THE UNIVERSITY OF CHICAGO

VULNERABILITY TO MOLECULAR NOISE CONSTRAINS THE ARCHITECTURE OF BACTERIAL CLOCKS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM: GENETICS, GENOMICS,

AND SYSTEMS BIOLOGY

BY

JUSTIN CHEW

CHICAGO, ILLINOIS

DECEMBER 2017

Table of Contents

Abstract

Circadian rhythms are a remarkable feature found in many disparate organisms across the planet, driving 24-hour oscillations in gene expression and behavior to align organism physiology with the day-night cycle. Perhaps two of the most fundamental questions in circadian biology are: how do these endogenous biological rhythms maintain a robust 24-hour periodicity, and what are the consequences if the clock fails to function correctly? To address these basic questions, in this work I and my colleagues turn to the model photosynthetic cyanobacterium Synechococcus elongatus, which possesses the simplest known circadian clock composed of only three proteins, KaiA, KaiB, and KaiC. In the main chapter of this dissertation, I investigate how molecular stochasticity impacts the precision of the cyanobacterial clock as a result of limited cellular Kai protein copy number, finding that high protein expression is required to suppress timing errors due to a noisy negative feedback loop in the oscillator. Additionally, I find that the molecular noise inherent in the feedback loop forces a smaller, related cyanobacterium to adopt a qualitatively different environmentally-driven timing strategy that is more optimal for a lower Kai copy number. The other two studies presented here are those in which I contributed as a co-author to investigate how the Kai oscillator maintains period robustness against fluctuations in protein stoichiometry as well as how timing mismatch between the clock and environment impacts the fitness of individual cyanobacterial cells. Lastly, I present an ongoing project that utilizes ancestral protein reconstruction techniques to determine whether the period of the Kai oscillator changed over geological time to match the changing period of the Earth's rotation.

List of Figures

List of Tables

Chapter 1: Introduction

The ubiquity and usefulness of circadian clocks

Circadian rhythms drive ~24-hour oscillations in physiology and gene expression that allow organisms to anticipate the day/night cycle, and they have been long known to exist across multiple kingdoms of life. Humans, insects, plants, and fungi all possess circadian clocks, and the fact that many of these timing systems do not share any degree of homology suggests that these rhythms evolved through convergent evolution, highlighting the importance of being able to tell time endogenously. While in principle it is possible to measure the time of day by resetting a daily timer with respect to either sunrise or sunset, circadian rhythms are distinguished from such "timer" mechanisms by the fact that they persist in the absence of any cyclic environmental stimuli.

Bona fide circadian clocks are themselves defined by three hallmark criteria: first, 24-hour rhythms must be endogenous and persist in constant conditions; second, rhythms must be entrainable, i.e. resettable by environmental cues; and third, rhythms must be temperature-compensated, maintaining correct periodicity across a range of temperatures. Here, I will briefly discuss each of these three criteria.

The first experiments to demonstrate the presence of endogenous rhythms in any organism were performed in the early 1700s, where French scientist Jean-Jacques d'Ortous de Mairan noted that leaf movements in the plant *Mimosa pudica* occurred on a 24-hour cycle even when the plants were incubated in constant darkness, shielded from the normal environmental light cues indicating day or night (*6*). More recent experiments demonstrated the same endogenous timing characteristic in the physiology of other organisms: in humans and mice, sleep patterns persist in constant light or constant dark (*7*, *8*), and in *Drosophila*, the circadian

timing of pupal eclosion also persists in constant conditions (*9*). The rhythms in these organisms are thus termed “free-running”.

Like a mechanical watch that only maintains a finite level of precision, the timing of biological clocks can drift from the true time of day. It is thus crucial that circadian clocks possess some form of entrainment, or a mechanism by which endogenously generated rhythms can synchronize with the day/night cycle. Otherwise, the information about the time of day captured by the clock would decay over time, and the clock would have no utility for the organism. Common cues that have been demonstrated to entrain circadian clocks include light (*10*, *11*), temperature (*12*, *13*), or metabolic cues from feeding (*14*).

Lastly, all biological clocks are biochemical in nature. Thus, they are subject to Arrhenius’s Law, which predicts that chemical reactions proceed more rapidly at higher temperatures. To counter this, circadian clocks have evolved to minimize the impact of temperature on the speed of clock reactions, a characteristic known as temperature compensation. Temperature compensation is especially critical for organisms that lack self-regulated temperature in order to cope with the temperature changes associated not only with the day/night cycle but also with weather and seasons. A clock that runs faster or slower on warmer or colder days would be a poor timekeeper. Temperature compensation is quantified by a measure known as Q10, or the fold change in chemical reaction rate given a 10°C change in temperature. Many biological reactions double or triple their reaction rates for a 10°C increase and thus have a Q10 value of ~2-3 (*15*), but the periodic frequency of circadian clocks show a Q10 value closer to 1, indicating that the circadian period remains close to 24 hours over a range of temperatures. Circadian rhythms in leaf movements in *Arabidopsis* show a Q10 value of 1.0-

1.1 (*16*), and circadian rhythms in *Drosophila* locomotion show a Q10 value of approximately 1 (*17*).

What utility does a circadian clock provide for an organism? The clock allows an organism to optimize its physiology for changes that accompany the day/night cycle. For a plant, this might involve increasing expression of genes involved in photosynthesis during the day and conserving energy during the night. For mice, the clock regulates sleep/wake behavior, allowing the mouse to become active only at night in order to avoid predators. It is even thought that in rodent skin cells, the circadian clock regulates the timing of cell division and DNA synthesis such that maximum sensitivity to UV-induced DNA damage occurs during the night and is avoided during the day (*18*).

The physical limits of biology: molecular noise

For many decades, biologists have thought of the inner workings of the cell as a series of pathways containing multitudes of genes and proteins that interact in a well-defined and deterministic manner. For example, a membrane receptor might bind to its ligand, causing phosphorylation of a downstream transcription factor which in turn binds to a further downstream gene promoter to induce expression of that gene. However, over the last fifteen years, another perspective has emerged that has profound implications for how we think of cells—the idea of biological noise. Within a population of genetically identical cells, individual cells may express specific proteins in randomly varying amounts, leading to heterogeneity in behavior, phenotype, and response to stimuli. We define biological noise as the random fluctuations that lead to such heterogeneity.

The landmark study that definitely illustrated the effects of biological noise in single cells was performed by Michael Elowitz, in which he designed a strain of *E. coli* to express two different fluorescent proteins (CFP and YFP) from identical promoters (*2*). Using time lapse fluorescent microscopy, Elowitz observed that in individual cells, the expression levels of the fluorescent proteins were not constant with respect to time and in fact fluctuated by a significant amount as cells grew and divided such that there was a wide range of expression ratios between the two proteins (Figure 1.1). Closer inspection revealed that the noise could be classified into two types: extrinsic noise and intrinsic noise (Figure 1.2). In this experiment, extrinsic noise was defined as being responsible for highly correlated fluctuations in CFP and YFP expression, and intrinsic noise was defined as being responsible for the uncorrelated component of these fluctuations.

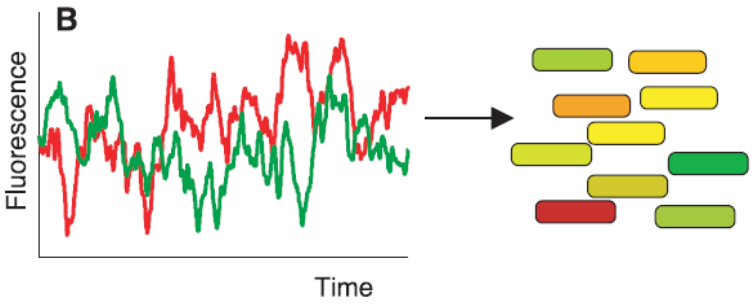

Figure 1.1. Schematic of gene expression noise in individual E. coli cells expressing two fluorescent proteins (CFP and YFP) from two identical promoters. Noise in gene expression causes a population of isogenic cells to express the two proteins in heterogenous fashion. Figure adapted from (*2*).

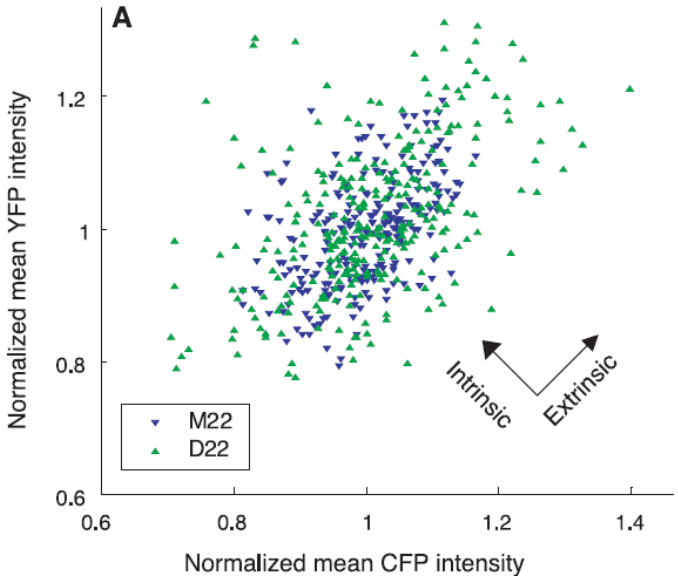

Figure 1.2. Quantifying extrinsic vs. intrinsic noise. Extrinsic noise causes correlated changes in expression for both fluorescent proteins, and intrinsic noise causes uncorrelated changes. Figure adapted from (*2*).

Mechanistically, extrinsic noise can be thought of as arising from non-genetic differences between cells that can impact cellular processes such as protein expression. Some examples of sources of extrinsic noise include local differences in the microenvironment (e.g. temperature or the local concentration of a signaling ligand), the unevenness of biomolecule distribution in daughter cells following cell division, or even variation in the number of ribosomes present in each cell. Each of these sources can introduce variability into a pool of genetically identical cells to produce population heterogeneity. In contrast, intrinsic noise originates from a more physically fundamental source, i.e. the randomness of molecular collisions and diffusion. This stochasticity is brought to light in the context of single cells, which have only finite copy numbers of biomolecules. Molecules present at low cellular copy numbers (e.g. mRNAs, measured to be < 10 copies per gene in *E. coli* (*19*)) are especially susceptible to intrinsic noise, as there are fewer copies over which to average biochemical stochasticity. As a consequence, even if it were possible to initialize cells in identical starting conditions (thus eliminating extrinsic noise), the biochemical randomness that defines intrinsic noise could set two cells off on differing trajectories of gene expression, contributing to population heterogeneity. In practice, extrinsic noise and intrinsic noise are not two independently operating phenomena, and in fact intrinsic noise can contribute to the amount of extrinsic noise (e.g. randomness of transcription factor binding for a cell division gene may lead a cell to divide earlier than its counterparts, leading to greater uneven partitioning of ribosomes after division due to a lower overall ribosome copy number at cell division). Our categorization of noise into either extrinsic or intrinsic sources is thus a convention that can help to organize sources of variability in a population of cells.

Unlike mechanical watches, circadian clocks operate within a cellular context and are composed of a series of biochemical reactions that generate oscillations. Like all other biological processes within the cell, circadian clock function is thus subject to the same physical constraints of biological noise. Given that the utility of clocks derives from their ability to keep precise time, a central question arises: how do biological clocks generate precise, 24-hour deterministic rhythms despite being composed of fundamentally stochastic biochemical reactions?

Synthetic biological oscillators and attempts to improve their precision

The design and construction of the first biological oscillator by Michael Elowitz in 2000 (the so-called "repressilator") highlights the difficulty in designing a biological oscillator with a precise and consistent period. This "repressilator" is composed of a network of three genes, in which each gene represses the expression of another, forming a feedback loop with three negatively regulated components (*1*) (Figure 1.3). The network was designed based on predictions from mathematical modeling that such a network would be capable of generating oscillations. To test the prediction, this gene network was constructed on a plasmid that was transformed into *E. coli*, and time lapse fluorescent microscopy confirmed that this network could indeed generate oscillations in single cells, although with complications. Perhaps one of the most obvious

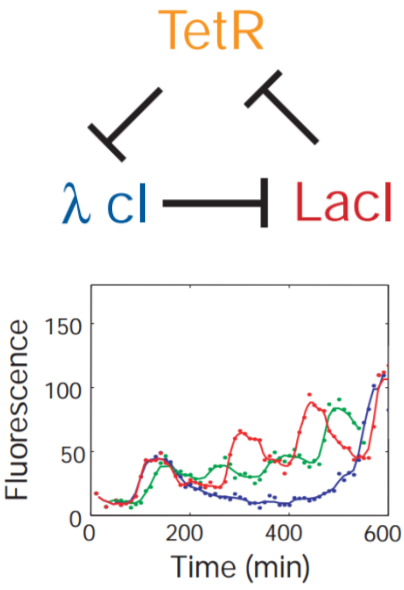

Figure 1.3. Repressilator schematic and behavior. Top: the repressilator oscillator is composed of three transcription factors that form a loop of negative feedback. Bottom: oscillations in individual cells are highly variable in period and amplitude. Figures adapted from (*1*).

characteristics of the repressilator was that oscillations in single cells were quite erratic, and that significant variation between cells existed in both oscillation amplitude and period (Figure 1.3).

Others have since designed and constructed other synthetic biological oscillators with varying degrees of precision. The two-gene oscillator network designed by Stricker et al. is slightly more reliable than the repressilator, but still falls short of producing the highly precise rhythms characteristic of circadian oscillations (*4*) (Figure 1.4). The mammalian synthetic oscillator engineered by Tigges et al. faces similar problems (*20*). It was not until recently, sixteen years after the initial repressilator, that a synthetic oscillator was engineered that could maintain an oscillatory period consistent enough to rival circadian oscillators.

In 2016, Potvin-Trottier et al. improved the original repressilator design, in which they systematically analyzed sources of molecular noise in the repressilator and eliminated them, improving the timing error per cycle (defined as period standard deviation/mean) from 28% to 14% (*21*). One of the most dominant sources of noise in the entire cycle lay within the threshold for negative repression of λ CI by TetR. The λ CI promoter switches from a repressed state to a non-repressed state at around only ~5 copies of TetR, making it extremely susceptible to low copy number noise effects. To remedy this, the authors introduced additional TetR binding sites on the plasmid that served as "sponges" that increased

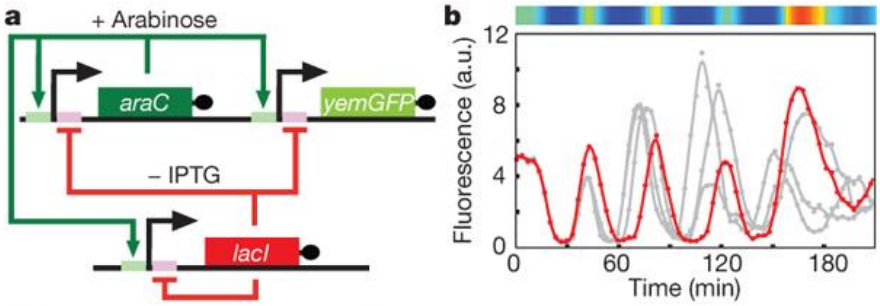

Figure 1.4. Two-gene oscillator schematic and behavior. Left: the oscillator is composed of two gene elements. *araC* positively regulates itself and *lacI*, while *lacI* negatively regulates itself and *araC*. Right: Oscillations in individual cells are variable in amplitude and period (gray traces, with one example trace highlighted in red). Top colorbar shows average fluorescent intensity of multiple cells. Figure adapted from (*4*).

this repressive threshold several-fold to ~20 copies of TetR. This study demonstrates that molecular noise can have a significant impact on the precision of biological oscillators, and that efforts must be taken to mitigate its impact if precision is an important design criterion.

The cyanobacterial clock is an extremely precise natural oscillator

Once thought to be too simple to ever possess something so complex, cyanobacteria regulate their physiology with circadian clocks that achieve a level of precision greater than any synthetic biological oscillator devised thus far (as shown in Chapter 2). The cyanobacterial clock is the simplest known circadian clock, and the core oscillator consists of three proteins: KaiA, KaiB, and KaiC. In contrast to every other known circadian clock, the core mechanism of the cyanobacterial clock is post-translational—remarkably, the clock reaction can be completely reconstituted *in vitro* with recombinant Kai protein (*22*). In fact, the demonstration of all three criteria for defining circadian clocks (endogenously generated 24-hour rhythms, entrainability, and temperature compensation) is

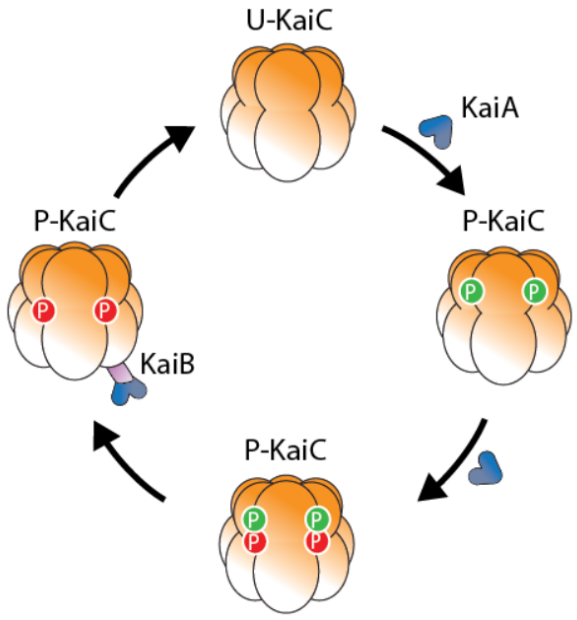

Figure 1.5. Diagram of the post-translational mechanism of the Kai clock cycle. KaiC encodes the time of day through 24-hour rhythmic ordered phosphorylation and dephosphorylation on two key residues, T432 (green) and S431 (red). Unphosphorylated KaiC (U-KaiC) binds to KaiA, stimulating KaiC's autokinase activity. Phosphorylated KaiC (P-KaiC) with a sufficient amount of S431 phosphorylation binds to KaiB, which sequesters KaiA and inhibits its activity. This allows KaiC to undergo its default phosphatase activity to return to unphosphorylated KaiC.

encapsulated in the biochemistry underlying the three Kai proteins (*22*, *23*).

The reconstitutability of the Kai system makes it a powerful platform for dissecting the biochemical mechanisms driving the core timekeeping process. Time of day is encoded in the hexameric protein KaiC, which undergoes rhythmic 24-hour cycles of autocatalyzed phosphorylation and dephosphroylation (Figure 1.5). KaiC's enzymatic activity is in turn modulated by KaiA and KaiB. Initially, unphosphorylated KaiC is competent to bind to KaiA, which stimulates KaiC's autokinase activity. Phosphorylated KaiC then adopts a conformation that allows it to bind to KaiB. This newly formed KaiBC complex can sequester KaiA, inhibiting its phosphorylation-stimulating activity to form the delayed negative feedback loop of the oscillator. This allows KaiC to then undergo its default autophosphatase activity. Once KaiC dephosphorylates, it binds to KaiA to start the cycle again.

Previous work has demonstrated that entrainment of the cyanobacterial clock can be explained by the effect of ATP levels on the Kai system. In cells, the ATP/(ATP + ADP) ratio cycles between a relatively high value of ~85% in the light and a lower value of ~40% in the dark (*24*). Simulation of this day/night cycle in the in vitro system (achieved by a buffer exchange protocol to alter the ATP ratio) results in the in vitro system entraining to the phase of ATP cycling, demonstrating that the capability for entrainment is present within the Kai proteins themselves (*23*).

Importantly, the cyanobacterial clock appears to operate independently within single cells, and there is currently no evidence of any intercellular communication to enhance the precision of the clock period. In fact, experiments demonstrate that physical contact between two cells that are in opposite clock phases has no effect on the oscillation phase in either cell (*25*). These results highlight that unlike the circadian pacemaker in higher organisms, in which

communication between individual neurons generates a high amplitude and deterministic circadian oscillation (*26*), the circadian clock in cyanobacteria must achieve a high level of precision on the single cell level. Experiments presented within this dissertation in Chapter 2 indicate that the cyanobacterial clock accumulates only ~5% timing error per cycle, well below that of the synthetic oscillators mentioned above. This level of timing precision can be observed in that a population of cyanobacteria can maintain synchronized oscillations for over two weeks (*27*), an impressive feat considering that each cell acts as an independent oscillator. One of the central questions that this dissertation will address is thus: how does the cyanobacterial clock overcome molecular noise to create such a precise oscillator, and what constraints does molecular noise place on circadian clock design?

Why should clocks be free-running?

One of the most outstanding questions in circadian biology is: what benefits does a free-running clock provide over an environmentally-driven timer in anticipating the day/night cycle? A fundamental requirement for rhythms to be considered "circadian" is that they continue to run in constant conditions. However, constant conditions almost never occur in nature, and the reality is that organisms evolved in the context of the day/night cycle, presumably since the dawn of life. Additionally, a biological timer that resets at sunrise or sunset (but does not generate free-running oscillations) should theoretically also be sufficient to track the passage of time. All of these considerations lead to the question of why circadian clocks evolved to become free-running, and in what conditions they are advantageous.

One hypothesis that has only been tested computationally thus far is that free-running clocks evolved in response to external noise in environmental entraining cues. Organisms that

rely on sunlight to synchronize with the day/night cycle must do so despite the presence of weather that may cause fluctuations in light levels. Troein et al. demonstrated *in silico* that gene networks that give rise to free-running oscillations are much more likely to evolve in the presence of environmental weather (*28*). These results suggest that a free-running clock may be beneficial by acting as a low-bandpass filter, allowing an organism to filter out the relatively high-frequency noise that may come from weather while retaining sensitivity to the low-frequency stimulus of the day/night cycle, necessary for entrainment. In other words, the internal timekeeping mechanism that a free-running clock provides can allow an organism to ignore environmental stimuli that do not correlate with the day/night cycle, e.g. avoiding the confusion of reduced light levels from a passing thunderstorm as impending nighttime, which timer-like systems might be more susceptible to. In addition to the effects of weather, Troein et al. showed that free-running oscillations evolved more frequently in the presence of a variable photoperiod, as it is in different seasons (e.g. the daylight period is longer in the summer and shorter in the winter in the Northern hemisphere). Thus, having an internal timekeeping mechanism may also allow organisms to better optimize their physiology for variable onset of day/night. However, while some studies have examined the mechanism by which clocks track seasons (e.g. in plants (*29*) and cyanobacteria (*23*)), the exact mechanisms by which circadian clocks generally measure seasonality still constitute an active area of research.

The question of why organisms may utilize clocks instead of timers is relevant to the work presented in this dissertation: I show that while the well-established model cyanobacterium *Synechococcus elongatus* PCC 7942 possesses a free-running circadian clock, a related picocyanobacterium *Prochlorococcus marinus* MED4 has an environmentally-driven timer. The specific questions that I investigate are outlined in the main objectives below.

Main objectives

What are the consequences for how biological oscillators are designed given the constraints from molecular noise, and how do these constraints arise? The work performed to investigate this question is elaborated upon in Chapter 2. Specifically, I experimentally test whether molecular noise stemming from limited Kai protein copy number is a determinant for clock precision in *S. elongatus* by engineering a strain in which Kai copy number is tunable. I then investigate whether a "noise bottleneck" exists in the system, or whether a specific part of the clock network is most susceptible to the effects of molecular noise. Lastly, I compare the two different Kai systems in *S. elongatus* and *P. marinus* to determine the precision of both systems with varying amounts of molecular noise and whether noise influences whether it is more optimal to keep time with a free-running clock or an environmentally-driven timer.

In Chapters 3 and 4, I present two studies on which I am second author. Chapter 3 explores how two opposing phosphorylation sites on KaiC generate an ultrasensitive switch governing KaiB/KaiC binding, ultimately generating robustness with respect to changes in Kai protein stoichiometry. Chapter 4 investigates the consequences for organism fitness from clock/environment mismatch in *Synechococcus* on the single cell level.

Lastly, in the appendix, I outline efforts to perform ancestral reconstruction of the Kai proteins to investigate questions about Kai protein evolution.

Chapter 2: Demand for high protein copy number can favor timers over clocks in bacteria

Foreword

Here, I present my first author work, currently in submission as of the writing of this dissertation. In brief, I investigate the biophysical constraints that molecular noise imposes on the precision of the cyanobacterial clock and conclude that environmentally driven timers are a more optimal timekeeping mechanism at low clock protein copy number in cyanobacteria. A discussion and future directions section at the end of this chapter provides further interpretation and insight of this work, and it details further experiments that may expand upon the work presented here.

Author contributions and acknowledgements

This study has been submitted for publication with the following authorship list: Justin Chew, Jenny Lin, Arvind Murugan, and Michael Rust. J.C. and M.J.R. designed experiments and prepared the manuscript. J.C. carried out experiments and data analysis. J.L. designed and performed protocol for purifying *Prochlorococcus* KaiC and edited the manuscript. J.C., M.J.R., and A.M. designed and analyzed the mathematical model, and J.C. implemented the model and analyzed output data.

In addition, we thank Eugene Leypunskiy and Connie Phong for assistance with purified proteins, Guillaume Lambert for assistance with microscopy, and Ed Munro, Joe Markson, and the Rust Lab for useful discussions. Maureen Coleman and Jesse Black provided assistance with flow cytometry and culturing *Prochlorococcus*. This work was supported by NIH training grant T32-GM007281, NIH F30 fellowship F30-GM117962 (to JC), NIH R01-GM107369 and a Pew Biomedical Scholars award (to MJR).

Abstract

Circadian clocks generate deterministic 24-hour rhythms to anticipate the day-night cycle, and they must accomplish this despite the fact that biological oscillators are based on fundamentally stochastic biochemical reactions. The model circadian clock in the cyanobacterium *Synechococcus elongatus* is based on the Kai proteins, a post-translational oscillator that can sustain precise rhythms for weeks in a test tube. However, a single bacterial cell has far fewer copies of the Kai proteins than a macroscopic reaction, raising the question of how bacteria produce deterministic behavior in spite of molecular stochasticity arising from finite protein copy number. Here, we show experimentally in *S. elongatus* that oscillations in single cells become erratic at low Kai copy number and that cells must express >10,000 copies of the Kai proteins to effectively suppress timing errors. Stochastic modeling shows that this need for many protein copies results from noise amplification in the post-translational feedback loop necessary for oscillations. We find that the much smaller cyanobacterium *Prochlorococcus*, a minimal photosynthetic cell, expresses only ~600 copies of the Kai proteins and has lost the crucial feedback loop, resulting in a timer-like Kai system that no longer free-runs. Information theoretic analysis shows that this timer strategy can outperform a free-running clock when stochastic effects are important. Thus, bacteria utilize two alternative time-keeping strategies: a free-running clock that uses many protein copies to achieve high timing precision, and a non-free-running timer that is less predictive but can function well when protein copy numbers are low. This conclusion has implications for the evolution of circadian rhythms, the design of synthetic biological timekeeping systems, and may point to a currently unexplored world of timer-like behavior in microbes in dynamic environments, such as the mammalian gut.

Main Text

Circadian clocks are biochemical oscillators that enable organisms to anticipate the day-night cycle. Their utility depends on the ability to make accurate predictions about the future (*30, 31*) and thus requires precise, deterministic timing. This precision must be achieved despite the fact that biochemical processes are composed of elementary reaction events, each of which occurs with stochastic timing. Indeed, most synthetic cellular oscillators produce noticeably irregular rhythms (*1, 4, 20*). In contrast, natural circadian clocks can be extremely precise (*27, 32, 33*). It is generally not known how biological clocks create deterministic rhythms from their stochastic components, or how the architecture of clock networks responds to the constraints of molecular noise.

To address these questions, we turned to the cyanobacterial circadian clock. Cyanobacteria are a diverse clade of photosynthetic prokaryotes that carry *kai* clock genes that generate daily oscillations in physiology (*34-36*). The core mechanism of oscillation in the cyanobacterial clock is post-translational and can be reconstituted using purified proteins (*22*). KaiA and KaiB modulate the autocatalytic activity of KaiC, producing self-sustaining rhythms of multisite phosphorylation on KaiC (*37*).

Because the volume of a bacterial cell is smaller than the volume of a test-tube reaction by many orders of magnitude, we suspected that stochasticity due to finite numbers of clock proteins might be an important constraint in cells. To study this effect, we engineered a strain of the model cyanobacterium *Synechococcus elongatus* PCC 7942 where the copy numbers of the Kai proteins are under experimental control. We replaced the native copies of the *kai* genes with copies containing a theophylline-inducible riboswitch previously shown to modulate translational efficiency (*38, 39*), allowing us to tune Kai protein expression (Fig. 2.1A, 2.1B). In this strain, *kaiB* and *kaiC* are transcribed from a constitutive promoter and *kaiA* from an IPTG-

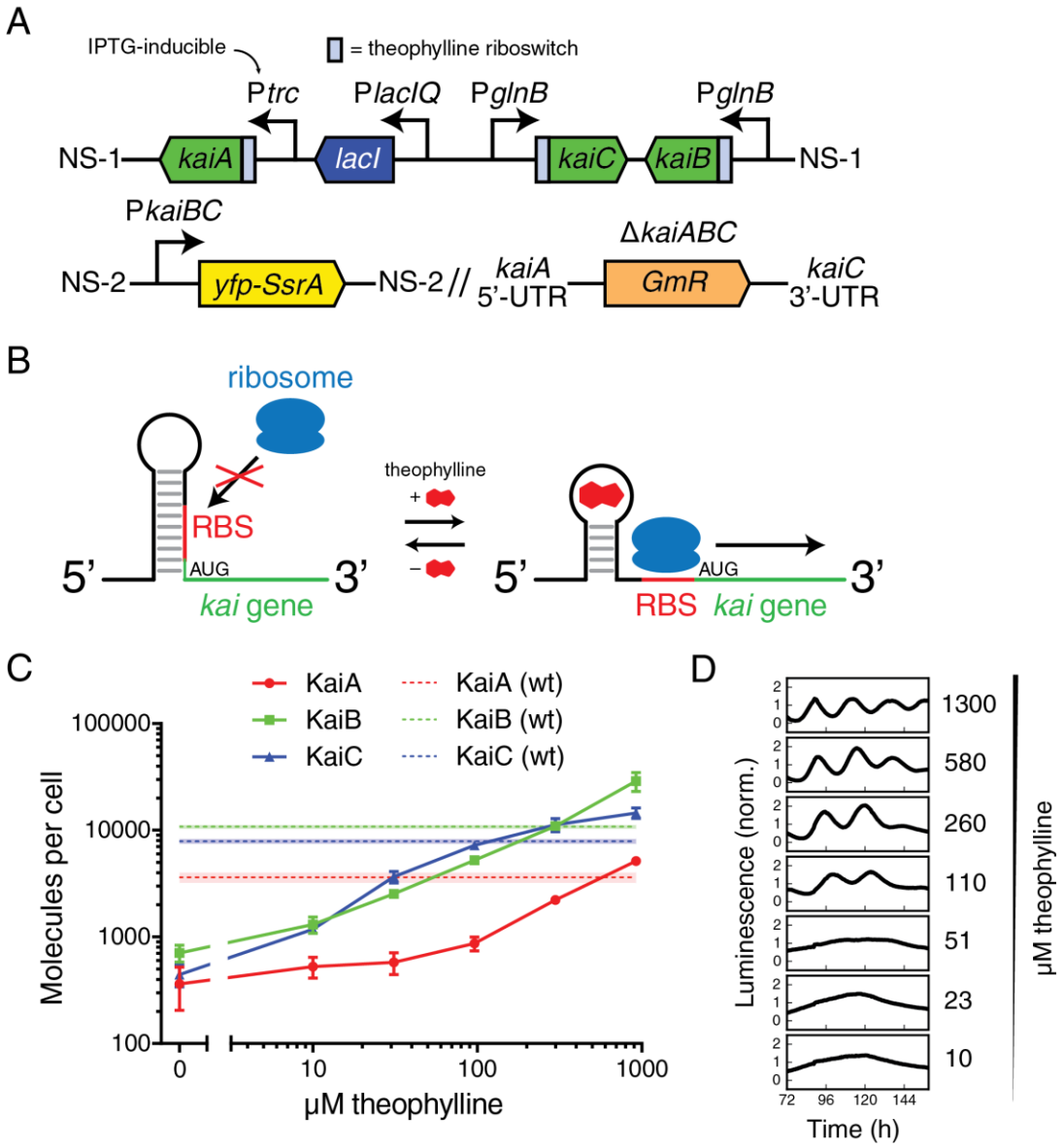

Fig 2.1. Characterization of the Kai copy-number tunable strain. (A) A theophylline riboswitch regulates translation efficiency of all three *kai* genes, and transcriptional regulation of *kaiA* is controlled by an IPTG-inducible promoter. Clock state is reported by EYFP-SsrA expressed from the *kaiBC* promoter. (B) Theophylline regulates translation by freeing the ribosome binding site upstream of each *kai* gene. (C) Kai copy numbers as a function of theophylline concentration with 1 μM IPTG (*solid line*), and Kai copy numbers in wild type cells (*dotted line*). Bars or shaded area indicate standard error of the mean from three replicates. (D) Average oscillations in colonies detected with a bioluminescent reporter in the tunable strain with 1 μM IPTG and various theophylline concentrations.

inducible promoter (Fig. 2.1A). This system removes the natural transcriptional feedback in the system and allows us to focus on the core post-translational oscillator.

Using quantitative western blotting, we found that wild-type cells express ~4,000 KaiA, ~11,000 KaiB, and ~8,000 KaiC copies per cell—a stoichiometry similar to that needed to support oscillations in vitro (*37*). We then determined that our engineered strain is capable of expressing Kai proteins in a range spanning from 100s up to 10,000s of copies per cell (Fig. 2.1C and fig. S2.1). To characterize the ability of this inducible system to produce circadian rhythms, we used a luciferase assay to report on population-level gene expression rhythms. We found that while high levels of theophylline induction produced wildtype-like rhythms, oscillations at the population level weakened or vanished at lower levels of induction even though Kai proteins were still expressed (Fig. 2.1D and fig. S2.2).

We reasoned that loss of population-level oscillations at lower Kai protein expression levels could be explained by two possibilities—rhythms could either be lost in individual cells, or they could persist in single cells but with significant desynchronization between cells. To distinguish between these scenarios, we used time-lapse fluorescence microscopy to observe single-cell rhythms in constant conditions (fig. S2.3). Consistent with previous reports (*25, 40*), we observed that circadian rhythms in single wild-type cells are remarkably precise with < 5% timing error per clock cycle (standard deviation / mean of peak-to-peak times). When we analyzed our tunable expression strain, we found that single cells in fact maintained high-amplitude rhythms even at low levels of theophylline (fig. S2.4), but these rhythms desynchronized over time between cells in a theophylline-dependent manner (Fig. 2.2A, B). 390 μM theophylline (~12,000 copies KaiC / cell) produced coherent single-cell rhythms comparable to wild-type that maintained synchrony over one week, while 92 μM theophylline (~7,000 copies KaiC / cell) led to rhythms that were markedly noisier, and 23 μM theophylline (~2,600 copies

KaiC / cell) produced very noisy rhythms where cells in the same microcolony appeared to adopt nearly random phases after a few days (Fig. 2.2).

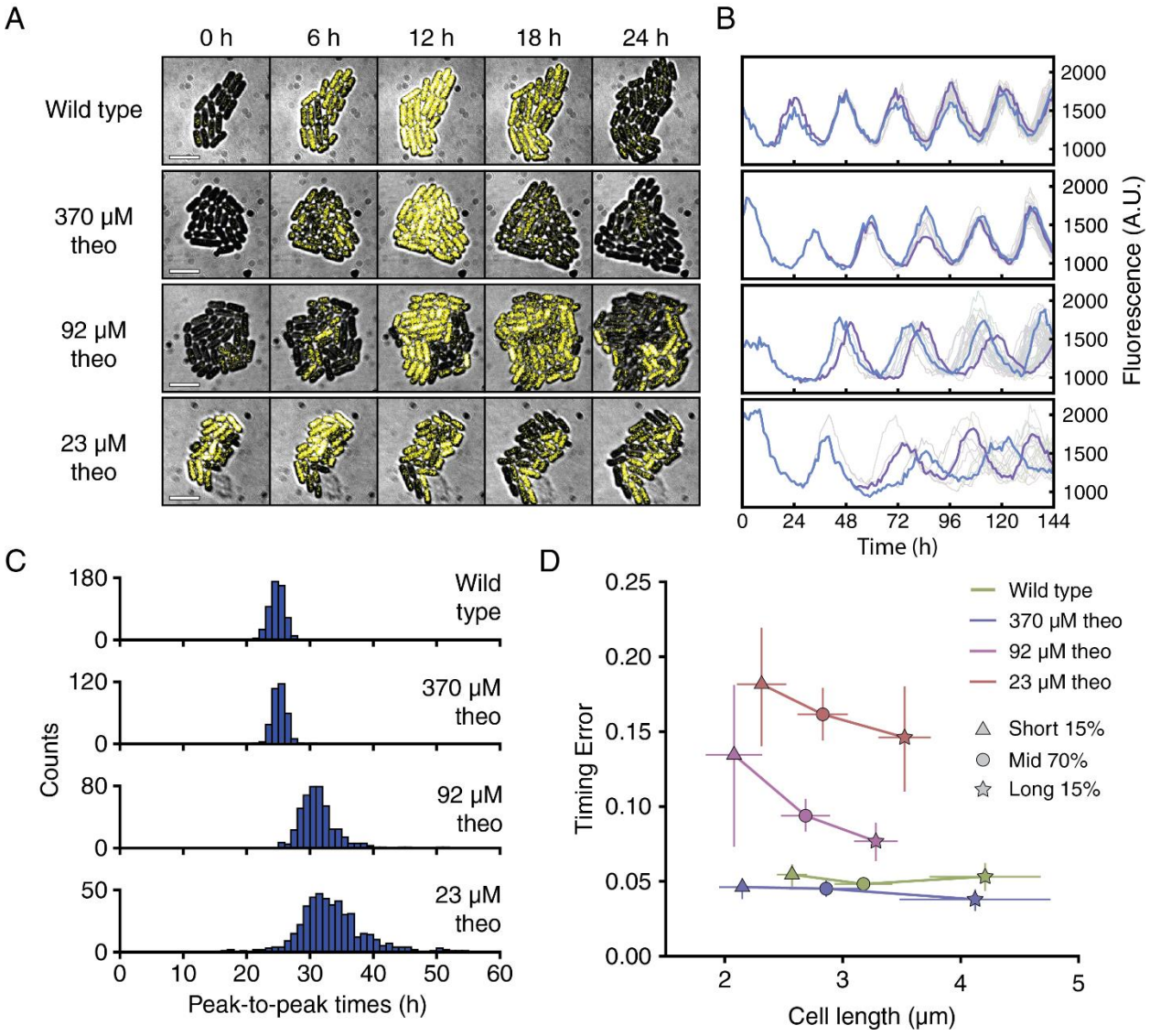

Fig 2.2. Single cell microscopy reveals desynchronized oscillations at low Kai copy number. (A) Filmstrips of YFP oscillations in wild type cells and the copy number tunable strain induced with 1 μM IPTG and various theophylline concentrations (brightfield and YFP fluorescence overlaid). Scale bar: 5 μm. (B) Single cell oscillator trajectories *(gray*) with two example cell lineages highlighted (*blue and purple*). (C) Distributions of peak-to-peak times in wild type and copy number tunable cells; n = 536 (wild type), 336 (370 μM), 455 (92 μM), 616 (23 μM). (D) Cell length vs. timing error (standard deviation/mean of peak-to-peak intervals) in the 15% shortest cells (triangles), middle 70% cells (circles), and 15% longest cells (stars) for each condition. Vertical bars indicate 95% confidence intervals from bootstrapping (5000 iterations), and horizontal bars indicate standard deviation in cell length.

When protein expression level is reduced in these experiments, both protein copy number and concentration are reduced. Because the post-translational oscillator is highly robust to protein concentration, we expect that copy number changes are the main driver of stochasticity (*37, 41*). To experimentally disentangle these effects, we used natural variability in cell size to stratify our analysis and focus on cells with unusually small or large volumes. Since protein concentration is relatively constant across cell sizes (*42, 43*), we used cell volume as a proxy for copy number within each induction condition, estimating that longer cells have higher protein copy number than the average. We quantified relative peak-to-peak timing errors in these cells at different induction levels and found that shorter cells had significantly noisier rhythms compared to longer cells (Fig. 2.2D). From these results, we conclude that high copy numbers of the Kai proteins are required to effectively suppress stochasticity in the circadian rhythm.

How does the presence of many copies of the Kai proteins suppress timing errors, and what features of the oscillator circuit are most vulnerable to noise at low copy number? To address these questions, we constructed a simplified mathematical model of the post-translational Kai oscillator based on (*44*) and (*45*) (Fig. 2.3A and fig. S2.5). This model incorporates experimentally observed Kai protein interactions that lead to oscillatory dynamics: KaiA promotes phosphorylation of individual KaiC hexamers, and without KaiA, KaiC dephosphorylates (*37*). When KaiC reaches a critical phosphorylation state, it switches into a KaiA-resistant, dephosphorylating mode (*46*). Because phosphorylation is ordered (*37*), the sequence of states KaiC visits during the phosphorylation phase (yellow box in Fig. 2.3A) is distinct from the dephosphorylation phase (blue box in Fig. 2.3A). Finally, the dephosphorylating form of KaiC binds KaiB which then captures and inhibits KaiA, forming a delayed negative feedback loop.

