protein: crystal structure of KaiC and functional insights," *Molecular Cell*, vol. 15, no. 3, pp. 375–88, 2004.

[89] S. Ye, I. Vakonakis, T. R. Ioerger, A. C. LiWang, and J. C. Sacchettini, "Crystal Structure of Circadian Clock Protein KaiA from *Synechococcus elongatus*," *Journal of Biological Chemistry*, vol. 279, no. 19, pp. 20511–20518, 2004.

[90] R. Murakami, R. Mutoh, R. Iwase, Y. Furukawa, K. Imada, K. Onai, M. Morishita, S. Yasui, K. Ishii, J. O. Valencia Swain, T. Uzumaki, K. Namba, and M. Ishiura, "The Roles of the Dimeric and Tetrameric Structures of the Clock Protein KaiB in the Generation of Circadian Oscillations in Cyanobacteria," *Journal of Biological Chemistry*, vol. 287, no. 35, pp. 29506–29515, 2012.

[91] N. M. Schmelling, R. Lehmann, P. Chaudhury, C. Beck, S.-V. Albers, I. M. Axmann, and A. Wiegard, "Minimal tool set for a prokaryotic circadian clock," *BMC Evolutionary Biology*, vol. 17, p. 169, Jul 2017.

[92] M. J. Rust, J. S. Markson, W. S. Lane, D. S. Fisher, and E. K. O'Shea, "Ordered Phosphorylation Governs Oscillation of a Three-Protein Circadian Clock," *Science*, vol. 318, no. 5851, pp. 809–812, 2007.

[93] J. Paijmans, D. K. Lubensky, and P. Rein ten Wolde, "Period Robustness and Entrainability of the Kai System to Changing Nucleotide Concentrations," *Biophysical Journal*, vol. 113, pp. 157–173, 07 2017.

[94] G. Lambert, J. Chew, and M. J. Rust, "Costs of Clock-Environment Misalignment in Individual Cyanobacterial Cells," *Biophysical Journal*, vol. 111, pp. 883–891, 08 2016.

[95] E. Leypunskiy, J. Lin, H. Yoo, U. Lee, A. R. Dinner, and M. J. Rust, "The cyanobacterial circadian clock follows midday *in vivo* and *in vitro*," *eLife*, 07 2017.

[96] M. Monti and P. R. ten Wolde, "The accuracy of telling time via oscillatory signals," *Physical Biology*, vol. 13, no. 3, 2016.

[97] W. Pittayakanchit, Z. Lu, J. Chew, M. J. Rust, and A. Murugan, "Biophysical clocks face a trade-off between internal and external noise resistance," *eLife*, vol. 7, p. e37624, 7 2018.

[98] J. Paijmans, M. Bosman, P. R. ten Wolde, and D. K. Lubensky, "Discrete gene replication events drive coupling between the cell cycle and circadian clocks," *PNAS*, vol. 113, no. 15, pp. 4063–4068, 2016.

[99] J. S. O'Neill and A. B. Reddy, "Circadian clocks in human red blood cells," *Nature*, vol. 469, no. 7331, pp. 498–503, 2011.

[100] M. Ishiura, S. Kutsuna, S. Aoki, H. Iwasaki, C. R. Andersson, A. Tanabe, S. S. Golden, C. H. Johnson, and T. Kondo, "Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobacteria," *Science*, vol. 281, no. 5382, pp. 1519–23, 1998.

[101] C. Brettschneider, R. J. Rose, S. Hertel, I. M. Axmann, A. J. R. Heck, and M. Kollmann, "A sequestration feedback determines dynamics and temperature entrainment of the KaiABC circadian clock," *Molecular Systems Biology*, vol. 6, no. 1, 2010.

[102] H. Ito, H. Kageyama, M. Mutsuda, M. Nakajima, T. Oyama, and T. Kondo, "Autonomous synchronization of the circadian KaiC phosphorylation rhythm," *Nature Structural & Molecular Biology*, vol. 14, p. 1084, 10 2007.

[103] M. Egli, R. Pattanayek, J. H. Sheehan, Y. Xu, T. Mori, J. A. Smith, and C. H. Johnson, "Loop-loop interactions regulate KaiA-stimulated KaiC phosphorylation in the cyanobacterial KaiABC circadian clock," *Biochemistry*, vol. 52, no. 7, pp. 1208–20, 2013.

[104] T. Nishiwaki, Y. Satomi, Y. Kitayama, K. Terauchi, R. Kiyohara, T. Takao, and T. Kondo, "A sequential program of dual phosphorylation of KaiC as a basis for circadian rhythm in cyanobacteria," *The EMBO Journal*, vol. 26, no. 17, pp. 4029–4037, 2007.

[105] X. Qin, M. Byrne, T. Mori, P. Zou, D. R. Williams, H. Mchaourab, and C. H. Johnson, "Intermolecular associations determine the dynamics of the circadian KaiABC oscillator," *PNAS*, vol. 107, no. 33, pp. 14805–14810, 2010.

[106] R. Tseng, Y.-G. Chang, I. Bravo, R. Latham, A. Chaudhary, N.-W. Kuo, and A. LiWang, "Cooperative KaiA–KaiB–KaiC Interactions Affect KaiB/SasA Competition in the Circadian Clock of Cyanobacteria," *Journal of Molecular Biology*, vol. 426, no. 2, pp. 389–402, 2014.

[107] S. A. Villarreal, R. Pattanayek, D. R. Williams, T. Mori, X. Qin, C. H. Johnson, M. Egli, and P. L. Stewart, "CryoEM and Molecular Dynamics of the Circadian KaiB–KaiC Complex Indicates That KaiB Monomers Interact with KaiC and Block ATP Binding Clefts," *Journal of Molecular Biology*, vol. 425, no. 18, pp. 3311 – 3324, 2013.

[108] S. Akiyama, A. Nohara, K. Ito, and Y. Maéda, "Assembly and Disassembly Dynamics of the Cyanobacterial Periodosome," *Molecular Cell*, vol. 29, no. 6, pp. 703–716, 2008.