Consistent with previous modeling work (*37*, *46*, *47*), these mechanisms can produce

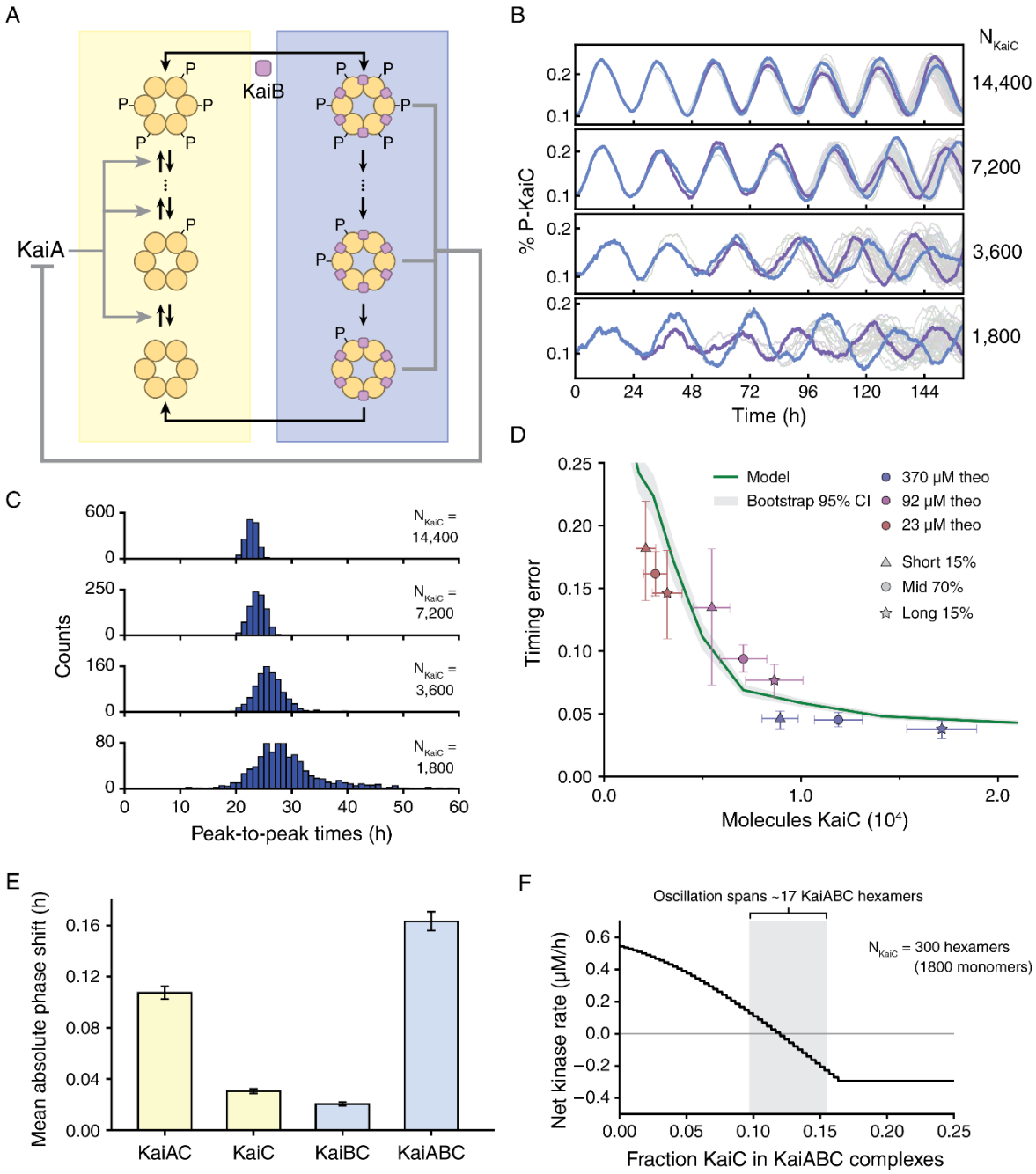

Fig 2.3. KaiA-dependent negative feedback loop is the noise bottleneck in a stochastic model of the Kai system. (A) Model of post-translational oscillator. KaiC hexamers undergo ordered phosphorylation (*yellow box*) and dephosphorylation (*blue box*). KaiA is required for KaiC phosphorylation, and dephosphorylating KaiC binds to KaiB to sequester and inhibit KaiA. (see fig. S5). (B) Simulated stochastic single cell trajectories (*gray*) at various Kai copy numbers with two example traces highlighted (*blue and purple*). (C) Distributions of peak-to-peak time intervals in the stochastic model. (D) Comparison of model and experimental data. Vertical bars indicate 95% confidence interval from bootstrapping. Horizontal bars indicate standard error of the mean (n = 3). Gray interval indicates the 95% bootstrapping confidence…

(Fig 2.3, continued) interval for the model. (E) Mean phase shift caused by Poisson noise perturbations to molecular species in the model (n = 500 trials). Bars indicate 95% confidence interval from bootstrapping. (F) Instantaneous KaiC phosphorylation rate vs. fraction of KaiC in KaiABC complexes in the stochastic model for N_{KaiC} = 300 hexamers. Shaded area indicates the range over which KaiABC complexes oscillate.

free-running oscillations in the deterministic limit, corresponding to infinite numbers of protein molecules (fig. S2.6). To simulate the circadian clock at copy numbers relevant to single bacterial cells, we implemented stochastic simulations of this reaction network. Similar to our experimental results, as Kai protein copy number is decreased, oscillations become noisier and the timing between cycles becomes variable (Fig. 2.3B, 2.3C). Though many models of the Kai oscillator can produce equivalent circadian rhythms when the role of molecular noise is ignored, we find that the impact of noise on the oscillator depends on the number of steps in the phosphorylation cycle required to switch between phosphorylation and dephosphorylation, with five steps giving the best fit (fig. S2.7, Fig. 2.3D).

At very high copy numbers of Kai proteins, stochastic fluctuations will be suppressed because the reaction averages over many molecules. Surprisingly, our results indicate that even with 1000s of Kai protein copies, timing error in the model may still be > 10% per cycle. Since the negative feedback loop synchronizes individual KaiC hexamers through sequestration of a shared pool of KaiA (*44*, *45*), we hypothesized that oscillator timing would be most sensitive to molecular noise in complexes that mediate negative feedback.

To test this, we systematically introduced pulses of molecular noise into the reaction network to find molecular species where noise caused the largest changes in oscillator phase. We found that the molecular complexes most susceptible to noise contain KaiA, precisely the molecules involved in the delayed negative feedback loop (Fig. 2.3E). The vulnerability of the oscillator to fluctuations in KaiA-containing complexes can be understood in terms of the sensitivity of KaiC phosphorylation rates to the amount of active KaiA. At our low inducer

conditions, the number of KaiA-sequestering complexes needed to shift the entire reaction from phosphorylation to dephosphorylation is only ~10 copies (Fig. 2.3F). These KaiABC complexes represent only a small fraction of total KaiC (Fig. 2.3F and (*41*))—thus the stochastic fluctuations from small numbers of KaiABC complexes can be sufficient to cause significant fluctuations in KaiC enzymatic rates. Together, these results suggest that the negative feedback loop is a dominant source of noise in the post-translational oscillator.

Although the KaiA-dependent negative feedback loop is the step most vulnerable to molecular noise, it also performs the crucial function of synchronizing individual KaiC hexamers within a single cell (*44*). Left uncoupled, individual KaiC hexamers would progress through phosphorylation cycles with irregular timing, and the circadian rhythm would rapidly die out. In this way, our results suggest that the negative feedback loop is both a strength and a liability: while it is needed to sustain free-running oscillations, Kai proteins must be expressed at ~10,000 copies per cell to suppress the noise amplification inherent in the negative feedback loop and keep time accurately over several days.

This finding has provocative implications. Circadian rhythms are a well-known strategy that allows organisms to robustly anticipate future events using internally generated oscillations. However, our analysis suggests that there is a minimum biosynthetic investment needed to create a reliable oscillator. Microbial cells span a wide range of sizes, and for very small cells, expressing many thousands of copies of clock proteins may not be tenable. This suggests that tiny cells may use alternative dynamical strategies to keep time.

To investigate this possibility, we focused on the small cyanobacterium *Prochlorococcus marinus*, whose cell volume is over twenty times smaller than *S. elongatus* (*48*). We found that *P. marinus* has ~600 copies of KaiC per cell (Fig. 2.4A), which is in the regime where the *S.*

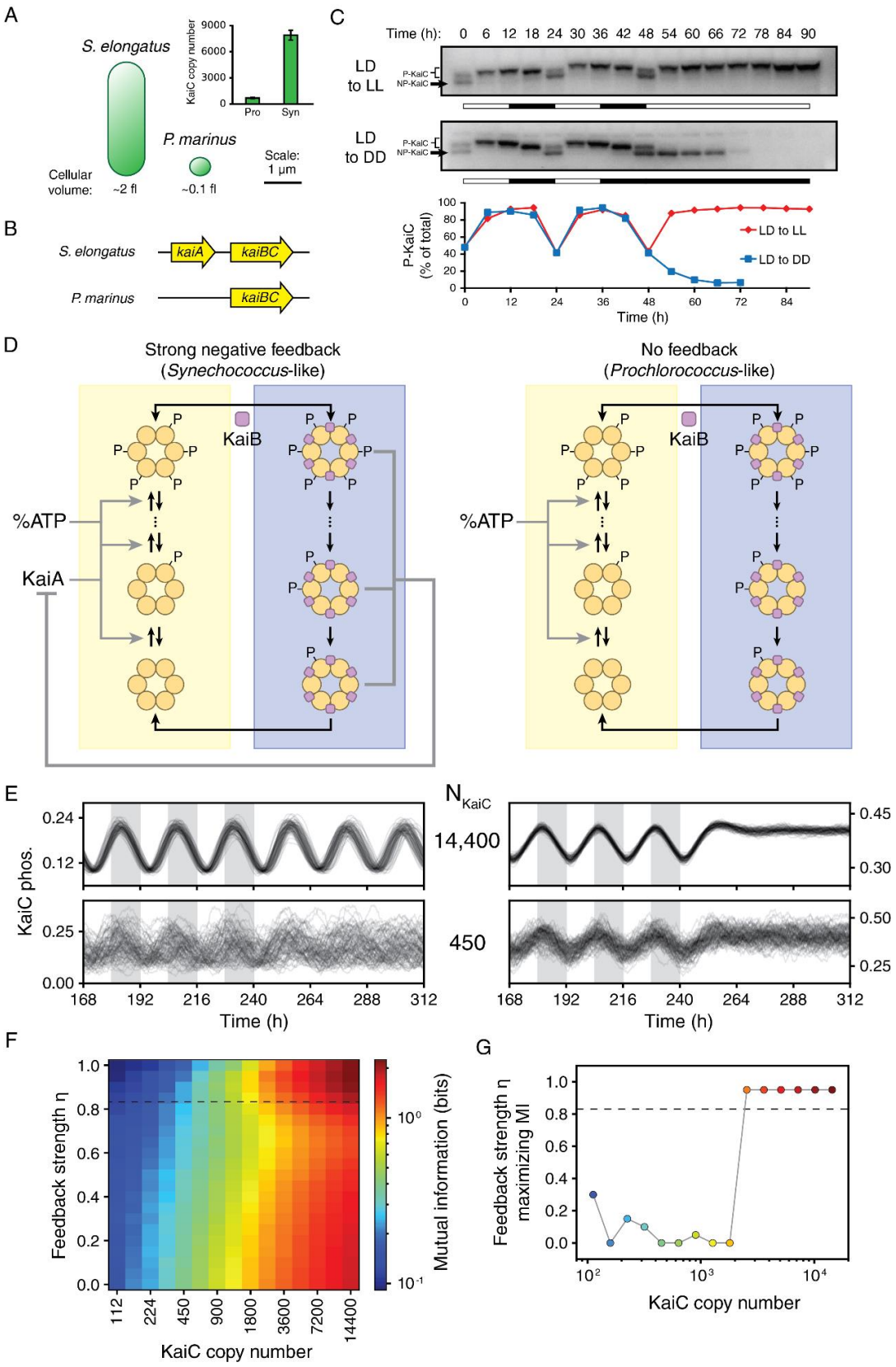

Fig 2.4. Removing the negative feedback loop creates a noise-resistant environmental timer in a *Prochlorococcus*-like system.

(Fig 2.4, continued) (A) Comparison of cell volume and KaiC copy number in *Prochlorococcus marinus* vs *Synechococcus elongatus*. Copy number (inset) determined by quantitative western blot (n = 3). (B) *Prochlorococcus* has a simplified Kai architecture that lacks *kaiA*. (C) *Top*: western blot time course showing *Prochlorococcus* KaiC (ProKaiC) phosphorylation in cultures incubated in light-dark cycles followed by constant light or constant dark. *Bottom*: quantification of ProKaiC phosphorylation. (D) Comparisons of model architectures corresponding to *Synechococcus* (*left*, strong feedback) and *Prochlorococcus* (*right*, no feedback). (E) Simulations of the strong feedback (*left*) and no feedback (*right*) Kai systems in light-dark cycles (shaded regions), followed by constant light at high copy number (*top*, 14,400 KaiC copies) and low copy number (*bottom*, 450 KaiC copies). (F) Mutual information between the clock and time of day during light-dark cycles in the presence of environmental fluctuations (see SI). Stable oscillations occur for feedback strength above 0.83 (*dashed line*). (G) Feedback strength that maximizes mutual information as a function of KaiC copy number. Above the dashed line, the system shows self-sustaining circadian rhythms. Marker colors correspond to the colorbar in (F).

elongatus oscillator becomes extremely error-prone (cf. Fig. 2.3D). Expressing 10,000s of Kai proteins to achieve noise suppression, as in *S. elongatus*, may not be feasible in *P. marinus* given that this investment in protein synthesis would represent ~20% of the proteome (see SI).

The *kaiA* gene at the heart of the negative feedback loop is missing in *P. marinus*, suggesting a qualitatively different time-keeping mechanism ((*49*) and Fig. 2.4B). We measured KaiC phosphorylation in both light-dark cycles and constant conditions, and found that the Kai system in *P. marinus* functions as an environmentally driven timer—KaiC phosphorylation increases in the light and decreases in the dark, but, unlike a circadian rhythm, ceases to cycle when the environment is held constant (Fig. 2.4C, compare to fig. S2.9). The lack of self-sustained oscillations in KaiC phosphorylation is likely the molecular explanation for the lack of free-running rhythms in gene expression in this microbe (*49*).

Does the alternative strategy of a driven timer without a feedback loop offer resistance to molecular noise? To computationally test this hypothesis, we extended our stochastic model of the Kai system, allowing us to vary the strength of the KaiA-dependent negative feedback loop to interpolate between a circadian rhythm and an environmentally-driven timer (Fig. 2.4D and

fig. S2.5). We modeled the input signal from the environment as the effect of the ATP/ADP ratio on KaiC phosphorylation (*24*).

To quantify the performance of these systems, we calculated mutual information between KaiC phosphorylation and the time of day in an environment with both a regular day-night cycle and random input fluctuations simulating weather. At high copy number, mutual information is maximized by a strong negative feedback loop that produces free-running oscillations. In contrast, at low copy number, the system that maximizes mutual information has a very weak or non-existent feedback loop, corresponding to an environmentally-driven timer (Fig. 2.4E-G and fig. S2.10).

This study reveals that the delayed negative feedback loop that sustains circadian rhythms can itself be a liability that amplifies molecular noise. Our experimental and computational analyses suggest that an alternative time-keeping strategy can be employed when protein copy numbers are low: a timer without a feedback loop can outperform a free-running circadian clock when molecular noise is substantial. In this view, the non-free-running Kai system in *Prochlorococcus* is not a degenerate circadian system, but rather an optimal adaptation to low protein copy number.

This result may be of broad significance to microbial physiology. The classical study of circadian rhythms focuses on oscillators that free-run in constant conditions, but our analysis suggests that for cells whose internal biochemistry is unreliable, non-free-running systems may perform better as time-keepers. This may be of particular relevance in niches with some environmental rhythmicity, such as the mammalian gut. Population oscillations have been observed in the gut microbiome (*50, 51*), but there is currently little evidence for free-running rhythms in gut microbes themselves. By broadening our perspective away from the precise, free-

running rhythms of *S. elongatus*, we may uncover a broader world of environmentally-driven timing systems in prokaryotes.

Materials and Methods

Cloning and strain construction

The copy number tunable strain was constructed by transforming a *kaiABC* knockout plasmid (pJC003, gentamycin resistance) into the *kaiABC* locus of wild-type *S. elongatus* carrying either an EYFP-SsrA fluorescence reporter driven by the *kaiBC* promoter (strain MRC1006, reporter first used in *(52)*) or the *luxABCDE* cassette driven by the *psbAI* promoter (strain MRC1005, reporter strain first used in *(53)*), followed by transformation of a plasmid carrying the three *kai* genes and *lacI* (pJC073-2, spectinomycin resistance) into neutral site I. The two versions of the copy number tunable strain carrying the YFP reporter or luciferase reporter are denoted as MRC1139 and MRC1138, respectively.

pJC003 was constructed from a pBSK+ backbone with restriction digest and ligation by flanking a gentamycin resistance cassette with sequences upstream of *kaiA* and downstream of *kaiC*. Specifically, the upstream sequence spans the 331 bp upstream of *kaiA* up to the start codon of *kaiA* flanked with 5' HindIII and 3' SphI sites. The downstream sequence spans from the end of the *kaiC* stop codon to 300 bp downstream, flanked with 5' BssHII and 3' BamHI sites.

pJC073-2 was constructed from a pAM2314 backbone through multiple rounds of Gibson assembly. The *lacI* cassette (*lacIQ* promoter, *lacI* coding sequence, and terminator) was cloned from pAM2991, and the IPTG-inducible *trc* promoter driving *kaiA* transcription was also cloned from pAM2991. Transcription of *kaiB* and *kaiC* are driven by two separate copies of the

S. elongatus *glnB* genomic promoter. The *glnB* promoter sequence was defined as the 750 bp upstream of the putative transcriptional start site of *glnB.* A synthetic theophylline-inducible riboswitch (riboswitch F from (*39*)) was placed immediately upstream of the start codons of all three *kai* genes. The riboswitch was placed at the transcriptional start sites of their respective promoters to reduce the possibility of interference of 5'-UTR sequences with riboswitch function. Additionally, a synthetic terminator, Bba_B0015 from the iGEM parts registry, was placed downstream of each *kai* gene (http://parts.igem.org/Part:BBa_B0015). 30 bp randomized linker sequences were placed upstream of the *glnB* promoter sequences for *kaiB* and *kaiC* to allow for proper Gibson assembly and plasmid sequencing.

Culture conditions

Synechococcus cultures were grown and maintained at 30°C in BG11 medium supplemented with 20 mM HEPES (pH 8.0) with shaking at 180 rpm under constant illumination of 75 μmol photons m^{-2} s^{-1}, and *Prochlorococcus* cultures were grown and maintained at 22°C in Pro99 medium (*54*) based on natural seawater (Woods Hole, MA) supplemented with 0.59 M $NaHCO_3$ under constant illumination of 16 μmol photons m^{-2} s^{-1} without shaking. Culture conditions for specific experiments are described in their respective sections.

To guard against potential genetic instability in the copy number tunable strain, all experiments were performed on cultures propagated for two weeks or less from the original freezer stock. We verified that no genomic loss of our engineered *kaiB* or *kaiC* expression system was detectable by genomic PCR or western blot in these cultures (data not shown).

Time lapse microscopy

To prepare cells for time lapse microscopy, cultures of either wild-type cells expressing the YFP reporter (MRC1006) or the copy number tunable cells expressing the YPF reporter (MRC1139) were grown in black, opaque 96-well plates and illuminated with custom-build LED arrays powered by an Arduino, which delivered 1.33 V across each LED (627 nm wavelength), illuminating cells with ~8.8 μmol photons m^{-2} s^{-1}. The cells were seeded at an initial OD750 of 0.1 and were entrained with two 12h:12h light/dark cycles. 48 hours after initial seeding, wells containing duplicate culture conditions were combined into single tubes.

After pooling cultures into single tubes, 1 μl of culture was pipetted into individual wells of a glass coverslip-bottomed 96-well plate (Mat-tek corporation). For each well, a BG11-agar pad (1 mm x 2 mm x 2 mm) was placed on top of each droplet of culture. 225 μl of molten BG11-agar cooled to 37°C and containing appropriate concentrations of IPTG and theophylline was pipetted into each well and left to cool and solidify.

Time lapse microscopy was performed with an Olympus IX-71 inverted microscope with motorized stage and focus control, and automation of image acquisition was implemented with the Micromanager software package. Images were captured with a 100x Olympus oil immersion objective and a Luca EMCCD camera (Andor). The microscope was housed in a custom-built incubator that maintained temperature at 30°C and insulated the apparatus from external light sources. The cells were exposed to a continuous light source of 2 μmol photons m^{-2} s^{-1} of light (660 nm wavelength), and the illumination condenser was removed in order to widen the light beam to sufficiently illuminate multiple wells evenly. Over the course of one hour, the microscope imaged 24 unique fields of view with brightfield, chlorophyll, and YFP filter sets (exc. 500 nm/20 nm bandpass, emm. 535 nm/30 nm bandpass, dichroic 515 nm long bandpass).

Single cell image processing and data analysis

Cell masks for image processing were obtained using custom-written Python software to allow the user to manually draw a "mask estimate" over individual cells in brightfield images. Using the estimated masks as initial guesses, the software optimized mask areas to fit the underlying cells such that the total pixel intensity was minimized, utilizing the relatively dark cell interior to do so. The fitted cell masks from one movie frame were then used as initial guesses for cells in the next frame, and any errors were corrected manually. After the cell masks were labeled for the duration of the experiment, the CellTracker software suite (*55*) was used to construct cell lineages based on these cell masks to measure cellular YFP fluorescence intensity over time for individual lineages.

Peak to peak intervals were detected using a Python implementation of the Matlab peak detection algorithm, and algorithm parameters were tuned to find all local maxima without any restrictions on minimum distance between maxima. All cell lineages within a single microcolony (i.e. all of the descendants of a single mother cell at the start of the experiment) share varying degrees of overlap due to common ancestry, so to avoid counting the same data multiple times, we only considered unique peak to peak intervals, defined as intervals that occurred within unique pairs of mother/daughter cells or at unique times in the experiment if the peaks occurred within the same cell. To prevent peak identification from identifying spurious peaks originating from high frequency measurement noise, the lineage data was smoothed with the Savitsky-Golay algorithm using a window size of 11 timepoints and third order polynomial fitting before peak finding. Statistics were then calculated for the peak to peak distribution mean, standard deviation, and coefficient of variation (defined as standard deviation divided by the mean). Error bars were estimated as the 95% confidence intervals from 5000 iterations of bootstrapping

analysis of the experimental data. Oscillation amplitude was measured by quantifying the difference between peaks and the troughs that immediately preceded them (troughs were identified using the same peak detection algorithm described above).

Western blotting analysis

To prepare cells for western blotting to quantify cellular Kai copy number, cultures were grown in black 96-well plates illuminated with the Arduino-controlled LED array under constant illumination, and cells were not subjected to any prior entrainment protocols before growth in 96-well plates. Cultures of the copy number tunable strain were supplemented with 1 μM IPTG and varying concentrations of theophylline. The cells were seeded at an initial OD750 of 0.3 with each sample distributed across 12 wells (200 μl culture/well), and they were allowed to grow for 48 hours. At this point, 1.75 ml of culture was taken per sample, pelleted at 3000 x g, flash frozen in liquid nitrogen, and stored at -80°C.

To prepare *Prochlorococcus* cultures for western blotting to quantify cellular KaiC copy number, cells were seeded at an initial OD750 of ~0.003, and they were grown to a final OD750 of 0.09, at which point 22.2 ml cells per sample were pelleted at 3000g, flash frozen in liquid nitrogen, and stored at -80°C.

Frozen cell pellets were resuspended in lysis buffer containing 8 M urea, 20 mM HEPES pH 8.0, 1 mM $MgCl_2$, and 0.5 μl benzonase (EMD Millipore). Samples were then lysed with 10 cycles of vortex bead beating using 0.1 mm glass beads. Complete lysis was verified by microscopy (*Synechococcus*) or flow cytometry (*Prochlorococcus*). Sample protein concentration was measured by Bradford assay, using BSA as a protein standard. Samples were then mixed with 3x SDS-PAGE sample buffer (150 mM Tris-Cl pH 6.8, 6% SDS, 300 mM DTT,

30% glycerol, 0.1% bromophenol blue) and immediately loaded into polyacrylamide gels for SDS-PAGE.

For quantification of either cellular Kai expression or KaiC phosphorylation dynamics, samples were resolved in 4-20% TGX gels or 7.5% Tris-HCl gels, respectively (Biorad). Gels were transferred onto PVDF membrane (Biorad) and blocked in 2% milk + TBST (137 mM NaCl, 2.7 mM KCl, 20 mM Tris, 0.05% Tween-20, pH 7.4). Membranes were then incubated in primary antibody, washed in TBST, and incubated in secondary antibody. Antibody information is listed below:

Table 2.1. Western blot antibody information

Antibody target	Antibody host	Dilution
KaiA	Rabbit	1:2500
KaiB	Rabbit	1:500
KaiC	Rabbit	1:5000
anti Rb	Goat	1:10000

Membranes were visualized with SuperSignal West Femto substrate (Thermo Fisher) and imaged (Biorad ChemiDoc MP). Bands were quantified by densitometry in ImageJ against purified recombinant protein standards, and the intensities of non-specific bands (determined from *kaiABC* null samples) were subtracted. *Synechococcus* recombinant protein was prepared as previously described (*47*), and *Prochlorococcus* KaiC was prepared as described below. Quantification of recombinant standards was performed by running dilution series of standards in SDS-PAGE gels against a BSA standard dilution series followed by staining and imaging with SimplyBlue SafeStain (Life Technologies).

Preparation of recombinant Prochlorococcus KaiC

Prochlorococcus kaiC was cloned from the MED4 genome with an N-terminal 6X His-tag followed by a HRV-3C protease site and inserted in between the KpnI and EcoRI restriction sites of the pMAL-c5e vector. The addition of the Maltose binding protein (MBP) tag from this vector was used to improve solubility of the KaiC protein in *E. coli* during expression. This plasmid was transformed in BL-21 *E. coli* cells and expressed at 18°C for 48 hours without induction. Cells were lysed by high pressure homogenization using an Emulsiflex homogenizer, and the lysate was clarified by centrifugation at 30,000g for 1 hour. This clarified lysate was applied through a Ni-NTA column and *Procholorococcus* KaiC (ProKaiC) was eluted with an imidazole gradient. HRV-3C protease (ThermoFisher Scientific) was added to eluted fractions containing ProKaiC to cleave off the MPB-6X His tags by incubation overnight at 4°C. The post-cleavage fractions were concentrated and further purified via size exclusion chromatography using a Hiprep 16/60 S300 Column and an elution buffer containing 150 mM NaCl, 20 mM Tris pH 8.0 and 1 mM ATP. It was determined by SDS-PAGE that the MPB-6X His tags were incompletely cleaved after the size exclusion step. Hence, fractions containing Pro-KaiC were again incubated overnight with HRV-3C at 4°C to achieve complete tag cleavage. The added HRV-3C protease, which contained a 6X-His-tag, and uncleaved ProKaiC was then removed via incubation with Ni-NTA resin. KaiC concentration in the resulting supernatant was estimated by gel densitometry. This solution was used as a recombinant protein standard for quantitative Western blotting.

Protein copy number measurement

Kai protein copy number per cell was quantified by determining the amount of Kai protein in cell pellet samples by quantitative western blotting (described above) and dividing by the number of cells in the pellet.

Synechococcus cell counts were determined by pipetting 1 µl of diluted sample into a 96-well plate and adding an agar pad and BG11-agar as described in the time lapse microscopy methods section. The microscope was then programmed to tile the entire well to image all the cells by chlorophyll autofluorescence, and the cells were then manually counted in ImageJ. *Prochlorococcus* cells were too small to count reliably on the microscope, and thus absolute cell counts were obtained with an Attune Acoustic Focusing Cytometer, which has been previously used to quantify cell counts for *Prochlorococcus* (*56, 57*).

Reported values for Kai protein copy numbers are the averages of three independent biological replicates for each condition. Uncertainties were calculated as the standard error of the mean for each estimate.

Circadian bioluminescence measurements

Bioluminescence measurements of either the wild-type strain carrying a luciferase reporter (MRC1005) or the copy number tunable strain carrying a luciferase reporter (MRC1138) were obtained using a PerkinElmer TopCount Microplate Scintillation and Luminescence Counter. Black, opaque 96-well plates were prepared by pipetting 250 µl of BG11-agar into each well. After the agar solidified, 10 µl of 25x inducer at varying concentrations (IPTG + theophylline) was pipetted onto the top of the agar, and the plate was left overnight to allow the inducer to diffuse uniformly throughout the agar. Cells were pipetted onto the plate and illuminated with the Arduino-controlled LED arrays as described above, and they were subjected

to two 12h:12h light/dark cycles followed by release into constant light, at which point the plate reader began taking bioluminescence measurements every 30 minutes. Data curves shown for each condition and strain represent the average of data recorded from four replicate wells.
To quantify the period and amplitude of circadian oscillations, the data were first detrended by dividing by a best fit line for the duration of the experiment after release into constant light (111 hours total). This detrended data was then fit to a sinusoid.

Supplementary Text

Identification of appropriate genomic promoters to drive kaiB and kaiC transcription

A major design criterion for the copy number tunable strain is to be able to express the Kai proteins in an expression range that spans from far below wild-type levels up through wild-type levels or above. The synthetic riboswitch has been shown previously at maximal induction to express genes at only ~25% of the expression levels obtained without the riboswitch (*39*), necessitating the use of a promoter that is stronger than the native *kaiBC* promoter. Additionally, for this study, we wished to investigate the effects of molecular noise on clock function in the absence of the transcriptional-translational feedback loop, which has been shown to also contribute to clock period robustness (*40*).

We identified candidate promoters that met these criteria by combining global transcriptomic datasets from (*58*) and (*59*), and we examined genes whose mRNA was expressed from 4-40 times the expression level of *kaiBC* mRNA and that were also arrhythmic. To identify promoter sequences for these genes, we examined regions 500-750 bp upstream of their 5'-UTR, taking care to consider the 5'-UTR of the upstream-most gene if the transcript encoded a multi-

gene operon. Taking the sequence only upstream of the 5'-UTR ensured that any untranslated regions would have minimal effect on the folded structure of the riboswitch.

As a result of this analysis, we identified six candidate promoters from the following genes: Synpcc7942_0065, Synpcc7942_0089, Synpcc7942_1152, Synpcc7942_0321, Synpcc7942_0416, and Synpcc7942_1048. To test which of these promoters expressed Kai protein at the appropriate levels, we created strains of *Synechococcus* where *kaiA* expression is under its native promoter but *kaiB* and *kaiC* expression is controlled by both the candidate promoter and the synthetic riboswitch (in the bioluminescent circadian reporter background). We successfully obtained clones carrying promoters Synpcc7942_0321, Synpcc7942_0416, and Synpcc7942_0089, and we compared the expression levels of KaiB and KaiC in these strains to wild-type levels at either 0 μM theophylline or 2 mM theophylline. From this blot, we identified Synpcc7942_0321 (*PglnB*) as a promoter of the appropriate strength.

Identifying functionally equivalent light levels across experimental setups

Due to the impact of light levels on overall protein expression in cyanobacteria (and therefore Kai copy number), we designed the experiments to ensure that cells were exposed to functionally equivalent light levels between the setups used to incubate cells for TopCount, western blot, and microscope experiments. Absolutely equal overhead illumination between conditions does not necessarily result in functionally equivalent light levels due to physical differences in experimental apparatus, such as the absorbance of light by the walls in the black 96-well plates, or the placement of BG11 agar above cells on the microscope. Additional differences may result from whether the cells are incubated in liquid media (western blots) or on top of/underneath solid media (TopCount/microscope experiments, respectively).

To identify functionally equivalent light levels, we incubated cells in the appropriate experimental setups for TopCount, western blot, and microscope experiments as described in the methods above. For the TopCount and western blot setups (powered by the Arduino-controlled LEDs), we exposed cells expressing YFP under control of the IPTG-inducible *trc* promoter (MRC1036) to varying levels of light while simultaneously incubating cells on the microscope at a fixed reference light level. For all experimental setups, cells were maximally induced with 1 mM IPTG. After 48 hours of incubation, cells were washed off the agar surface (for TopCount experiments) or taken directly from liquid culture (for western blot experiments) and placed on the microscope and imaged alongside cells grown on the microscope. The average single cell intensity under each condition was quantified with the CellProfiler software suite (*60*), and we determined light intensities for the TopCount and western blot setups that gave rise to cell fluorescence that matched cells grown on the microscope. The functionally equivalent light intensities were ~2 μmol photons m^{-2} s^{-1} for cells on the microscope, ~8.8 μmol photons m^{-2} s^{-1} for cells in liquid culture in 96-well black plates (e.g. Western blot cultures), and ~20 μmol photons m^{-2} s^{-1} for cells on solid media in 96-well black plates (e.g. TopCount experiments). These light intensities were used for all subsequent experiments.

Details of model implementation

The mathematical model we use here (fig. S5) is a simplified model which treats some elements of the biochemistry of the system abstractly, but captures known aspects of the Kai system that can produce oscillations. It has similarities to previously published models, where the functional units are KaiC hexamers (*44*, *46*). First, the sequence of phosphorylation and dephosphorylation of KaiC occurs in an ordered fashion, reflected in an initial phase where

phosphorylation occurs (yellow box, analogous to Thr432 phosphorylation dominant) until a threshold level of peak phosphorylation, after which a dephosphorylation phase initiates (blue box, analogous to Ser431 phosphorylation dominant). The ordering of the phosphorylation sites prevents the system from crossing between the intermediate states in the yellow and blue boxes. Second, KaiC only phosphorylates when bound to KaiA and otherwise dephosphorylates. This results in effective phosphorylation rates that depend on the concentration of KaiA, as observed experimentally (*37*). Third, KaiC enters the dephosphorylation phase (blue box), is coincident with adopting a state that is bound to KaiB (likely ADP bound in the N-terminal domain of KaiC) (*61*). These KaiBC complexes can bind and sequester KaiA. We assume that binding of KaiB is highly cooperative, so that 6 KaiB monomers bind to a single KaiC hexamer, and that each of these KaiB monomers is capable of inhibiting one KaiA dimer, and does so with very high affinity so that KaiA is not released until that KaiC hexamer exits the dephosphorylation phase (*62*). Additionally, we assume KaiB is present in excess and that KaiBC complex formation does not significantly deplete cellular KaiB, and thus we do not explicitly account for KaiB amounts in the model.

We designed the model to be flexible so that the number of reaction steps in the phosphorylation cycle m can be varied. Each elementary reaction that changes the state of a KaiC hexamer occurs with a rate constant *k* which can then be rescaled so that the oscillatory period remains independent of *m.* For a given value of m, the deterministic model is defined by the ordinary differential equations below.

$[C]$ Concentration of KaiC hexamers during phosphorylation cycle that are unbound to KaiA

$[AC]$	Concentration of KaiC hexamers during phosphorylation cycle that are bound to KaiA
$[BC]$	Concentration of KaiC hexamers during dephosphorylation cycle that are bound to KaiB and unbound to KaiA
$[ABC]$	Concentration of KaiC hexamers during dephosphorylation cycle bound to KaiB and sequestering KaiA
$[A]$	Concentration of KaiA dimers with free, unbound KaiA denoted as $[A_{free}]$ and total amount of initial KaiA denoted as $[A_{tot}]$

A given KaiC hexamer can exist in varying states of phosphorylation, from a minimum of 0 to a maximum of m. Here, i represents phosphorylation state during the phosphorylation cycle and j represents phosphorylation state during the dephosphorylation cycle, and we first consider cases where $0 < i < m - 1$ and $0 < j < m$:

$$\frac{d[C_i]}{dt} = k_{dephos} \cdot ([C_{i+1}] - [C_i]) + k_{A_{off}} \cdot [AC_i] - k_{A_{on}} \cdot [C_i] \cdot [A_{free}] \quad (1)$$

$$\frac{d[AC_i]}{dt} = k_{phos} \cdot ([AC_{i-1}] - [AC_i]) + k_{A_{on}} \cdot [C_i] \cdot [A_{free}] - k_{A_{off}} \cdot [AC_i] \quad (2)$$

$$\frac{d[BC_j]}{dt} = k_{dephos} \cdot ([BC_{j+1}] - [BC_j]) - k_{ABC} \cdot [BC_j] \quad (3)$$

$$\frac{d[ABC_j]}{dt} = k_{dephos} \cdot \left([ABC_{j+1}] - [ABC_j]\right) + k_{ABC} \cdot [BC_j] \cdot [A_{free}] \quad (4)$$

$$\left[A_{free}\right] = \max\left(0, [A_{tot}] - \sum_{i=0}^{m-1} [AC_i] - 6\sum_{j=1}^{m} \left[ABC_j\right]\right) \quad (5)$$

For special cases $i = 0, m - 1$:

$$\frac{d[C_0]}{dt} = k_{dephos} \cdot ([C_1] + [BC_1] + [ABC_1]) + [C_1](k_{A_{off}} - k_{A_{on}}) \quad (6)$$

$$\frac{d[AC_0]}{dt} = -k_{phos} \cdot [AC_0] + k_{A_{on}} \cdot [C_0] \cdot [A_{free}] - k_{A_{off}} \cdot [AC_0] \quad (7)$$

$$\frac{d[C_{m-1}]}{dt} = k_{dephos} \cdot ([BC_m] - [C_{m-1}]) + k_{A_{off}} \cdot [AC_{m-1}] - k_{A_{on}} \cdot [C_{m-1}] \cdot [A_{free}] \quad (8)$$

$$\frac{d[AC_{m-1}]}{dt} = k_{phos} \cdot ([AC_{m-2}] - [AC_{m-1}]) + k_{A_{on}} \cdot [C_{m-1}] \cdot [A_{free}] - k_{A_{off}} \cdot [AC_{m-1}] \quad (9)$$

And for $j = m$:

$$\frac{d[BC_m]}{dt} = k_{phos} \cdot [AC_{m-1}] - 2 \cdot k_{dephos} \cdot [BC_m] - k_{ABC} \cdot [BC_m] \quad (10)$$

$$\frac{d[ABC_m]}{dt} = -k_{dephos} \cdot [ABC_m] + k_{ABC} \cdot [BC_m] \cdot [A_{free}] \quad (11)$$

To interpolate between full strength negative feedback (*Synechococcus*-like) and no negative feedback (*Prochlorococcus*-like), we make the following modifications to the model. First, we introduce an additional state of KaiC in the phosphorylating cycle, C^*, which can phosphorylate in a KaiA-independent manner, consistent with experimental evidence indicating that Prochlorococcus KaiC can autophosphorylate in the absence of KaiA (*63*). KaiC can thus transition to one of two states that promote phosphorylation: one in which it binds to KaiA (AC), and one in which it does not (C^*). The relative probability of these states is determined by a parameter η which can take values between 0 and 1. When $\eta = 1$, all phosphorylation is KaiA-dependent (the Synechococcus-like model). When $\eta = 0$, all phosphorylation occurs via the

KaiA-independent state. Lastly, we introduce environmental input in the form of the cellular ATP/(ATP + ADP) ratio, which has been shown to alter KaiC phosphorylation, leading to entrainment to metabolic signals (*24*). Here, we model this effect by allowing the ATP/ADP ratio to influence the rate at which KaiC enters a phosphorylation-competent state, and we denote the fraction of total (ATP + ADP) that is ATP as f_{ATP}. This hypothetical mechanism allows phosphorylation of *Prochlorococcus* KaiC to depend on the light-dark cycle through metabolism.

Specifically, the rate at which KaiC enters the KaiA-bound phosphorylating state changes:

$$k_{A_{on}} \rightarrow k_{A_{on}} \cdot \eta \cdot f_{ATP}$$

Similarly, the total rate at which KaiC switches to the KaiA-independent phosphorylating state can be expressed as:

$$k_{C \rightarrow C^*} \cdot (1 - \eta) \cdot f_{ATP}$$

In total, the entire modified system of equations can be written, modifying equations (1) and (2) to produce equations (12) and (13). First, we consider cases where $0 < i < m - 1$ (the reactions on the dephosphorylating cycle remain unchanged):

$$\frac{d[C_i]}{dt} = k_{dephos} \cdot ([C_{i+1}] - [C_i]) + k_{A_{off}} \cdot [AC_i] - k_{A_{on}} \cdot \eta \cdot f_{ATP} \cdot [C_i] \cdot [A_{free}] - k_{C \rightarrow C^*} \cdot (1 - \eta) \cdot f_{ATP} \cdot [C_i] \quad (12)$$

$$\frac{d[AC_i]}{dt} = k_{phos} \cdot ([AC_{i-1}] - [AC_i]) + k_{A_{on}} \cdot \eta \cdot f_{ATP} \cdot [C_i] \cdot [A_{free}] - k_{A_{off}} \cdot [AC_i] \quad (13)$$

$$\frac{d[C^*{}_i]}{dt} = k_{phos} \cdot ([C^*{}_{i-1}] - [C^*{}_i]) + k_{C \to C^*} \cdot (1 - \eta) \cdot f_{ATP} \cdot [C_i] - k_{C^* \to C} \cdot [C^*{}_i] \quad (14)$$

Equations (6)-(9) are rewritten to produce equations (15)-(18) for $i = 0, m-1$:

$$\frac{d[C_0]}{dt} = k_{dephos} \cdot ([C_1] + [BC_1] + [ABC_1]) + k_{A_{off}} \cdot [AC_0] - k_{A_{on}} \cdot \eta \cdot f_{ATP} \cdot [C_0] \cdot [A_{free}] + k_{C \to C^*} \cdot (1 - \eta) \cdot f_{ATP} \cdot [C_0] \quad (15)$$

$$\frac{d[AC_0]}{dt} = -k_{phos} \cdot [AC_0] + k_{A_{on}} \cdot \eta \cdot f_{ATP} \cdot [C_0] \cdot [A_{free}] - k_{A_{off}} \cdot [AC_0] \quad (16)$$

$$\frac{d[C_{m-1}]}{dt} = k_{dephos} \cdot ([BC_m] - [C_{m-1}]) + k_{A_{off}} \cdot [AC_{m-1}] - k_{A_{on}} \cdot \eta \cdot f_{ATP} \cdot [C_{m-1}] \cdot [A_{free}] + k_{C \to C^*} \cdot (1 - \eta) \cdot f_{ATP} \cdot [C_{m-1}] \quad (17)$$

$$\frac{d[AC_{m-1}]}{dt} = k_{phos} \cdot ([AC_{m-2}] - [AC_{m-1}]) + k_{A_{on}} \cdot \eta \cdot f_{ATP} \cdot [C_{m-1}] \cdot [A_{free}] - k_{A_{off}} \cdot [AC_{m-1}] \quad (18)$$

Deterministic simulations of the system were implemented by using fourth-order Runge-Kutta numerical integration with a step size of 0.01 hours.

To simulate the system stochastically, we reimplemented the reactions shown above using the Gillespie algorithm (*64*), denoting the copy number of a particular species as $N_{species}$. Thus, the rate at which KaiC binds to KaiA during the phosphorylation cycle can be rewritten:

$$k_{A_{on}} \cdot \eta \cdot f_{ATP} \cdot N_{C_i} \cdot N_{A_{free}} \rightarrow k_{A_{on}} \cdot \eta \cdot f_{ATP} \cdot N_{C_i} \cdot [A_{tot}] \cdot \frac{N_{A_{free}}}{N_{A_{tot}}}$$

And the rate at which KaiBC binds to KaiA to sequester it during the dephosphorylation cycle can be rewritten:

$$k_{ABC} \cdot N_{BC_j} \cdot N_{A_{free}} \rightarrow k_{ABC} \cdot N_{BC_j} \cdot [A_{tot}] \cdot \frac{N_{A_{free}}}{N_{A_{tot}}}$$

In experiments, changing theophylline concentrations changes copy number but also changes concentration of Kai proteins. To reflect this in the stochastic implementation where the cell volume remains constant over varying copy numbers, the concentration $[A_{tot}]$ is scaled by the overall Kai copy number relative to wild-type, where wild-type copy number of KaiC is specified to be 1200 hexamers (derived from experiments in this study measuring Kai copy number to be ~8000).

A table of simulation parameter values for this study are listed in Table S2.2. Rate constants in the model were selected such that their relative magnitudes were reasonably consistent with previous findings (e.g. the KaiA binding rate is faster than the disassociation rate, and important slow steps in the cycle are reactions within a KaiC hexamer: phosphorylation, dephosphoryation, and switching into a KaiB-binding competent state (*65*)). Values for KaiA-KaiC stoichiometries for stimulating phosphorylation (1:1 stoichiometry) and for sequestration (6:1 stoichiometry) were chosen based on reported literature values (*62*, *66*). As a final step, all model rate constants were scaled proportionally such that the period of oscillations was 24 hours.

Determining model timing error

To quantify the amount of oscillatory noise in the model at varying copy numbers (as shown in Fig. 3), the simulation was first initiated with all KaiC completely dephosphorylated and unbound to KaiA (C_0). From this initial condition, the simulation was simulated deterministically, and the system quickly converged to a limit cycle. To allow the system to reach steady state oscillations, the simulation was run for 1000 hours. The relative amounts of each KaiC species was then recorded and saved at the KaiC phosphorylation trough most immediately before the 1000 hour simulation endpoint. KaiC phosphorylation was calculated by summing up the total number of phosphorylated sites on all KaiC molecules in the simulation.

The save point was then used as initial conditions for all subsequent simulations, and concentrations of Kai proteins were converted to absolute copy number assuming a conversion factor of 1200 molecules$/1.56\ \mu M$. To more closely mimic the experimental data in which single cells grow and divide, the simulation was first run for 24 hours (the "mother cell"), at which point the simulation state was cloned into two simulations which were left to run independently thereafter (as the "daughter cells"). 24 hours later, the simulations were cloned and split again into four, and this was repeated for a total of 7 complete "generations", or 168 hours of simulation time for each individual simulation lineage. The 24 hour doubling time chosen here is consistent with the ~24 hour doubling time in our experiments (data not shown). KaiC phosphorylation data over time was compiled for individual lineages, and to facilitate further downstream processing, the data were resampled at a time interval of 0.01 h with linear interpolation. Peak-to-peak measurements were then obtained in identical manner to that used for experimental data (outlined in the Methods and Materials) with the exception that a window size of 211 was used for the Savitsky-Golay smoothing filter on the model data before peak finding.

This simulation was repeated for varying copy numbers of Kai protein and for varying values of m as shown in Fig. S6, with $m = 5$ providing the best fit to the experimental data as shown in Fig. 3D.

Noise sensitivity analysis of all molecular species in the model

We performed a noise sensitivity analysis to determine which molecular species in the model were most susceptible to noise. Specifically, we defined noise susceptibility as the average magnitude of phase shift induced by an instantaneous single perturbation of molecular noise in the amount of a given molecular species. Perturbations were simulated to be Poisson-like in nature in order to reflect the expected $1/\sqrt{N}$ noise scaling behavior, where N represents the copy number of a given species. Thus, we expected that molecular species that were present in smaller amounts should experience greater fluctuations in their amounts.

To precisely isolate the effect of a single perturbation of molecular noise, the model was first initialized from steady state oscillations (specifically after 1994.49 hours of deterministic simulation starting from dephosphorylated KaiC) and run deterministically for $24 + \Delta t$ hours, where Δt represents a random interval of time picked uniformly between 0 and 24 hours. At this point a molecular noise perturbation was introduced by converting the concentration of a given molecular species to a discrete copy number (corresponding to 25% wild-type copy number and using a conversion factor of 1200 molecules$/1.56\ \mu M$), at which point the copy number was resampled from a Poisson distribution whose mean was the initial copy number. This resampled copy number was converted back into a concentration value, and from here the simulation was run deterministically for an additional 48 hours.

The phase shift between the pre- and post-perturbation trajectories was calculated by taking the absolute value of the difference in the phases obtained from sinusoid fitting to both trajectories.

For each molecular species in the model, we repeated this process a total of n = 500 trials. We grouped molecular species into four categories: KaiAC complexes, unbound KaiC, KaiBC complexes, KaiABC complexes. For each category, we then calculated the average magnitude of phase shift, which is shown in Fig. 3E.

Determining the range of KaiABC complexes that switch KaiC between kinase and phosphatase modes

To calculate the dependence of KaiC net kinase rates on the number of KaiABC complexes present at a given moment in time, the average amount of each KaiC species was first calculated by averaging KaiC amounts over a 24 hour period sampled at 1 hour intervals. The 24 hour period over which individual KaiC amounts were recorded ranged from t = 1958.07 h (trough of phosphorylation) to t = 1982.07 in a deterministic simulation in which KaiC was initialized in the unphosphorylated state and unbound to KaiA. Once the average KaiC amounts were calculated, these amounts were used as the initial conditions for subsequent simulation.

Because the amount of free KaiA available to bind to KaiC depends on the amount of KaiABC complex, the simulation was allowed to reach an equilibrium of KaiA-KaiC binding by first allowing only the KaiA-KaiC binding/unbinding reactions to run for 20 hours, disabling any phosphorylation and dephosphorylation reactions. After KaiA-KaiC binding reached equilibrium, the net KaiC kinase rate was calculated by taking the sum of all rate constants for

phosphorylating reactions and subtracting the sum of all rate constants for dephosphorylating reactions. This process was then repeated for varying amounts of KaiABC complexes.

The curve shown in Fig. 3F was produced by sampling the net kinase vs. KaiABC complex curve in discrete amounts corresponding to incrementing the copy number of KaiABC complexes by 1 and calculating the corresponding change in net kinase rate, and this was simulated for a KaiC copy number of 1,800.

Calculating clock/environment mutual information for varying feedback loop strengths

To determine the optimal feedback loop strength at varying Kai copy numbers, we varied the feedback loop strength in the presence of environmental fluctuations. Feedback loop strength was tuned by varying the parameter η to take values between 0 (no feedback loop) to 1 (full feedback strength). For each value of η, the simulation was first run deterministically for 50 cycles of 12h:12h light/dark cycles without environmental noise to reach steady state, initialized from unphosphorylated KaiC not bound to KaiA. Light/dark cycles were emulated by cycling ATP levels between 80% (day) and 40% (night), reflecting experimental values observed previously. These steady states were then used as the starting points for subsequent simulations.

For stochastic simulations, the simulations were run at constant concentration, where Kai copy number was changed by changing the effective simulation volume (thus cells with smaller Kai copy number will have proportionally smaller volumes). This reflects our experimental findings that while *Synechococcus* and *Prochlorococcus* have vastly different Kai copy numbers, the concentrations of Kai proteins are roughly equivalent in both organisms due to their difference in volume. The simulations were then run at varying copy numbers for 10 cycles of 12h:12h light/dark cycles with fluctuations.

Environmental fluctuations were simulated by adding fluctuations to a deterministic 12h:12h light/dark cycle as outlined above. In the case without external fluctuations, f_{ATP} is 80% during the day and 40% at night. When fluctuations are present, we instead generate f_{ATP} via a Markov process where the input signal switches to a new level with a mean waiting time of 1.33 hours. For each environmental transition, f_{ATP} is drawn from a normal distribution centered on 80% (day) or 40% (night) with a standard deviation of 10%. An example plot of such ATP fluctuations is shown in figure S10. To facilitate downstream data processing, the data from each simulation was resampled at a rate of 0.01 hours using linear interpolation. For each copy number and value of η, the simulation was repeated a total of 900 times.

For mutual information calculations, the data from the last 5 cycles of 12h:12h light/dark was used. The multidimensional clock state was projected into 2 dimensions by quantifying the total amount of phosphorylation in the phosphorylating half of the cycle (P, yellow box in fig. S5) vs. the dephosphorylating half of the cycle (D, blue box in fig. S5).

Time data was binned into 24 bins, and phosphorylation data was binned into 100 x 100 bins. Mutual information (I) was calculated for each Kai copy number and value of η:

$$I(T, \langle P, D \rangle) = \sum_{t \in T} \sum_{\langle p,d \rangle \in \langle P,D \rangle} Pr(t, \langle p, d \rangle) \cdot \log_2 \frac{Pr(t, \langle p, d \rangle)}{Pr(t) Pr(\langle p, d \rangle)}$$

Here, T denotes the entire set of time points, and t denotes a specific time point. Similarly, $\langle P, D \rangle$ denotes the entire set of observed phosphorylation states, and $\langle p, d \rangle$ denotes a specific phosphorylation state observation. Pr denotes the probability function.

Determining the bifurcation point where oscillations lose stability as feedback strength decreases

The bifurcation point in Figures 4F-G (denoted by the horizontal dashed line) was determined for various values of feedback loop strength η and ATP/(ATP + ADP) ratios by the following method. For a given, constant ATP ratio and starting at $\eta = 1.0$, a deterministic simulation was run for 1000 hours, and a binary search was executed over varying values of η to determine the level of η where oscillations would disappear or become unstable. Oscillations were considered to be stable and present if the amplitude did not drop by more than 2% over the last 100 hours of the simulation and if the amplitude over the last 50 hours was greater than 10% of the oscillatory amplitude obtained at $\eta = 1.0$.When the input signal is constant, the bifurcation point depends on the ATP ratio value, so in the case of fluctuating environmental input, we defined the bifurcation point of the system to be the value of η above which oscillations were present and stable for all ATP ratio levels between 80% and 40%.

Estimate of the fraction of the total proteome occupied by Kai proteins for Synechococcus*-like expression levels in* Prochlorococcus

Given a cellular diameter of ~0.5 μm (*48*), an estimate for the cell volume of *Prochlorococcus* is ~0.1 fl assuming a spherical cell shape, a reasonable assumption based on electron micrographs (*67*). Cellular protein amount was measured to be ~15 fg/cell in *Prochlorococcus* (corresponding to ~150 mg/ml concentration), which was quantified by measuring the total protein yield from cell lysate by Bradford assay and dividing by the total number of cells in the sample as measured by flow cytometry. Protein concentration was estimated by dividing mg/ml by cell volume.

Given a protein amount of ~15 fg/cell and assuming an average protein molecular weight of 26.7 kDa in bacteria (*68*), we can estimate that there are ~10^5 copies of proteins/cell in *Prochlorococcus*. Wild-type *Synechococcus* expresses a total of ~20,000 Kai proteins, which would constitute ~20% of the total proteome in *Prochlorococcus*, a large amount for an already minimal photosynthetic organism.

Supplemental Figures

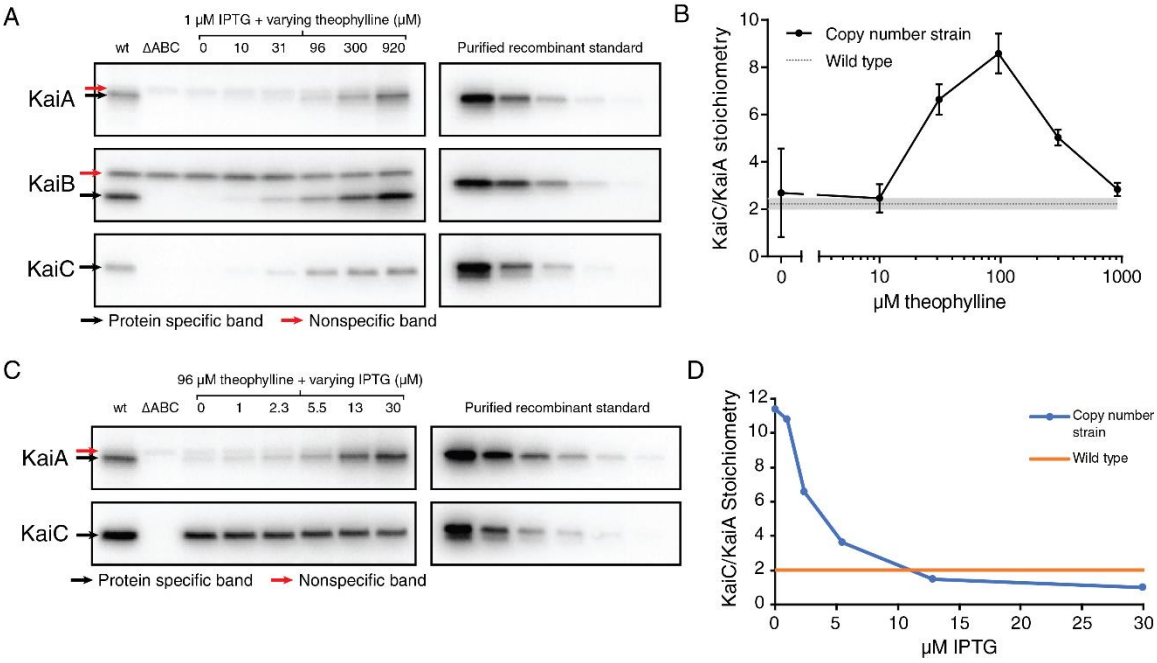

Fig. S2.1. Characterization of copy number tunable strain. (**A**) Representative western blot images for data presented in Fig. 1C (n = 3 biological replicates). For each protein KaiA, KaiB, and KaiC, cell lysate and recombinant standards were blotted simultaneously on the same membrane. *Left*: cell lysate from wild-type, *kaiABC* null, or copy number tunable cells incubated in 1 μM IPTG and various amounts of theophylline. For KaiA and KaiB western blots, 3 μg total protein was loaded per well. For KaiC western blot, 0.75 μg total protein was loaded per well. Protein specific bands are highlighted with black arrows, and nonspecific bands are highlighted with red arrows. *Right*: recombinant protein standards were all carried out in a dilution series, with the maximum and minimum amounts loaded per standard as follows: 8 ng to 0.16 ng for KaiA, 5 ng to 0.10 ng for KaiB, and 12 ng to 0.24 ng for KaiC. (**B**) Relative stoichiometry of KaiC to KaiA cellular copy number using data from (**A**). Error bars and shaded interval indicate the standard error of the mean. (**C**) *Left*: cell lysate from wild-type, *kaiABC* null, or copy number tunable cells incubated in 96 μM theophylline and various amounts of IPTG. For both KaiA and

(Fig. S2.1 continued) KaiC, 3 μg total protein was loaded per well. *Right*: recombinant protein standards were similarly carried out in a 2-fold dilution series, with the maximum and minimum amounts loaded per standard as follows: 8 ng to 0.25 ng for KaiA and 12 ng to 0.38 ng for KaiC. (**D**) Relative stoichiometry of KaiC to KaiA cellular copy number using data from (**C**).

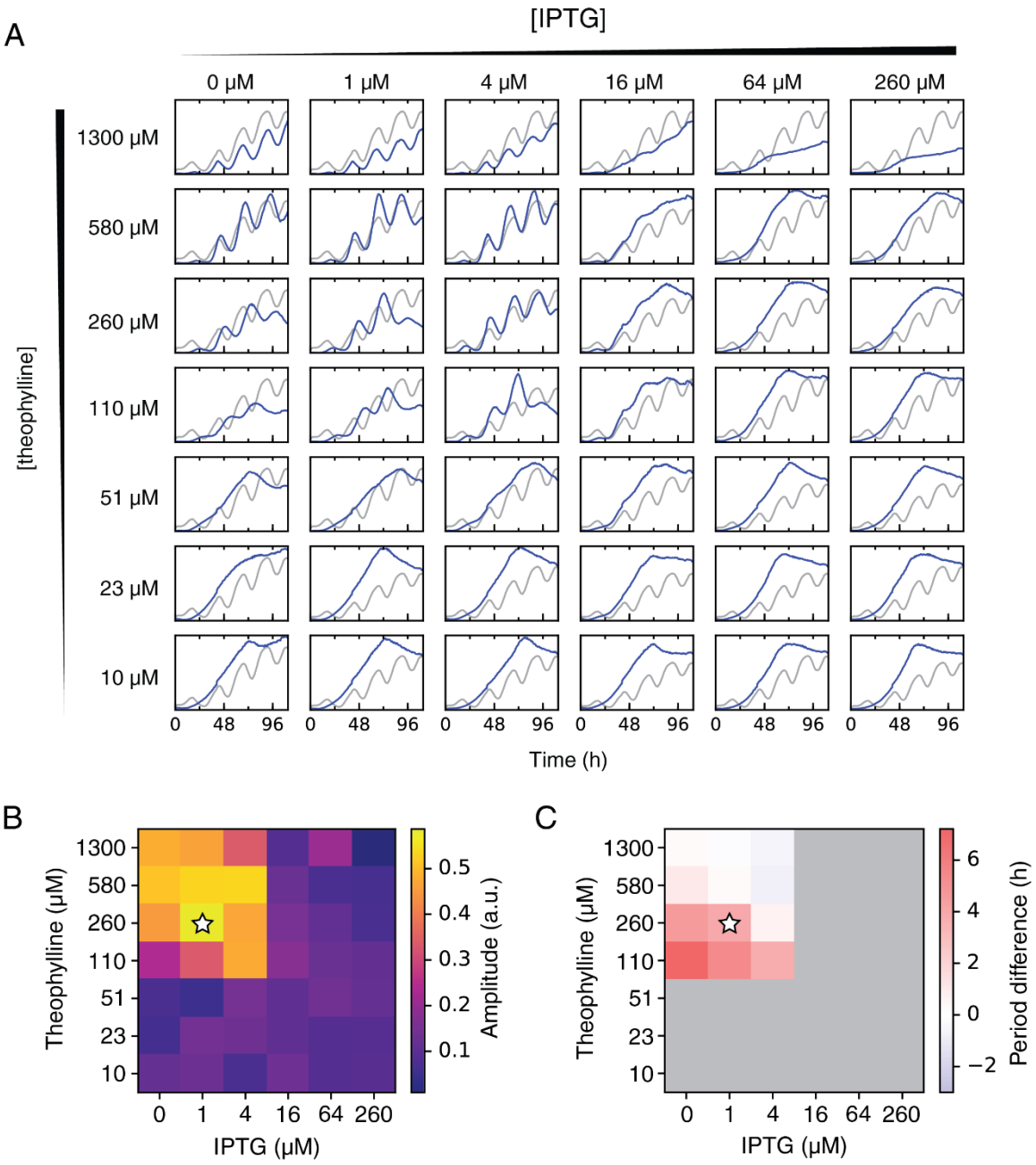

Fig. S2.2. Bulk level oscillation data from the TopCount plate reader assay. (**A**) Bioluminescence traces over time in either wild-type cells (*gray*) or copy number tunable cells plated on BG11-agar containing various amounts of inducer (*blue*). Shown are the averages of data from n = 12 individual wells for wild-type and n = 4 individual wells for copy number tunable cells. Upward trend in bioluminescence is due to cell growth/division over the course of the experiment

(Fig. S2.2 continued) (data shown is pre-normalization). (**B**) Heatmap of oscillation amplitude at various levels of inducer. Amplitudes are obtained from fitting data to a sinusoid after normalizing data for cell growth/division. White star indicates conditions that approximate wild-type levels of Kai protein expression. (**C**) Heatmap of oscillation period relative to wild-type cells after similar fitting to sine curves after normalization. Gray areas indicate conditions where the oscillation amplitude was too weak to determine period in a satisfactory manner. White star indicates conditions that approximate wild-type levels of Kai protein expression.

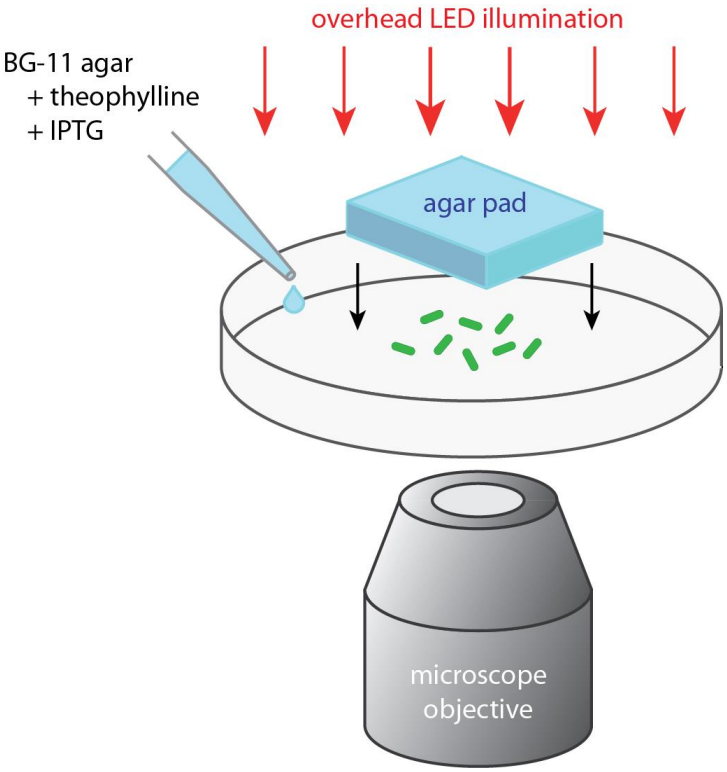

Fig. S2.3. Experimental setup for fluorescent time-lapse microscopy. Cells are incubated in 96-well plates (only a single well is shown, not drawn to scale) underneath an agar pad with uniform illumination. Additional BG11-agar containing inducers is added on top of the agar pad. The entire setup is contained within a light and temperature-controlled box held at 30°C.

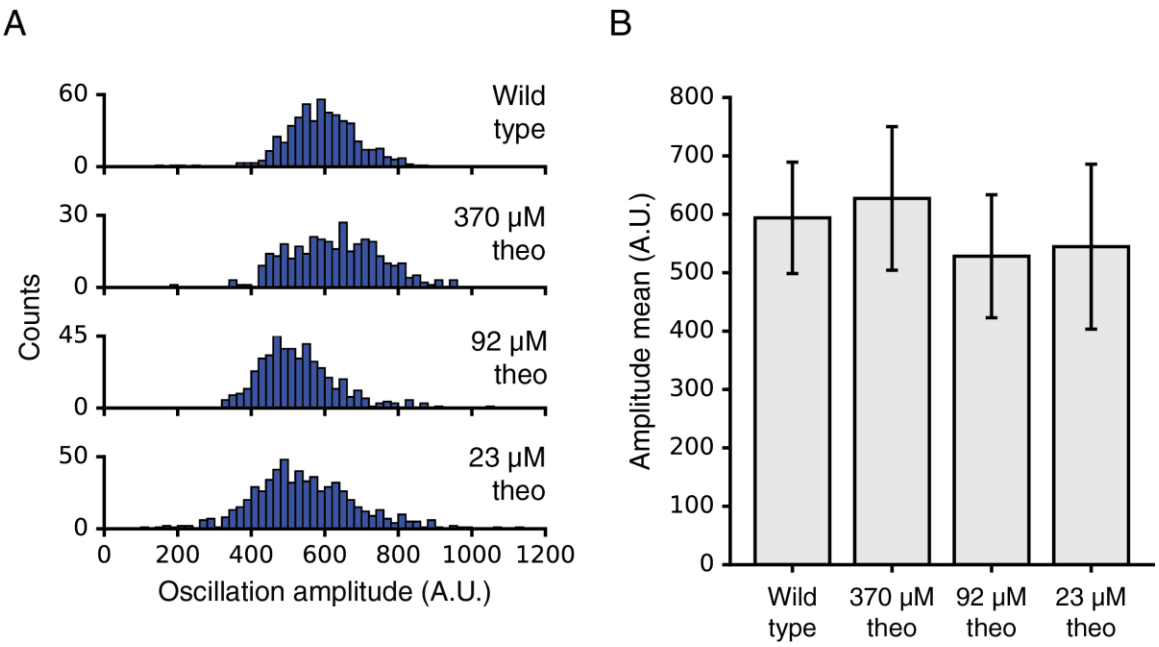

Fig. S2.4. Oscillation amplitude in wild-type cells and in the copy number tunable strain at various theophylline concentrations and 1 μM IPTG. (**A**) Histograms of oscillation amplitudes. (**B**) Quantification of mean oscillation amplitudes shown in (A) with error bars denoting standard deviations of amplitude distributions.

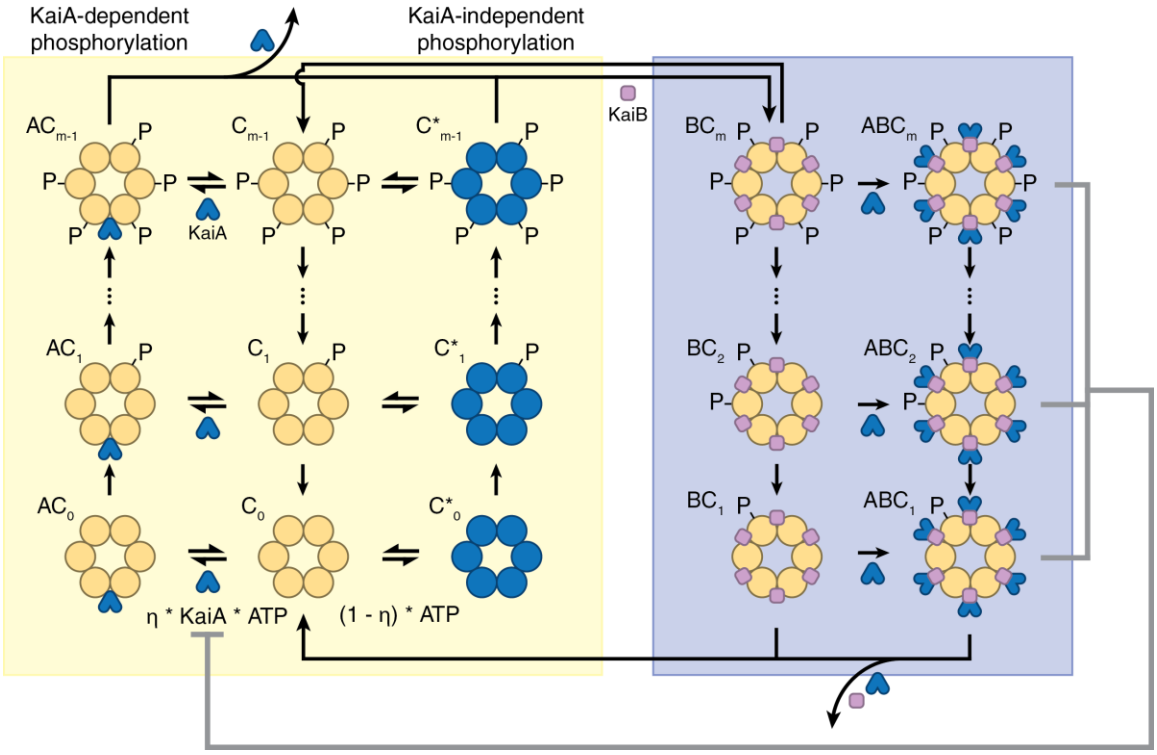

Fig. S2.5. Model of Kai system. KaiC is either in a phosphorylation-competent mode (*yellow box*) or a dephosphorylation mode (*blue box*), which corresponds to the ordered steps of phosphorylation and dephosphorylation observed experimentally. In the phosphorylation-competent mode, ground-state KaiC (*middle of yellow box*) can enter one of two states that allow for phosphorylation: KaiA-bound (*left*) and an equivalent state that is not KaiA-bound (*right*). These two states correspond to the scenario in which the KaiC C-terminal tails are exposed, promoting phosphorylation (*69*). Feedback loop strength η determines the relative rates at which KaiC enters the KaiA-bound or KaiA-unbound phosphorylating state. KaiC undergoes a series of phosphorylation steps (a total of m steps, $m = 6$ shown here for illustrative purposes) until it becomes fully phosphorylated, unbinding KaiA if previously bound and binding to KaiB. Dephosphorylating KaiC can either dephosphorylate as part of a KaiBC complex or it can do so as part of a ternary KaiABC complex which sequesters KaiA and forms the oscillator negative

(Fig. S2.5 continued) feedback loop. KaiC that becomes completely dephosphorylated unbinds KaiB and KaiA if previously bound, and the cycle restarts.

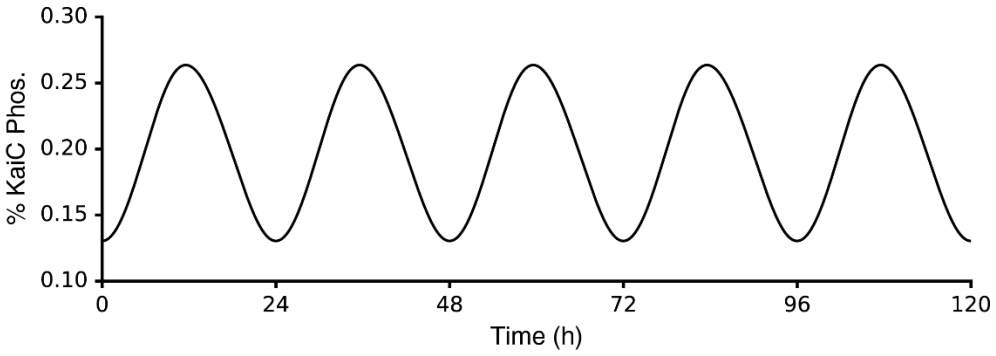

Fig. S2.6. Deterministic simulation of model depicted in fig. S5 with parameter m = 5 and numerically integrated with the fourth order Runge-Kutta method. The simulation was first initialized from 2.0 μM dephosphorylated KaiC hexamers and 2.0 μM KaiA dimers, and it was allowed to reach steady state oscillations by running the simulation for 1994.49 hours and saving the simulation state at that time. The simulation was then run for an additional 120 hours from the saved state, and the resulting trajectory is plotted here.

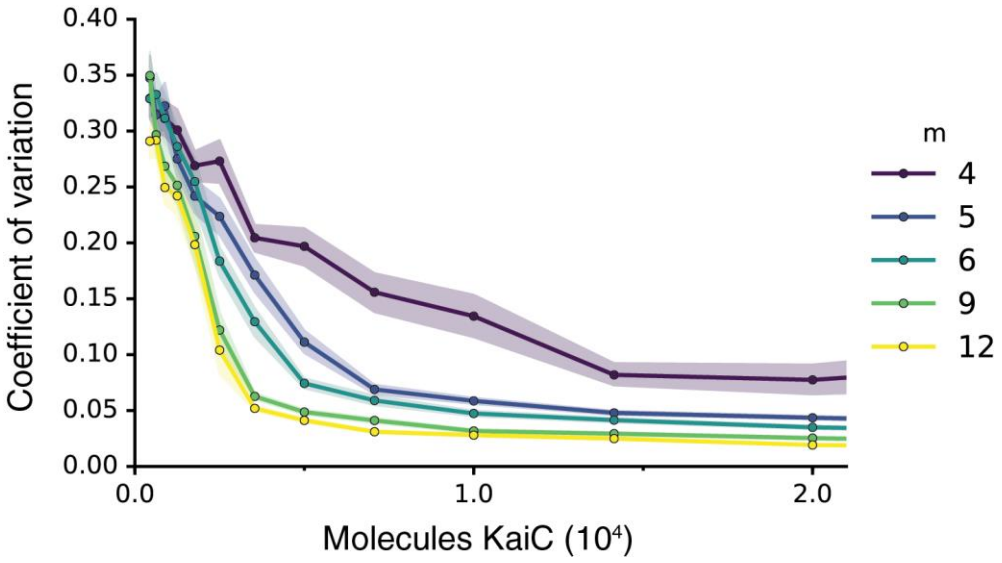

Fig. S2.7. Different values of m (number of phosphorylations on KaiC required to switch from phosphorylating to dephosphorylating) show different noise scaling properties. Shown above is Kai copy number plotted against the coefficient of variation for various values of *m*. Shaded areas indicate 95% confidence intervals from bootstrapping (5000 iterations). The number of peak-to-peak intervals in each data point in each curve ranges from n = ~800 to n = ~1400.

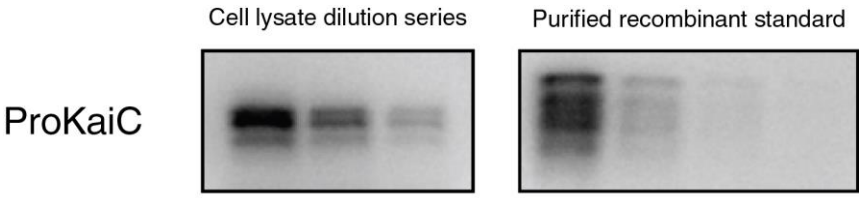

Fig. S2.8. Representative western blot image of Prochlorococcus KaiC used for quantifying the cellular copy number (n = 3 biological replicates). *Left*: cell lysate from *Prochlorococcus* cells were loaded in a dilution series with 7.5 μg, 3.8 μg, and 1.9 μg of total protein loaded in each well from left to right. *Right*: purified recombinant standard of *Prochlorococcus* KaiC with 29 ng, 15 ng, 7.1 ng, and 3.6 ng loaded in each well from left to right.

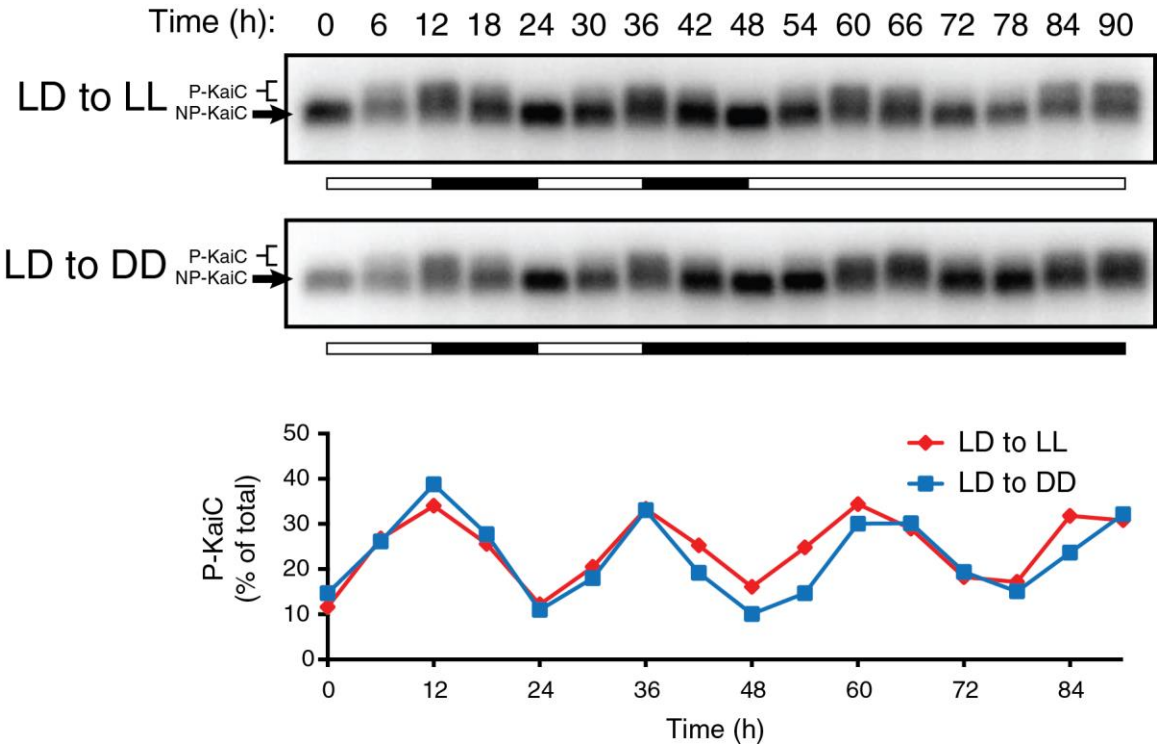

Fig. S2.9. KaiC phosphorylation dynamics in Synechococcus. *Top*: western blot time course showing *Synechococcus* KaiC phosphorylation in cultures incubated in light-dark cycles followed by either constant light or constant dark. *Bottom*: quantification of phosphorylated KaiC over time by densitometry.