[109] R. Pattanayek, D. R. Williams, G. Rossi, S. Weigand, T. Mori, C. H. Johnson, P. L. Stewart, and M. Egli, "Combined SAXS/EM Based Models of the *S. elongatus* Post-Translational Circadian Oscillator and its Interactions with the Output His-Kinase SasA," *PLoS ONE*, vol. 6, pp. 1–19, 08 2011.

[110] J. Snijder, R. J. Burnley, A. Wiegard, A. S. J. Melquiond, A. M. J. J. Bonvin, I. M. Axmann, and A. J. R. Heck, "Insight into cyanobacterial circadian timing from structural details of the KaiB–KaiC interaction," *PNAS*, vol. 111, pp. 1379–1384, 1 2014.

[111] R. Pattanayek, D. R. Williams, S. Pattanayek, T. Mori, C. H. Johnson, P. L. Stewart, and M. Egli, "Structural model of the circadian clock KaiB–KaiC complex and mechanism for modulation of KaiC phosphorylation," *The EMBO Journal*, vol. 27, no. 12, pp. 1767–1778, 2008.

[112] Y.-G. Chang, R. Tseng, N.-W. Kuo, and A. LiWang, "Rhythmic ring–ring stacking drives the circadian oscillator clockwise," *PNAS*, vol. 109, no. 42, pp. 16847–16851, 2012.

[113] M. Nakao, S. Okamoto, M. Kohara, T. Fujishiro, T. Fujisawa, S. Sato, S. Tabata, T. Kaneko, and Y. Nakamura, "CyanoBase: the cyanobacteria genome database update 2010," *Nucleic Acids Research*, vol. 38, pp. 379–381, 2010.

[114] D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler, "GenBank," *Nucleic Acids Research*, vol. 36, pp. 25–30, 2008.

[115] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton, "Jalview Version 2—a multiple sequence alignment editor and analysis workbench," *Bioinformatics*, vol. 25, no. 9, pp. 1189–1191, 2009.

[116] G. E. Crooks, G. Hon, J.-M. Chandonia, and S. E. Brenner, "WebLogo: A Sequence Logo Generator," *Genome Research*, vol. 14, no. 6, pp. 1188–1190, 2004.

[117] S. Mohammed and A. J. Heck, "Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications," *Current Opinion in Biotechnology*, vol. 22, no. 1, pp. 9–16, 2011.

[118] F. Liu, D. T. S. Rijkers, H. Post, and A. J. R. Heck, "Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry," *Nature Methods*, vol. 12, p. 1179, 9 2015.

[119] A. Korinek, F. Beck, W. Baumeister, S. Nickell, and J. M. Plitzko, "Computer controlled cryo-electron microscopy – TOM² a software package for high-throughput applications," *Journal of Structural Biology*, vol. 175, no. 3, pp. 394–405, 2011.

[120] X. Li, P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard, and Y. Cheng, "Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM," *Nature Methods*, vol. 10, p. 584, 5 2013.

[121] A. Rohou and N. Grigorieff, "CTFFIND4: Fast and accurate defocus estimation from electron micrographs," *Journal of Structural Biology*, vol. 192, no. 2, pp. 216–221, 2015.

[122] S. H. Scheres, "RELION: Implementation of a Bayesian approach to cryo-EM structure determination," *Journal of Structural Biology*, vol. 180, no. 3, pp. 519–530, 2012.

[123] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke, "EMAN2: An extensible image processing suite for electron microscopy," *Journal of Structural Biology*, vol. 157, no. 1, pp. 38–46, 2007.

[124] S. Nickell, F. Förster, A. Linaroudis, W. Del Net, F. Beck, R. Hegerl, W. Baumeister, and J. M. Plitzko, "TOM software toolbox: acquisition and analysis for electron tomography," *Journal of Structural Biology*, vol. 149, no. 3, pp. 227–234, 2005.

[125] J. M. Schuller, F. Beck, P. Lössl, A. J. R. Heck, and F. Förster, "Nucleotide-dependent conformational changes of the AAA+ ATPase p97 revisited," *FEBS Letters*, vol. 590, no. 5, pp. 595–604, 2016.

[126] S. H. Scheres, "Beam-induced motion correction for sub-megadalton cryo-EM particles," *eLife*, vol. 3, p. e03665, 8 2014.

[127] G. Cardone, J. B. Heymann, and A. C. Steven, "One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions," *Journal of Structural Biology*, vol. 184, no. 2, pp. 226–236, 2013.

[128] N. Eswar, B. Webb, M. A. Marti-Renom, M. Madhusudhan, D. Eramian, M.-Y. Shen, U. Pieper, and A. Sali, "Comparative Protein Structure Modeling Using MODELLER," *Current Protocols in Protein Science*, vol. 50, no. 1, pp. 2.9.1–2.9.31, 2007.

[129] A. Meier and J. Söding, "Automatic Prediction of Protein 3D Structures by Probabilistic Multi-template Homology Modeling," *PLoS Computational Biology*, vol. 11, pp. 1–20, 10 2015.

[130] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "UCSF Chimera—A visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.

[131] J. J. Headd, N. Echols, P. V. Afonine, R. W. Grosse-Kunstleve, V. B. Chen, N. W. Moriarty, D. C. Richardson, J. S. Richardson, and P. D. Adams, "Use of knowledge-based restraints in *phenix.refine* to improve macromolecular refinement at low resolution," *Acta Crystallographica Section D*, vol. 68, pp. 381–390, 4 2012.

[132] C. H. S. Aylett, E. Sauer, S. Imseng, D. Boehringer, M. N. Hall, N. Ban, and T. Maier, "Architecture of human mTOR complex 1," *Science*, vol. 351, no. 6268, pp. 48–52, 2016.

[133] J. Snijder and A. J. Heck, "Analytical Approaches for Size and Mass Analysis of Large Protein Assemblies," *Annual Review of Analytical Chemistry*, vol. 7, pp. 43–64, 6 2014.

[134] J. Abe, T. B. Hiyama, A. Mukaiyama, S. Son, T. Mori, S. Saito, M. Osako, J. Wolanin, E. Yamashita, T. Kondo, and S. Akiyama, "Atomic-scale origins of slowness in the cyanobacterial circadian clock," *Science*, vol. 349, no. 6245, pp. 312–316, 2015.

[135] Y.-G. Chang, S. E. Cohen, C. Phong, W. K. Myers, Y.-I. Kim, R. Tseng, J. Lin, L. Zhang, J. S. Boyd, Y. Lee, S. Kang, D. Lee, S. Li, R. D. Britt, M. J. Rust, S. S. Golden, and A. LiWang, "A protein fold switch joins the circadian oscillator to clock output in cyanobacteria," *Science*, vol. 349, no. 6245, pp. 324–328, 2015.

[136] R. Iwase, K. Imada, F. Hayashi, T. Uzumaki, M. Morishita, K. Onai, Y. Furukawa, K. Namba, and M. Ishiura, "Functionally Important Substructures of Circadian Clock Protein KaiB in a Unique Tetramer Complex," *Journal of Biological Chemistry*, vol. 280, no. 52, pp. 43141–43149, 2005.

[137] K. Hitomi, T. Oyama, S. Han, A. S. Arvai, and E. D. Getzoff, "Tetrameric Architecture of the Circadian Clock Protein KaiB: A novel interface for intermolecular interactions and its impact on the circadian rhythm," *Journal of Biological Chemistry*, vol. 280, no. 19, pp. 19127–19135, 2005.

[138] R. Pattanayek and M. Egli, "Protein–Protein Interactions in the Cyanobacterial Circadian Clock: Structure of KaiA Dimer in Complex with C-Terminal KaiC Peptides at 2.8 Å Resolution," *Biochemistry*, vol. 54, no. 30, pp. 4575–4578, 2015.

[139] M. J. Simons, "The evolution of the cyanobacterial posttranslational clock from a primitive "phoscillator"," *Journal of Biological Rhythms*, vol. 24, no. 3, pp. 175–82, 2009.

[140] J. L. Ditty, S. B. Williams, and S. S. Golden, "A Cyanobacterial Circadian Timing Mechanism," *Annual Review of Genetics*, vol. 37, no. 1, pp. 513–543, 2003.

[141] S. S. Golden and S. R. Canales, "Cyanobacterial circadian clocks–timing is everything," *Nature Review Microbiology*, vol. 1, no. 3, pp. 191–9, 2003.

[142] C. S. Pittendrigh, "Temporal organization: reflections of a Darwinian clock-watcher," *Annual Review of Physiology*, vol. 55, pp. 16–54, 1993.

[143] B. M. Sweeney and M. B. Borgese, "A circadian rhythm in cell division in a prokaryote, the cyanobacterium *Synechococcus* WH7803," *Journal of Phycology*, vol. 25, no. 1, pp. 183–186, 1989.

[144] A. Mitsui, S. Kumazawa, A. Takahashi, H. Ikemoto, S. Cao, and T. Arai, "Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically," *Nature*, vol. 323, no. 6090, pp. 720–722, 1986.

[145] T. C. Huang and T. J. Chow, "New Type of N-2-Fixing Unicellular Cyanobacterium (Blue-Green-Alga)," *FEMS Microbiology Letters*, vol. 36, no. 1, pp. 109–110, 1986.

[146] N. Grobbelaar, T. C. Huang, H. Y. Lin, and T. J. Chow, "Dinitrogen-Fixing Endogenous Rhythm in *Synechococcus* Rf-1," *FEMS Microbiology Letters*, vol. 37, no. 2, pp. 173–177, 1986.

[147] T. C. Huang and T. J. Chow, "Characterization of the Rhythmic Nitrogen-Fixing Activity of *Synechococcus* sp. Rf-1 at the Transcription Level," *Current Microbiology*, vol. 20, no. 1, pp. 23–26, 1990.

[148] J. Tomita, M. Nakajima, T. Kondo, and H. Iwasaki, "No transcription-translation feedback in circadian rhythm of KaiC phosphorylation," *Science*, vol. 307, no. 5707, pp. 251–4, 2005.

[149] J. L. Ditty, S. R. Canales, B. E. Anderson, S. B. Williams, and S. S. Golden, "Stability of the *Synechococcus elongatus* PCC 7942 circadian clock under directed anti-phase expression of the *kai* genes," *Microbiology*, vol. 151, no. 8, pp. 2605–2613, 2005.

[150] R. S. Edgar, E. W. Green, Y. Zhao, G. van Ooijen, M. Olmedo, X. Qin, Y. Xu, M. Pan, U. K. Valekunja, K. A. Feeney, E. S. Maywood, M. H. Hastings, N. S. Baliga, M. Merrow, A. J. Millar, C. H. Johnson, C. P. Kyriacou, J. S. O'Neill, and A. B. Reddy, "Peroxiredoxins are conserved markers of circadian rhythms," *Nature*, vol. 485, no. 7399, pp. 459–64, 2012.

[151] J. S. O'Neill, G. van Ooijen, L. E. Dixon, C. Troein, F. Corellou, F.-Y. Bouget, A. B. Reddy, and A. J. Millar, "Circadian rhythms persist without transcription in a eukaryote," *Nature*, vol. 469, no. 7331, pp. 554–8, 2011.

[152] C. R. McClung, "Circadian Rhythms in Plants," *Annual Review of Plant Physiology and Plant Molecular Biology*, vol. 52, pp. 139–162, 2001.

[153] Y. Hihara, K. Sonoike, M. Kanehisa, and M. Ikeuchi, "DNA microarray analysis of redox-responsive genes in the genome of the cyanobacterium *Synechocystis* sp. strain PCC 6803," *Journal of Bacteriology*, vol. 185, no. 5, pp. 1719–25, 2003.

[154] T. Mori, S. V. Saveliev, Y. Xu, W. F. Stafford, M. M. Cox, R. B. Inman, and C. H. Johnson, "Circadian clock protein KaiC forms ATP-dependent hexameric rings and binds DNA," *PNAS*, vol. 99, no. 26, pp. 17203–8, 2002.

[155] S. R. Mackey, J. S. Choi, Y. Kitayama, H. Iwasaki, G. Dong, and S. S. Golden, "Proteins found in a CikA interaction assay link the circadian clock, metabolism, and cell division in *Synechococcus elongatus*," *Journal of Bacteriology*, vol. 190, no. 10, pp. 3738–3746, 2008.