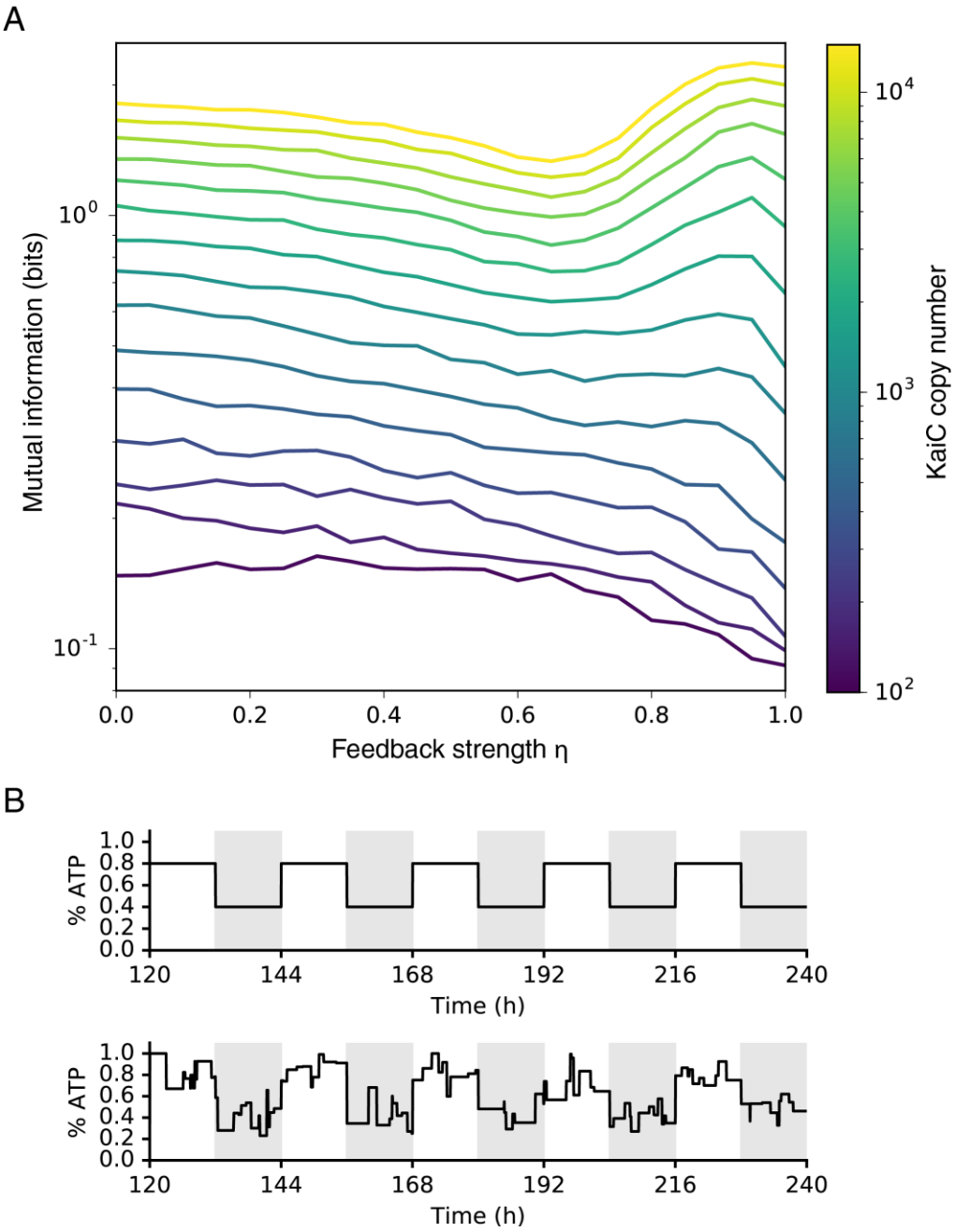

Fig. S2.10. Additional data from mutual information calculations. (**A**) Clock/environment mutual information from model simulations at varying feedback strength and Kai copy number. Each curve corresponds to a single column in the heatmap in Fig. 4F, and the feedback strengths that maximize the mutual information of each curve corresponds to the points plotted in Fig. 4G. (**B**)

(Fig. S2.10 continued) ATP ratio environmental input without fluctuations (*top*) and with fluctuations (*bottom*). Alternating light and shaded backgrounds indicate 12 hour day-night cycle.

Supplemental Tables

Table S2.2. Parameters used in the Kai model.

Category	Parameter	Value	Description
Common parameters for all simulations	m	5 unless otherwise specified	Number of phosphorylation steps
	η	1 unless otherwise specified, varies from 0-1	Feedback loop strength
	f_{ATP}	1 unless otherwise specified, varies from 0-1	Environmental input
	dt	$0.01\ h^{-1}$	Step size for deterministic simulation
Rate constants	k_{phos}	$2m \cdot 0.04902\ h^{-1}$	Phosphorylation rate
	k_{dephos}	$2m \cdot 0.04902\ h^{-1}$	Dephosphorylation rate
	$k_{A_{on}}$	$0.2451\ \mu M^{-1}\ h^{-1}$	Rate of KaiA/KaiC binding to switch KaiC to phosphorylating state
	$k_{A_{off}}$	$0.02451\ h^{-1}$	Rate of KaiA/KaiC disassociation from phosphorylating state
	$k_{C\rightarrow C^*}$	$0.2451\ h^{-1}$	Rate of KaiC switching to KaiA-independent phosphorylating state
	$k_{C^*\rightarrow C}$	$0.02451\ h^{-1}$	Rate of KaiC switching from KaiA-independent phosphorylating state
	k_{ABC}	$110.80\ h^{-1}$	Rate of KaiC binding to KaiB to sequester KaiA
Deterministic simulation initial conditions	$[C_0]_{init}$	$1.56\ \mu M$	Initial concentration of KaiC hexamers
	$[A_{tot}]$	$2.0\ \mu M$	Initial concentration of KaiA dimers
	dt	$0.01\ h$	Step size for numerical integration
Stochastic simulation initial conditions	scale_factor	Varies	Scaling factor for copy number, where 1 indicates wild-type-like levels
	$n_{C_{0_{init}}}$	$1200 \cdot$ scale_factor	Initial copy number of KaiC hexamers
	$n_{A_{tot}}$	$1200 \cdot$ scale_factor	Initial copy number of KaiA dimers

Table S2.3. Calculated concentrations of Kai proteins in *S. elongatus* and *P. marinus*. Concentrations were calculated using western blotting copy number estimates and estimated cell volumes of 2 fl and 0.1 fl for *S. elongatus* and *P. marinus*, respectively.

	[KaiA]	[KaiB]	[KaiC]
Wild-type *S. elongatus*	3.0 μM	9.0 μM	6.6 μM
P. marinus	N/A	n.d.	11.4 μM

Discussion and Future Directions

Background context of the study

Although it has been long established that circadian clocks are able to oscillate with a robust, precise 24-hour period, one of the fundamental questions in circadian biology is how these biochemical oscillators maintain such consistent periodicity in the face of both environmental and internal fluctuations. The work presented here utilizes a combined experimental and theoretical approach to probe the effect of molecular stochasticity in the cyanobacterial clock by varying clock protein copy number in single cells and observing the effect on timing precision. We find that reduction of Kai copy number to below ~10,000 copies leads to erratic oscillations, and our modeling suggests that this is due to a noise bottleneck in the negative feedback loop necessary for oscillations. While many others have theoretically investigated the effects of molecular noise on biological oscillators, there are relatively few studies that investigate molecular noise experimentally.

One previous study theoretically investigated molecular noise by stochastically simulating the mammalian circadian clock (*70*), which relies on a transcriptional feedback loop to sustain oscillations, unlike the cyanobacterial oscillator. A key result from this work was that a noise bottleneck in the system occurred at the interface of transcription factors binding to their

promoters, consisting of discrete binding and unbinding events. Slow binding and unbinding led to greater variability in oscillation period, and increasing the rate of binding kinetics in turn decreased timing variability, perhaps due to the fact that more binding events per hour will average out noise in binding kinetics to a greater degree. Additionally, increasing the number of promoters in the system (reflecting a gene duplication event) also decreased the amount of oscillation variability, mirroring the results from our work in which increasing the total copy number of molecules suppresses the contribution to noise from the negative feedback loop of the Kai oscillator.

The effect of increasing gene dosage to reduce oscillatory noise was also observed experimentally in the timing of the yeast cell cycle, in which increasing the genomic copy number of G1 cyclins (covering 1N/2N/4N ploidy) decreased the variability in timing of G1 duration, or the period between the initiation of cell division and budding (*71*). Additionally, the variability in G1 duration was observed to decrease by $\sqrt{2}$ for each doubling of ploidy, indicating that noise likely scaled with a $1/\sqrt{N}$ dependence, which is consistent with the predicted noise scaling if molecular noise was the main driver of noise.

The main conclusions of our work are consistent with the findings of these previous studies, which together find that molecular noise stemming from finite copy number constraints can increase variability in the timing of cellular processes. In addition, our work expands on these conclusions by introducing the idea that molecular noise is not only present in biological systems but can also present a substantial evolutionary pressure due to the conflict between noise suppression from high biomolecular copy number and the biosynthetic cost of synthesizing large numbers of molecules. In our work, we do not solely test the theoretical prediction that biochemical reactions become more stochastic at low molecular copy number. Our observation

that a smaller cyanobacteria possessing a naturally lower Kai copy number also possesses a qualitatively different timing system suggests that across the entire range of microbial cell sizes, cells may face different biophysical constraints in how to optimally suppress molecular noise. This demonstration that molecular noise can have significant consequences for organism function will be important for future efforts to engineer synthetic biological oscillators, for example, and it may also be relevant to our continued study of the microbiome, especially as we investigate how the gut microbiome may anticipate circadian changes in the mammalian gut environment.

Study limitations

While our work expands upon previous ideas of how molecular noise presents itself in biological systems, it is also important to note its limitations. The first to note is that high Kai copy number expression almost certainly plays only a partial role in noise suppression in the cyanobacterial oscillator. Expression of the *kaiBC* operon is itself under circadian control, forming an additional transcriptional feedback loop on top of the mechanisms driving the post-translational oscillator. In placing the expression of KaiB and KaiC under control of constitutive genomic promoters in this study, we focus on noise in the post-translational oscillator but neglect the contribution of the feedback loop to noise suppression. Indeed, one previous study demonstrated that the transcriptional feedback loop in the Kai oscillator does make a contribution to noise suppression, and that replacement of the *kaiBC* promoter with a constitutive promoter increases the variability in period of the oscillator (*40*). Our results are consistent with this previous study—the copy number tunable strain induced with 92 μM theophylline shows much noisier oscillations compared to wild-type cells, despite having a similar copy number of

KaiC (~8,000 KaiC/cell in the wild-type vs ~7,000 KaiC/cell in the copy number tunable strain at 92 μM theophylline).

An interesting follow-up study that could be performed is to further dissect the mechanism by which the transcriptional feedback loop reduces noise in the oscillator. How many extra copies of Kai proteins is equivalent to the presence of transcriptional feedback in terms of improving oscillator precision? Perhaps transcriptional feedback is an alternate way to reduce oscillatory noise without invoking the extra biosynthetic burden that would accompany expression of Kai proteins at a higher level. How does transcriptional feedback reduce clock noise exactly? An attractive hypothesis is that perhaps certain points in the oscillator cycle are more sensitive to low copy molecular noise, and that increased expression of KaiB and KaiC at this point in the cycle effectively buffers these vulnerable points in the cycle against molecular noise. Indeed, expression levels of KaiC have been reported to oscillate over a three-fold range from ~5,000 copies per cell at the minimum up to a maximum of ~15,000 copies, with the maximum expression levels occurring close to the onset of subjective dusk, or CT 12 hours (*72*). One potential time of day when the oscillator may be susceptible to low copy number noise is at the time of cell division, which is significant because cell division reduces cellular protein copy number by roughly two-fold. The clock has been shown to gate the timing of cell division, and under certain conditions in which cells divide roughly once per day, cell division is most likely to occur during the period roughly corresponding to increased levels of KaiB and KaiC expression (*73*). It is thus possible that the transcriptional feedback loop buffers the oscillator against the reduction in copy number accompanying cell division, meriting further investigation both experimentally and theoretically.

Another limitation of our study is in identifying the relative contribution of molecular noise to oscillatory period variability compared to other sources of noise in the system. In addition to molecular noise from low copy number effects, other possible contributors of noise include: fluctuations in Kai protein expression, fluctuations in Kai protein stoichiometry, noise in the clock output pathway (e.g. the YFP clock reporter), noise in image acquisition, and noise in image analysis. While it is possible to measure whether Kai protein expression noise changes significantly as theophylline induction is altered in our system (using a strain in which expression of a YFP reporter is placed under control of the riboswitch), it is more difficult to determine the relative contribution to oscillatory noise from fluctuations in protein expression versus biochemical stochasticity of the clock reaction.

A final limitation of our study lies within our comparison of the Kai systems within *Synechococcus* and *Prochlorococcus*. Here, we make the assumption that the Kai system in *Prochlorococcus* is responsible for global regulation of gene expression as it is in *Synechococcus*. Currently, this hypothesis is untestable due to the difficulty of performing genetics in *Prochlorococcus*, and it is thus untenable to knock out the *kaiBC* locus in Prochlorococcus to assess the effects on previously reported global rhythms in gene transcription. However, the presence of the clock output protein SasA in *Prochlorococcus* suggests that it is still at least plausible that the Kai system continues to influence gene expression. The inability to perform genetics in *Prochlorococcus* (and thus the inability to engineer a clock reporter) also precludes us from performing experiments that directly test our theoretical predictions that the *Prochlorococcus* timer-like system can track the time of day better at low copy number compared to *Synechococcus* given that there is minimal or no

environmental input fluctuation. These experiments may become possible in the future should methods be developed to allow genetic manipulation of *Prochlorococcus*.

Additional thoughts and interpretations of data

Although our study is unable to genetically probe the role of the Kai proteins in *Prochlorococcus*, we are the first to document the phosphorylation dynamics of KaiC. One previous study established that in *Prochlorococcus*, transcription of *psbA* oscillates in light-dark cycles but not in constant light (*49*), but it was unknown whether this reflected gene-specific dynamics or whether the clock system itself also behaved as a timer. In showing that KaiC phosphorylation dynamics are solely driven by light-dark cycles (phosphorylation in the day and dephosphorylation in the dark), our work suggests that KaiC in *Prochlorococcus* may still regulate global gene expression, assuming that the clock output pathway in *Prochlorococcus* functions similarly to that in *Synechococcus*. Additionally, our work indicates that unlike *Synechococcus*, in which the net phosphorylation and dephosphorylation phases of the cycle proceed at roughly comparable speeds to produce a symmetric oscillatory waveform (see Figure S2.9), phosphorylation of *Prochlorococcus* KaiC occurs much more rapidly than dephosphorylation (see Figure 2.4C). Indeed, KaiC becomes fully phosphorylated within 6 hours in light (and potentially even more rapidly due to the coarse time resolution of the experiment), but takes ~12 hours to dephosphorylate from 100% to 50% phosphorylation, and it takes an additional ~12 hours to become completely dephosphorylated. This raises an interesting question. It may be possible that timers are more adversely affected by environmental noise compared to free-running clocks because they lack an internal timekeeping mechanism that helps to filter environmental noise, and indeed our information theoretic analysis indicates that free-

running clocks can track the time of day better in the presence of environmental fluctuations (ignoring the effects of molecular noise at the deterministic limit of high protein copy number). If this is true, might *Prochlorococcus* KaiC have evolved to dephosphorylate slowly as a way to filter environmental noise during the day, acting as a low bandpass filter? It should be noted that the presence of environmental fluctuations in light levels during the day/night cycle is asymmetrical—while there may be unexpected pulses of darkness during the day due to weather, it is much less likely for there to be unexpected brightness during the night. Could the slow dephosphorylation dynamics of *Prochlorococcus* KaiC ensure that dephosphorylation is only initiated when cells undergo an extended period of darkness (indicative of true night and less likely to occur from weather), thus preventing cells from spuriously perceiving the onset of night? In this view, the asymmetrical phosphorylation dynamics of *Prochlorococcus* KaiC may represent an evolutionary adaptation to filtering out the asymmetrically distributed light fluctuations during the day vs. the night given that the cell no longer has a free-running clock.

Our work also raises the question of how and why *Prochlorococcus*, specifically *P. marinus* MED4, lost KaiA and thus the ability to sustain free-running oscillations. Phylogenetic studies indicate that as the *Prochlorococcus* clade is traversed from the more basal cyanobacteria (which have copies of all three Kai genes) to the more recently diverged cyanobacteria, KaiA is gradually truncated from the N-terminus until it is lost entirely in *P. marinus* MED4 (*49*). How is the loss of KaiA related to the decrease in cell size over evolutionary time? It is thought that small cells are part of a cell-wide adaptation to nutrient-deplete waters in order to increase the efficiency of nutrient uptake, and the decrease in cell size may be one component of an overall process that includes progressive genome streamlining and reduction of genomic GC content to decrease cellular demand for nitrogen (*48*). Interestingly, one of the intermediate species of

Prochlorococcus (MIT9313) has a truncated form of KaiA that retains only the C-terminus (*49*), and its cell size appears to be roughly twice that of MED4 (*74*). Could the Kai copy number in MIT9313 represent the minimum copy number at which a free-running clock is more optimal than a timer? To answer this question more fully, more work will be needed to more thoroughly catalogue the cell sizes, Kai copy numbers, and KaiC phosphorylation dynamics of these intermediate *Prochlorococcus* species. A complementary approach to the question of how *Prochlorococcus* lost KaiA utilizing ancestral protein reconstruction is also briefly discussed in the appendix of this dissertation.

Additional future directions

One of the main predictions of this work is that cells with noisy clocks will experience fitness defects due to failure to accurately predict the onset of light or dark. An experiment that could be performed to test this prediction is to induce the copy number tunable strain to express various copy numbers of Kai proteins while incubating the cells in 12h:12h light/dark cycles. To determine the effects on fitness, cell growth can be measured on the microscope at the single cell level, which will additionally provide information on whether individual cells that experience growth defects do so because of a clock that is set to the wrong time due to molecular noise. If cells with noisier clocks experience growth defects in a cyclic environment, these findings would also confirm the work that is discussed in Chapter 4 of this dissertation, which specifically investigates the consequences of clock-environment mismatch on cell fitness in cyanobacteria.

The other major theoretical prediction from the work in this chapter is that removal of the noisy negative feedback loop in the Kai oscillator will lead to an environmentally driven timer that can better track the time of day at low protein copy number. Testing this prediction

experimentally would require tracking the clock state in single *Synechococcus* cells expressing Kai protein at *Prochlorococcus*-like copy numbers, and comparing the noise in clock output in *Prochlorococcus* itself. Currently, such an experiment would be difficult to execute, again due to the fact that *Prochlorococcus* is not a genetically tractable organism, making it infeasible to track clock state in single *Prochlorococcus* cells with a genetically-encoded fluorescent clock reporter.

An alternative to genetically engineering *Prochlorococcus* with a clock reporter is to instead transplant the *Prochlorococcus* Kai system into *Synechococcus*, replacing the Kai system native to *Synechococcus*. This approach has been attempted in the Rust lab by Gopal Pattanayak, and initial experiments appeared successful, producing a strain in which oscillations dampened in constant light. However, issues remain regarding the reproducibility of these results, casting doubt on the initial observations. In order to successfully transplant ProKai, there are several issues that must be dealt with, including finding an appropriate expression level for both proteins as well as ensuring that ProKai can interface correctly with the clock output proteins in *Synechococcus*, SasA and CikA. Interestingly, preliminary results indicate that *Prochlorococcus sasA* (ProSasA) can rescue a *sasA* null *Synechococcus* mutant (data not shown, experiment performed by Gopal Pattanayak), suggesting that transplanting both the ProKai system as well as ProSasA might help solve problems that arise if ProKai and *Synechococcus* clock output proteins are found to be incompatible.

Chapter 3: Mixtures of opposing phosphorylations within hexamers precisely time feedback in the cyanobacterial circadian clock

Foreword

Here, I present work published by Jenny Lin and myself (on which I am a second author), and all supplementary material can be found in (*46*). This work investigates the mechanism by which the Kai oscillator can generate oscillations over a large range of protein stoichiometries. In brief, we found that the hexameric nature of KaiC combined with the opposing effects of two phosphorylation residues can generate an ultrasensitive switch that mediates KaiB-KaiC binding, and that this is sufficient to explain the robustness of the Kai system to changes in KaiA/KaiC stoichiometry. Jenny Lin designed and performed all of the experiments and designed and implemented the vast majority of the model, and I contributed to modeling in Figure 3.5 that further explored the role of ultrasensitivity in generating robustness to changes in protein stoichiometry. This chapter concludes with a "Perspective" section that discusses possible implications of this study for the work published in Chapter 2.

Abstract

Circadian oscillations are generated by the purified cyanobacterial clock proteins, KaiA, KaiB, and KaiC, through rhythmic interactions that depend on multisite phosphorylation of KaiC. However, the mechanisms that allow these phosphorylation reactions to robustly control the timing of oscillations over a range of protein stoichiometries are not clear. We show that when KaiC hexamers consist of a mixture of differentially phosphorylated subunits, the two phosphorylation sites have opposing effects on the ability of each hexamer to bind to the

negative regulator KaiB. We likewise show that the ability of the positive regulator KaiA to act on KaiC depends on the phosphorylation state of the hexamer and that KaiA and KaiB recognize alternative allosteric states of the KaiC ring. Using mathematical models with kinetic parameters taken from experimental data, we find that antagonism of the two KaiC phosphorylation sites generates an ultrasensitive switch in negative feedback strength necessary for stable circadian oscillations over a range of component concentrations. Similar strategies based on opposing modifications may be used to support robustness in other timing systems and in cellular signaling more generally.

Introduction

Circadian clocks are biological timing systems that allow organisms to anticipate and prepare for daily changes in the environment. A hallmark of a circadian oscillator is its ability to drive self-sustained rhythms in gene expression and behavior with a period close to 24 hours even in the absence of environmental cues (*75*). A general challenge for the biochemical machinery that generates rhythms is to precisely define the duration of the day in the face of perturbations, including fluctuations in the cellular abundance of the molecular components. The importance of maintaining precise circadian timing is underscored by experiments that show that mismatch between the clock period and the rhythms in the external environment results in health problems and fitness defects (*76, 77*).

Though circadian clocks are found across all kingdoms of life, the Kai oscillator from cyanobacteria presents a uniquely powerful model system to study the design principles inherent in the molecular interactions that generate rhythms. A mixture of the purified proteins KaiA, KaiB, and KaiC results in stable oscillations in the phosphorylation state of KaiC *in vitro* that

persist for many days and share many of the properties of circadian clocks *in vivo* (*22, 24, 41*). In particular, the oscillator is able to successfully generate near-24 hour rhythms over a range of concentrations of the clock proteins both *in vivo* and *in vitro* (*78-80*), so that fine-tuning of gene expression is not needed to support a functional clock. Much has been learned about the behavior of the isolated Kai proteins, including the determination of high-resolution crystal structures of all three components (*81-83*). A critical challenge that remains is to understand how the properties of the Kai proteins are integrated together in the full system to generate precisely timed rhythms.

KaiC appears to be the central hub of timing information in the oscillator. Each KaiC molecule consists of two AAA+ family ATPase domains which consume the free energy of ATP hydrolysis to drive oscillations. Like many other members of this family, KaiC forms hexamers, and the enzymatic active sites are formed at the subunit interfaces where nucleotides are bound. The C-terminal, or CII, domain of KaiC has additional phosphotransferase activities that are unusual for the AAA+ family: it can phosphorylate and dephosphorylate two residues near the subunit interface, Ser431 and Thr432 (*84*). KaiC autokinase and autophosphatase activities occur at the same active site (*85, 86*). In isolation, KaiC has high phosphatase activity, but the enzyme is pushed towards kinase activity by the activator protein KaiA which interacts directly with the KaiC C-terminal tail (*87, 88*). Roughly speaking, kinase activity predominates during the day, and phosphatase activity predominates during the night (*89*). Thus, understanding the feedback mechanisms that generate a precise time delay between these modes is crucial to understanding timing in the oscillator (*69*).

Inactivation of KaiA and a transition from kinase to phosphatase mode occur when KaiB•KaiC complexes form, closing a negative feedback loop by sequestering KaiA in a ternary

complex and leaving it unable to act on other KaiC molecules (*37, 90*). By temporarily removing KaiA molecules from their activating role, this molecular titration mechanism can act to synchronize the activity of all KaiC hexamers in the reaction (*37, 91, 92*). Phosphorylation and dephosphorylation proceed in a strongly ordered fashion so that, in response to a change in KaiA activity, Thr432 is (de)phosphorylated first, followed later by Ser431 (*37, 89, 90*). It is known that phosphorylated Ser431 is important for allowing the formation of KaiB•KaiC complexes. However, recent work has made it clear that the binding of KaiB involves both KaiC domains—in particular, the slow ATPase activity of the N-terminal CI domain, which is not phosphorylated, is required for KaiB interaction (*47, 93*).

Because of the importance of precisely timing negative feedback via KaiB•KaiC complex formation for generating appropriate rhythms (*91*), we wanted to understand the role of phosphorylation of the KaiC hexamer in controlling this process. The involvement of both KaiC domains suggests that information about phosphorylation in CII is communicated allosterically through changes in hexamer structure to the CI domain, potentially through ring-ring stacking interactions (*93, 94*). We therefore hypothesized that the KaiC phosphorylation sites on each subunit might act as allosteric regulators in the context of a hexameric ring, so that phosphorylation of one subunit would alter the ability of all other subunits in the ring to engage with KaiA and KaiB, providing a cooperative mechanism to control the timing of these interactions.

We conducted a series of biochemical experiments and perturbations to study the effect of altering the status of each phosphorylation site on the KaiC hexamer. We then developed a mathematical model to interpret these results analogous to classical models of allosteric transitions in multimeric proteins. We constrain the kinetic parameters in this model using

experimental measurements of rate constants, allowing us to compare the predictions of the model directly with data. We conclude that maintenance of circadian timing over a range of protein concentrations requires an effectively ultrasensitive switch in each KaiC hexamer from an exclusively KaiA-binding state to a state that can bind to KaiB as phosphorylation proceeds. This effect requires that KaiC hexamers consist of mixtures of differentially phosphorylated subunits, as would be produced by stochastic autophosphorylation of a hexamer. Ultrasensitivity results from opposing effects of phosphorylation on Thr432 and Ser431 in controlling a concerted transition within a given KaiC hexamer. Including this mechanism in the model is necessary to explain the experimentally observed tolerance of the system to altered protein concentrations.

Results

KaiC hexamers are composed of subunits in a mixture of phosphorylation states

To experimentally interrogate the role of phosphorylation in regulating interaction with KaiB, we co-immunoprecipitated KaiC bound to KaiB during oscillating reactions, then analyzed the phosphorylation state of KaiC using electrophoresis conditions that resolve the modification status of Ser431 and Thr432 (Fig. 3.1A). This allowed us to sample a wide range of phosphoform abundances as both KaiC phosphorylation and the formation of KaiB•KaiC complexes oscillate over time (*37, 89*). As standards, we prepared mutants of KaiC either where a phosphorylation site was mutated to Ala to prevent phosphorylation or to Glu to mimic phosphorylation. When prepared as homogeneous hexamers, these mutants interact very strongly with KaiB if Ser431 is phosphomimetic, but weakly if not (*24, 37, 89, 91*). In contrast, in the

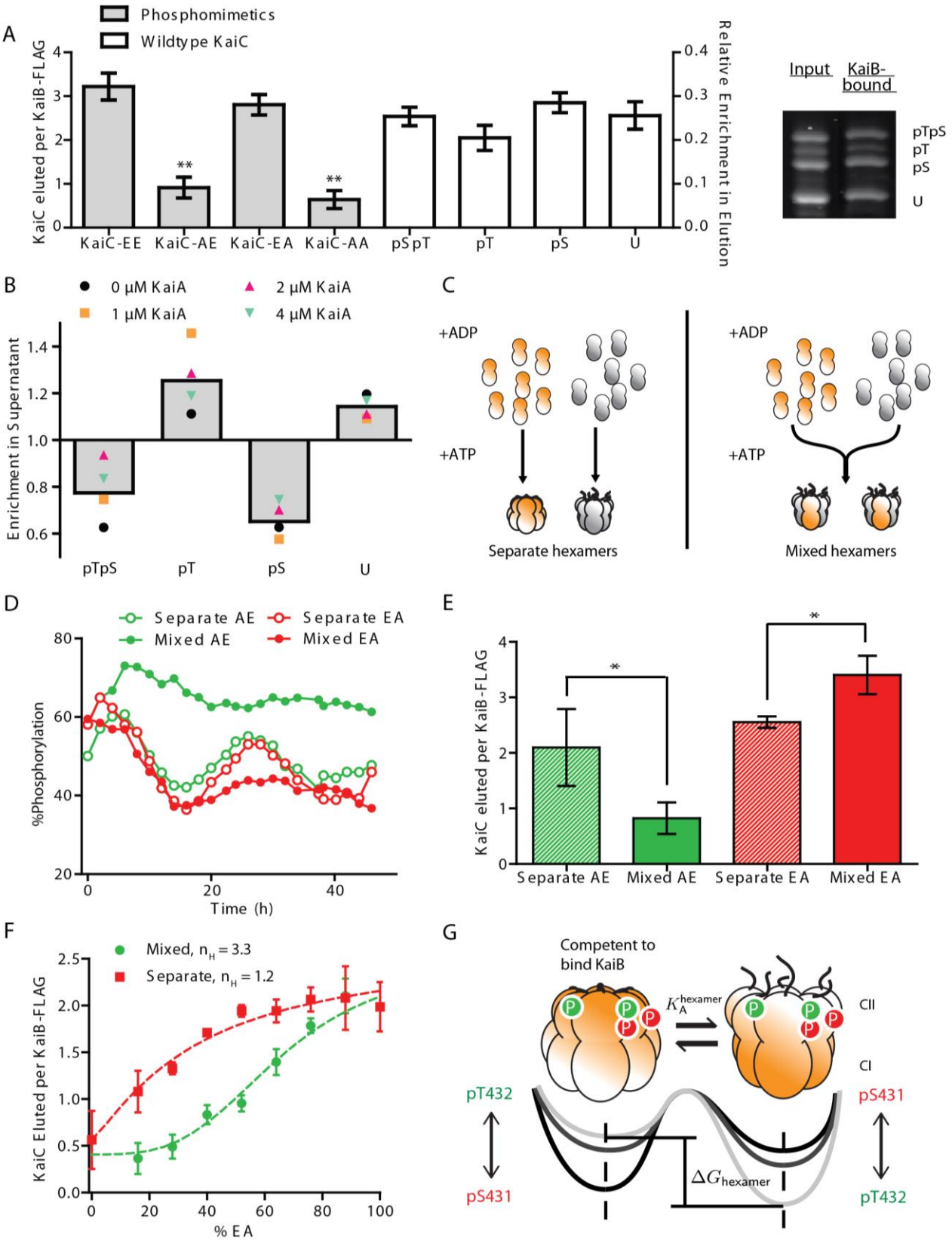

Fig. 3.1. KaiB-KaiC interaction favors KaiC hexamers with appropriate mixtures of phosphorylated subunits.

(Fig. 3.1 continued) (A) Left: Phosphorylation site mutants in homogeneous hexamers co-IPed by KaiB-FLAG (grey bars, right axis). Error bars represent standard error of three replicates after 24 h incubation. Average amount (n = 4) of each KaiC phosphoform co-IPed by KaiB-FLAG from starting from a highly phosphorylated state (white bars, left axis). Error bars represent standard deviation over a 4 h time course. Values were determined by gel densitometry as the ratio of the KaiC band intensity to the KaiB-FLAG band. Right: a representative SDS-PAGE gel image of the input and elution phosphoform compositions of KaiC co-IPed by KaiB-FLAG. ** indicates that t-tests versus KaiC-EA and KaiC-EE both gave $p < 0.01$ (S1). (B) Enrichment of each KaiC phosphoform in the supernatant relative to the material bound to KaiB-FLAG in clock reactions. Colored symbols show an average from 7-9 timepoints taken over 24 or 34 h. Grey bars indicate averages over all KaiA concentrations and timepoints. (C) Schematic for preparation of mixed hexamers and separate hexamers. KaiC phosphomimetics (orange) and His6-tagged wildtype KaiC (His6-KaiC, grey) are monomerized by the replacement of ATP with ADP, and mixed in a 1:1 ratio before ("mixed") or after ("separate") rehexamerization with ATP. (D) Phosphorylation dynamics of wildtype KaiC in clock reactions in the presence of phosphomimetics either as mixed or separate hexamers. (E) Amounts of total KaiC co-IPed by KaiB-FLAG during the dephosphorylation phase (6 h-12 h) in clock reactions with KaiC-AE and wildtype KaiC (green bars) or during the phosphorylation phase (22 h-28 h) in clock reactions with KaiC-EA and wildtype KaiC (red bars). Bar height show averages of 3-4 timepoints. Error bars indicate standard deviation. * indicates $p < 0.05$ by Student's t-test. (F) Total amount of KaiC co-IPed by KaiB-FLAG as a function of the percentage of KaiC-EA combined with KaiC-AE in either mixed or separate hexamer preparations. Points show the averages of three measurements and the error bars indicate standard deviation. Dotted lines show fits to Hill functions $y = y_{\max} \frac{1}{1+(\frac{K}{\%EA})^{n_H}}$, with nH = Hill coefficient. (G) An allosteric framework for modeling the KaiC hexamer. The phosphorylation state of each subunit contributes to the free energy difference between two conformational states of the hexamer: one that is competent to bind KaiB and one that is not. Tails extending from the CII domain suggest changes in A-loop conformation associated with each conformational state. Arrowheads indicate proposed influence of the phosphorylation sites on the stability of the two hexameric states.

case of wildtype KaiC hexamers, all forms of the KaiC subunits can be found bound to KaiB, including unphosphorylated KaiC and KaiC phosphorylated only on Thr432 (Fig. 3.1A).

We interpret these data to indicate that KaiC subunits that do not favor KaiB interaction are often co-immunoprecipitated in the context of a hexameric ring that is nevertheless bound to KaiB. This suggests that wildtype KaiC hexamers consist of mixtures of subunits in various phosphorylation states. However, because the phosphorylation site mutations indicate that Ser431 phosphorylation is required for KaiB interaction, we hypothesized that, although wildtype KaiC hexamers may contain subunits in all possible states, the relative abundance of

each phosphorylation state within a hexamer should bias the probability of that hexamer binding to KaiB.

We therefore asked whether there are systematic trends in the enrichment of the various possible phosphorylation states of the KaiC that are bound to KaiB. To detect systematic trends across many KaiC phosphorylation conditions, we sampled several reactions with different concentrations of KaiA at various time points throughout the oscillator cycle. (Fig. S1 in (*46*)). As expected, KaiC phosphorylated only on Ser431 was strongly enriched in the material bound to KaiB relative to the unbound material. However, KaiC phosphorylated only on Thr432 was preferentially excluded from the KaiB-KaiC interaction, and enriched in the unbound material (Fig. 3.1B, S1 in (*46*)). These results suggest a working hypothesis where the ability of KaiC to interact with KaiB indeed depends on the relative abundance of each phosphorylation state within a given KaiC hexamer.

Two KaiC phosphorylation sites have opposing effects on the ability of mixed hexamers to interact with KaiB

According to this hypothesis, the phosphorylation state of one subunit will alter the ability of the entire hexamer to interact with KaiB through allosteric communication within the KaiC ring. Therefore, experimentally forming mixed hexamers that contain both wildtype KaiC and phosphomimetic mutants should alter the ability of the wildtype KaiC to interact with KaiB and disrupt the function of the oscillator. In contrast, if each subunit acts independently of its hexameric context, producing mixed rings would result in no greater effect than leaving the mutant and wildtype segregated into separate hexamers. To distinguish between these alternatives, we used an ATP depletion protocol to prepare pools of largely monomeric KaiC

S431A;T432E (KaiC-AE, a mimic of pT432-only), KaiC S431E;T432A (KaiC-EA, a mimic of pS431-only), and His_6-tagged wildtype protein (*86*). To create mixtures of KaiC mutants and wildtype monomers within the same hexamers, we combined pools of monomers together and reintroduced ATP to hexamerize the mixture. As a control, we reversed the order of this procedure so that the proteins were rehexamerized without mixing before being later combined (Fig. 3.1C). This monomerization and rehexamerization procedure does not compromise the ability of the wildtype protein to oscillate (Fig. S2 in in (*46*)). We used the His_6 tag on wildtype KaiC to verify that our procedure succeeded in creating forced mixtures of mutant and wildtype where a large majority of hexamers are composite. When we rehexamerized the pools of mutant and wildtype protein separately, they remained largely segregated for at least 48 hours. In contrast, our forced mixing procedure succeeded in creating a population of hexamers that was largely composite (Fig. S2 in in (*46*)).

To test the oscillator function of these mixed hexamers, we then added KaiA and KaiB to initiate clock reactions. Consistent with our hypothesis, oscillations fail when KaiC-AE is forced to mix into wildtype hexamers, resulting in highly phosphorylated KaiC, the expected phenotype if KaiB cannot act. Mixing KaiC-EA into wildtype hexamers causes oscillations to fail with the opposite phenotype—weakly phosphorylated KaiC. However, in both cases, circadian oscillations are maintained when the mutants are present but segregated into separate hexamers (Fig. 3.1D). These failure modes of the oscillator correspond to disrupted interaction with KaiB induced by the mixing of KaiC-AE (or KaiC-AA) into wildtype KaiC hexamers, or enhanced interaction with KaiB induced by the mixing of KaiC-EA into wildtype hexamers (Fig. 3.1E, S2 in in (*46*)). These results are also consistent with a recently published report from Kitayama et al. showing that the activity of KaiC hexamers depends on their subunit composition (*95*).

To quantitatively assess how hexameric mixtures of Ser431- and Thr432-phosphorylated subunits regulate binding to KaiB, we prepared hexamers using various percentages of KaiC-AE and KaiC-EA phosphomimetics. We found a preparation of hexamers containing a mixture of KaiC-AE with KaiC-EA subunits suppressed the total amount of KaiB-KaiC interaction relative to a control where the same proteins were present, but segregated into separate hexamers (Fig. 3.1F, S2 in in (*46*)). This indicates that the presence of pThr432 subunits within the same hexamer is able to prevent the interaction of pSer431 subunits with KaiB, consistent with the correlations we observed in the wildtype oscillator.

Crucially, hexameric mixtures of pSer431 and pThr432 mimics show a sigmoidal dependence of KaiB interaction strength on the fraction of pSer431 mimic present in the mixture (effective Hill coefficient ≈ 3.3), an effect which was absent (effective Hill coefficient ≈ 1.2) when the two phosphomimetics were kept in separate hexamers (Fig. 3.1F). Because of the kinetic ordering of phosphorylation reactions in KaiC, oscillations are characterized by periods where either pThr432 or pSer431 alternately dominate in relative abundance (*37*, *89*). Considering the switch-like transition in KaiB-KaiC interaction we observed as the balance within hexamers is shifted to favor pSer431 over pThr432, we hypothesized that dynamic changes in the mixture of phosphorylation states in a hexamer might be key to understanding the timing of the transition between the phosphorylation and dephosphorylation phases of the circadian rhythm.

To mathematically model these effects, where the binding affinity of a KaiC hexamer for KaiB depends on the mixture of post-translational modifications on all of the KaiC subunits, we introduce a simple allosteric framework. In classical models of allostery in oligomeric proteins, such as the Monod-Wyman-Changeux treatment of hemoglobin, it is assumed that each subunit

can adopt different conformational states, but that, because breaking the symmetry of the molecule is disfavored, all of the subunits in a given oligomer must be in the same conformational state at any moment in time. The role of ligand binding is then to bias the free energies of the possible subunit conformations, resulting in a cooperative switch in the conformation of the oligomers as ligand concentration increases (*96*).

We extend this treatment to describe allosteric effects in KaiC, by hypothesizing that KaiC hexamers can exist in two conformational states: one that allows interaction with KaiB and one that does not. These two conformational states are likely related to changes in the stacking interactions between the CII and CI rings and the exposure or burial of the hydrogen-bonded network of KaiA-binding activation loops recently identified by structural studies (*69, 93, 94, 97*). Consistent with this picture of allostery, recent structural work has identified changes in solvent accessibility across both domains of KaiC when it is bound to KaiB (*98*).

As in Monod-Wyman-Changeux, we assume that the dynamic interconversion between these states are rapid and must be all-or-none within a given hexamer. We take phosphorylation of KaiC to play a role similar to that of ligand binding in hemoglobin, so that the free energy difference between the subunit conformations, and thus the probability of each state occurring at equilibrium, is biased by the pattern of multisite phosphorylation in a given KaiC hexamer (Fig. 3.1G). In order to describe this effect mathematically, we introduce an energetic cost for the conformational interconversion of each subunit that depends on its phosphorylation state. Because each subunit has two phosphorylation sites (Ser431 and Thr432), it can exist as 4 possible phosphoforms. This introduces 4 unknown thermodynamic parameters to our model, $\Delta G_U, \Delta G_{pS}, \Delta G_{pT}, \Delta G_{pSpT}$ .Under these assumptions, the total free energy difference between the two allosteric states of a hexamer is simply a linear combination $\Delta G_{\text{hexamer}} = \sum_{i=1}^{6} \Delta G_i$. In this

model, the ultrasensitivity in KaiB interaction (Fig. 3.1F) arises from the exponential dependence of the equilibrium occupancy of each allosteric state on hexamer phosphorylation. We now proceed to test the validity of this allosteric framework, and place constraints on the free energies associated with the phosphorylation state of a KaiC subunit.

Binding to KaiB is allosterically incompatible with stimulation by KaiA

Given our data showing that mixtures of KaiC phosphorylation states regulate the ability of a KaiC hexamer to interact with KaiB, we speculated that the ability of KaiA to stimulate KaiC might also depend on the composite phosphorylation state of an entire hexamer. To examine the influence of phosphorylation on the sensitivity of KaiC to KaiA, we prepared wild-type KaiC in different initial phosphorylation states, then added various concentrations of KaiA and measured initial rates of phosphorylation for the unphosphorylated KaiC molecules (Fig. 3.2A-C, S3 in (*46*)). In all cases, the effective Michaelis constant for KaiA-stimulated autophosphorylation increased with increasing phosphorylation on Ser431, and is more than a factor of 4 higher when KaiC is heavily phosphorylated on Ser431 (Fig. 3.2D, S3 in (*46*)).

To isolate the allosteric effect of pSer431 on the function of a KaiC hexamer, we then measured the ability of KaiA to drive phosphorylation of unphosphorylated wildtype KaiC in the presence of varying amounts of the pSer431 phosphomimetic mutant. We observed a dose-dependent increase in the effective Michaelis constant for KaiA acting on KaiC, similar in magnitude to the effects we observed with differentially phosphorylated wildtype protein. Importantly, these effects are only present when the pSer431 mimic is mixed into the wildtype hexamers and not when it is present as separate hexamers (Fig. 3.2E, S4 in (*46*), S5 in (*46*)). This mixing-dependent effect indicates that phosphorylation on Ser431 acts allosterically in the KaiC

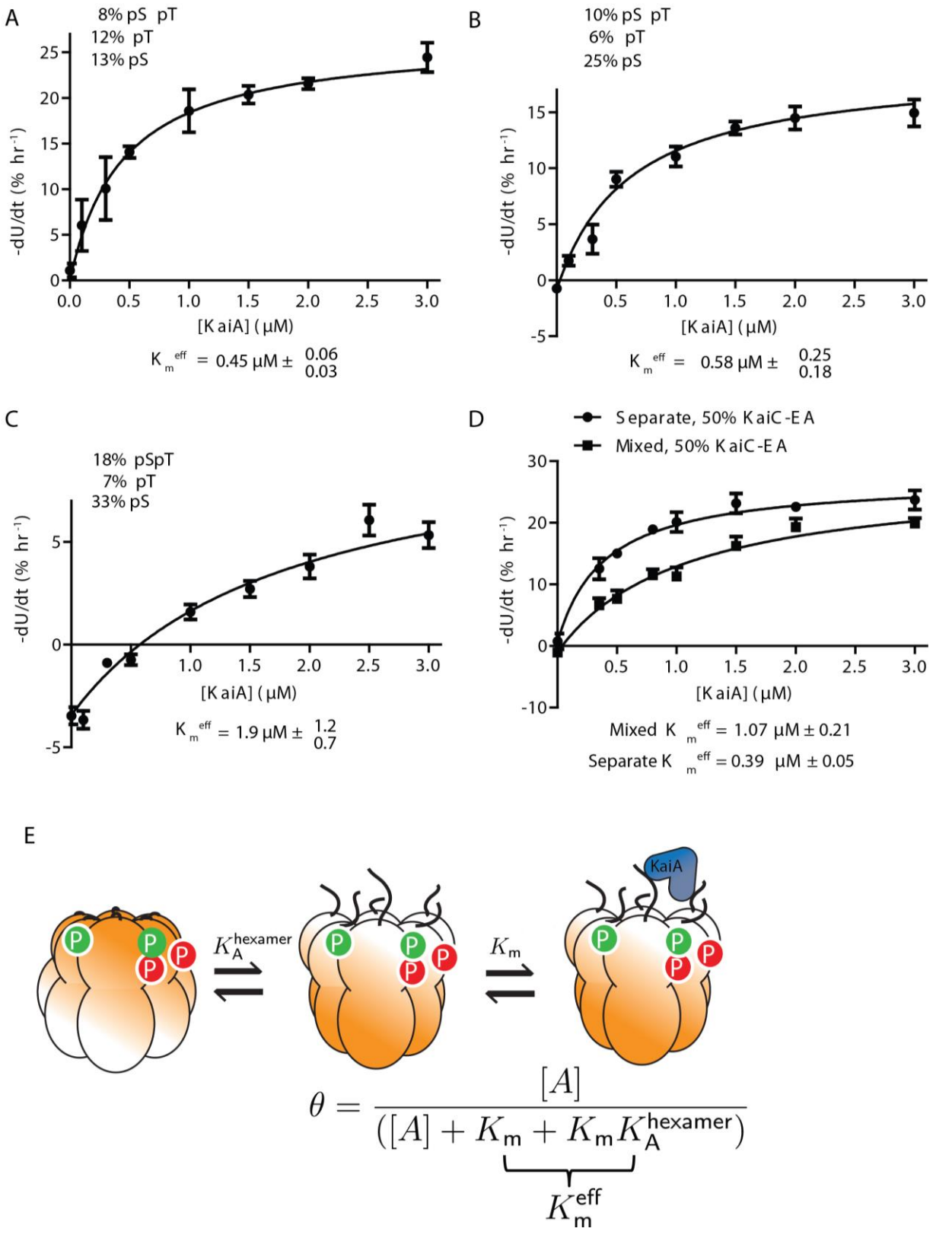

Fig. 3.2. KaiC hexamers with heavy Ser431 phosphorylation are less sensitive to KaiA.

(Fig. 3.2 continued) (A)-(C) Rates of KaiA-stimulated KaiC autokinase activity as a function of [KaiA] for various initial phosphorylation states. Fits are to a modified Michaelis-Menten equation with baseline: Vi = Vmax [KaiA]/([KaiA] + Kmeff) + Vdephos, to account for KaiA independent dephosphorylation. (D) Same as in (A)-(C) with 1:1 KaiC-EA and dephosphorylated wildtype KaiC (total = 3.5μM), either as mixed hexamers or separate hexamers. (E) Cartoon of KaiC kinase activation by KaiA in the allosteric framework: KaiA selectively binds and activates KaiC hexamers in the non-KaiB binding state. Changes in the phosphorylation state of a hexamer changes KA, the allosteric equilibrium constant, and hence also Kmeff = Km (1 + KA).

hexamer to lower the sensitivity of the other subunits to KaiA. These results are consistent with recent observations that high concentrations of KaiA are needed to sustain KaiC phosphorylation (*99*) and that phosphomimetic mutation at Ser431 makes the KaiA-binding "A loops" inaccessible to proteolytic cleavage (*100*).

Since phosphorylation on Ser431 promotes an allosteric transition towards KaiB binding, the increase in K_m^{eff} associated with higher Ser431 phosphorylation levels suggests that KaiA selectively binds and activates KaiC in an allosteric state that KaiB cannot bind. In other words, activation by KaiA is incompatible with the state of KaiC that triggers KaiB binding. We can mathematically describe this effect using a quasi-steady state approximation valid in the limit that both interconversion between the allosteric states of KaiC and interaction with KaiA occur much faster than changes in phosphorylation. Then the probability of a hexamer being activated for autophosphorylation by KaiA is: $\frac{[\text{KaiA}]}{[\text{KaiA}]+ K_m(1+K_A(pS,pT,pSpT))}$ where K_A is a phosphorylation-dependent allostery constant (see Supporting Appendix for derivation). Consistent with the data, this describes a Michaelis-Menten-like dependence of the autokinase rate on [KaiA] starting from a given phosphorylation state, and the higher effective Michaelis constant $K_m(1+K_A)$ results from higher Ser431 phosphorylation which increases K_A (Fig. 3.2F).

This model further predicts that, because KaiA is stabilizing the kinase-active state, sufficient stimulation by KaiA should shift the allosteric equilibrium away from the state that can

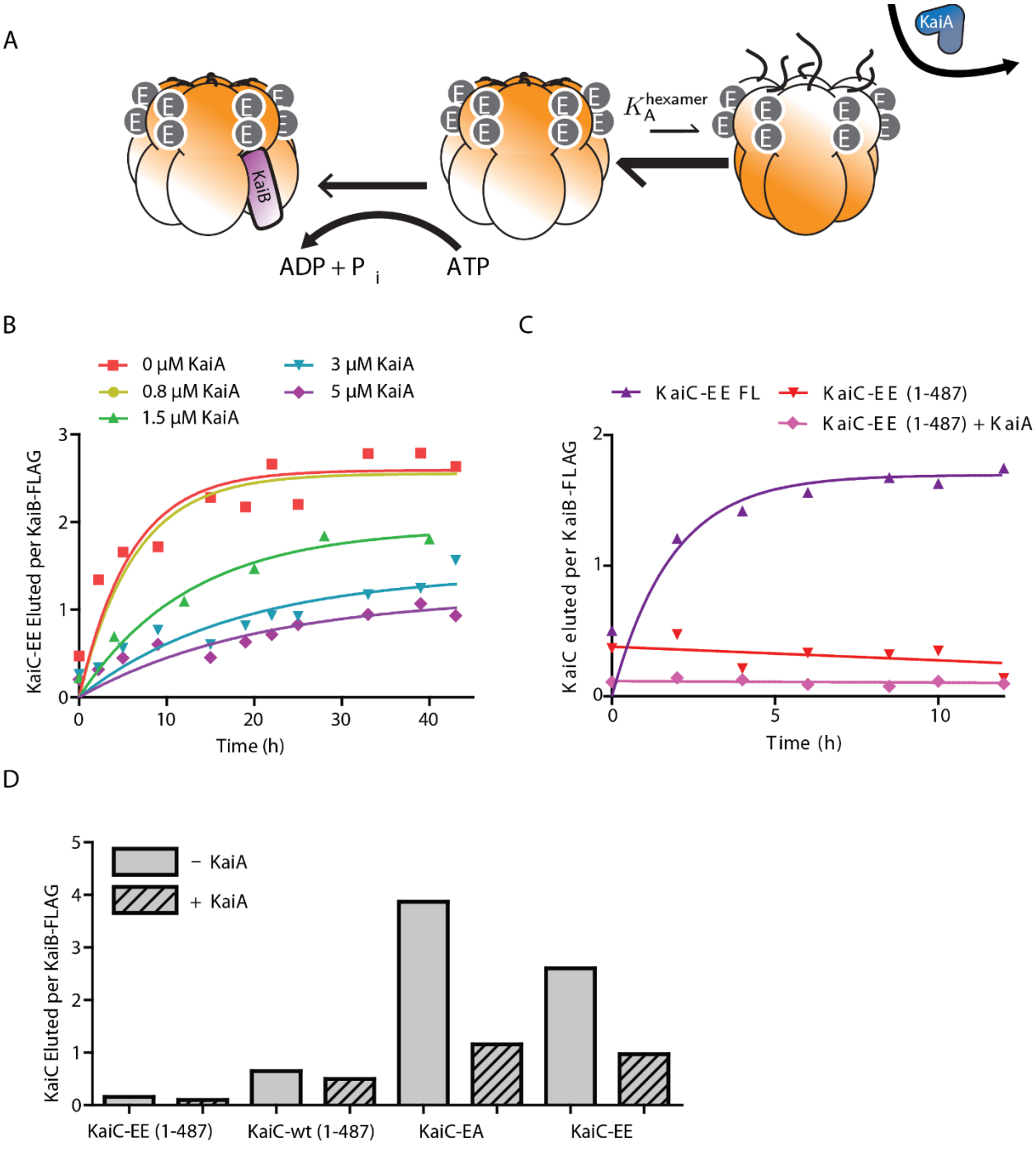

Fig. 3.3. KaiA allosterically stabilizes a KaiC state that KaiB cannot bind.

(Fig. 3.3. continued) (A) Cartoon of allosteric state selection by KaiA and KaiB. Stimulation of the kinase active mode by KaiA shifts the allosteric equilibrium away from KaiB binding. (**B**) Timecourse of KaiC-EE co-IPed by KaiB-FLAG in the presence of various concentrations of KaiA. Normalized co-IP amounts were calculated as the ratio of gel densitometry measurements of KaiC-EE to KaiB-FLAG in the eluate. Fits (*solid lines*) are to a first order exponential. (**C**) Timecourse of KaiB interaction assessed by co-IP with KaiB-FLAG for either full-length KaiC-EE (KaiC-EE FL) or a mutant (KaiC-EE 1-487) that mimics the KaiA-activated state, in the presence or absence of 1.5 μM KaiA. (**D**) Normalized co-IP of KaiC by KaiB-FLAG at 24 h for various KaiC mutants, with or without KaiA. The concentration of KaiA is 5 μM for the KaiC-EE case and 1.5 μM for all other reactions.

bind KaiB even when the phosphorylation state is held fixed, causing KaiC to resist interaction with KaiB (Fig. 3.3A). To test this prediction, we used a mimic of the doubly phosphorylated form, KaiC S431E;T432E (KaiC-EE), and measured kinetics of the formation of KaiB•KaiC complexes in the presence of various amounts of KaiA. Despite the fact that kinase activation by KaiA cannot alter the phosphorylation state of these mutant residues, we found that high concentrations of KaiA could disrupt the interaction with KaiB, consistent with a model where KaiA is stabilizing an allosteric state of KaiC incompatible with KaiB binding (Fig. 3.3B). The very slow (longer than a day) kinetics of binding that result from the antagonistic effect of KaiA on this mutant are likely related to the long period transcriptional oscillations that have been reported in the KaiC-EE mutant strain (*80*). Similar results held for the KaiC-EA mutant (Fig. S6 in in (*46*)).

To investigate the structural basis of this effect, we deleted the C-terminal tail of KaiC, a manipulation that mimics hyperactivation by KaiA and permanently locks the enzyme into the kinase mode (*69*). As predicted, this mutation causes severe defects in KaiB interaction (Fig. 3.3C-D). The extent to which KaiB binding is disrupted is correlated with KaiA's ability to stimulate KaiC kinase activity: a CII domain catalytically impaired mutant (E318Q) mutation allows KaiC-EE to bind to KaiB even in the presence of KaiA and an N-terminal deletion

produces hyperactive KaiA that can inhibit the KaiB-KaiC interaction at a lower dosage than wildtype KaiA (Fig. S6 in (*46*)).

The mutations made here and the known KaiA binding site on KaiC are distant from proposed KaiB-KaiC interaction sites (*93, 101*). Manipulating KaiC kinase activity either mutationally or by increasing the KaiA concentration, affects the strength of KaiB-KaiC interaction. We thus interpret our results as indicating an allosteric conflict between the action of KaiA and KaiB-KaiC binding. However, our data cannot exclude the possibility of an unknown mode of interaction where KaiA might physically occlude a KaiB binding site.

Allosteric models constrained by experimental data can reproduce circadian rhythms that adapt to altered protein concentrations

Taken together, our experimental data indicate that a role of the KaiC phosphorylation sites is to regulate an allosteric transition in the KaiC hexamer that permits KaiB binding. A simpler alternative scenario is that the phosphorylation sites on each KaiC subunit independently present a binding surface for KaiB, as in some previously studied mathematical models (*37*). To analyze the consequences of these two possible scenarios and gain insight into the role of each phosphorylation site, we constructed two mathematical models: an allosteric model where the ability of KaiC to interact with KaiB and KaiA is determined by an allostery constant set by the phosphorylation state of a given hexamer (Fig. 3.4A), and an "independent subunits" model where each KaiC subunit can interact independently with KaiB with a binding affinity that depends on its phosphorylation state.

In both models, KaiA stimulates KaiC phosphorylation which occurs in an uncoordinated fashion once a hexamer is activated. KaiA is subsequently inhibited globally by sequestration

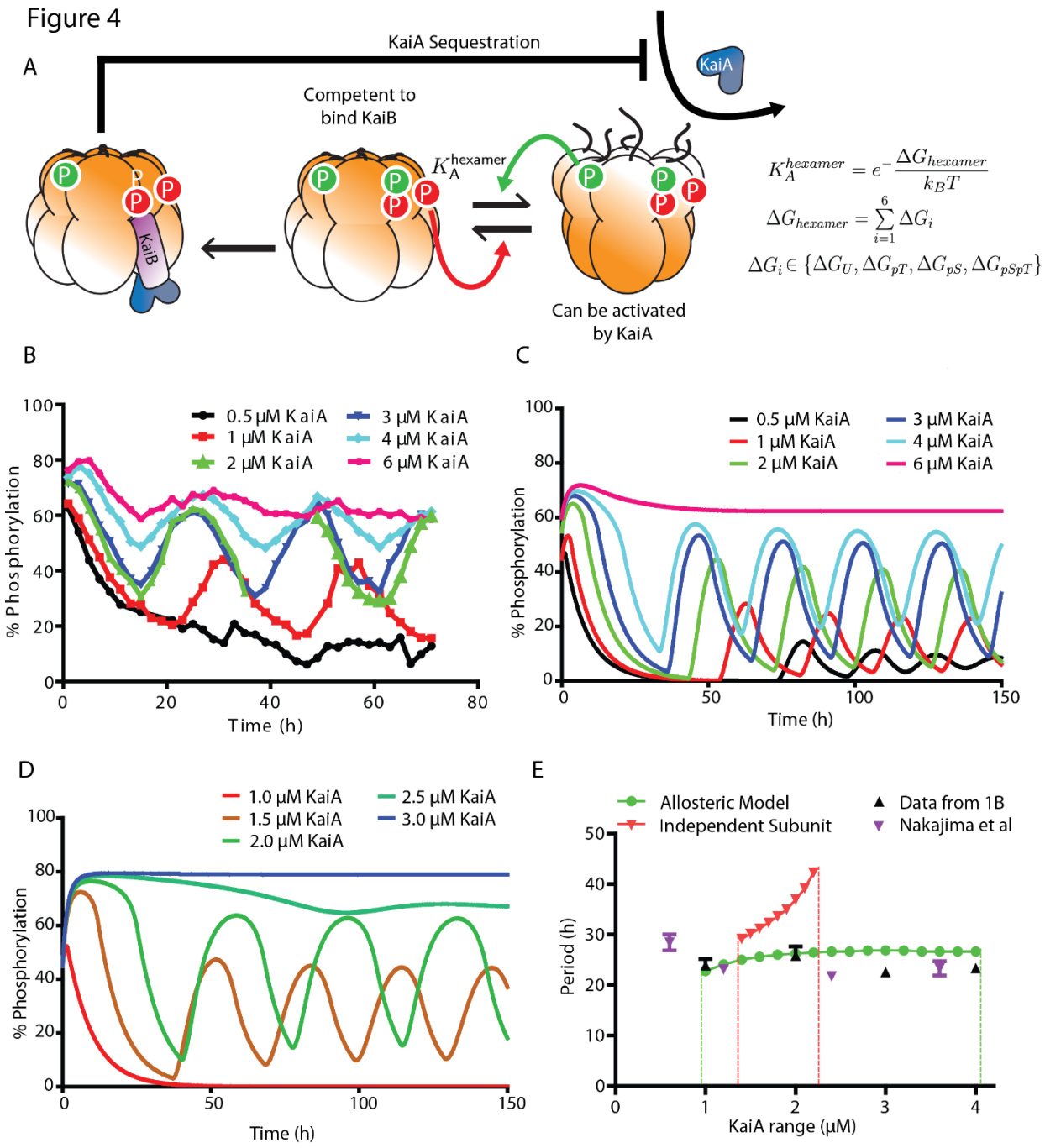

Fig. 3.4. The allosteric model predicts the experimentally observed robustness of the oscillator period to changes in component concentrations.

(Fig. 3.4, continued) (**A**) Allosteric multisite phosphorylation clock model: KaiC hexamers switch in a concerted fashion between a state competent for KaiB binding and a state that can be activated by KaiA. The probability of a given hexamer occupying either state is determined by the thermodynamic equilibrium set by a linear combination of phosphorylation-dependent subunit free energies. Phosphorylation on Thr432 (*green*) and Ser431 (*red*) have opposing effects on the allosteric transition. Kinase activation by KaiA allosterically stabilizes the non-KaiB binding state. Both allosteric transitions and interaction with KaiA are at quasi-steady state relative to the slow phosphorylation changes and CI ATPase-mediated KaiB binding reactions. KaiB•KaiC complexes sequester KaiA to drive a global negative feedback loop on KaiA-dependent phosphorylation. (**B**) Experimental timecourse of KaiC phosphorylation in purified clock reactions with various concentrations of KaiA. (**C**) Simulated reactions in the optimized allosteric model at various concentrations of KaiA. (**D**) Simulated reactions in an optimized model where KaiC subunits interact with KaiB independently, at various concentrations of KaiA. (**E**) Experimental oscillator period estimated from data from this study (*black triangles*) or data from Nakajima *et al*. (2010) (*purple triangles*) compared to the optimized allosteric multisite phosphorylation model (*green curve*) and the optimized independent subunits model (*red curve*). Dashed lines estimate the boundaries where stable oscillations fail. Error bars on the experimental period indicate fitting error from least-squares regression to a sinusoid.

into KaiA•KaiB•KaiC complexes at a stoichiometry of one KaiA dimer to one KaiC subunit, consistent with recent structural work (*102*). The rate constants for phosphorylation and dephosphorylation on each site and for the slow, CI ATPase-mediated assembly of KaiB•KaiC are constrained by experimental kinetic studies (*37*, *47*). Thus, our allosteric oscillator model shares features with previous treatments of allostery in the KaiABC system (*92*), but now explicitly includes distinct roles for the two KaiC phosphorylation sites and uses experimentally derived kinetic parameters to allow us to make direct comparisons with experimental data.

Because we do not have direct measurements of the influence of each KaiC subunit's phosphorylation state on KaiB binding, we initially left the models agnostic about the influence of phosphorylation on protein-protein interaction. In the allosteric model, this is represented by unknown free energy contributions, $\Delta G_U, \Delta G_{pS}, \Delta G_{pT}, \Delta G_{pSpT}$, to the allostery constant. In the independent subunits model, these are replaced by phosphorylation-dependent binding constants for KaiB binding to each subunit.

To compare the two models, we randomly sampled parameter space for these unknown parameters and attempted to optimize each model for its ability to simulate oscillations measured

in the experimental system with a period near 24 hours over a wide range of KaiA concentrations, a feature of the system that has been difficult for modeling studies to correctly describe (*78*) (Fig. 3.4B). With appropriate thermodynamic parameters, the allosteric model qualitatively recapitulates the tolerance of the system to varying protein concentrations, including the increase in the abundance of specific phosphoforms as [KaiA] increases (Fig. 3.4C, S7 in (*46*)).

Remarkably, the range of protein concentrations over which this model can generate circadian rhythms is nearly the same as the experimental system (Fig. 3.4D). The role of KaiA in stabilizing the allosteric state that cannot bind KaiB helps to enhance the robustness of the period in this model (Fig. S8 in (*46*)). In contrast, the independent subunits model is only able to generate oscillations over a narrow range of conditions, and the period of that model is much more sensitive to KaiA concentration than the experimental system (Fig. 3.4E). These conclusions still hold when the rate constants in the two models are randomly varied near the best-fit values, indicating that the improved robustness of the allosteric model is a property of the fundamentally different role for the phosphorylation sites in that model, rather than a consequence of a particular choice of kinetic parameters (Fig. S9 in (*46*)). We conclude that models that describe subunit phosphorylation as mediating a concerted allosteric transition in the KaiC hexamer are much more successful at recapitulating experimentally observed circadian rhythms than models without these mechanisms.

Robust timing requires that the two phosphorylation sites have opposing effects, creating an ultrasensitive switch in KaiC activity

We then asked if there were common features of the parameter sets in the allosteric model that successfully generated circadian oscillations over the range of protein concentrations observed experimentally. First we analyzed the values of the free energy parameters from our search that produced stable oscillations with a standard deviation in the resulting oscillator period of < 10% over a ~3 μM range of KaiA concentrations, as seen in the experimental system. Strikingly, these results predict that for robust oscillations, pSer431 must always favor KaiB interaction $(\Delta G_{pS} < 0)$ and pThr432 must always oppose it $(\Delta G_{pT} > 0)$. While being opposite in sign, these two energetic parameters have the largest magnitudes, hence changes in the balance of pSer431-only and pThr432-only subunits, as occurs when the oscillator shifts from the phosphorylation phase to the dephosphorylation phase, most critically determine the KaiB-

binding capacity of the system (Fig. 3.5A). These findings parallel the enrichment and depletion

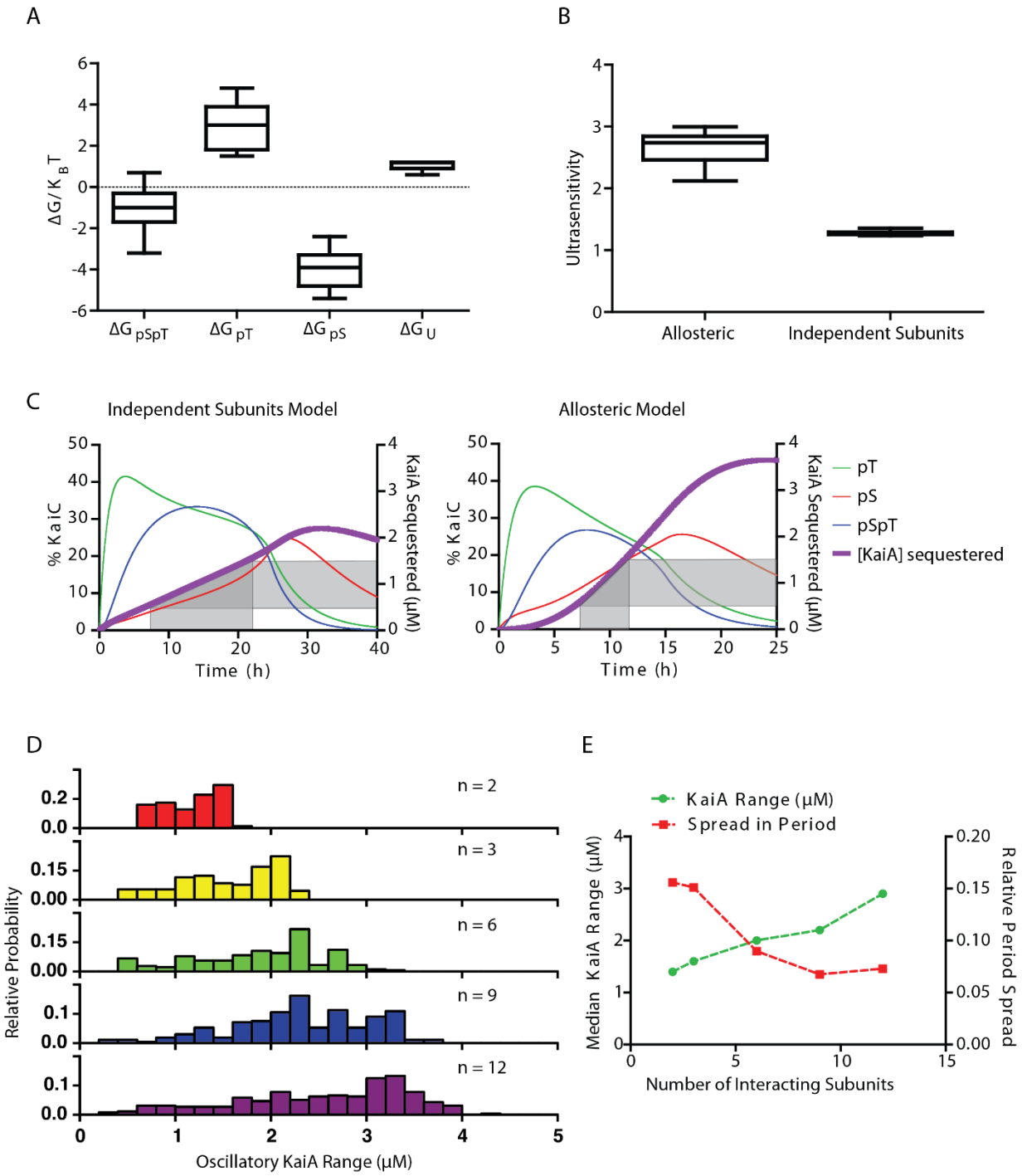

Fig. 3.5. Opposing effects of pSer431 and pThr432 on the allosteric equilibrium produces an ultrasensitive switch in negative feedback necessary for a robust period.

(Fig. 3.5, continued) (**A**) Box and whisker plot of the free energy (ΔG) distribution associated with each subunit phosphorylation state for parameter sets that produce oscillations over the experimental range of KaiA concentrations with < 10% standard deviation in period and a circadian period (22-29 h) at 1.5 μM KaiA. (**B**) Distribution of the effective Hill coefficient (measure of ultrasensitivity) describing the sigmoidal increase in KaiB•KaiC complexes over time in simulated clock reactions in either the allosteric multisite model or the independent subunits model. Same criteria as described in (A), except the requirement on the standard deviation of the period was relaxed for the independent subunits model. (**C**) Representative time courses of simulated KaiC phosphorylation and the amount of KaiA sequestered in the allosteric multisite model (*left*) and the independent subunits model. Shaded regions show the spread in time delays required to achieve inactivation of a three-fold range of KaiA concentrations in both models. (**D**) Distribution of the range of KaiA concentrations that produce stable oscillations from 10,000 randomly sampled free energy parameters for allosteric clock models with varying numbers (n) of allosterically linked subunits. (**E**) Dependence of the median standard deviation of the oscillator period (*red squares*) and the median range of KaiA concentrations that support stable oscillations (*green diamonds*) on the number of allosterically linked subunits.

for pSer431 and pThr432 respectively that we observed experimentally for KaiC hexamers interacting with KaiB (*cf.* Fig 3.1B). Further, the median free energy parameters for pThr432 and pSer431 that allow the model to meet these criteria have a magnitude on the order of k_BT, the energy scale of thermal fluctuations, implying that changing a single subunit's phosphorylation state has a large but not overwhelming effect on the allosteric state of the KaiC hexamer.

Why do these parameter sets allow the model to work well? We reasoned that the ordered phosphorylation of Ser431 and Thr432 and their opposing effects on KaiC conformation could cause an effectively ultrasensitive switch from a KaiA-activated state to a KaiB-binding competent state as the degree of the phosphorylation within a hexamer is increased (*cf.* Fig 3.1B). The threshold in this switch is crossed after a specific time due to kinetic ordering of phosphorylation in KaiC: because Thr432 phosphorylation occurs first, KaiB interaction is initially inhibited, and this inhibition is only overcome at late times when Ser431 phosphorylation has risen enough to cancel out the effect of pThr432. To quantify this effect we determined how the amount of KaiB•KaiC complexes changes in the models as KaiC phosphorylation increases over time by fitting a simulated time course of KaiB•KaiC complex

formation to a Hill function: $1/(1+\left(\frac{K}{t}\right)^{n_H})$, where n_H is a Hill coefficient that quantifies the sigmoidal, switch-like character to the kinetics, and $n_H > 1$ indicates ultrasensitivity. The parameter sets that allow circadian rhythms over a wide range of KaiA concentrations all show a Hill coefficient of at least 2 (Fig. 3.5B).

We then sought to understand in qualitative terms why an ultrasensitive dependence of the KaiB interaction on phosphorylation state can allow the oscillator to function properly over a wide range of conditions. We compared the optimally tuned allosteric model and the independent subunits model by simulating a time course of phosphorylation when an oscillator reaction is first initiated from the unphosphorylated state. On these plots we overlaid the capacity of the pool of hexamers to inhibit various amounts of KaiA by forming KaiB•KaiC complexes (Fig. 3.5C). In the allosteric model, ultrasensitivity from opposition between the phosphorylation sites allows the inhibitory strength of the reaction against KaiA to rise sharply after an initial lag. The result is that for different amounts of KaiA, the onset of inhibition happens at a similar time and the timing of the oscillation is therefore similar (Fig. S10 in (*46*)). For the independent subunits model, the reaction requires substantially different delays to inhibit different amounts of KaiA, resulting in a period of oscillation that changes markedly as [KaiA] increases, and complete failure of rhythms outside of a narrow range of conditions.

The success of the allosteric model comes from a cooperative mechanism where the phosphorylation states of the six subunits in a hexamer are weighed together to compute a single functional output manifested as the KaiB-binding state of the entire ring. Because of the importance of having all six interacting subunits linked as a concerted allosteric unit, we asked how the model would perform if the number of subunits that could interact allosterically was altered. We found that the range of KaiA concentrations over which the system shows

oscillations grows rapidly as the number of subunits participating in the allosteric switch increases (Fig. 3.5D, S10 in (*46*)), while the spread in period seen over this KaiA range decreases (Fig. 3.5E, S10 in (*46*)). Both of these effects are correlated with increased ultrasensitivity of KaiB-KaiC complex formation (Fig. S10 in (*46*)). Once hexameric interactions are included, the model has the potential to reach the full oscillatory range seen in the experimental data with minimal spread in period. The oscillator can function over an even wider range of KaiA concentrations if even higher order (unrealistic) oligomeric interactions are present in the model, underscoring the importance of coupling between many oligomeric subunits for this mechanism (Fig. 3.5D-E, S10 in (*46*)).

Discussion

The simplicity of the purified KaiABC oscillator makes it a remarkably powerful model system to investigate the mechanistic origins of circadian rhythms and to study the robustness of biochemical circuits generally. We wanted to understand which biochemical features of the proteins are crucial for generating oscillations with a precisely defined period, with the goal of producing a mechanistic mathematical model that can account for the behavior of the purified components.

The key negative feedback process that allows sustained oscillation in this system is the sequestration of the activator KaiA into inactive KaiB-dependent complexes. This kind of molecular titration is widely used throughout biology including control of morphogens in development (*103*), regulation of transcription through sigma/anti-sigma interactions (*104*), and microRNA-mRNA buffering (*105*). The dynamics of stochiometric titration mechanisms are typically quite sensitive to the relative concentrations of the components involved, which is why

it has been argued that, when they are used in timing systems, tight controls must be placed on gene expression (*106*).

Our modeling-based analysis of the KaiABC system shows that one way to make the dynamics of such a negative feedback loop less dependent on component concentration is to make the ability of molecules that can participate in sequestering the activator an ultrasensitive function of their activation state. In other words, if the system abruptly switches from very little sequestration to its full capacity, timing can be maintained precisely even if the activator concentration is not tightly controlled, extending the range of conditions over which a biochemical circuit can function.

We found that this ultrasensitivity can only be realized in the hexameric architecture of KaiC when the two phosphorylation sites, Ser431 and Thr432, oppose each other's influence. Effectively, each KaiC hexamer acts as a comparator, switching its state when modification on Ser431 outweighs modification on Thr432. This helps to explain the role of Thr432 in the clock; because Thr432 is quickly phosphorylated after a hexamer is stimulated by KaiA and then favors an allosteric state of KaiC that can be further activated, it effectively forms a fast positive feedback loop on KaiA activity. Subsequent Ser431 phosphorylation then acts as a dominant slow negative feedback loop. This fast positive-slow negative network motif is a common means of generating oscillations (*107*).

KaiC is related to the AAA+ ATPases, many of which exhibit strong long-range allosteric communication effects, both within a ring when subunits change *e.g.* their nucleotide-bound state, and through ring-ring stacking interactions (*108*). We propose that KaiC has evolved to make use of these communication mechanisms to ensure precise timing: antagonism between differentially phosphorylated subunits within the CII ring to precisely define timing and

then ring-ring stacking interactions to transduce this signal to the CI ring to allow KaiB-dependent feedback. The result is that while the strength of negative feedback can adapt dynamically to accommodate changes in protein concentration, the timing of the response is precise.

Because of the ease of using opposed post-translational modifications to drive an allosteric switch, we suspect that other biological timing circuits may have analogous mechanisms to achieve precise timing. In eukaryotic clock systems, the analysis is complicated by the presence of many phosphorylation sites. However, it appears that the FRQ protein in *N. crassa* has distinct clusters of phosphorylation sites which have opposing effects on the clock period when mutated (*109*). Allosteric response of protein structure to multisite post-translational modification may allow clock proteins to cooperatively communicate the opposing effects of phosphorylation sites throughout the protein, effecting an ultrasensitive switch in activity, a key mechanism for precise timing that the cyanobacteria have implemented via the KaiC hexameric ring structure.

Materials and Methods

Protein purification and in vitro protein reactions

All proteins were recombinantly expressed and purified from *E. coli*, and protein reactions were prepared as previously described (*47*). Unless otherwise specified, all reactions were performed using 3.5 μM KaiB and 3.5 μM KaiC at 30 °C in a reaction buffer containing 10% glycerol, 150 mM NaCl, 20 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 50 μM EDTA, 5 mM ATP. For full details, see the Supporting Appendix.

Measuring KaiB-KaiC Interaction

Clock reactions at varying KaiA concentrations (1-4 μM), 3.5 μM KaiC and 3.5 μM KaiB-FLAG were first pre-incubated for 16 hours to allow the initial transient behavior to decay. Samples were then taken every 4 hours over a 24 hour cycle by flash freezing in liquid nitrogen. For the no KaiA condition, KaiC was first hyperphosphorylated using HA-tagged KaiA which was removed by immunoprecipitation prior to adding KaiB-FLAG (*37*). The input, supernatant, and eluate samples from the immunoprecipitation were analyzed by SDS-PAGE electrophoresis to resolve each of the four KaiC subunit species. Percentages of each subunit phosphorylation state relative to the total KaiC loaded per lane were extracted by densitometry of the scanned SDS-PAGE gels. We propagated an estimated 2% absolute error in our gel densitometry measurements through the calculation of these enrichment ratios and then excluded points from the final average with a relative error > 1.0. For full details, see the Supporting Appendix.

Reactions with artificially mixed KaiC hexamers

Monomerization of KaiC was carried out following Nishiwaki et al. (*86*). Briefly, KaiC was buffer exchanged into a buffer with 0.5 mM ADP and incubated at 4 °C to disrupt hexamer structure. To prepare "mixed" hexamers, monomerized KaiC mutants or wildtype were mixed 1:1 prior to rehexamerization via the addition of ATP, as detailed in the Supporting Appendix. To prepare "separate" hexamers, KaiC mutants or wildtype were first rehexamerized separately and then combined.

Standard clock reactions were prepared with 3.5 μM total of the phosphomimetic and wildtype KaiC preparations. Identical reactions with KaiB-FLAG in place of KaiB were sampled every 2-6 hours to assay KaiB•KaiC interaction by anti-FLAG immunoprecipitation. To assess

the extent of mixing in these experiments, one of the KaiC mutants or wildtype carried an N-terminal His_6 tag. Mixing was determined as the extent to which untagged KaiC could be coprecipitated with tagged KaiC. Detailed protocols for His-tag pulldowns and anti-FLAG immunoprecipitation can be found in the Supporting Appendix.

Michaelis constant (K_m) determination for KaiA acting on KaiC

KaiC with various levels of phosphorylation was produced by dephosphorylation of hyperphosphorylated KaiC for 33, 9.5, or 4 hours at 30°C (Fig. 2A-C, respectively). Various concentrations of KaiA were then reintroduced to a 3.5 μM solution of these partially dephosphorylated KaiC samples. Subsequent KaiC phosphorylation was analyzed by SDS-PAGE. Initial rates of change of unphosphorylated KaiC were determined by linear regression to the early portions of the time course (Fig. S3). Initial rates were plotted with respect to KaiA concentration and fit to a Michaelis-Menten function with baseline ($V_i = V_{\text{dephos}} + V_{\max}[\text{KaiA}]/([\text{KaiA}] + K_m^{\text{eff}})$ to determine the effective Michaelis constant, or K_m^{eff}. Upper and lower error bounds on K_m^{eff} were determined by the distribution of fits using bootstrapped datasets.

To determine the effective K_m for KaiC mixed with KaiC-EA, wildtype His_6-KaiC was first fully dephosphorylated by incubation at 30 °C for 36 h, and then prepared as "separate" or "mixed" hexamers with KaiC-EA at various molar ratios. The preparations were then diluted to 3.5μM, mixed with various concentrations of KaiA, and then assayed as described above. We assayed degree of mixing by His_6-tag copreciptation. Complete details are in the Supporting Appendix.

Mathematical modeling of the KaiABC oscillator

A detailed derivation and analysis of the allosteric model and the independent subunits model can be found in the Supporting Appendix. Differential equations governing the rate of change for each possible KaiC hexamer phosphorylation state were numerically integrated over time using the ode45 algorithm in MATLAB. Kinetics rates for KaiB binding, KaiC phosphorylation and dephosphorylation were constrained by fits to experimental kinetics as previously reported (*37, 47*). Each KaiC subunit state (U, pT, pS, pTpS) was assigned a free energy parameter defining its influence on the equilibrium between the two allosteric hexamer states. Negative ΔG favors the KaiB-binding competent state.

Perspective

Interestingly, preliminary stochastic simulation of the hexamer model presented here (using the Gillespie algorithm) is much noisier than stochastic simulation of the original independent monomer model published by Rust in (*37*) (data not shown). Previous studies suggest that ultrasensitivity can amplify molecular noise in gene networks (*110, 111*), in which small fluctuations in an input signal can be magnified due to the sharply increasing nature of ultrasensitive switches. The major difference between the hexamer model and the independent monomer model is that the hexamer model relies on ultrasensitivity to generate sufficient non-linearity required to drive oscillations. In fact, as shown above, inclusion of more subunits per KaiC ring can increase the robustness of oscillations across different concentrations of KaiA because this increases the cooperativity of KaiC switching between a KaiA-binding state and a KaiB-binding state.

However, does the increased ultrasensitivity of a concerted hexamer model decrease the resistance to molecular noise? While a sharper switch may increase the robustness of oscillations in one dimension (e.g. the total range of KaiA concentrations that supports oscillations), it may also amplify molecular noise to a greater degree. Ultimately, this may represent an interesting generalized tradeoff that biological oscillators must contend with. Like the work presented in Chapter 2 in which a simplified clock architecture traded decreased robustness to environmental noise for increased robustness to molecular noise, the amount of ultrasensitivity present in an oscillatory system may represent a balance between generating sufficiently strong oscillations over a wide array of protein stoichiometries vs. minimizing the degree to which molecular noise is amplified. This represents an intriguing future direction that should be explored further.