[156] V. Dvornyk, *The Circadian Clock Gear in Cyanobacteria: Assembled by Evolution*, book section 14, pp. 241–258. Berlin Heidelberg: Springer, 2009.

[157] P. Ma, T. Mori, C. Zhao, T. Thiel, and C. H. Johnson, "Evolution of KaiC-Dependent Timekeepers: A Proto-circadian Timing Mechanism Confers Adaptive Fitness in the Purple Bacterium *Rhodopseudomonas palustris*," *PLoS Genetics*, vol. 12, no. 3, p. e1005922, 2016.

[158] H. J. Kang, K. Kubota, H. Ming, K. Miyazono, and M. Tanokura, "Crystal structure of KaiC-like protein PH0186 from hyperthermophilic archaea *Pyrococcus horikoshii* OT3," *Proteins*, vol. 75, no. 4, pp. 1035–9, 2009.

[159] M. Maniscalco, J. Nannen, V. Sodi, G. Silver, P. L. Lowrey, and K. A. Bidle, "Light-dependent expression of four cryptic archaeal circadian gene homologs," *Frontiers in Microbiology*, vol. 5, p. 79, 2014.

[160] I. Baca, D. Sprockett, and V. Dvornyk, "Circadian input kinases and their homologs in cyanobacteria: Evolutionary constraints versus architectural diversification," *Journal of Molecular Evolution*, vol. 70, no. 5, pp. 453–465, 2010.

[161] V. Dvornyk, H. W. Deng, and E. Nevo, "Structure and molecular phylogeny of *sasA* genes in cyanobacteria: Insights into evolution of the prokaryotic circadian system," *Molecular Biology and Evolution*, vol. 21, no. 8, pp. 1468–1476, 2004.

[162] P. J. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, and M. J. de Hoon, "Biopython: freely available Python tools for computational molecular biology and bioinformatics," *Bioinformatics*, vol. 25, no. 11, pp. 1422–3, 2009.

[163] F. Pérez and E. Granger, "IPython: A System for Interactive Scientific Computing, Computing in Science and Engineering," *Computing in Science and Engineering*, vol. 9, no. 3, pp. 21–29, 2007.

[164] S. van der Walt, S. C. Colbert, and G. Varoquaux, "The NumPy Array: A Structure for Efficient Numerical Computation," *Computing in Science and Engineering*, vol. 13, no. 2, pp. 22–30, 2011.

[165] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Computing in Science and Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[166] W. McKinney, "Data structures for statistical computing in Python," in *Proceedings of the 9th Python in Science Conference* (S. van der Walt and J. Millman, eds.), vol. 445, pp. 51–56, 2010.

[167] N. M. Schmelling, "Code from: Minimal Tool Set for a Prokaryotic Circadian Clock. zenodo Repository.." http://doi.org/10.5281/zenodo.229910, 2017.

[168] N. M. Schmelling, "Raw Data from: Minimal Tool Set for a Prokaryotic Circadian Clock. figshare Repository.." https://doi.org/10.6084/m9.figshare.3823902.v3, 2017.

[169] R. A. Fisher, "On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P," *Journal of the Royal Statistical Society*, vol. 85, no. 1, pp. 87–94, 1922.

[170] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[171] R. G. Labiosa, K. R. Arrigo, C. J. Tu, D. Bhaya, S. Bay, A. R. Grossman, and J. Shrager, "Examination of diel changes in global transcript accumulation in *Synechocystis* (Cyanobacteria)," *Journal of Phycology*, vol. 42, no. 3, pp. 622–636, 2006.

[172] K.-I. Kucho, K. Okamoto, Y. Tsuchiya, S. Nomura, M. Nango, M. Kanehisa, and M. Ishiura, "Global analysis of circadian expression in the cyanobacterium *Synechocystis* sp. strain PCC 6803," *Journal of Bacteriology*, vol. 187, no. 6, pp. 2190–9, 2005.

[173] J. R. Toepel, J. E. McDermott, T. C. Summerfield, and L. A. Sherman, "Transcriptional analysis of the unicellular, diazotrophic cyanobacterium *Cyanothece* sp. 51142 grown under short day/night cycles," *Journal of Phycology*, vol. 45, no. 3, pp. 610–20, 2009.

[174] Q. Yang, B. F. Pando, G. Dong, S. S. Golden, and A. van Oudenaarden, "Circadian gating of the cell cycle revealed in single cyanobacterial cells," *Science*, vol. 327, no. 5972, pp. 1522–1526, 2010.

[175] P. O. Westermark and H. Herzel, "Mechanism for 12 hr rhythm generation by the circadian clock," *Cell Reports*, vol. 3, no. 4, pp. 1228–38, 2013.

[176] U. K. Laemmli, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4," *Nature*, vol. 227, no. 5259, pp. 680–5, 1970.

[177] M. Loza-Correa, L. Gomez-Valero, and C. Buchrieser, "Circadian clock proteins in prokaryotes: hidden rhythms?," *Frontiers in Microbiology*, vol. 1, p. 130, 2010.

[178] T. Elvitigala, J. Stöckel, B. K. Ghosh, and H. B. Pakrasi, "Effect of continuous light on diurnal rhythms in *Cyanothece* sp. ATCC 51142," *BMC Genomics*, vol. 10, p. 226, 2009.

[179] T. Shi, I. Ilikchyan, S. Rabouille, and J. P. Zehr, "Genome-wide analysis of diel gene expression in the unicellular N_2-fixing cyanobacterium *Crocosphaera watsonii* WH 8501," *The ISME Journal*, vol. 4, no. 5, pp. 621–632, 2010.

[180] S. S. Golden, M. Ishiura, C. H. Johnson, and T. Kondo, "Cyanobacterial Circadian Rhythms," *Annual Review of Plant Physiology and Plant Molecular Biology*, vol. 48, pp. 327–354, 1997.

[181] Y.-B. Chen, J. P. Zehr, and M. Mellon, "Growth and nitrogen fixation of the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium sp.* IMS 101 in defined media: Evidence for a circadian rhythm," *Journal of Phycology*, vol. 32, no. 6, pp. 916–923, 1996.