Chapter 4: Costs of clock-environment misalignment in individual cyanobacterial cells

Foreword

In Chapter 2, I outlined how molecular noise can cause the cyanobacterial circadian clock to become inaccurate, making it poor at predicting the correct time of day. What are the consequences for having an incorrect clock? While it is a generally accepted idea that the clock benefits the organism in anticipating the day/night cycle, this is a hypothesis that lacks rigorous exploration of the impact of the clock on fitness. Most research in the field has instead focused on elucidating the mechanism of clock function. The only evidence thus far in cyanobacteria that demonstrates the benefit of the clock are competition experiments showing that clock period mutants or cells lacking a clock suffer from fitness defects in cyclic light/dark environments (*30, 76*). These measurements were performed in bulk cultures of mixed populations of cells, and the details remained unknown of exactly how the mismatch between the clock and environment caused these defects in fitness. Therefore, the work presented in this chapter seeks to investigate at the single cell level how failure to correctly predict the time of day impacts cellular fitness. Here, I present the work published in (*5*) on which I am a second author, with a "Perspective" afterwards that discusses the relevance to my work in Chapter 2. All supplemental materials may be found in (*5*). The experiments and fitness model were designed by Guillaume Lambert, and both Guillaume Lambert and I performed the experiments. Specifically, I performed the experiments in which *kaiBC*-null cells and *kaiBC*-overexpression cells were subjected to dark pulses, after which the growth arrest probability was determined for each strain (results shown in Figure 4.3.C). Data analysis was performed by Guillaume Lambert.

Abstract

Circadian rhythms are endogenously generated daily oscillations in physiology found in all kingdoms of life. Experimental studies have shown that the fitness of Synechococcus elongatus, a photosynthetic microorganism, is severely affected in non-24h environments. However, it has been difficult to study the effects of clock-environment mismatch on cellular physiology because such measurements require the precise determination of both clock state and growth rates in the same cell. Here, we designed a microscopy platform that allows us to expose cyanobacterial cells to pulses of light and dark while quantitatively measuring their growth, division rate, and circadian clock state over many days. Our measurements reveal that decreased fitness can result from a catastrophic growth arrest caused by unexpected darkness in a small subset of cells with incorrect clock times corresponding to the subjective morning. We find that the clock generates rhythms in the instantaneous growth rate of the cell, and that time of darkness vulnerability coincides with the time of most rapid growth. Thus, the clock mediates a fundamental trade-off between growth and starvation tolerance in cycling environments. By measuring the response of the circadian rhythm to dark pulses of varying lengths, we constrain a mathematical model of a population's fitness under arbitrary light/dark schedules. This model predicts that the circadian clock is only advantageous in highly regular cycling environments with frequencies sufficiently close to the natural frequency of the clock.

Introduction

Synechococcus elongatus PCC 7942 (S. elongatus) is a photosynthetic, unicellular cyanobacterium that has been extensively used as a model system for the study of circadian rhythms (*76, 112*). Each cell contains a remarkably precise oscillator based on the kai genes

(*113*). KaiA, KaiB, and KaiC work together to generate near-24 hour rhythms in the phosphorylation of the core clock protein KaiC, forming a biochemical oscillator that can be reconstituted in vitro (*37, 114*). In the cell, rhythmic changes in KaiC signal through histidine kinases to exert genome-wide control of transcription (*58, 115, 116*) and metabolism (*117, 118*).

Much is known about the behavior of this system under conditions of constant illumination, where robust cell-autonomous oscillations are easiest to observe (*25, 40, 52, 119, 120*). However, under constant conditions, S. elongatus can grow robustly even without a functioning clock (*30, 120*), leading us to suspect that the importance of the clock would be revealed by monitoring cellular physiology under conditions that fluctuate between light and dark. Landmark work by the Johnson lab established that fitness defects occur in fluctuating environments with schedules that do not match the circadian clock period, but the underlying mechanisms for these effects are still unclear (*30, 76*). Because environmental challenges may reveal heterogeneous behavior in a population, we designed a microscopy system that allows us to quantitatively measure clock state, growth rate, and cell division in individual cyanobacterial cells over several days in an environment that fluctuates between light and dark (Fig. 4.1, Movie S1 in (*5*)). Using these single-cell measurements, we then develop a phenomenological model where growth rate and the probability of surviving the night are determined by the current clock state, which is itself updated following each light-dark transition. This model provides a framework to calculate the impact on organismal fitness from a circadian clock driven by an arbitrary fluctuating environment.

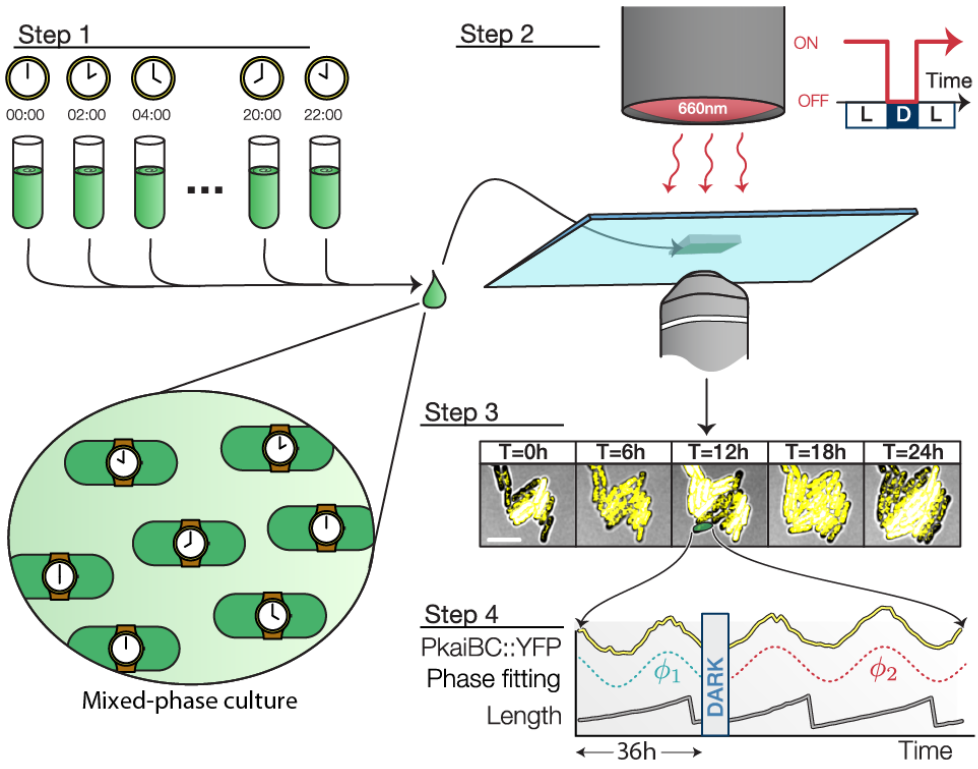

Fig. 4.1. Experimental setup. 12 populations were entrained under staggered LD 12:12 regimes and combined into a single experiment. A multiplexed measurement of phase shift or growth rate modulation was achieved by exposing the mixed-phase population to a single pulse of darkness (scale bar = 5 µm). Fluorescence and brightfield micrographs recorded every hour were used to extract every cell's physiological parameters (e.g. length, clock reporter).

Results and Discussion

A subset of cells with misaligned clocks do not survive the night

Some photosynthetic organisms that rely exclusively on light for growth are known to halt DNA synthesis and enter a dormant state (*121*), or even die (*122*), in the absence of light. Since control of gene expression in the dark and consumption of energy metabolites are both under active control of the circadian clock in cyanobacteria (*117*, *118*, *123*), we hypothesized that unanticipated nightfall at clock times when energy reserves are low and metabolic rates are high could have deleterious effects. To test this hypothesis, we exposed a mixed-phase population (Fig. S1 in (*5*)) to a period of darkness corresponding to a long night (18 h). Surprisingly, we found that a subset of cells experienced a catastrophic growth arrest after the

simulated night: growth of these cells ceased and did not resume even after 36 hours of subsequent light exposure (Figs. 4.2A-B, Movie S2 in (*5*)). Importantly, this effect required prolonged darkness—we did not observe arrested cells following 5-hour dark pulses.

To determine whether the ability to tolerate darkness-induced starvation is influenced by the circadian clock, we assigned a clock time to each cell by measuring rhythms in a fluorescent reporter of clock gene expression prior to the dark pulse. We found that the fraction of cells that failed to resume growth was strongly enriched for cells with clock states corresponding to the early day, when nightfall is not anticipated. Indeed, the probability of dark-induced growth arrest oscillates with clock time, reaching a minimum at subjective dusk when nightfall is expected to occur (Figs. 4.2C-D). Thus, the ability of individual

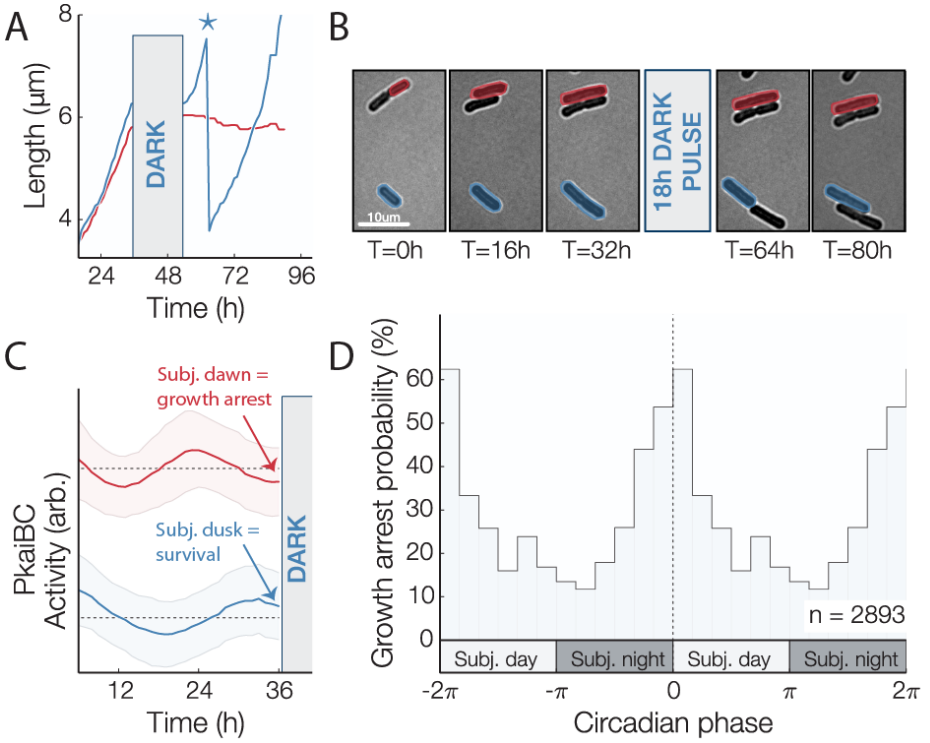

Fig. 4.2. Clock-dependent growth arrest following unexpected darkness. (A) Individual growth curves showing survival (blue) and growth arrest (red) of two neighboring cells (* = cell division). (B) Example of cells that entered a state of arrested growth following an 18 h pulse of darkness. Cells remained dormant after > 36 h in constant light. (C) Average ± std. dev. of the pre-darkness clock reporter signal for surviving (blue) and arrested (red) populations. An arbitrary vertical shift has been added to the surviving/arrested sub-populations to assist in comparison. (D) Phase-dependent probability of growth arrest for cells grown under constant light conditions for 36 h before being subjected to a 18 h pulse of darkness (n=2983). The maximal growth arrest probability occurs when a dark pulse occurs near subjective dawn (transition between subjective night and day). Growth arrest probability is double plotted to illustrate its periodicity.

S. elongatus cells to tolerate prolonged starvation is clock-dependent, with cells displaying enhanced starvation tolerance when the onset of darkness coincides with subjective dusk.

The clock allows rapid growth early in the day

In many microbes, stress tolerance is generally anticorrelated with growth rate (*124*). A classic example is the bacterial stringent response to amino acid starvation: mutants that cannot mount the stringent response can grow faster than the wildtype as nutrients are being depleted, but these mutants cannot survive conditions of prolonged starvation (*125, 126*). We therefore asked whether the rhythmic dark tolerance we observed in cyanobacteria is similarly linked to a change in growth rate during the circadian cycle.

By tracking morphological changes in single cells, we assigned an instantaneous growth rate to each cell and identified cell division events. We found that subjective dusk, the time when

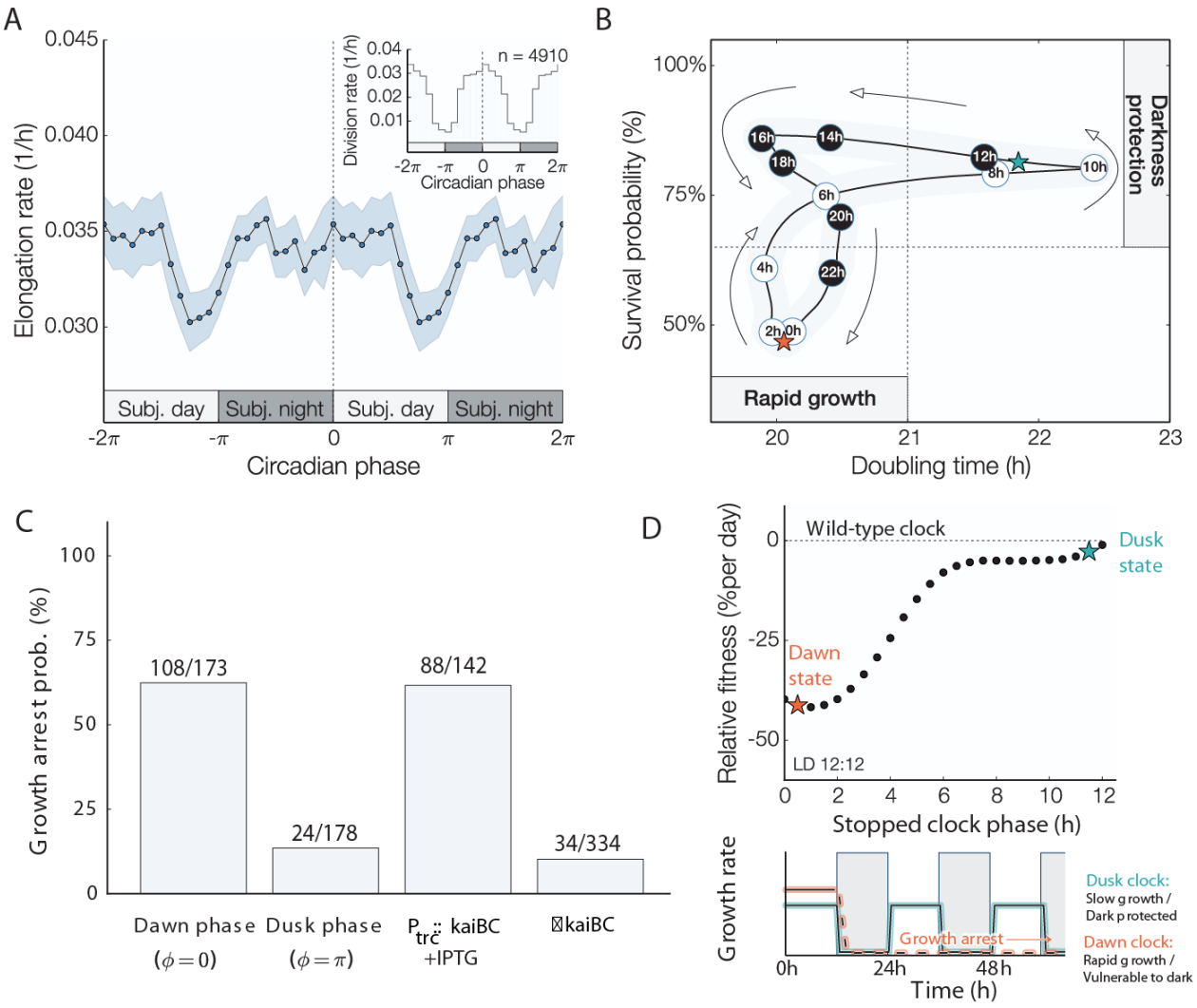

Fig. 4.3. Clock dependent fitness trade-offs. (A) Elongation rate measurements (mean ±3×s.e.m) for cells grown under constant light conditions display a transient decrease at subjective dusk. Inset: Clock-dependent division rates computed from 4910 individual division events. Elongation and division rate measurements are double plotted to illustrate their periodicity. (B) Phenomenological model of the time-evolution of the circadian fitness trade-off. The subjective circadian time is inscribed inside each datapoint to show how a cell's phenotype cycles between rapid growth and a starvation-protected state. The properties of the dusk and dawn phenotypes is marked with a cyan and orange star, respectively. (C) Growth arrest probabilities for wildtype and *kaiBC* mutant cells following an 18 h pulse of darkness. (D) Comparison between a clock stopped at dusk (cyan) or a dawn (orange). Since no fixed daytime strategy exhibits superior fitness at all times, the performance of stopped-clock daytime strategies is lower than a circadian phenotype under circadian (LD 12:12) environments. Although we measure the growth-arrest and elongation rates for all circadian times, physiological states corresponding to subjective night (circadian times between 12-24 h) were excluded from this analysis because they are unattainable during the night in cycling LD 12:12 conditions. Bottom: Cells in a dusk-like phenotypic state grow more slowly but are protected against the dark. Cells in a dawn-like phenotypic state grow more rapidly but are vulnerable to darkness.

starvation resistance is highest, is also a time of slowed biomass incorporation (Fig. 4.3A). This time of slowed growth approximately coincides with the previously reported (*73, 127*) clock-controlled inhibition of cell division (Fig. 4.3A, inset). This reduction in cell growth and division is anticorrelated in time with the vulnerability of cells to darkness, suggesting the existence of a fundamental trade-off between the capacity for rapid growth, active division, and the ability to tolerate starvation (Fig. 4.3B). Interestingly, cells early in the subjective night (i.e. circadian time between 14 – 18 h) are able to both grow rapidly *and* survive prolonged darkness. This indicates that darkness protection is not caused entirely by the instantaneous growth rate. One possible explanation is that a key determinant of darkness protection is not slow growth per se, but the accumulation of starvation-tolerance factors produced at subjective dusk in anticipation of prolonged darkness that transiently persist in the cell after rapid growth resumes. A concrete example is glycogen storage—glycogen has been shown to accumulate during the portion of the cycle when we find slowed growth (*118*), potentially protecting the cell against starvation. This proposal is an example of the general phenomenon of phenotypic memory (*128*), wherein previous adaptations are retained to confer an adaptive phenotype after the source of stress or stimulus has been removed.

To determine if these changes in dark tolerance are indeed caused by signaling from the circadian clock, we repeated these experiments using cells with either the *kaiBC* genes deleted or overexpressed under the control of an IPTG-inducible promoter. Based on previous studies, we expect deletion of *kaiBC* to result in arrhythmic high expression of dusk-expressed genes and elevated glycogen levels, mimicking a dusk-like state (*113, 118*). Conversely, we expect overexpression of *kaiBC* to cause arrhythmia while repressing dusk genes (*129*).

Consistent with the expectation that their physiology is dusk-like, we find that *kaiBC*-null cells fail to efficiently undergo cytokinesis and some cells exhibit filamentous growth under the microscope (Movie S3 in (*5*)). Further, the *kaiBC*-null mutant is quite dark-tolerant and shows a slightly lower elongation rate but a much higher survival rate, independent of the timing of darkness (Fig. 4.3C and Fig S2 in (*5*)). In contrast, *kaiBC* overexpression makes cells highly vulnerable to a light-dark transition, and the majority of these cells do not survive our dark pulse treatment (Fig. 4.3C). When grown on the microscope, *kaiBC* overexpression leads to some cell death prior to the dark pulse, and it also leads to a surprising morphological defect where the cytoplasm appears to expand at a rate that is not properly balanced by elongation, causing cells to lose their rod-like shape (Movie S4 in (*5*)).

We used these growth and survival data to calculate the expected fitness for a simulated population of cells with a constant growth rate and constant dark tolerance, according to the inverse relationship we observed for the wildtype. This calculation predicts that oscillating growth outperforms all fixed daytime growth strategies in 12h:12h light-dark cycles (Fig. 4.3D). Recent theoretical work suggests that organisms typically optimize evolutionary trade-offs by adopting a compromise phenotype that interpolates between "archetypes" that represent the extreme demands on the system (*130*). Our findings represent a dynamical version of this phenomenon where cyanobacteria are able to achieve higher fitness by cycling between incompatible states of growth and starvation protection.

Response of single cells to dark pulses is nearly all-or-none

Having characterized the impact of a pulse of prolonged darkness on clock-dependent growth, we sought to determine how the clock state in single cells is reset by pulses of darkness

in order to build a model describing how cyanobacterial cells grow in arbitrary fluctuating environments. External cues, such as dark pulses, cause the cell to reset its phase in an attempt to bring the clock into alignment with the environment (*24*, *131*). This phenomenon has been studied using bulk cultures of cyanobacteria (*118*, *132*, *133*), but such population-wide measurements may mask important features such as loss of coherence and amplitude attenuation because they are based on signals that represent the average of the oscillations coming from many independent cells.

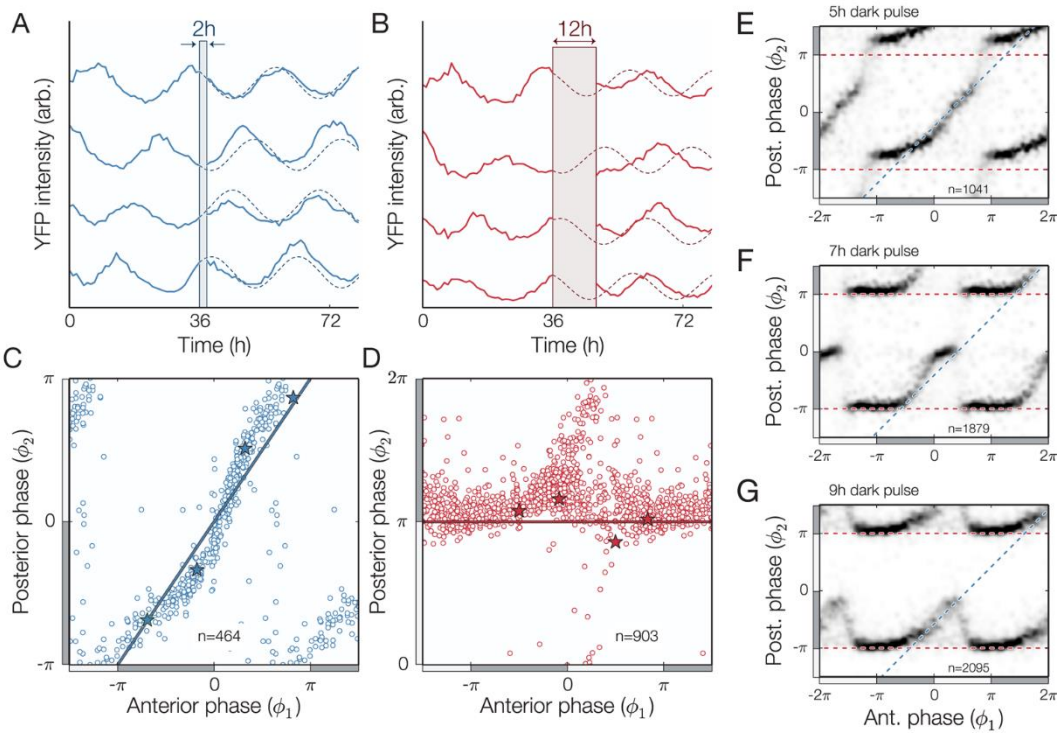

Fig. 4.4. Single cell clock response to dark pulse perturbations. (A-B) Individual traces (P_{kaiBC}::eyfp-ssrA) showing phase-shifts caused by weak (2 h dark pulse) and strong (12 h dark pulse) perturbations. (C-D) Phase resetting of individual cells grown under constant light conditions for 36 h before being subjected to a 2 h (blue) or 12 h (red) pulse of darkness, with specific examples from panels A-B are highlighted (*). Two distinct resetting behaviors are observed: a robust response (blue line in panel C) or a full phase reset (red line in panel D). (E-G) Density plot showing the phase response to 5 h, 7 h, and 9 h dark pulses, with resetting that switches between no response and full-reset (blue and red dashed lines, respectively). Cells were grown under constant light conditions for 36 h prior to each dark pulse. In all panels, the location of subjective day (night) is marked with a light (dark) gray bar. See also Figure S3 in (*5*).

We thus exposed populations of hundreds of cells in a spectrum of clock states to dark pulses of varying lengths and determined the clock phase both before (φ_1) and after (φ_2) the dark pulse. When the dark pulse was brief (2 hours) cells at all clock times were largely resistant to the perturbation, and we observed only small phase changes (Fig. 4A, C). In contrast, long dark pulses corresponding to the length of the night (12 hours) were capable of effecting nearly a full reset where most cells were synchronized to the onset of darkness (Fig. 4.4B, D).

Surprisingly, dark pulses of intermediate length produced a discontinuous combination of these responses. When the clock time was far from subjective dusk, the response of the system to a dark pulse was very small (Fig. 4.4E-G, data near the blue 1:1 lines). In a critical range of clock times, however, the response changes abruptly so that the system strongly synchronizes to the onset of darkness. The range of times when this nearly complete reset occurs grows as the dark pulse becomes longer (Fig. 4.4E-G, data near the $\varphi_2=\pi$ red lines). Biochemical studies of the Kai proteins have implicated changes in levels of metabolites during the night, such as the ATP/ADP ratio, with the ability of the circadian clock to reset its phase (*24, 134*). The change in resetting behavior we observe here for intermediate-length dark pulses may in part be caused by a timescale associated with depletion of key metabolites in the cell.

This abrupt change from insensitivity to strong sensitivity as clock time progresses may represent an optimal strategy for dealing with unexpected fluctuations in the environment, as it can ignore short dark pulses at times of day when they are unlikely (Fig. S3B in (*5*)). This behavior would be difficult to detect without high-resolution single-cell measurements, because averaging over many cells with slightly different phases would tend to blur out the true sharpness of the response (Fig. S3C in (*5*)).

Mathematical model of fitness in fluctuating environments

We combined our measurements of clock-dependent growth and darkness-induced growth arrest (Fig. 4.5A) with darkness-induced clock resetting to build a mathematical model (*135*) of cyanobacteria growing under arbitrary schedules of light and dark. We first asked how successfully the circadian clock could synchronize to a 24-hour day (12h L:12h D) if the clock period were altered. In particular, the phase resetting information obtained in Fig. 4.4E-G was used to generate the mapping for different clock periods. The relationship between φ_i and φ_{i+1} was given by the recurrence relation $\varphi_{i+1} = f_n^P(\varphi_i)$, where the map $f_n^P(\varphi_i)$ describes the phase at dusk following a dark pulse of duration *n* subjected to a clock of period *P*. The precise shape of $f_n^P(\varphi_i)$ was found by interpolating the phase resetting curves by assuming that the effect of a dark pulse of the phase scales with the period of the clock (for instance, a 5 h dark pulse would have the same effect on a 24 h clock that a 10 h dark pulse would have on a 48 h clock). The shift between the phase before (φ_i) and after (φ_{i+1}) was used to construct an expression for f_n^P that accurately captures the features of the 5 h, 7 h, 9 h, and 12 h dark pulse measurements. In particular, we used the following expression to model f_n^P:

$$f_n^P(\phi_i) = \begin{cases} \pi & \text{if } \frac{6\pi(n-9)}{P} < \phi_1 < \frac{\pi}{2} \text{ or } \phi_1 > \frac{6\pi(n-1)}{P} \\ \phi_i - \frac{6\pi(n-1)}{P} + A \cdot sin(\phi_i - \frac{6\pi(n-1)}{P}) & \text{otherwise} \end{cases}$$

where *A* was given by min(n/P, π/4). Plots of this function for various dark pulse lengths are shown in Fig. S4 in (*5*).

Using this expression for f_n^P , we interpolated from our measured dark pulse response data to find a stable recurrence relation corresponding to clock entrainment by plotting the value(s) of φ_i which converged for $i \gg 1$ for a 12 h dark pulse each clock period. (Fig. 4.5B). When the clock period is less than 40 h, the model predicts that the clock will stably entrain to the environment, but a period mismatch results in an entrained phase that is generally incorrect

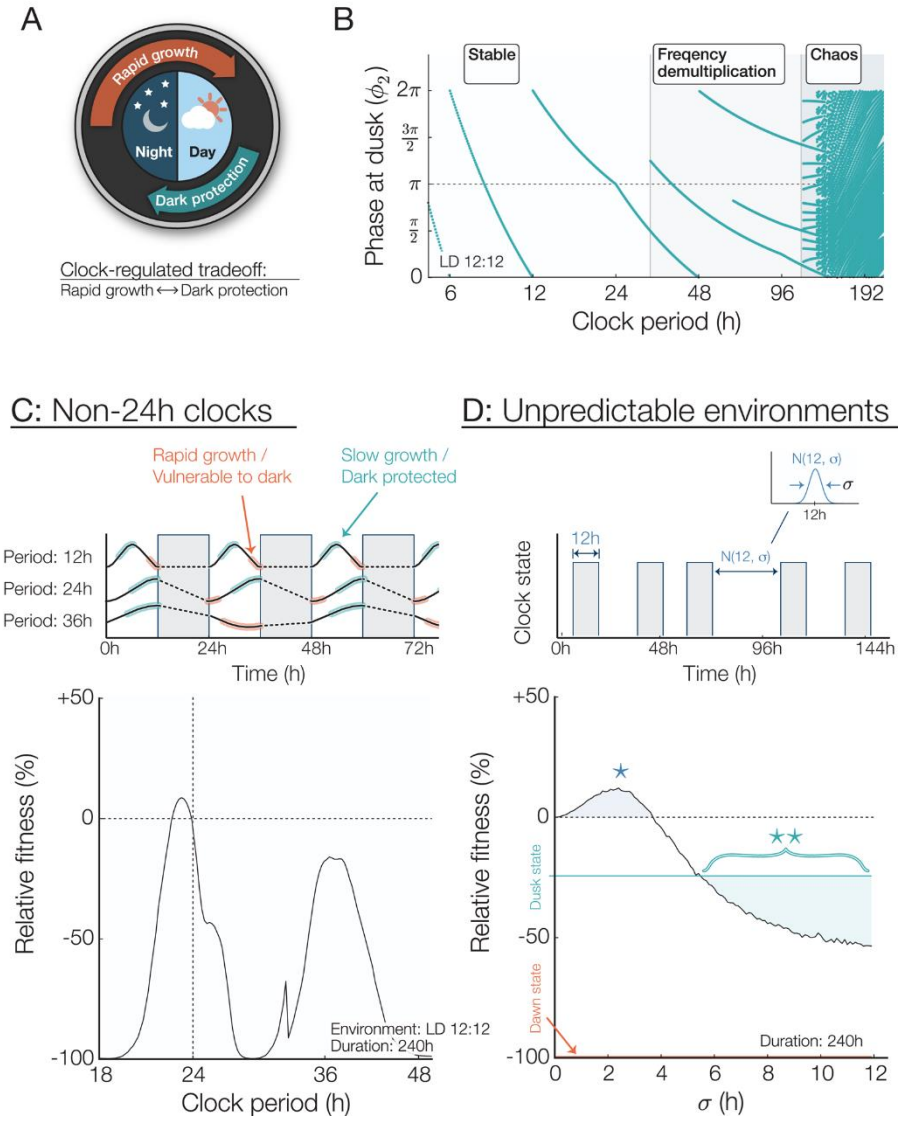

Fig. 4.5. Mathematical model of clock-regulated growth under noisy and period mismatched schedules. (A) Circadian trade-off between rapid growth and stress protection. Each phenotype occupies a specific region of the circadian cycle. (B) Map showing phase entrainment, or lack thereof, under LD 12:12 for different clock periods. A perfectly entrained clock would result in a phase at dusk equal to π. Clock periods longer than 40 h result in faulty phase tracking caused by frequency demultiplication and those longer than 100 h lead to chaos. (C) Top: Effect of a non-circadian clock period. Only the 24-h clock is stable and accurate under LD 12:12 environments: a 12h clock is robustly entrained but is unable to accurately anticipate nightfall, and the entrainment of a 36h clock is unstable and leads to an incorrect nightfall prediction 50% of the time. Bottom: The historical fitness H_P of clock mutants with a period P subjected to a LD 12:12 circadian environment for 240h is compared with the historical fitness H_{24h} of a

(i.e. subjective dusk does not fall near actual nightfall). For longer periods, entrainment can occur to subharmonics of the environmental period; for still longer periods, chaotic dynamics can occur.

We next investigated how circadian misalignment would affect the long-term fitness of a population of cells in the model by using the growth and survival functions measured in Figs. 1-2 to extract the fitness of a single cells under a given light/dark regime. By assuming that growth occurs only

(Fig. 4.5, continued) 24h-clock (relative fitness = $\frac{H_P}{H_{24h}}$). Fitness advantage is the greatest when the clock and the environment are in constructive resonance (i.e. 24 h and 36 h clocks, which leads to a correct nightfall prediction 100% and 50% of the time, respectively) and the lowest for destructive resonance (i.e. 18h, 30h, 48h clocks, which results in nightfall occurring during subjective morning). (D) Top: Fitness under 12 h nights but variable day lengths. Day durations are normally distributed with a 12 h average and a variance σ. Bottom: The presence of a low level of noise in the day length distribution increases fitness for wild-type clocks (region highlighted with a *). As the day length becomes more unpredictable (σ > 5 h), an "always-protected" dusk state is more beneficial than a circadian phenotype (** region). Fitness is the average of 10,000 simulations.

under light conditions and growth arrest probability is dependent on the phase at the onset of darkness, we calculated growth and survival of cells with a range of clock periods and find that the model predicts that the long-term fitness is highest when the clock period is near 24 hours. This result indicates that the measured effects on growth and darkness tolerance may be sufficient selective pressures to explain the precision of the circadian clock. However, when the clock period is far from 24 h, large fitness costs can occur because the clock synchronizes inappropriately so that the morning clock state occurs at nightfall every day (Fig. 4.5C). These model results show similar trends to those previously reported from competition experiments by the Johnson lab (*30, 76*): a long period (30 h) clock mutant is severely disadvantaged relative to wildtype in 12:12 LD cycles, while a short period (23 h) mutant is more mildly affected. Our model calculation shows an asymmetry in fitness as a function of clock period where periods slightly shorter than 24 h outperform longer periods. This follows from an asymmetry in the fitness cost associated with the window of protection to nightfall. If the clock period is slightly short, the protected window arrives early and the cost is unnecessarily slow growth. However, if the clock period is long, the protected window is delayed causing the much more severe cost of cell death.

How does the circadian rhythm optimize the fitness of a cell? In our model, cells must grow slowly near nightfall to avoid the possibility of metabolic catastrophe when darkness falls.

The clock enforces this growth slowdown late in the day while allowing cells to grow rapidly in the morning. This suggests that the advantages conferred by a circadian clock are a result of a finely tuned match with the temporal structure in the environment. Such a strategy might become detrimental in unnatural conditions where the environment cycles irregularly between light and dark. To test this hypothesis, we simulated cyanobacteria growing in days with random variation by selecting the duration of each light period from a normal distribution with a mean of 12 h. The model predicts that when the variability in the light-dark schedule exceeds $\sigma = 5$ h, an arrhythmic, slow-growing strategy similar to deletion of the *kaiBC* genes becomes more successful than the wildtype (Fig. 4.5C). Surprisingly, our model predicts that wildtype cells may grow faster and achieve a higher fitness in the presence of some timing variability (Fig. 5C, $\sigma <$ 3.75 h) in the duration of the day.

Conclusions

Despite the ubiquity of circadian clocks, it has remained challenging to pinpoint the benefits of rhythmic physiology (*136*). Our ability to detect the costs and benefits of clock function at the single cell level provides a framework to answer these questions. We found that considerable fitness penalties result from the failure of cells to correctly predict the withdrawal of energy associated with nightfall. We thus conclude that a major function of the cyanobacterial circadian clock is to provide a safeguard against darkness-induced starvation, giving the cell permission to grow rapidly early in the day.

A possible explanation for the failure of a subset of cells to survive the night is that these cells might have been unable to properly manage their energy consumption over the length of the night. Analysis of microarray expression data of *S. elongatus* (*116*) provides some mechanistic

insight into the origin of clock-dependent starvation tolerance (Supplementary Table I). Expression of genes involved in photosynthesis and the biosynthesis of essential compounds peaks at clock times corresponding to the early morning, suggesting that the clock tunes metabolism to allow rapid growth early in the day. On the other hand, genes involved in DNA replication, DNA repair, and metabolism under nutrient limitation, peak late in the day, suggesting clock-dependent activation of mechanisms needed to tolerate nightfall. Furthermore, we previously found that the clock controls storage and consumption of energy storage metabolites, with reserves of glycogen at their lowest near the beginning of the day (*118*). Coupled with gene expression data, our results suggest that proper temporal regulation of energy storage and circadian regulation of growth and division in anticipation of dusk may play a critical role in allowing the cell to survive the night.

These results suggest an alternative to the hypothesis that circadian rhythms evolved primarily as a means to anticipate and avoid light-induced photodamage, i.e. a "flight-from-light" scenario. If our experimental conditions approximate the challenges faced by the ancient ancestors of modern cyanobacteria, the daily threat posed by an extended time of resource limitation during the night may have been a major selective pressure on primordial clock systems. That is, a key function of the circadian clock is to direct preparations ahead of nightfall, i.e. a "prepare-for-night" scenario. We note that our microscopy growth conditions may well intensify the stresses associated with darkness relative to growth in flasks, allowing us to observe growth defects following a single light-dark cycle while liquid culture studies have required many days before measurable effects emerge (*30, 76*).

Our dark pulse experiments show that the circadian clock has robustness properties that allow it to track the 24-hour cycle in the environment even in the face of random fluctuations.

The flipside of this robustness is a remarkable fragility in environments that fail to have a 24-hour periodicity. Although such environments are unlikely to occur in nature, poor performance of clocks in these conditions may have relevance to other organisms that display a clock control of cellular divisions (*137, 138*) and to the irregular work schedules and patterns of light exposure typical of modern society.

Materials and Methods

Cyanobacterial strains

The clock phase was tracked using the yfp-ssrA reporter strain wild-type (WT)/JRCS35 (MRC1006), which carries a PkaiBC::eyfp-ssrA fluorescence reporter. The JRCS35 plasmid integrated PkaiBC::eyfp-ssrA into NS2 (neutral site 2) with a kanamycin resistance cassette (12). To create the *ΔkaiBC* strain (MRC1009), the WT/JRCS35 strain was transformed with plasmid MR0091, replacing the endogenous *kaiBC* locus (from the *kaiB* start codon to ~200 bp upstream of the *kaiC* stop codon) with a gentamicin resistance cassette. The KaiBC overexpression strain (MRC1010) was created by transforming the WT/JRCS35 strain with plasmid MR0095, integrating *kaiBC* under control of the isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible *trc* promoter into neutral site 1 (NS1).

Culture conditions

In all the experiments, cyanobacterial strains were grown in BG11 liquid medium supplemented with 20 mM HEPES (pH 8.0) at 30 °C. To create the mixed-phase population, 200 uL of a cell culture grown under continuous illumination (LL) of 75 μmol photons $m^{-2}s^{-1}$ were pipetted into each well of a black (opaque) 96-well plate. For the experiments that included

either the kaiBC-null or kaiBC-overexpression strains, these strains were grown in separate wells from the wild-type cells within the same plate as to expose them to the same culture and illumination conditions. For the kaiBC-overexpression experiment, the media was supplemented with IPTG at a final concentration of 1 mM within the plate. A custom-made Arduino driven LED array was used to illuminate each well. Each output pin of the Arduino supplied 23 mA of current to 8 red LEDs and each pin corresponded to one column of the 96-well plate. The Arduino was programmed to generate 2 days of symmetric light/dark conditions (light conditions: 10 μmol photons $m^{-2}s^{-1}$ (23 mA); Dark conditions: 0 mA) preceded by at least 12 h of continuous light conditions so that each population was subjected to 2 entrainment cycles. Light levels were maintained at ~10 μmol photons $m^{-2}s^{-1}$ for an additional 24 h before cells were collected for microscopy. Each culture well of the entrained 96-well plate was collected and combined into a single test tube. The distribution of phases produced by this protocol is broad, but is not precisely uniform. The deviations from a uniform distribution might be caused by differential growth of the wells subjected to differently phased cycles, or mild phase shifts caused by transferring the cultures to the microscope.

For experiments that included *kaiBC*-null or *kaiBC*-overexpression strains, the cells were combined in equal proportions, determined by OD750 measurements after entrainment. This provided a mixed population of wildtype and mutant cells within a single experiment.

Timelapse microscopy

The mixed-phase culture was diluted to an optical density OD750 = 0.1 using BG11 medium and 1 μL of the cell solution was pipetted onto a glass-bottom 6 well plate (MatTek Inc.). A small (1 mm X 1 mm X 0.5 mm) pad of BG11 + 2% low-melting point agarose (LMPA)

was placed atop the cell suspension. 10 mL of liquid BG11 + 2% LMPA which had been cooled to 37 °C was then poured inside the well to cover the LMPA pad. For the *kaiBC*-overexpression experiment, the BG-11/agar mixture was supplemented with IPTG at a final concentration of 1 mM before being poured into the well. Once the LMPA solidified, the 6-well plate was then moved to a motorized microscope (IX71, Olympus) and fluorescence and brightfield images were recorded every 60 minutes.

Control of the microscope was carried out using micromanager (*139*). Every 60 minutes, a motorized microscope stage (Prior) visited 24 pre-assigned locations containing at least 10 cells and bright-field (exposure: 100 ms), chlorophyll (exposure: 200 ms; excitation: 501 nm; emission: 590 nm) and YFP fluorescence (exposure: 2 s; excitation: 501 nm; emission: 550 nm) micrographs were then recorded using an EMCCD camera (Luca, Andor).

The "simple-autofocus" routine provided by the Micro-manager suite used the chlorophyll autofluorescence of the population to identify the focal plane before each set of micrographs was recorded. A collimated LED light (Thorlabs; wavelength: 625 nm) was used to illuminate the cells throughout the experiment and a microcontroller (Arduino) controlled the output level of the LED light (Light conditions: ∼10 μmol photons $m^{-2}s^{-1}$ (23 mA); Dark conditions: 0 mA).

Single-cell analysis and phase information extraction.

The outline of every cell in the brightfield image was traced using a watershed algorithm and the physiological properties of each cell (length and YFP fluorescence intensity) were recorded. The celltracker image processing (*55*) suite was then used to reconstruct the lineage

history of each cell, assigning an age to each pole and computing the instantaneous elongation rate.

The complete lineage of every cell present at the onset of the dark pulse was reconstituted and the YFP signal of each lineage was then subjected to a Fourier transform. The (complex) factor multiplying the 24 h frequency component (henceforth called c_{24}) was computed using the last 36 h of data leading to the dark pulse. The phase of the cell before the dark pulse (called φ_1 in the main text) was found by extracting the angle of c_{24} using the arctan2 branching function – ie. $\varphi = \text{arctan2}(\text{imag}(c_{24}), \text{real}(c_{24}))$.

To find the phase of the clock after the dark pulse (φ_2), the YFP-intensity of the "old-pole" lineage (i.e. the lineage which inherited the oldest pole after each division) was extracted and the phase information was found by computing the angle of the c_{24} factor of the YFP signal. After the dark pulse, only the first 36 h of data were considered (to ensure that c_{24} existed). Non-oscillatory cells (such as the Δ*kaiBC* and *kaiBC*-overexpression strains) were identified by monitoring intensity traces which varied by less than 30% over the duration of the experiment.

Since the production and maturation rate of YFP proteins have a finite timescale that is determined, among other factors, by the growth rate of the cells, the clock phases we report are shifted by 4 h relative to the extracted peak YFP phase information to bring them more closely in line with the estimated peak transcriptional activity. This value is similar to other values reported in the literature (*52*).

Growth arrest probability

"Arrested" cells were identified by tracking the cumulative increase in the total length of the cell after the dark pulse. If the total length of a cell and its progeny increased by less than

33% after 24 h, the cell was scored as arrested. Because of the altered morphology of the *kaiBC* overexpression strain, we used an alternative test for these cells: if the relative elongation rate was less than 1% / hour over the 4 hours following the dark pulse, the cell was scored as arrested.

Elongation rate measurements

The elongation rate was computed from 6 experimental replicates of mixed-phase populations grown under constant light conditions. Cells were grown for a total of 36 h under constant light conditions and the last 12 h (T=24–36 h) were used to compute the elongation rate. The instantaneous elongation rate was found by computing the relative increase in cell size between two consecutive frames. Specifically,

$$g = \frac{\ell_{t+1} - \ell_t}{\ell_t} \sim \frac{dlog(\ell)}{dt}$$

The growth rate $g(t)$ was then binned according to the cell's circadian phase and averaged over a 1 h window. In Fig. 3B, a 3-pt moving average was used to smooth the data used to derive the phenomenological model.

Fitness advantage measurement

To measure the performance of various clock periods under a sustained LD 12:12 schedule, it was necessary to compute 1) the probability μ that a cell would enter a state of growth arrest following the 12 h dark pulse and 2) the number of cell doublings that happened during the 12 h of light. μ was determined using f_{12h}^{P} to find the phase of the cell at nightfall to identify the survival probability at that phase using Fig. 2D (that is, it was assumed for the sake of simplicity that growth arrest occurs at the same rate for 12 h and 18 h nights). The number of

cell doublings per day was found by first using f_{12h}^{P} to identify the phase at the beginning of the day. We then advanced this phase variable through the light portion of the day to compute the average elongation rate $\bar{g}$. The historical fitness (*140*) was used to quantify the fitness of the population at a given period. Historical fitness differs from conventional fitness measurements in that it considers the cumulative (or integrated) number of doublings of a single cell over its complete life-history. Consequently, the historical fitness of a population will contain two factors: $(1-\mu)^{\left\lfloor \frac{T}{2} \right\rfloor}$, which accounts for the fraction of the population that enter a state of growth arrest during the dark pulse, and $e^{\frac{\bar{g}T}{2}}$, which tracks the number of cell doublings that happen under light conditions.

The historical fitness for a simulation that lasted for a time *T* for a given clock period *P* was given by the product of these two factors:

$$H_P = (1-\mu)^{\lfloor \frac{T}{2} \rfloor} e^{\frac{\bar{g}T}{2}}$$

for a simulation that lasted for a time *T* for a given clock period *P*. The values of H_P were plotted relative to H_{24h}.

Perspective

The results presented here provide insight into the selective pressures that contribute to the level of oscillatory precision in single cells presented in Chapter 2. A noisy clock that predicts the time of day poorly is more likely to experience a mismatch between cellular physiology and environmental conditions. The most obvious penalty for a clock-environment mismatch as shown here is growth arrest if darkness occurs when the cell expects daytime. This amount of mismatch likely requires an extreme amount of noise in the clock to occur, but fitness defects may also arise from smaller clock-environment mismatches. For instance, cells that

predict darkness several hours before the actual onset of darkness may slow growth and division at a time when rapid growth is still permissible, falling short of the maximal amount of growth that can happen over the day. In conclusion, the results presented here provide a powerful explanation for why such high precision is required in the cyanobacterial circadian clock such that Kai proteins are expressed in high amounts to minimize the impact of molecular noise.

Appendix: Ancestral reconstruction of the Kai system using phylogenetic analysis by maximum likelihood (PAML)

Introduction

Ancestral reconstruction is a recently developed statistical method that utilizes phylogenetics to estimate DNA or protein sequences of a given ancestor of a clade of extant, modern sequences. It has been used to successfully investigate the evolution of transcription factors in yeast (*141*), proteins responsible for the rise to multicellularity (*142*), and ethanol tolerance in *Drosphila* alcohol dehydrogenase (*143*). For any reconstruction, the basic workflow is as follows (*3*) (Figure A.1). First, a set of extant sequences must be obtained, which might be a series of homologous protein sequences from related organisms (although in theory the reconstruction should also allow for reconstruction of ancestors of protein families that might share multiple members within a single organism). From this set of sequences, a multiple sequence alignment should be performed to construct a phylogenetic tree that specifies the ancestral relationships between each sequence. Finally, information from the multiple sequence alignment and phylogenetic tree can be combined to estimate the ancestral sequence using maximum likelihood methods on a residue-by-residue basis to predict the most likely ancestral residue at each site in the reconstructed sequence.

My initial efforts to reconstruct the Kai system were aimed at addressing two separate questions related to the evolution of the cyanobacterial clock. First, assuming that the Earth's rotation was much faster in the past, is there a "biological timekeeping fossil" hidden in modern cyanobacterial clock sequences that indicate that the ancestral cyanobacterial clock was also faster? Second, how did the abbreviated Kai system in *Prochlorococcus* evolve? Here, I will first provide background information for the first question followed by the second.

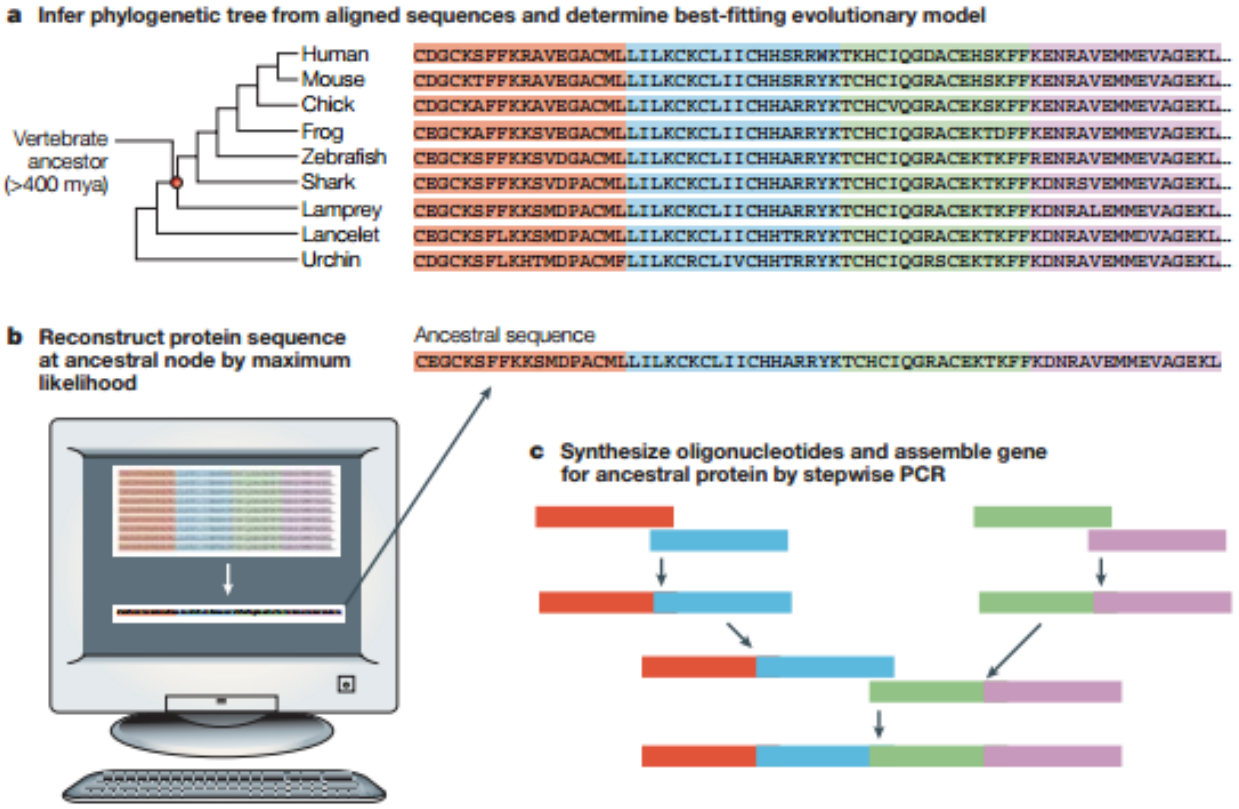

Figure A.1. Ancestral reconstruction process from (*3*). A phylogeny is first constructed from a multiple sequence alignment of extant protein sequences, and ancestral reconstruction is performed on a residue-by-residue basis using maximum likelihood statistical methods. Finally, reconstructed genes can be assembled from oligonucleotides and expressed in vivo.

Does the ancient cyanobacterial clock hold a record of the Earth's rotational speed?

Physical measurements currently indicate that tidal friction between the oceans and the Earth's surface is causing the Earth's rotation to slow by approximately 1.8 milliseconds per century (*144*). This finding is consistent with the fact that the Moon is slowly receding from Earth in its orbit by about 3.8 centimeters per year (*145*) due to the conservation of angular momentum. The observed slowing of the Earth's rotation thus implies that in the past, the Earth rotated much more rapidly. A quick calculation shows that if the current rate of slowing is extrapolated backwards in time by 1 billion years, the rotational period was 18.4 hours instead of

24 hours. This estimate of the past Earth's period is consistent with both geological and fossil evidence. Alternating patterns of silt deposition between the spring and neap tides suggest that 900 million years ago (MYA), the Earth's rotational period was ~18 hours (*146*). Daily oscillations in coral growth rates, visualized as growth ring bands, give rise to estimations that 400 MYA the Earth's rotational period was ~20 hours (*147*).

The modern clock in *S. elongatus* has a free-running period close to 24 hours, and the earliest fossil evidence of cyanobacteria date to over 2 billion years ago (GYA) (*148*), meaning that the earliest cyanobacteria would have presumably lived in an environment with a day/night cycle much shorter than 24 hours. Competition experiments demonstrate that a clock with a period that is mismatched with the environment has negative consequences for organism fitness. In a mixed population of cyanobacteria with either a 24-hour wild-type clock or a 20-hour mutant clock, the strain that dominates the population after several weeks of growth has a clock that matches the period of environmental light/dark cycles (*30, 76*). This suggests that there is a strong selective pressure to match the period of the clock to the environment. This inspires the question: did ancient cyanobacteria have a circadian clock with a shorter period, and did that period lengthen over time as the Earth's rotation slowed?

How confident are we that cyanobacteria were present when the Earth's rotation was faster? In addition to fossil evidence that suggests an ancient origin for cyanobacteria, other studies hypothesize that the origin of cyanobacteria is related to the Great Oxygenation Event, an event that occurred roughly 2.5 billion years ago (GYA) as indicated by geological and chemical evidence (*149*). Cyanobacteria are hypothesized to be the first photosynthetic organisms to generate atmospheric oxygen in significant amounts, increasing atmospheric oxygen from

negligible amounts to their modern levels. The timing of the Great Oxygenation Event is thus consistent with the idea of an ancient origin for cyanobacteria when the day was much shorter.

Thus, the main objective in investigating this question is to ideally reconstruct multiple ancestral versions of the Kai system, with the hypothesis that nodes that more ancestral will have a faster free-running period. The results of the efforts to reconstruct the ancient Kai system are hereafter described in the Results section. For now, we move on to the motivating factors for the second question that can be investigated with ancestral reconstruction of the Kai system.

How did Prochlorococcus evolve into an environmentally-driven timer?

As noted in the dissertation introduction and in Chapter 2, *P. marinus* has an abbreviated Kai system in which the gene for *kaiA* is completely missing (*49*), and as a result, KaiC can no longer generate autocatalytic rhythms in phosophorylation (shown in Chapter 2). Instead, *Prochlorococcus* must be incubated in a cyclic light/dark environment to generate phosphorylation rhythms in KaiC, with KaiC settling into a hyperphosphorylated state in the light and a hypophosphorylated state in the dark. Previous in vitro studies suggest that *Prochlorococcus* KaiC (ProKaiC) has gained a KaiA-independent autokinase activity (*63*), which is consistent with mutations in the C-terminus of ProKaiC in which the residues necessary for burying the A-loop to suppress kinase activity are no longer conserved (*49*). While it is possible that other factors in *Prochlorococcus* have supplanted the role of KaiA, the simplest model for the Kai system in *Prochlorococcus* is that the kinase rate of ProKaiC is metabolically regulated by ATP/ADP levels in the cell, which change with the day/night cycle.

The question of how the *Prochlorococcus* clock lost free-running capability assumes that its ancestor had a free-running clock. How do we know this? First, *P. marinus* MED4 is only

one of two cyanobacterial species that do not have *kaiA* (the other is a tiny symbiotic cyanobacteria, UCYN-A) (*150*). All other cyanobacteria contain copies of all three Kai proteins. Considering that *P. marinus* is thought to have phylogenetically diverged somewhat recently, this makes it extremely unlikely that the *kaiA* gene evolved independently in all other cyanobacteria and that the original, ancient cyanobacterium had a *Prochlorococcus*-like *kaiB* and *kaiC* system. Second, as one traverses the cyanobacterial phylogenetic tree from a relative that has all three Kai proteins (*Synechococcus WH 8102*) towards *P. marinus* MED4, one finds that the *kaiA* gene gets progressively shorter from the N-terminus until only the C-terminal third of *kaiA* is remaining in *Prochlorococcus sp.* MIT 9313 and *Prochlorococcus sp.* MIT 9303 (*49*). This view suggests that the *kaiA* gene was truncated over evolutionary history as *Prochlorococcus* differentiated into its niche, and that the lack of a *kaiA* gene in *P. marinus* is the terminal result in a series of piecewise deletions in the *kaiA* gene.

The main question to investigate here is the evolutionary path that the Kai system traversed to lose *kaiA*. Did the (presumed) original free-running clock become a dampened oscillator before losing *kaiA* entirely, analogous to a continuous decrease in feedback loop strength described in the model in Chapter 2? As *kaiA* was shortened, what functions did it lose, and what functions did it retain? How did KaiC switch to having a KaiA-independent kinase activity? While ancestral reconstruction can contribute to our investigation of these questions, it is also necessary to perform a multitude of biochemical experiments to dissect which residues may have contributed to these changes, which requires additional planning from what is presented here.

I did not focus my ancestral reconstruction efforts towards the evolution of the *Prochlorococcus* system, instead focusing on the first posed question of whether ancestral clock

systems had faster periods to match the Earth's rotation. As such, all of the following sections will focus on the ancestral reconstruction problem with that perspective in mind.

Methods and Results

Cyanobacterial protein sequences for all three Kai proteins were downloaded from GenBank using the following search term (example shown with KaiA):

(KaiA) AND "cyanobacteria"[porgn:__txid1117]

"KaiA" was substituted with "KaiB" and "KaiC" for their respective sequences. The sequences for each protein were curated by hand to remove irrelevant sequences. For instance, KaiB sequences appeared in the search results for KaiA and vice versa. Other irrelevant results included partial sequences from crystal structures as well as hypothetical proteins and non-Kai proteins. Results date to May 2014; as such, those seeking to continue this work should re-run the search as it is likely that more cyanobacterial genomes have been sequenced since.

Only organisms that contained all three Kai proteins were considered for reconstruction, and a Python program was written to remove species that did not have all three Kai proteins. For organisms with duplicate sequences present, the first sequence in the list for that particular organism was taken. Completion of this step yielded 146 distinct sets of Kai protein sequences. As a first step in constructing a phylogenetic tree, the KaiC sequences underwent multiple sequence alignment using the MUSCLE algorithm in the MEGA bioinformatics software platform. A phylogenetic tree for KaiC was constructed using PhyML, which constructs phylogenetic trees based on maximum likelihood, and a root for the tree was chosen such that

Synechococcus sp. JA-3-3Ab and Synechococcus sp. JA-2-3-Ba formed the outgroup for the remaining species, which is based on phylogenetics of 16s rRNA sequencing that show the JA-3-3-Ab and JA-2-3-Ba species as the most basal cyanobacteria out of the 146 cyanobacteria sampled.

The resulting tree with all 146 sets of KaiC sequences was compared to previous phylogenetic studies that used cyanobacterial fossils to calibrate their estimate of the evolutionary molecular clock in order to estimate the age of ancestral nodes within the phylogenetic tree (*151, 152*), see Figures A.2 and A.3. Based on these previous studies, three nodes were chosen for ancestral reconstruction that were predicted to span a suitable range of times to test whether a trend towards shorter periods existed as nodes became more ancient. The estimated age of these nodes is listed in the table below:

Table A.1. Ages of selected ancestral nodes

	Node and age in MYA		
Study	Red	Yellow	Green
Baca et al. (ML)	2223 ± 145	1031 ± 105	588 ± 38
Baca et al. (Bayesian)	2583 (2186-2973)	1943 (1372-2524)	1035 (576-1531)
Dvorak et al.	~2700	~2300	~1500

The red node is the ancestor of most cyanobacteria, the yellow node is the most recent common ancestor of *S. elongatus* and the marine picocyanobacteria (such as *Synechococcus* WH 8102 and *P. marinus* MED 4), and the green node is the most recent common ancestor of the marine picocyanobacterial.

It should be noted that the following analysis uses only a subset of the sequences for all 146 species for reasons that are now outdated, and any future analysis should be repeated with all 146 species as it would increase the amount of information available for reconstruction. Originally, the three nodes selected for reconstruction lay entirely within a clade that only comprised a small part of the complete tree with all 146 species, and to reduce the complexity of

reconstruction, the phylogenetic tree of 146 species was trimmed to contain only 36 species.

These 36 species included all species that were descendants of the three original target nodes (19

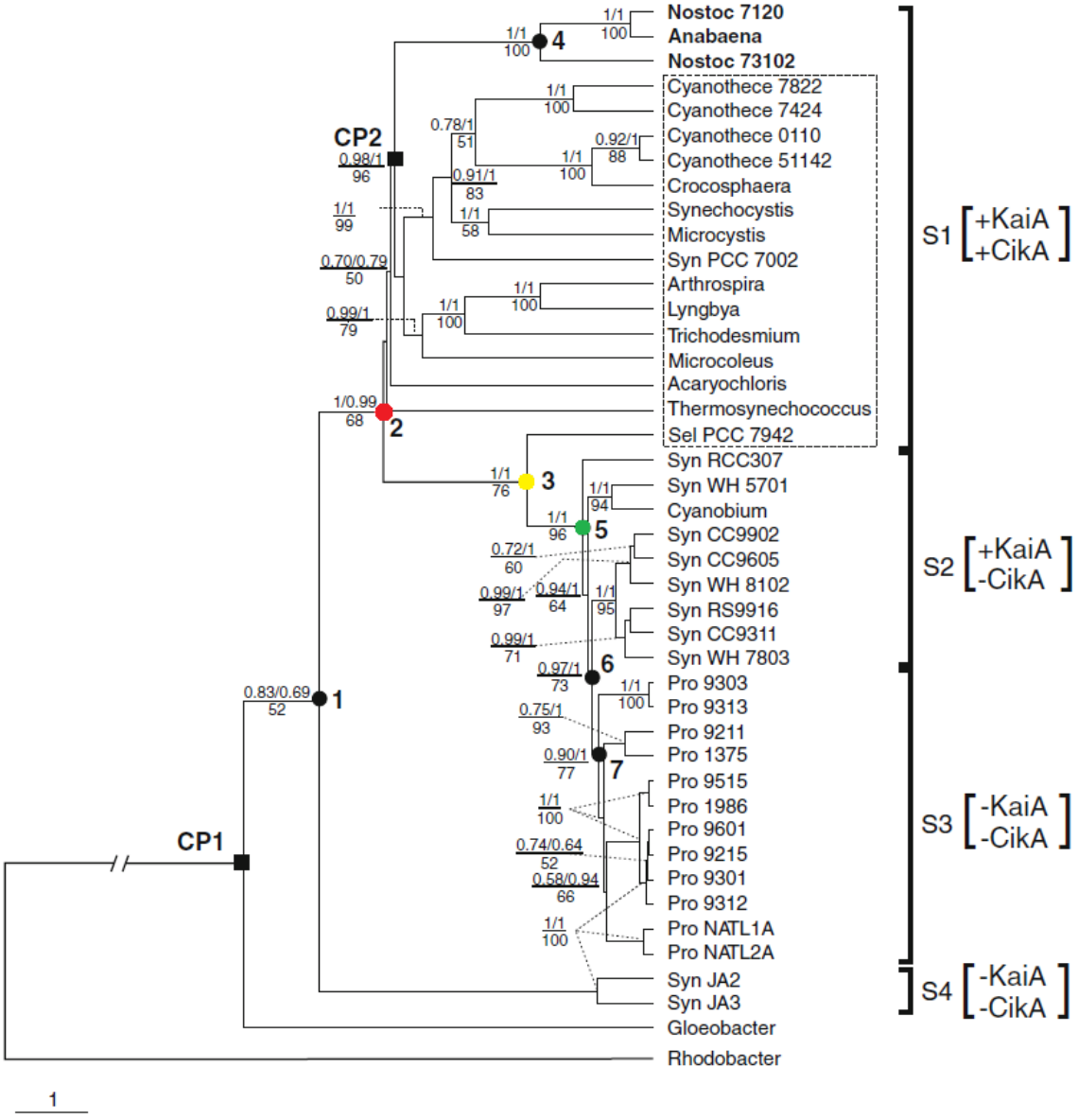

Figure A.2. Phylogenetic tree constructed from concatenated 16S rRNA and 23S rRNA sequences in cyanobacteria, adapted from (*16*). Bar indicates 1 substitution per site. Estimates of node ages were derived based on this tree. The three target ancestral nodes listed in table A.1 are shown in red, yellow, and green.

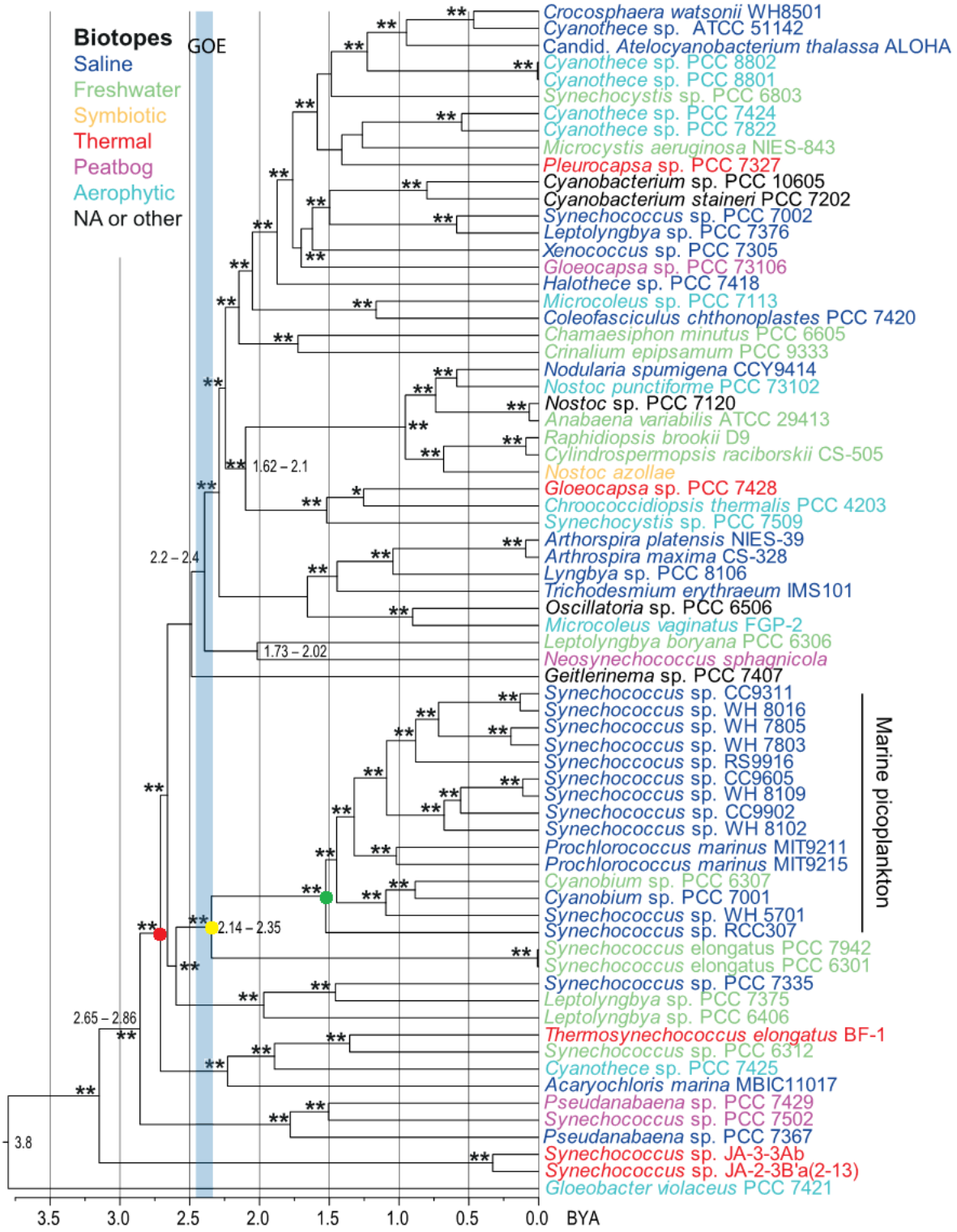

Figure A.3. Phylogenetic tree constructed from 16s rRNA sequencing from (*152*). This tree provided an alternate estimate of node age. Targeted ancestral nodes indicated with red, yellow, and green circles.

species), and species in the outgroup (17 species) were selected by hand in a manner that represented the remaining cyanobacterial species somewhat equally. These 36 species were used in all subsequent analyses presented below, and the three selected nodes for study were later revised to the three presented in the table above, which will also be used in all subsequent analyses below.

Before obtaining ancestral sequences, multiple sequence alignments were performed for the 36 species for each individual set of Kai protein sequences (for KaiA, KaiB, and KaiC), and phylogenetic trees were generated again in the manner described above. The multiple sequence alignments and phylogenetic trees were processed by the Lazarus software package developed by the Thornton lab (*153*), which uses the PAML algorithm to infer ancestral sequences based on maximum likelihood approaches, and ancestral sequences were obtained for the target nodes. It is important to note that ancestral reconstruction evaluates each residue in a protein sequence independently and that PAML treats blank residues in the multiple sequence alignment as missing information. This can be problematic for sites in which one species in particular has a unique insertion of an extra residue that does not align to any other species. Rather than assuming that the inserted residue does not exist in other species, PAML instead assumes that the corresponding residue is unknown. To deal with cases like these, the ancestral sequences were realigned with the extant sequences, and sites that were unique insertions were manually removed from the ancestral sequence. It is for this reason that the initial 146 species tree was trimmed to 36 species, as additional sequences increases the number of spurious insertions that will exist in the final ancestral sequences.

As a side comment, it is remarkable to note that the degree of conservation of Kai sequences across the entire cyanobacterial clade is strikingly high for proteins that originally

diverged billions of years ago. However, despite this overall level of conservation, there are also small, specific regions of the Kai proteins that are less well conserved, such as the N- and C-termini of KaiC. A lack of conservation could be attributed to a bona fide change in protein function (such as the changes in the KaiA-binding part of the C-terminus from *S. elongatus* to *P. marinus* that are presumably responsible for the constitutive autokinase activity in ProKaiC), but it may also indicate that the evolutionary constraints in poorly conserved regions are much lower, indicating that the function of the region is less dependent on its specific sequence (e.g. a linker sequence).

These poorly conserved sites in the Kai proteins present a problem for ancestral reconstruction, in that the statistical confidence in the ancestral residues tends to be lower than for highly conserved sites. How should this uncertainty be dealt with? If only a few sites have low statistical confidence, then it may be possible to construct several variants of the protein using permutations of the second-most likely residues for each site with confidence below a given threshold (e.g. < 85% posterior probability). However, this approach becomes untenable if the number of uncertain sites becomes large. At this point, some judgement will need to be exercised for each uncertain site based on currently available biochemical studies as to whether one should simply use the highest confidence estimate or whether the sites should be replaced with a modern sequence. These considerations were used in the approach I took in finalizing ancestral sequences for experimental testing.

Reconstructing the entire Kai system in vitro for three ancestral nodes is a moderate undertaking, involving the purification of nine individual proteins. Thus, as a first step, I aimed to test whether ancestral KaiC could function with modern KaiA and KaiB from *S. elongatus*, reducing the number of proteins for reconstruction to three (see Figure A.4 for phylogenetic tree

of KaiC with target nodes highlighted). In all three ancestral nodes, the site in KaiC that had the lowest estimation confidence was the N-terminus, which was highly variable across all

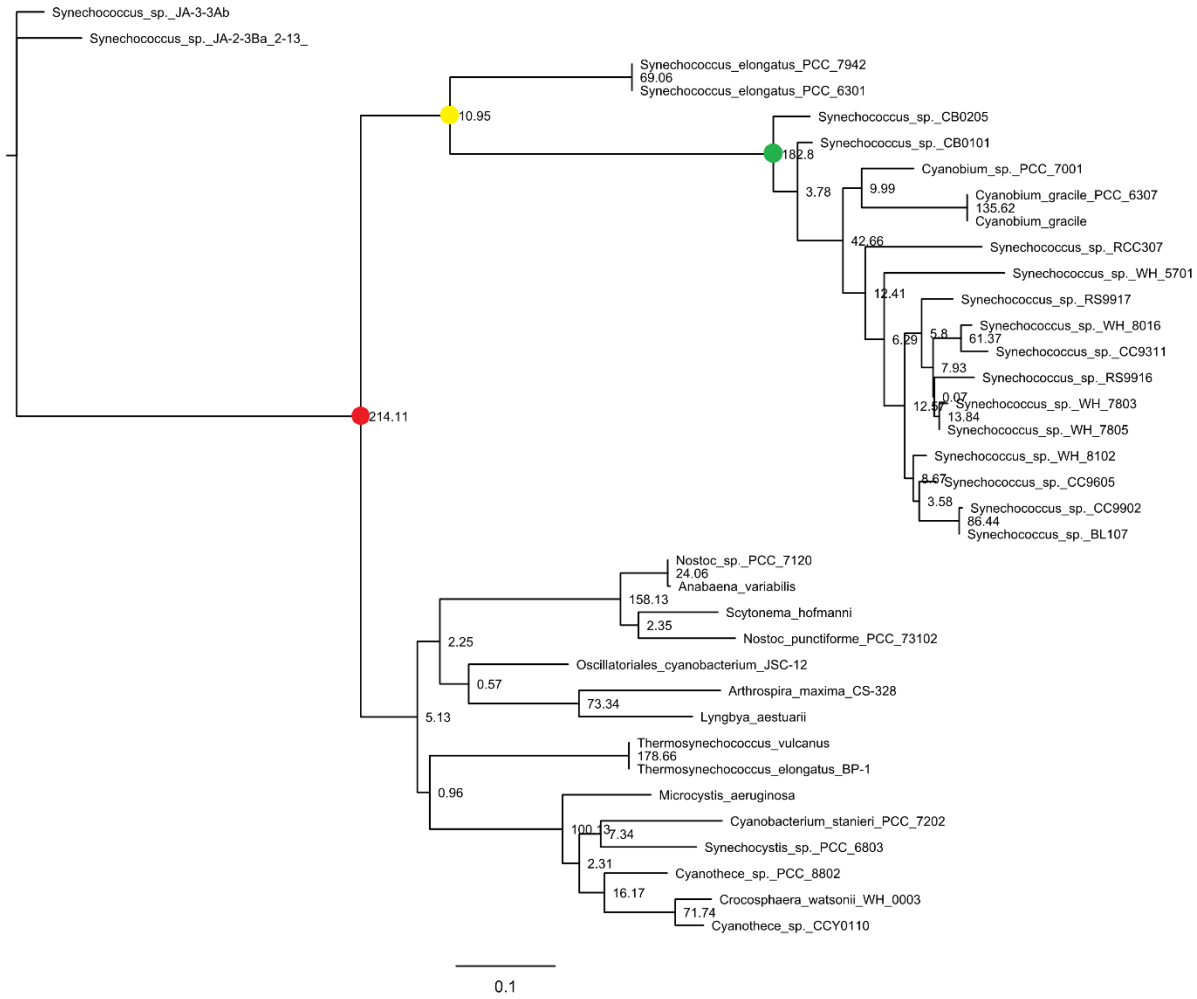

Figure A.4. KaiC phylogenetic tree constructed in this study with ancestral target nodes labeled. Bar indicates 0.1 substitutions per site, and numbers at nodes indicate maximum likelihood support numbers for each node.

cyanobacteria. A scanning mutagenesis study performed by Jenny Lin showed that in general, random single residue mutations in the N-terminus (residues 1-16) were less likely to impact oscillation amplitude or period than mutations after residue 16, suggesting that the N-terminus is poorly conserved because its function is not as strictly dependent on sequence (*154*). Additionally, the vast majority of individual residue changes between the ancestral sequences and modern *S. elongatus* KaiC between residues 1-16 were shown not to have a significant

impact on oscillation amplitude or period, further decreasing the likelihood that the ancestral sequences contained a predicted residue that would greatly impact KaiC function. Nevertheless, as a precaution, the first 16 residues of the ancestral sequences were replaced by those from modern KaiC. The remaining sequence was left as the generated ancestral sequences, and the final sequences are listed below (see Figure A.5 for a multiple sequence alignment of modern KaiC and the ancestral nodes):

Red node (predicted age ~2200-2700 MYA):

Reconstructed sequence
MNQPEQQSSNNGPNSAGVQKIRTMIEGFDDISHGGLPVGRTTLVSGTSGTGKTLFAVQF
LYNGITYFDEPGIFVTFEESPTDIIKNASSFGWDLQKLIDEGKLFILDASPDPEGQDVVGNF
DLSALIERIQYAIRKYKAKRVSIDSVTAVFQQYDAASVVRREIFRLVARLKQLGVTTIMT
TERVEEYGPVARFGVEEFVSDNVVILRNVLEGERRRRTIEILKLRGTTHMKGEYPFTITN
NGINIFPLGAMRLTQRSSNVRVSSGVKTLDEMCGGGFFKDSIILATGATGTGKTLLVSKF
LENGCQNGERAILFAYEESRAQLSRNASSWGIDFEELERQGLLKIICAYPESAGLEDHLQI
IKSEIAEFKPSRIAIDSLSALARGVSNNAFRQFVIGVTGYAKQEEITGFFTNTTDQFMGSHS
ITDSHISTITDTILMLQYVEIRGEMSRAINVFKMRGSWHDKGIREYTITENGPEIKDSFRNY
ERIISGSPTRISVDEKSELSRIVRGVQDKEEEIDE

Sequence with replaced N-terminus from *S. elongatus* KaiC
MTSAEMTSPNNNSEHQGVQKIRTMIEGFDDISHGGLPVGRTTLVSGTSGTGKTLFAVQF
LYNGITYFDEPGIFVTFEESPTDIIKNASSFGWDLQKLIDEGKLFILDASPDPEGQDVVGNF
DLSALIERIQYAIRKYKAKRVSIDSVTAVFQQYDAASVVRREIFRLVARLKQLGVTTIMT
TERVEEYGPVARFGVEEFVSDNVVILRNVLEGERRRRTIEILKLRGTTHMKGEYPFTITN
NGINIFPLGAMRLTQRSSNVRVSSGVKTLDEMCGGGFFKDSIILATGATGTGKTLLVSKF
LENGCQNGERAILFAYEESRAQLSRNASSWGIDFEELERQGLLKIICAYPESAGLEDHLQI
IKSEIAEFKPSRIAIDSLSALARGVSNNAFRQFVIGVTGYAKQEEITGFFTNTTDQFMGSHS
ITDSHISTITDTILMLQYVEIRGEMSRAINVFKMRGSWHDKGIREYTITENGPEIKDSFRNY
ERIISGSPTRISVDEKSELSRIVRGVQDKEEEIDE

Yellow node (predicted age ~1000-2000 MYA):

Reconstructed sequence
MTQPEQQSPNNNSNLAGVQKIRTMIEGFDDISHGGLPIGRSTLVSGTSGTGKTLFSVQFL
YNGITQFDEPGIFVTFEESPQDIIKNASSFGWDLQKLVDEGKLFILDASPDPEGQDVVGNF
DLSALIERINYAIRKYKARRVSIDSVTAVFQQYDAASVVRREIFRLVARLKQIGVTTVMT
TERIEEYGPIARYGVEEFVSDNVVILRNVLEGERRRRTIEILKLRGTTHMKGEYPFTITNH
GINIFPLGAMRLTQRSSNVRVSSGVKRLDEMCGGGFFKDSIILATGATGTGKTLLVSKFV
ENACANKERAILFAYEESRAQLLRNASSWGIDFEEMERQGLLKIICAYPESAGLEDHLQII

KSEISEFKPSRIAIDSLSALARGVSNNAFRQFVIGVTGYAKQEEITGFFTNTSDQFMGSHSI
TDSHISTITDTILLLQYVEIRGEMSRAINVFKMRGSWHDKGIREYMITDKGPEIKDSFRNF
ERIISGSPTRISVDEKSELSRIVRGVQEKEPEIEE

Sequence with replaced N-terminus from *S. elongatus* KaiC
MTSAEMTSPNNNSEHQGVQKIRTMIEGFDDISHGGLPIGRSTLVSGTSGTGKTLFSVQFL
YNGITQFDEPGIFVTFEESPQDIIKNASSFGWDLQKLVDEGKLFILDASPDPEGQDVVGNF
DLSALIERINYAIRKYKARRVSIDSVTAVFQQYDAASVVRREIFRLVARLKQIGVTTVMT
TERIEEYGPIARYGVEEFVSDNVVILRNVLEGERRRRTIEILKLRGTTHMKGEYPFTITNH
GINIFPLGAMRLTQRSSNVRVSSGVKRLDEMCGGGFFKDSIILATGATGTGKTLLVSKFV
ENACANKERAILFAYEESRAQLLRNASSWGIDFEEMERQGLLKIICAYPESAGLEDHLQII
KSEISEFKPSRIAIDSLSALARGVSNNAFRQFVIGVTGYAKQEEITGFFTNTSDQFMGSHSI
TDSHISTITDTILLLQYVEIRGEMSRAINVFKMRGSWHDKGIREYMITDKGPEIKDSFRNF
ERIISGSPTRISVDEKSELSRIVRGVQEKEPEIEE

Green node (predicted age ~600-1500 MYA):