[182] T. Roenneberg and E. J. Carpenter, "Daily rhythm of O_2-evolution in the cyanobacterium *Trichodesmium thiebautii* under natural and constant conditions," *Marine Biology*, vol. 117, no. 4, pp. 693–697, 1993.

[183] K. Onai, M. Morishita, S. Itoh, K. Okamoto, and M. Ishiura, "Circadian Rhythms in the Thermophilic Cyanobacterium *Thermosynechococcus elongatus*: Compensation of Period Length over a Wide Temperature Range," *Journal of Bacteriology*, vol. 186, no. 15, pp. 4972–4977, 2004.

[184] T. Nakayama, R. Kamikawa, G. Tanifuji, Y. Kashiyama, N. Ohkouchi, J. M. Archibald, and Y. Inagaki, "Complete genome of a nonphotosynthetic cyanobacterium in a diatom reveals recent adaptations to an intracellular lifestyle," *PNAS*, vol. 111, no. 31, pp. 11407–11412, 2014.

[185] Y. Nakamura, T. Kaneko, S. Sato, M. Mimuro, H. Miyashita, T. Tsuchiya, S. Sasamoto, A. Watanabe, K. Kawashima, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, N. Nakazaki, S. Shimpo, C. Takeuchi, M. Yamada, and S. Tabata, "Complete Genome Structure of *Gloeobacter violaceus* PCC 7421, a Cyanobacterium that Lacks Thylakoids," *DNA Research*, vol. 10, no. 4, pp. 137–145, 2003.

[186] R. K. Shultzaberger, J. S. Boyd, T. Katsuki, S. S. Golden, and R. J. Greenspan, "Single mutations in *sasA* enable a simpler Δ*cikA* gene network architecture with equivalent circadian properties.," *PNAS*, vol. 111, no. 47, pp. E5069–75, 2014.

[187] V. Dvornyk, "Molecular evolution of *ldpA*, a gene mediating the circadian input signal in cyanobacteria," *Journal of Molecular Evolution*, vol. 60, no. 1, pp. 105–112, 2005.

[188] H. Min, H. Guo, and J. Xiong, "Rhythmic gene expression in a purple photosynthetic bacterium, *Rhodobacter sphaeroides*," *FEBS Letters*, vol. 579, no. 3, pp. 808–12, 2005.

[189] T. Nishiwaki-Ohkawa, Y. Kitayama, E. Ochiai, and T. Kondo, "Exchange of ADP with ATP in the CII ATPase domain promotes autophosphorylation of cyanobacterial clock protein KaiC," *PNAS*, vol. 111, no. 12, pp. 4455–60, 2014.

[190] Y. Taniguchi, N. Takai, M. Katayama, T. Kondo, and T. Oyama, "Three major output pathways from the KaiABC-based oscillator cooperate to generate robust circadian *kaiBC* expression in cyanobacteria," *PNAS*, vol. 107, no. 7, pp. 3263–3268, 2010.

[191] I. Vakonakis, D. A. Klewer, S. B. Williams, S. S. Golden, and A. C. LiWang, "Structure of the N-terminal Domain of the Circadian Clock-associated Histidine Kinase SasA," *Journal of Molecular Biology*, vol. 342, no. 1, pp. 9–17, 2004.

[192] Y. Shi, G. W. Tyson, J. M. Eppley, and E. F. DeLong, "Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean.," *The ISME Journal*, vol. 5, no. 6, pp. 999–1013, 2011.

[193] A. Wilde and Y. Hihara, "Transcriptional and posttranscriptional regulation of cyanobacterial photosynthesis," *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, vol. 1857, no. 3, pp. 296–308, 2016.

[194] R. K. Shultzaberger, J. S. Boyd, S. Diamonad, R. J. Greenspan, and S. S. Golden, "Giving Time Purpose: The *Synechococcus elongatus* Clock in a Broader Network Context," *Annual Review of Genetics*, vol. 49, no. 1, pp. 485–505, 2015.

[195] M. A. Woelfle and C. H. Johnson, "No Promoter Left Behind: Global Circadian Gene Expression in Cyanobacteria," *Journal of Biological Rhythms*, vol. 21, no. 6, pp. 419–431, 2006.

[196] S. R. Jammalamadaka and A. SenGupta, *Topics in Circular Statistics*, vol. 5. World Scientific, 2001.

[197] O. Cheregi, C. Sicora, P. B. Kos, P. J. Nixon, and I. Vass, "The FtsH protease is required for the repair of Photosystem II in the cyanbacterium *Synechocystis sp.* PCC 6803 damaged UV-B radiation," *BMC Plant Biology*, vol. 5, no. 1, pp. 1–2, 2005.

[198] S. Jacobshagen, J. R. Whetstine, and J. M. Boling, "Many but not all genes in *Chlamydomonas reinhardtii* are regulated by the Circadian Clock," *Plant Biology*, vol. 3, no. 6, pp. 592–597, 2001.

[199] S. L. Harmer, J. B. Hogenesch, M. Straume, H.-S. Chang, B. Han, T. Zhu, X. Wang, J. A. Kreps, and S. A. Kay, "Orchestrated Transcription of Key Pathways in *Arabidopsis* by the Circadian Clock," *Science*, vol. 290, no. 5499, pp. 2110–2113, 2000.

[200] Y. Kabeya, H. Nakanishi, K. Suzuki, T. Ichikawa, Y. Kondou, M. Matsui, and S.-Y. Miyagishima, "The YlmG protein has a conserved function related to the distribution of nucleoids in chloroplasts and cyanobacteria," *BMC Plant Biology*, vol. 10, no. 1, pp. 1–13, 2010.

[201] K. Imai, T. Nishiwaki, T. Kondo, and H. Iwasaki, "Circadian rhythms in the synthesis and degradation of a master clock protein KaiC in cyanobacteria," *Journal of Biological Chemistry*, vol. 279, no. 35, pp. 36534–9, 2004.

[202] H. Iwasaki, T. Nishiwaki, Y. Kitayama, M. Nakajima, and T. Kondo, "KaiA-stimulated KaiC phosphorylation in circadian timing loops in cyanobacteria.," *PNAS*, vol. 99, no. 24, pp. 15788–93, 2002.