Reconstructed sequence
MTQPMQDPSPTNNHLASVQKLPTGIEGFDDVCQGGLPIGRSTLISGTSGTGKTVFSLNFL
YNGIRQFDEPGIFVTFEESPLDILRNAASFGWNLQEMVEQDKLFLLDASPDPEGQDVAGS
FDLSGLIERINYAIRKYKARRVAIDSITAVFQQYDAVSVVRREIFRLIARLKEIGVTTVMT
TERIDEYGPIARYGVEEFVSDNVVILRNVLEGERRRRTVEILKLRGTTHMKGEFPFTMGS
HGISIFPLGAMRLTQRSSNVRVSSGVPRLDEMCGGGFFKDSIILATGATGTGKTLLVSKF
VENACANKERAILFAYEESRAQLLRNATSWGIDFEEMERQGLLKIICAYPESTGLEDHLQ
IIKTEISQFKPSRMAIDSLSALARGVSHNAFRQFVIGVTGYAKQEEIAGFFTNTSEEFMGS
HSITDSHISTITDTILLLQYVEIRGEMARALNVFKMRGSWHDKGIREYIITSNGPEIKDSFS
NFERIISGVPHRINTDERSELSRIVKGVGEDQSLIEE

Sequence with replaced N-terminus from *S. elongatus* KaiC
MTSAEMTSPNNNSEHQSVQKLPTGIEGFDDVCQGGLPIGRSTLISGTSGTGKTVFSLNFL
YNGIRQFDEPGIFVTFEESPLDILRNAASFGWNLQEMVEQDKLFLLDASPDPEGQDVAGS
FDLSGLIERINYAIRKYKARRVAIDSITAVFQQYDAVSVVRREIFRLIARLKEIGVTTVMT
TERIDEYGPIARYGVEEFVSDNVVILRNVLEGERRRRTVEILKLRGTTHMKGEFPFTMGS
HGISIFPLGAMRLTQRSSNVRVSSGVPRLDEMCGGGFFKDSIILATGATGTGKTLLVSKF
VENACANKERAILFAYEESRAQLLRNATSWGIDFEEMERQGLLKIICAYPESTGLEDHLQ
IIKTEISQFKPSRMAIDSLSALARGVSHNAFRQFVIGVTGYAKQEEIAGFFTNTSEEFMGS
HSITDSHISTITDTILLLQYVEIRGEMARALNVFKMRGSWHDKGIREYIITSNGPEIKDSFS
NFERIISGVPHRINTDERSELSRIVKGVGEDQSLIEE

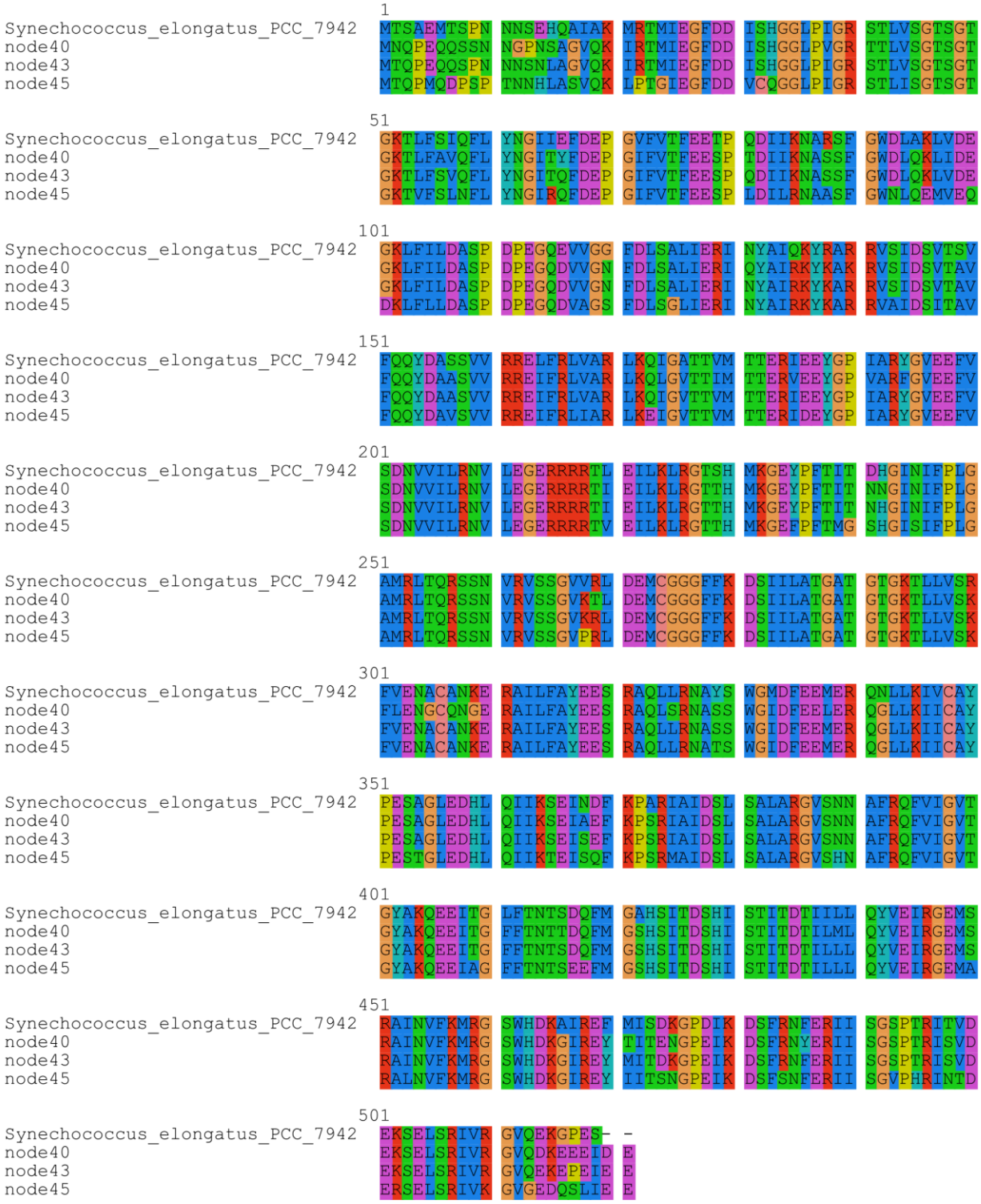

Figure A.5. Multiple sequence alignment of modern S. elongatus KaiC and reconstructed ancestral KaiC. Nodes 40, 43, and 45 correspond to the red, yellow, and green target nodes respectively.

Next, I examined the reconstructed sequences to determine whether residues known to be important for KaiC function were present in the ancestral sequences, and these residues and their functions are listed below from (*69*, *155*):

Table A.2. Residues important for KaiC function

		Present in ancestral sequence?		
Residue	Function	Red node	Yellow node	Green node
T432	Clock phosphorylation site	Yes	Yes	Yes
S431	Clock phosphorylation site	Yes	Yes	Yes
I497	Terminal residue in A-loop, stabilizes buried position of A-loop	Yes	Yes	Yes
E487	Stabilizes A-loop buried state with T495	Yes	Yes	Yes
T495	Stabilizes A-loop buried state with E487	Yes	Yes	No (T->H)
I472	Prevents ATP from approaching phosphorylation site (with D474)	Yes	Yes	Yes
D474	Prevents ATP from approaching phosphorylation site (with I472)	No (D->E)	Yes	No (D->S)
W331	Thought to stabilize ATP away from phosphorylation site (W331A leads to hyperphosphorylation)	Yes	Yes	Yes
E77	Catalytic carboxylate in CI	Yes	Yes	Yes
E78	Catalytic carboxylate in CI	Yes	Yes	Yes
E318	Catalytic carboxylate in CII	Yes	Yes	Yes
E319	Catalytic carboxylate in CII	Yes	Yes	Yes

Notably, most of these important residues are present in the ancestral sequences, with a few exceptions. Although the effect of the remaining mutations in *S. elongatus* KaiC are currently unknown, the predicted ancestral residues are present in other modern cyanobacterial KaiC sequences and are presumed to remain functional in their individual organismal contexts.

The ancestral sequences were also compared to known mutations that alter the period of oscillation, summarized below from (*156*):

Table A.3. KaiC period mutants				
		Residue present in ancestral sequence		
Mutation	Period	Red node	Yellow node	Green node
T42S	28 hours	T	T	T
A251V	46 hours	A	A	A
S157P	21 hours	A	A	V
F470Y	17 hours	Y	Y	Y
R393C	15 hours	R	R	R

While these comparisons are far from conclusive, it is interesting to note that the F470Y change, which shortens the circadian period to 17 hours, is present in all ancestral sequences. Residue S157 is also completely different in all ancestral sequences, but the effect of these mutations is unknown. However, a major caveat is that the period mutants were all characterized in the background of wild-type modern *S. elongatus* KaiC. Indeed, the residue at position 470 in roughly half of the cyanobacteria is a tyrosine, but those species presumably have a 24-hour clock, weakening the hypothesis that the tyrosine at position 470 in the ancestral sequences is predictive of the ancestral Kai system having a shorter period.

The sequences for all three nodes (with replaced N-termini) were ordered from IDT as gBlocks, and using Gibson assembly, they were combined with expression plasmid pRSET-B which was digested with BamHI and NcoI. The insertion of the KaiC sequences for the red, yellow, and green nodes into pRSET-B generated plasmids pJC074, pJC075, and pJC076. Attempts at preparing recombinant ancestral KaiC should follow the Rust lab protocol for preparing *S. elongatus* KaiC, as reported in (*37*).

Future Directions

Expression plasmids for ancestral KaiC at the three target nodes have been constructed; the next step is to express and purify recombinant ancestral KaiC and test their function in vitro

by incubating ancestral KaiC with modern KaiA and KaiB from *S. elongatus*. Two methods for assaying clock function include sampling the reaction periodically and running samples in an SDS-PAGE gel to resolve KaiC phosphorylation over time, or using fluorescence polarization to assay clock state in an automated fashion with high temporal resolution as described in (*23*). While the fluorescence polarization method requires less manual effort, running reaction samples on a gel may be more informative if the clock reaction fails to oscillate as the phosphorylation state of KaiC will be revealed at each point in time.

Many reasons exist for why the clock reaction may fail with ancestral KaiC, including intrinsic failure of the protein structure or failure to find correct reaction conditions or protein concentrations. In order to address condition-dependent failures, the fluorescence polarization assay is ideal for troubleshooting due to its high-throughput plate reader format. However, failure stemming from a protein sequence or structure level cannot be so easily rectified due to uncertainty in which parts of the sequence are problematic. Additionally, the ancestral reconstruction process provides the most likely estimate for each site in the sequence independently, ignoring context dependency between different residues. Hence, it is quite possible that multiple residues in the ancestral sequence never actually existed together in the same organism and that together they are unable to recapitulate the ancestral phenotype. To assess whether context dependency is important, an information theoretic approach can be used to determine the mutual information between residues in the Kai proteins across all known sequences, similar to the approach used in (*157*).

Another approach to address uncertainties in the ancestral protein sequences is to construct variant sequences that instead use the residues that have the second-highest statistical confidence. If there are many variants, a combinatorial approach may be devised to construct a

library, perhaps using some form of combinatorial Quikchange mutagenesis. It would be a great deal of work to test more than a handful of variant sequences in vitro with recombinantly purified protein—thus, an approach of this kind would necessitate transforming the variant Kai systems into cyanobacteria to perform a high-throughput screening of clock phenotype in vivo. A caveat to this approach that must also be considered is that the ancestral Kai proteins may not interface with downstream clock output proteins correctly, making it difficult to interpret a result that is arrhythmic.

Lastly, it is possible that the ancestral Kai proteins are able to interface with each other but not with modern variants. To investigate, an analysis should be performed to compare contact residues important for Kai protein binding in the ancestral proteins vs the modern proteins. If it becomes clear that the binding interfaces between Kai proteins have co-evolved over time, the entire ancestral Kai system should be reconstructed and tested in vitro, instead of testing a mixture of ancestral KaiC and modern KaiA and KaiB as initially presented here.

Acknowledgement of contributions

Here, I would like to recognize Haneul Yoo for assisting in finalizing the candidate ancestral KaiC sequences presented here. Haneul performed the KaiC multiple sequence alignment, phylogenetic tree building, and ancestral reconstruction. I provided the Kai protein sequence database, chose the nodes for reconstruction, chose which parts of ancestral KaiC to replace with modern *S. elongatus* KaiC, compared residues within the ancestral sequences to residues in modern KaiC known to be important for function, and built the expression plasmids.

References

1. M. B. Elowitz, S. Leibler, A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-338 (2000).

2. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, Stochastic gene expression in a single cell. *Science* **297**, 1183-1186 (2002).

3. J. W. Thornton, Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat Rev Genet* **5**, 366-375 (2004).

4. J. Stricker *et al.*, A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516-519 (2008).

5. G. Lambert, J. Chew, M. J. Rust, Costs of Clock-Environment Misalignment in Individual Cyanobacterial Cells. *Biophys J* **111**, 883-891 (2016).

6. M. J. Gardner, K. E. Hubbard, C. T. Hotta, A. N. Dodd, A. A. Webb, How plants tell the time. *Biochemical Journal* **397**, 15-24 (2006).

7. R. A. Wever, Properties of human sleep-wake cycles: parameters of internally synchronized free-running rhythms. *Sleep* **7**, 27-51 (1984).

8. M. H. Vitaterna *et al.*, Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. *Science (New York, NY)* **264**, 719 (1994).

9. R. J. Konopka, S. Benzer, Clock mutants of Drosophila melanogaster. *Proceedings of the National Academy of Sciences* **68**, 2112-2116 (1971).

10. M. P. Myers, K. Wager-Smith, A. Rothenfluh-Hilfiker, M. W. Young, Light-induced degradation of TIMELESS and entrainment of the Drosophila circadian clock. *Science* **271**, 1736 (1996).

11. D. B. Bolvin, J. F. Duffy, R. E. Kronauer, C. A. Czeisler, Dose-response relationships for resetting of human circadian clock by light. *Nature* **379**, 540 (1996).

12. R. K. Barrett, J. S. Takahashi, Temperature compensation and temperature entrainment of the chick pineal cell circadian clock. *Journal of Neuroscience* **15**, 5681-5692 (1995).

13. D. A. Wheeler, M. J. Hamblen-Coyle, M. S. Dushay, J. C. Hall, Behavior in light-dark cycles of Drosophila mutants that are arrhythmic, blind, or both. *Journal of biological rhythms* **8**, 67-94 (1993).

14. K.-A. Stokkan, S. Yamazaki, H. Tei, Y. Sakaki, M. Menaker, Entrainment of the circadian clock in the liver by feeding. *Science* **291**, 490-493 (2001).

15. B. A. Reyes, J. S. Pendergast, S. Yamazaki, Mammalian peripheral circadian oscillators are temperature compensated. *Journal of biological rhythms* **23**, 95-98 (2008).

16. D. E. Somers, A. Webb, M. Pearson, S. A. Kay, The short-period mutant, toc1-1, alters circadian clock regulation of multiple outputs throughout development in Arabidopsis thaliana. *Development* **125**, 485-494 (1998).

17. A. Matsumoto, K. Tomioka, Y. Chiba, T. Tanimura, timrit Lengthens circadian period in a temperature-dependent manner through suppression of PERIOD protein cycling and nuclear localization. *Molecular and cellular biology* **19**, 4343-4354 (1999).

18. P. L. Lowrey, J. S. Takahashi, Mammalian circadian biology: elucidating genome-wide levels of temporal organization. *Annu. Rev. Genomics Hum. Genet.* **5**, 407-441 (2004).

19. Y. Taniguchi *et al.*, Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *science* **329**, 533-538 (2010).

20. M. Tigges, T. T. Marquez-Lago, J. Stelling, M. Fussenegger, A tunable synthetic mammalian oscillator. *Nature* **457**, 309-312 (2009).

21. L. Potvin-Trottier, N. D. Lord, G. Vinnicombe, J. Paulsson, Synchronous long-term oscillations in a synthetic gene circuit. *Nature* **538**, 514-517 (2016).

22. M. Nakajima *et al.*, Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science* **308**, 414-415 (2005).

23. E. Leypunskiy *et al.*, The cyanobacterial circadian clock follows midday in vivo and in vitro. *eLife* **6**, e23539 (2017).

24. M. J. Rust, S. S. Golden, E. K. O'Shea, Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator. *Science* **331**, 220-223 (2011).

25. I. Mihalcescu, W. Hsing, S. Leibler, Resilient circadian oscillator revealed in individual cyanobacteria. *Nature* **430**, 81-85 (2004).

26. A. C. Liu *et al.*, Intercellular coupling confers robustness against mutations in the SCN circadian clock network. *Cell* **129**, 605-616 (2007).

27. S. S. Golden, S. R. Canales, Cyanobacterial circadian clocks--timing is everything. *Nat Rev Microbiol* **1**, 191-199 (2003).

28. C. Troein, J. C. Locke, M. S. Turner, A. J. Millar, Weather and seasons together demand complex biological clocks. *Current Biology* **19**, 1961-1964 (2009).

29. I. Searle, G. Coupland, Induction of flowering by seasonal changes in photoperiod. *The EMBO Journal* **23**, 1217-1222 (2004).

30. M. A. Woelfle, Y. Ouyang, K. Phanvijhitsiri, C. H. Johnson, The adaptive value of circadian clocks: an experimental assessment in cyanobacteria. *Current biology : CB* **14**, 1481-1486 (2004).

31. K. Spoelstra, M. Wikelski, S. Daan, A. S. Loudon, M. Hau, Natural selection against a circadian clock gene mutation in mice. *Proc Natl Acad Sci U S A* **113**, 686-691 (2016).

32. C. H. Lowe, D. S. Hinds, P. J. Lardner, K. E. Justice, Natural free-running period in vertebrate animal populations. *Science* **156**, 531-534 (1967).

33. C. A. Czeisler *et al.*, Stability, precision, and near-24-hour period of the human circadian pacemaker. *Science* **284**, 2177-2181 (1999).

34. V. Dvornyk, O. Vinogradova, E. Nevo, Origin and evolution of circadian clock genes in prokaryotes. *Proc Natl Acad Sci U S A* **100**, 2495-2500 (2003).

35. A. Wiegard *et al.*, Biochemical analysis of three putative KaiC clock proteins from Synechocystis sp. PCC 6803 suggests their functional divergence. *Microbiology (Reading, England)* **159**, 948-958 (2013).

36. H. Kushige *et al.*, Genome-wide and heterocyst-specific circadian gene expression in the filamentous Cyanobacterium Anabaena sp. strain PCC 7120. *J Bacteriol* **195**, 1276-1284 (2013).

37. M. J. Rust, J. S. Markson, W. S. Lane, D. S. Fisher, E. K. O'Shea, Ordered phosphorylation governs oscillation of a three-protein circadian clock. *Science* **318**, 809-812 (2007).

38. Y. Nakahira, A. Ogawa, H. Asano, T. Oyama, Y. Tozawa, Theophylline-dependent riboswitch as a novel genetic tool for strict regulation of protein expression in Cyanobacterium Synechococcus elongatus PCC 7942. *Plant & cell physiology* **54**, 1724-1735 (2013).

39. A. T. Ma, C. M. Schmidt, J. W. Golden, Regulation of gene expression in diverse cyanobacterial species by using theophylline-responsive riboswitches. *Applied and environmental microbiology* **80**, 6704-6713 (2014).

40. S. W. Teng, S. Mukherji, J. R. Moffitt, S. de Buyl, E. K. O'Shea, Robust circadian oscillations in growing cyanobacteria require transcriptional feedback. *Science* **340**, 737-740 (2013).

41. H. Kageyama *et al.*, Cyanobacterial circadian pacemaker: Kai protein complex dynamics in the KaiC phosphorylation cycle in vitro. *Molecular cell* **23**, 161-171 (2006).

42. X. Y. Zheng, E. K. O'Shea, Cyanobacteria Maintain Constant Protein Concentration despite Genome Copy-Number Variation. *Cell reports* **19**, 497-504 (2017).

43. G. M. Suel, R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, M. B. Elowitz, Tunability and noise dependence in differentiation dynamics. *Science* **315**, 1716-1719 (2007).

44. J. S. van Zon, D. K. Lubensky, P. R. Altena, P. R. ten Wolde, An allosteric model of circadian KaiC phosphorylation. *Proc Natl Acad Sci U S A* **104**, 7420-7425 (2007).

45. S. Clodong *et al.*, Functioning and robustness of a bacterial circadian clock. *Mol Syst Biol* **3**, 90 (2007).

46. J. Lin, J. Chew, U. Chockanathan, M. J. Rust, Mixtures of opposing phosphorylations within hexamers precisely time feedback in the cyanobacterial circadian clock. *Proc Natl Acad Sci U S A* **111**, E3937-3945 (2014).

47. C. Phong, J. S. Markson, C. M. Wilhoite, M. J. Rust, Robust and tunable circadian rhythms from differentially sensitive catalytic domains. *Proc Natl Acad Sci U S A* **110**, 1124-1129 (2013).

48. A. Dufresne *et al.*, Genome sequence of the cyanobacterium Prochlorococcus marinus SS120, a nearly minimal oxyphototrophic genome. *Proc Natl Acad Sci U S A* **100**, 10020-10025 (2003).

49. J. Holtzendorff *et al.*, Genome streamlining results in loss of robustness of the circadian clock in the marine cyanobacterium Prochlorococcus marinus PCC 9511. *Journal of biological rhythms* **23**, 187-199 (2008).

50. C. A. Thaiss *et al.*, Microbiota Diurnal Rhythmicity Programs Host Transcriptome Oscillations. *Cell* **167**, 1495-1510 e1412 (2016).

51. V. Leone *et al.*, Effects of diurnal variation of gut microbes and high-fat feeding on host circadian clock function and metabolism. *Cell Host Microbe* **17**, 681-689 (2015).

52. J. R. Chabot, J. M. Pedraza, P. Luitel, A. van Oudenaarden, Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock. *Nature* **450**, 1249-1252 (2007).

53. S. R. Mackey, J. L. Ditty, E. M. Clerico, S. S. Golden, Detection of rhythmic bioluminescence from luciferase reporters in cyanobacteria. *Methods Mol Biol* **362**, 115-129 (2007).

54. L. R. Moore *et al.*, Culturing the marine cyanobacterium Prochlorococcus. *Limnol Oceanogr Methods* **5**, 353-362 (2007).

55. G. Lambert, E. Kussell, Quantifying selective pressures driving bacterial evolution using lineage analysis. *Phys Rev X* **5**, (2015).

56. O. Raveh, N. David, G. Rilov, E. Rahav, The Temporal Dynamics of Coastal Phytoplankton and Bacterioplankton in the Eastern Mediterranean Sea. *PLoS One* **10**, e0140690 (2015).

57. E. Rahav, M. J. Giannetto, E. Bar-Zeev, Contribution of mono and polysaccharides to heterotrophic N2 fixation at the eastern Mediterranean coastline. *Sci Rep* **6**, 27858 (2016).

58. H. Ito *et al.*, Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in Synechococcus elongatus. *Proc Natl Acad Sci U S A* **106**, 14168-14173 (2009).

59. V. Vijayan, I. H. Jain, E. K. O'Shea, A high resolution map of a cyanobacterial transcriptome. *Genome Biol* **12**, R47 (2011).

60. A. E. Carpenter *et al.*, CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* **7**, R100 (2006).

61. R. Tseng *et al.*, Structural basis of the day-night transition in a bacterial circadian clock. *Science* **355**, 1174-1180 (2017).

62. J. Snijder *et al.*, Structures of the cyanobacterial circadian oscillator frozen in a fully assembled state. *Science* **355**, 1181-1184 (2017).

63. I. M. Axmann *et al.*, Biochemical evidence for a timing mechanism in prochlorococcus. *J Bacteriol* **191**, 5342-5347 (2009).

64. D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry* **81**, 2340-2361 (1977).

65. X. Qin *et al.*, Intermolecular associations determine the dynamics of the circadian KaiABC oscillator. *Proc Natl Acad Sci U S A* **107**, 14805-14810 (2010).

66. F. Hayashi *et al.*, Stoichiometric interactions between cyanobacterial clock proteins KaiA and KaiC. *Biochem Biophys Res Commun* **316**, 195-202 (2004).

67. S. Giovannoni, U. Stingl, The importance of culturing bacterioplankton in the 'omics' age. *Nat Rev Microbiol* **5**, 820-826 (2007).

68. L. Brocchieri, S. Karlin, Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res* **33**, 3390-3400 (2005).

69. Y. I. Kim, G. Dong, C. W. Carruthers, Jr., S. S. Golden, A. LiWang, The day/night switch in KaiC, a central oscillator component of the circadian clock of cyanobacteria. *Proc Natl Acad Sci U S A* **105**, 12825-12830 (2008).

70. D. B. Forger, C. S. Peskin, Stochastic simulation of the mammalian circadian clock. *Proc Natl Acad Sci U S A* **102**, 321-324 (2005).

71. S. Di Talia, J. M. Skotheim, J. M. Bean, E. D. Siggia, F. R. Cross, The effects of molecular noise and size control on variability in the budding yeast cell cycle. *Nature* **448**, 947-951 (2007).

72. Y. Kitayama, H. Iwasaki, T. Nishiwaki, T. Kondo, KaiB functions as an attenuator of KaiC phosphorylation in the cyanobacterial circadian clock system. *The EMBO Journal* **22**, 2127-2134 (2003).

73. Q. Yang, B. F. Pando, G. Dong, S. S. Golden, A. van Oudenaarden, Circadian gating of the cell cycle revealed in single cyanobacterial cells. *Science* **327**, 1522-1526 (2010).

74. C. S. Ting, C. Hsieh, S. Sundararaman, C. Mannella, M. Marko, Cryo-electron tomography reveals the comparative three-dimensional architecture of Prochlorococcus, a globally important marine cyanobacterium. *Journal of bacteriology* **189**, 4485-4493 (2007).

75. A. T. Winfree, *The Geometry of Biological Time*. (Springer-Verlag, New York, ed. 2nd, 2000), pp. 777.

76. Y. Ouyang, C. R. Andersson, T. Kondo, S. S. Golden, C. H. Johnson, Resonating circadian clocks enhance fitness in cyanobacteria. *Proc Natl Acad Sci U S A* **95**, 8660-8664 (1998).

77. F. A. Scheer, M. F. Hilton, C. S. Mantzoros, S. A. Shea, Adverse metabolic and cardiovascular consequences of circadian misalignment. *Proc Natl Acad Sci U S A* **106**, 4453-4458 (2009).

78. M. Nakajima, H. Ito, T. Kondo, In vitro regulation of circadian phosphorylation rhythm of cyanobacterial clock protein KaiC by KaiA and KaiB. *FEBS Lett* **584**, 898-902 (2010).

79. N. Hosokawa, H. Kushige, H. Iwasaki, Attenuation of the posttranslational oscillator via transcription-translation feedback enhances circadian-phase shifts in Synechococcus. *Proc Natl Acad Sci U S A* **110**, 14486-14491 (2013).

80. Y. Kitayama, T. Nishiwaki, K. Terauchi, T. Kondo, Dual KaiC-based oscillations constitute the circadian system of cyanobacteria. *Genes & development* **22**, 1513-1521 (2008).

81. R. Pattanayek *et al.*, Visualizing a circadian clock protein: crystal structure of KaiC and functional insights. *Mol Cell* **15**, 375-388 (2004).

82. S. Ye, I. Vakonakis, T. R. Ioerger, A. C. LiWang, J. C. Sacchettini, Crystal structure of circadian clock protein KaiA from Synechococcus elongatus. *The Journal of biological chemistry* **279**, 20511-20518 (2004).

83. K. Hitomi, T. Oyama, S. Han, A. S. Arvai, E. D. Getzoff, Tetrameric architecture of the circadian clock protein KaiB. A novel interface for intermolecular interactions and its impact on the circadian rhythm. *The Journal of biological chemistry* **280**, 19127-19135 (2005).

84. Y. Xu *et al.*, Identification of key phosphorylation sites in the circadian clock protein KaiC by crystallographic and mutagenetic analyses. *Proc Natl Acad Sci U S A* **101**, 13933-13938 (2004).

85. M. Egli *et al.*, Dephosphorylation of the core clock protein KaiC in the cyanobacterial KaiABC circadian oscillator proceeds via an ATP synthase mechanism. *Biochemistry* **51**, 1547-1558 (2012).

86. T. Nishiwaki, T. Kondo, Circadian autodephosphorylation of cyanobacterial clock protein KaiC occurs via formation of ATP as intermediate. *The Journal of biological chemistry* **287**, 18030-18035 (2012).

87. H. Iwasaki, T. Nishiwaki, Y. Kitayama, M. Nakajima, T. Kondo, KaiA-stimulated KaiC phosphorylation in circadian timing loops in cyanobacteria. *Proc Natl Acad Sci U S A* **99**, 15788-15793 (2002).

88. I. Vakonakis, 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. *Proc Natl Acad Sci U S A* **101**, 10925-10930 (2004).

89. T. Nishiwaki *et al.*, A sequential program of dual phosphorylation of KaiC as a basis for circadian rhythm in cyanobacteria. *EMBO J* **26**, 4029-4037 (2007).

90. C. Brettschneider *et al.*, A sequestration feedback determines dynamics and temperature entrainment of the KaiABC circadian clock. *Molecular systems biology* **6**, 389 (2010).

91. X. Qin *et al.*, Intermolecular associations determine the dynamics of the circadian KaiABC oscillator. *Proc Natl Acad Sci U S A*, (2010).

92. J. S. van Zon, D. K. Lubensky, P. R. H. Altena, P. R. ten Wolde, An allosteric model of circadian KaiC phosphorylation. *P Natl Acad Sci USA* **104**, 7420-7425 (2007).

93. Y. G. Chang, R. Tseng, N. W. Kuo, A. Liwang, Rhythmic ring-ring stacking drives the circadian oscillator clockwise. *Proc Natl Acad Sci U S A*, (2012).

94. J. Snijder *et al.*, Insight into cyanobacterial circadian timing from structural details of the KaiB–KaiC interaction. *Proceedings of the National Academy of Sciences*, (2014).

95. Y. Kitayama, T. Nishiwaki-Ohkawa, Y. Sugisawa, T. Kondo, KaiC intersubunit communication facilitates robustness of circadian rhythms in cyanobacteria. *Nat Commun* **5**, (2013).

96. J. Monod, J. Wyman, J. P. Changeux, On Nature of Allosteric Transitions - a Plausible Model. *J Mol Biol* **12**, 88-& (1965).

97. M. Egli *et al.*, Loop-Loop Interactions Regulate KaiA-Stimulated KaiC Phosphorylation in the Cyanobacterial KaiABC Circadian Clock. *Biochemistry* **52**, 1208-1220 (2013).

98. J. Snijder *et al.*, Insight into cyanobacterial circadian timing from structural details of the KaiB-KaiC interaction. *Proc Natl Acad Sci U S A* **111**, 1379-1384 (2014).

99. L. Ma, R. Ranganathan, Quantifying the rhythm of KaiB-C interaction for in vitro cyanobacterial circadian clock. *PLoS One* **7**, e42581 (2012).

100. R. Tseng *et al.*, KaiA Assists the KaiB-KaiC Interaction and KaiB/SasA Competition in the Circadian Clock of Cyanobacteria. *J Mol Biol*, (2013).

101. S. A. Villarreal *et al.*, CryoEM and molecular dynamics of the circadian KaiB-KaiC complex indicates that KaiB monomers interact with KaiC and block ATP binding clefts. *J Mol Biol* **425**, 3311-3324 (2013).

102. S. A. Villarreal *et al.*, CryoEM and Molecular Dynamics of the Circadian KaiB-KaiC Complex Indicates That KaiB Monomers Interact with KaiC and Block ATP Binding Clefts. *J Mol Biol* **425**, 3311-3324 (2013).

103. N. E. Buchler, M. Louis, Molecular Titration and Ultrasensitivity in Regulatory Networks. *J Mol Biol* **384**, 1106-1119 (2008).

104. K. L. Brown, K. T. Hughes, The Role of Anti-Sigma Factors in Gene-Regulation. *Mol Microbiol* **16**, 397-404 (1995).

105. S. Mukherji *et al.*, MicroRNAs can generate thresholds in target gene expression. *Nat Genet* **43**, 854-U860 (2011).

106. J. K. Kim, D. B. Forger, A mechanism for robust circadian timekeeping via stoichiometric balance. *Molecular systems biology* **8**, (2012).

107. T. Y. C. Tsai *et al.*, Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science* **321**, 126-129 (2008).

108. A. Gribun *et al.*, The ClpP double ring tetradecameric protease exhibits plastic ring-ring interactions, and the N termini of its subunits form flexible loops that are essential for ClpXP and ClpAP complex formation. *The Journal of biological chemistry* **280**, 16185-16196 (2005).

109. C. T. Tang *et al.*, Setting the pace of the Neurospora circadian clock by multiple independent FRQ phosphorylation events. *P Natl Acad Sci USA* **106**, 10722-10727 (2009).

110. S. Hooshangi, S. Thiberge, R. Weiss, Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci U S A* **102**, 3581-3586 (2005).

111. T. Shibata, K. Fujimoto, Noisy signal amplification in ultrasensitive signal transduction. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 331-336 (2005).

112. S. Yerushalmi, R. M. Green, Evidence for the adaptive significance of circadian rhythms. *Ecology letters* **12**, 970-981 (2009).

113. M. Ishiura *et al.*, Expression of a gene cluster kaiABC as a circadian feedback process in cyanobacteria. *Science* **281**, 1519-1523 (1998).

114. J. Tomita, M. Nakajima, T. Kondo, H. Iwasaki, No transcription-translation feedback in circadian rhythm of KaiC phosphorylation. *Science* **307**, 251-254 (2005).

115. A. Gutu, E. K. O'Shea, Two antagonistic clock-regulated histidine kinases time the activation of circadian gene expression. *Mol Cell* **50**, 288-294 (2013).

116. V. Vijayan, R. Zuzow, E. K. O'Shea, Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proc Natl Acad Sci U S A* **106**, 22564-22568 (2009).

117. S. Diamond, D. Jun, B. E. Rubin, S. S. Golden, The circadian oscillator in Synechococcus elongatus controls metabolite partitioning during diurnal growth. *Proc Natl Acad Sci U S A* **112**, E1916-1925 (2015).

118. G. K. Pattanayak, C. Phong, M. J. Rust, Rhythms in energy storage control the ability of the cyanobacterial circadian clock to reset. *Current biology : CB* **24**, 1934-1938 (2014).

119. M. Amdaoud, M. Vallade, C. Weiss-Schaber, I. Mihalcescu, Cyanobacterial clock, a stable phase oscillator with negligible intercellular coupling. *Proc Natl Acad Sci U S A* **104**, 7051-7056 (2007).

120. T. Kondo *et al.*, Circadian clock mutants of cyanobacteria. *Science* **266**, 1233-1236 (1994).

121. T. Mori, B. Binder, C. H. Johnson, Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. *Proc Natl Acad Sci U S A* **93**, 10183-10188 (1996).

122. F. Ribalet *et al.*, Light-driven synchrony of Prochlorococcus growth and mortality in the subtropical Pacific gyre. *Proc Natl Acad Sci U S A* **112**, 8008-8012 (2015).

123. S. Takano, J. Tomita, K. Sonoike, H. Iwasaki, The initiation of nocturnal dormancy in Synechococcus as an active process. *BMC biology* **13**, 36 (2015).

124. A. M. New *et al.*, Different levels of catabolite repression optimize growth in stable and variable environments. *PLoS biology* **12**, e1001764 (2014).

125. C. C. Boutte, J. T. Henry, S. Crosson, ppGpp and polyphosphate modulate cell cycle progression in Caulobacter crescentus. *J Bacteriol* **194**, 28-35 (2012).

126. D. Chatterji, A. K. Ojha, Revisiting the stringent response, ppGpp and starvation signaling. *Current opinion in microbiology* **4**, 160-165 (2001).

127. G. Dong *et al.*, Elevated ATPase activity of KaiC applies a circadian checkpoint on cell division in Synechococcus elongatus. *Cell* **140**, 529-539 (2010).

128. G. Lambert, E. Kussell, Memory and fitness optimization of bacteria under fluctuating environments. *PLoS genetics* **10**, e1004556 (2014).

129. Y. Nakahira *et al.*, Global gene repression by KaiC as a master process of prokaryotic circadian system. *Proc Natl Acad Sci U S A* **101**, 881-885 (2004).

130. O. Shoval *et al.*, Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science* **336**, 1157-1160 (2012).

131. Y. Xu, T. Mori, C. H. Johnson, Circadian clock-protein expression in cyanobacteria: rhythms and phase setting. *EMBO J* **19**, 3349-3357 (2000).

132. Y. B. Kiyohara, M. Katayama, T. Kondo, A novel mutation in kaiC affects resetting of the cyanobacterial circadian clock. *J Bacteriol* **187**, 2559-2564 (2005).

133. O. Schmitz, M. Katayama, S. B. Williams, T. Kondo, S. S. Golden, CikA, a bacteriophytochrome that resets the cyanobacterial circadian clock. *Science* **289**, 765-768 (2000).

134. Y. I. Kim, D. J. Vinyard, G. M. Ananyev, G. C. Dismukes, S. S. Golden, Oxidized quinones signal onset of darkness directly to the cyanobacterial circadian oscillator. *Proc Natl Acad Sci U S A* **109**, 17765-17769 (2012).

135. L. Glass, M. C. Mackey, *From clocks to chaos : the rhythms of life*. Princeton paperbacks (Princeton University Press, Princeton, N.J., 1988), pp. xvii, 248 p.

136. V. K. Sharma, Adaptive significance of circadian clocks. *Chronobiology international* **20**, 901-919 (2003).

137. J. Bieler *et al.*, Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. *Mol Syst Biol* **10**, 739 (2014).

138. C. I. Hong *et al.*, Circadian rhythms synchronize mitosis in Neurospora crassa. *Proc Natl Acad Sci U S A* **111**, 1397-1402 (2014).

139. A. D. Edelstein *et al.*, Advanced methods of microscope control using muManager software. *Journal of biological methods* **1**, (2014).

140. E. Kussell, S. Leibler, Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075-2078 (2005).

141. C. R. Baker, V. Hanson-Smith, A. D. Johnson, Following gene duplication, paralog interference constrains transcriptional circuit evolution. *Science* **342**, 104-108 (2013).

142. D. P. Anderson *et al.*, Evolution of an ancient protein function involved in organized multicellularity in animals. *Elife* **5**, e10147 (2016).

143. M. A. Siddiq, D. W. Loehlin, K. L. Montooth, J. W. Thornton, Experimental test and refutation of a classic case of molecular adaptation in Drosophila melanogaster. *Nature Ecology & Evolution* **1**, s41559-41016-40025 (2017).

144. F. R. Stephenson, L. V. Morrison, C. Y. Hohenkerk, Measurement of the Earth's rotation: 720 BC to AD 2015. *Proc Math Phys Eng Sci* **472**, 20160404 (2016).

145. A. Zlenko, A celestial-mechanical model for the tidal evolution of the Earth-Moon system treated as a double planet. *Astronomy Reports* **59**, 72 (2015).

146. C. P. Sonett, E. P. Kvale, A. Zakharian, M. A. Chan, T. M. Demko, Late Proterozoic and Paleozoic Tides, Retreat of the Moon, and Rotation of the Earth. *Science* **273**, 100-104 (1996).

147. J. W. Wells, Coral growth and geochronometry. *Nature* **197**, 948-950 (1963).

148. J. W. Schopf, A. B. Kudryavtsev, A. D. Czaja, A. B. Tripathi, Evidence of Archean life: stromatolites and microfossils. *Precambrian Research* **158**, 141-155 (2007).

149. S. A. Crowe *et al.*, Atmospheric oxygenation three billion years ago. *Nature* **501**, 535-538 (2013).

150. N. M. Schmelling *et al.*, Minimal tool set for a prokaryotic circadian clock. *BMC Evol Biol* **17**, 169 (2017).

151. I. Baca, D. Sprockett, V. Dvornyk, Circadian input kinases and their homologs in cyanobacteria: evolutionary constraints versus architectural diversification. *Journal of molecular evolution* **70**, 453-465 (2010).

152. P. Dvorak *et al.*, Synechococcus: 3 billion years of global dominance. *Mol Ecol* **23**, 5538-5551 (2014).

153. V. Hanson-Smith, B. Kolaczkowski, J. W. Thornton, Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Mol Biol Evol* **27**, 1988-1999 (2010).

154. J. Lin, The University of Chicago, Ann Arbor (2016).

155. F. Hayashi *et al.*, Roles of two ATPase-motif-containing domains in cyanobacterial circadian clock protein KaiC. *Journal of Biological Chemistry* **279**, 52331-52337 (2004).

156. K. Terauchi *et al.*, ATPase activity of KaiC determines the basic timing for circadian clock of cyanobacteria. *Proc Natl Acad Sci U S A* **104**, 16377-16381 (2007).

157. L. Martin, G. B. Gloor, S. Dunn, L. M. Wahl, Using information theory to search for co-evolving residues in proteins. *Bioinformatics* **21**, 4116-4124 (2005).