[203] D. Zwicker, D. K. Lubensky, and P. R. Ten Wolde, "Robust circadian clocks from coupled protein-modification and transcription-translation cycles," *PNAS*, vol. 107, no. 52, pp. 22540–5, 2010.

[204] D. Memon, A. K. Singh, H. B. Pakrasi, and P. P. Wangikar, "A global analysis of adaptive evolution of operons in cyanobacteria," *Antonie van Leeuwenhoek*, vol. 103, no. 2, pp. 331–346, 2013.

[205] T. Uzumaki, M. Fujita, T. Nakatsu, F. Hayashi, H. Shibata, N. Itoh, H. Kato, and M. Ishiura, "Crystal structure of the C-terminal clock-oscillator domain of the cyanobacterial KaiA protein.," *Nature Structural and Molecular Biology*, vol. 11, no. 7, pp. 623–631, 2004.

[206] H. Iwasaki, Y. Taniguchi, M. Ishiura, and T. Kondo, "Physical interactions among circadian clock proteins KaiA, KaiB and KaiC in cyanobacteria.," *The EMBO Journal*, vol. 18, no. 5, pp. 1137–1145, 1999.

[207] Y. Xu, T. Mori, R. Pattanayek, S. Pattanayek, M. Egli, and C. H. Johnson, "Identification of key phosphorylation sites in the circadian clock protein KaiC by crystallographic and mutagenetic analyses," *PNAS*, vol. 101, no. 38, pp. 13933–8, 2004.

[208] T. Nishiwaki, Y. Satomi, M. Nakajima, C. Lee, R. Kiyohara, H. Kageyama, Y. Kitayama, M. Temamoto, A. Yamaguchi, A. Hijikata, M. Go, H. Iwasaki, T. Takao, and T. Kondo, "Role of KaiC phosphorylation in the circadian clock system of *Synechococcus elongatus* PCC 7942," *PNAS*, vol. 101, no. 38, pp. 13927–32, 2004.

[209] I. Vakonakis and A. C. LiWang, "Structure of the C-terminal domain of the clock protein KaiA in complex with a KaiC-derived peptide: implications for KaiC regulation," *PNAS*, vol. 101, no. 30, pp. 10925–30, 2004.

[210] D. D. Leipe, L. Aravind, N. V. Grishin, and E. V. Koonin, "The Bacterial Replicative Helicase DnaB Evolved from a RecA Duplication," *Genome Research*, vol. 10, no. 1, pp. 5–16, 2000.

[211] J. E. Walker, M. Saraste, M. J. Runswick, and N. J. Gay, "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold," *The EMBO Journal*, vol. 1, no. 8, pp. 945–51, 1982.

[212] M. Egli, T. Mori, R. Pattanayek, Y. Xu, X. Qin, and C. H. Johnson, "Dephosphorylation of the core clock protein KaiC in the cyanobacterial KaiABC circadian oscillator proceeds via an ATP synthase mechanism," *Biochemistry*, vol. 51, no. 8, pp. 1547–58, 2012.

[213] Y. Kitayama, T. Nishiwaki-Ohkawa, Y. Sugisawa, and T. Kondo, "KaiC intersubunit communication facilitates robustness of circadian rhythms in cyanobacteria," *Nature Communications*, vol. 4, p. 2897, 2013.

[214] R. Pattanayek, Y. Xu, A. Lamichhane, C. H. Johnson, and M. Egli, "An arginine tetrad as mediator of input-dependent and input-independent ATPases in the clock protein KaiC," *Acta Crystallographica Section D: Biological Crystallography*, vol. 70, no. 5, pp. 1375–90, 2014.

[215] R. Pattanayek, T. Mori, Y. Xu, S. Pattanayek, C. H. Johnson, and M. Egli, "Structures of KaiC circadian clock mutant proteins: a new phosphorylation site at T426 and mechanisms of kinase, ATPase and phosphatase," *PLoS One*, vol. 4, no. 11, p. e7529, 2009.

[216] Y. Xu, T. Mori, X. Qin, H. Yan, M. Egli, and C. H. Johnson, "Intramolecular regulation of phosphorylation status of the circadian clock protein KaiC," *PLoS One*, vol. 4, no. 11, p. e7509, 2009.

[217] R. G. Garces, N. Wu, W. Gillon, and E. F. Pai, "Anabaena circadian clock proteins KaiA and KaiB reveal a potential common binding site to their partner KaiC," *The EMBO Journal*, vol. 23, no. 8, pp. 1688–1698, 2004.

[218] V. Dvornyk and B. Knudsen, "Functional divergence of the circadian clock proteins in prokaryotes," *Genetica*, vol. 124, pp. 247–254, 2005.

[219] A. Neuner, H. W. Jannasch, S. Belkin, and K. O. Stetter, "*Thermococcus litoralis* sp. nov.: A new species of extremely thermophilic marine archaebacteria," *Archives of Microbiology*, vol. 153, no. 2, pp. 205–207, 1990.

[220] J. M. Gonzalez, Y. Masuchi, F. T. Robb, J. W. Ammerman, D. L. Maeder, M. Yanagibayashi, J. Tamaoka, and C. Kato, "*Pyrococcus horikoshii* sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough," *Extremophiles*, vol. 2, pp. 123–130, 1998.

[221] H. Kato, S. Watanabe, N.-M. Kaori, T. Chibazakura, Y. Tozawa, and H. Yoshikawa, "Exploration of a Possible Partnership among Orphan Two-Component System Proteins in Cyanobacterium *Synechococcus elongatus* PCC 7942," *Bioscience, Biotechnology, and Biochemistry*, vol. 76, no. 8, pp. 1484–1491, 2012.

[222] N. Dalchau, G. Szép, R. Hernansaiz-Ballesteros, C. P. Barnes, L. Cardelli, A. Phillips, and A. Csikász-Nagy, "Computing with biological switches and clocks," *Natural Computing*, 6 2018.

[223] M. A. Woelfle, Y. Xu, X. Qin, and C. H. Johnson, "Circadian rhythms of superhelical status of DNA in cyanobacteria," *PNAS*, vol. 104, no. 47, pp. 18819–18824, 2007.

[224] R. R. Klevecz and C. M. Li, "Evolution of the Clock from Yeast to Man by Period-Doubling Folds in the Cellular Oscillator," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 72, pp. 421–429, 2007.

[225] R. Machné and D. B. Murray, "The Yin and Yang of Yeast Transcription: Elements of a Global Feedback System between Metabolism and Chromatin," *PLoS ONE*, vol. 7, pp. 1–22, 06 2012.

[226] E. Emberly and N. S. Wingreen, "Hourglass Model for a Protein-Based Circadian Oscillator," *Phys. Rev. Lett.*, vol. 96, pp. 38303–38307, Jan 2006.

[227] T. Mori, D. R. Williams, M. O. Byrne, X. Qin, M. Egli, H. S. Mchaourab, P. L. Stewart, and C. H. Johnson, "Elucidating the Ticking of an *In Vitro* Circadian Clockwork," *PLoS Biology*, vol. 5, pp. 1–13, 03 2007.

[228] M. Yoda, K. Eguchi, T. P. Terada, and M. Sasai, "Monomer-Shuffling and Allosteric Transition in KaiC Circadian Oscillation," *PLoS ONE*, vol. 2, pp. 1–8, 05 2007.

[229] S. Clodong, U. Dühring, L. Kronk, A. Wilde, I. M. Axmann, H. Herzel, and M. Kollmann, "Functioning and robustness of a bacterial circadian clock," *Molecular Systems Biology*, vol. 3, no. 1, 2007.

[230] J. S. van Zon, D. K. Lubensky, P. R. H. Altena, and P. R. ten Wolde, "An allosteric model of circadian KaiC phosphorylation," *PNAS*, vol. 104, no. 18, pp. 7420–7425, 2007.

[231] S. Legewie, B. Schoeberl, N. Blüthgen, and H. Herzel, "Competing Docking Interactions can Bring About Bistability in the MAPK Cascade," *Biophysical Journal*, vol. 93, no. 7, pp. 2279–2288, 2007.

[232] C. C. Jolley, K. L. Ode, and H. R. Ueda, "A Design Principle for a Post-translational Biochemical Oscillator," *Cell Reports*, vol. 2, no. 4, pp. 938–950, 2012.

[233] A. Mehra, C. I. Hong, M. Shi, J. J. Loros, J. C. Dunlap, and P. Ruoff, "Circadian Rhythmicity by Autocatalysis," *PLoS Computational Biology*, vol. 2, pp. 1–8, 07 2006.

[234] H. Takigawa-Imamura and A. Mochizuki, "Predicting Regulation of the Phosphorylation Cycle of KaiC Clock Protein Using Mathematical Analysis," *Journal of Biological Rhythms*, vol. 21, no. 5, pp. 405–416, 2006.

[235] I. M. Axmann, S. Legewie, and H. Herzel, "A Minimal Circadian Clock Model," *Genome Informatics*, vol. 18, pp. 54–64, 2007.

[236] E. E. Zhang and S. A. Kay, "Clocks not winding down: unravelling circadian networks," *Nature Reviews Molecular Cell Biology*, vol. 11, 10 2010.

[237] G. Kurosawa, K. Aihara, and Y. Iwasa, "A Model for the Circadian Rhythm of Cyanobacteria that Maintains Oscillation without Gene Expression," *Biophysical Journal*, vol. 91, pp. 2015–2023, 09 2006.

[238] S. Hertel, C. Brettschneider, and I. M. Axmann, "Revealing a Two-Loop Transcriptional Feedback Mechanism in the Cyanobacterial Circadian Clock," *PLoS Computational Biology*, vol. 9, pp. 1–16, 03 2013.

[239] R. D. Hernansaiz-Ballesteros, L. Cardelli, and A. Csikász-Nagy, "Single molecules can operate as primitive biological sensors, switches and oscillators," *BMC Systems Biology*, vol. 12, p. 70, Jun 2018.

[240] V. Dvornyk, "Evolution of the Circadian Clock Mechanism in Prokaryotes," *Israel Journal of Ecology & Evolution*, vol. 52, no. 3-4, pp. 343–357, 2006.

[241] C. Troein, J. C. Locke, M. S. Turner, and A. J. Millar, "Weather and Seasons Together Demand Complex Biological Clocks," *Current Biology*, vol. 19, pp. 1961–1964, 12 2009.

[242] G. E. Williams, "Geological constraints on the Precambrian history of Earth's rotation and the Moon's orbit," *Reviews of Geophysics*, vol. 38, pp. 37–59, 2 2000.

[243] W. F. Bottke, D. Vokrouhlický, S. Marchi, T. Swindle, E. R. D. Scott, J. R. Weirich, and H. Levison, "Dating the Moon-forming impact event with asteroidal meteorites," *Science*, vol. 348, no. 6232, pp. 321–323, 2015.

[244] M. Ćuk and S. T. Stewart, "Making the Moon from a Fast-Spinning Earth: A Giant Impact Followed by Resonant Despinning," *Science*, vol. 338, no. 6110, pp. 1047–1052, 2012.

[245] R. de Meijer, V. Anisichkin, and W. van Westrenen, "Forming the Moon from terrestrial silicate-rich material," *Chemical Geology*, vol. 345, pp. 40–49, 2013.

[246] S. Gan and E. K. O'Shea, "An Unstable Singularity Underlies Stochastic Phasing of the Circadian Clock in Individual Cyanobacterial Cells," *Molecular Cell*, vol. 67, pp. 659–672, 08 2017.

[247] S. J. Biller, A. Coe, S. E. Roggensack, and S. W. Chisholm, "Heterotroph Interactions Alter *Prochlorococcus* Transcriptome Dynamics during Extended Periods of Darkness," *mSystems*, vol. 3, no. 3, 2018.

[248] C. Schmal, J. Myung, H. Herzel, and G. Bordyugov, "A Theoretical Study on Seasonality," *Frontiers in Neurology*, vol. 6, p. 94, 2015.

[249] J. Paijmans, D. K. Lubensky, and P. R. ten Wolde, "A thermodynamically consistent model of the post-translational Kai circadian clock," *PLoS Computational Biology*, vol. 13, pp. 1–43, 03 2017.

[250] J. Paijmans, D. K. Lubensky, and P. Rein ten Wolde, "Robustness of synthetic oscillators in growing and dividing cells," *Phys. Rev. E*, vol. 95, pp. 52403–52415, 05 2017.

[251] Y. Kitayama, T. Nishiwaki, K. Terauchi, and T. Kondo, "Dual KaiC-based oscillations constitute the circadian system of cyanobacteria," *Genes & Development*, vol. 22, no. 11, pp. 1513–1521, 2008.

[252] C. W. Mullineaux and R. Stanewsky, "The Rolex and the Hourglass: a Simplified Circadian Clock in *Prochlorococcus*?," *Journal of Bacteriology*, vol. 191, no. 17, pp. 5333–5335, 2009.

[253] J. Ungerer, K. E. Wendt, J. I. Hendry, C. D. Maranas, and H. B. Pakrasi, "Comparative genomics reveals the molecular determinants of rapid growth of the cyanobacterium *Synechococcus elongatus* UTEX 2973," *PNAS*, 2018.

[254] B. Damer and D. Deamer, "Coupled Phases and Combinatorial Selection in Fluctuating Hydrothermal Pools: A Scenario to Guide Experimental Approaches to the Origin of Cellular Life," *Life*, vol. 5, no. 1, pp. 872–887, 2015.

[255] A. Y. Mulkidjanian, A. Y. Bychkov, D. V. Dibrova, M. Y. Galperin, and E. V. Koonin, "Origin of first cells at terrestrial, anoxic geothermal fields," *PNAS*, vol. 109, no. 14, pp. E821–E830, 2012.

[256] J. G. Forsythe, S.-S. Yu, I. Mamajanov, M. A. Grover, R. Krishnamurthy, F. M. Fernández, and N. V. Hud, "Ester-Mediated Amide Bond Formation Driven by Wet–Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth," *Angewandte Chemie International Edition*, vol. 54, no. 34, pp. 9871–9875, 2015.

[257] R. J. Rapf and V. Vaida, "Sunlight as an energetic driver in the synthesis of molecules necessary for life," *Phys. Chem. Chem. Phys.*, vol. 18, pp. 20067–20084, 2016.

[258] A. Y. Mulkidjanian, D. A. Cherepanov, and M. Y. Galperin, "Survival of the fittest before the beginning of life: selection of the first oligonucleotide-like polymers by UV light," *BMC Evolutionary Biology*, vol. 3, p. 12, 05 2003.

Author's Declaration of Contribution

The proportional contributions of the before listed publication, which are the subject matter of this thesis, are explained in detail in the following.

Publication I

Joost Snijder, Jan M. Schuller, Anika Wiegard, Philip Lössl, Nicolas Schmelling, Ilka M. Axmann, Jürgen M. Plitzko, Friedrich Förster, Albert J. R. Heck (2017) Structures of the cyanobacterial circadian oscillator frozen in a fully assembled state, *Science*, 355 (6330), 1181–1184, doi: *10.1126/science.aag3218*

Author Contributions

JS, JMS, JMP, FF, and AJRH designed the structural experiments. JS, JMS, JMP, and FF performed the cryo-EM analyses. JS and PL performed the mass spectrometry analyses. AW designed the protein expression experiments and cloned, expressed, and purified recombinant proteins. NS designed and carried out bioinformatic sequence analyses. AJRH, FF, JMP, and IMA supervised the study. JS and JMS prepared the manuscript together with AW, PL, JMP, FF, and AJRH. All authors read and approved the final manuscript.

Pro rata

Bioinformatic study design: 100%
Bioinformatic data analyses: 100%

Publication II

Nicolas M. Schmelling, Robert Lehmann, Paushali Chaudhury, Christian Beck, Sonja-Verena Albers, Ilka M. Axmann, and Anika Wiegard (2017) Minimal tool set for a prokaryotic circadian clock, *BMC Evolutionary Biology*, 17:169, doi: *10.1186/s12862-017-0999-7*

Author Contributions

IMA and AW supervised the study. NMS, CB, and RL designed and NMS and RL carried out bioinformatic data analyses. AW designed and performed sequence analysis and *in vitro* experiments. AW and PC purified recombinant proteins. NMS, RL, CB, AW, and IMA prepared the manuscript. PC and SVA contributed to the interpretation of the data and provided intellectual input. All authors read and approved the final manuscript.

Pro rata

Bioinformatic study design: 40%
Bioinformatic data analyses: 50%
Manuscript preparation: 40%

Publication III

Nicolas M. Schmelling and Ilka M. Axmann (2018) Computational modeling unravels the precise clockwork of cyanobacteria, *Interface Focus*, 20180038. doi: *10.1098/rsfs.2018.0038*

Author Contributions

NMS performed the literature search, designed the figures, and wrote the manuscript. IMA supervised the work, gave valuable comments on the manuscript throughout the writing process, and proofread the manuscript. Both authors gave final approval for publication.

Pro rata

Literature research: 100%
Manuscript preparation: 80%

Statutory declaration and statement

Hiermit versichere ich an Eides statt, dass die eingereichte Dissertation selbständig und ohne unzulässige fremde Hilfe unter Beachtung der "Ordnung über die Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf" verfasst worden ist. Weiter versichere ich andere als die in der Dissertation angegebene Literatur nicht benutzt zu haben und dass ich alle ganz oder annähernd übernommenen Textstellen sowie verwendete Grafiken, Tabellen und Auswertungsprogramme kenntlich gemacht habe. Außerdem versichere ich, dass die vorgelegte elektronische mit der schriftlichen Version der Dissertation übereinstimmt und die Abhandlung in dieser oder ähnlicher Form noch nicht anderweitig als Promotionsleistung vorgelegt und bewertet wurde.

English Version

I hereby declare an oath that the submitted dissertation has been written independently and without undue assistance, taking into account the "Rules and Principles for Saveguarding and Good Scientific Practice at HHU". Further, I assure that I did not use other literature than the ones specified in the dissertation and that I have highlighted complete or partly adopted texts as well as graphics, tables, and analysis programs used in this work. Furthermore, I assure that the submitted electronic version is in accordance with the written version of the dissertation and that the dissertation in this or similar form has not yet been presented and evaluated as a doctoral thesis.

Gezeichnet

Nicolas M. Schmelling
Düsseldorf, den XX. Dezember 2018