THE UNIVERSITY OF CHICAGO

BIOCHEMICAL INSIGHTS INTO THE CYANOBACTERIAL CIRCADIAN CLOCK

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

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

BY

JENNY LIN

CHICAGO, ILLINOIS

AUGUST 2016

Table of Contents

List of Figures

List of Tables

Chapter 1

An Introduction to Cyanobacterial Circadian Clock System

Introduction to Circadian Biology

Circadian clocks

Given the Earth's rotation, all life on Earth is subject to daily environment fluctuations such as in temperature, light, and humidity with a cycle of 24 hours. To anticipate and prepare for these predicable changes, organisms have evolved internal timing systems called circadian clocks. These clocks generate circadian, or near 24 hour rhythms in the organism's physiology, metabolism and gene expression and can do so even in a constant environment(5). This ability to temporally organize biological activities is largely universal in all organisms with diverse examples found from unicellular organisms like cyanobacteria to fungi, algae as well as higher organisms such as plants, birds, mice, and

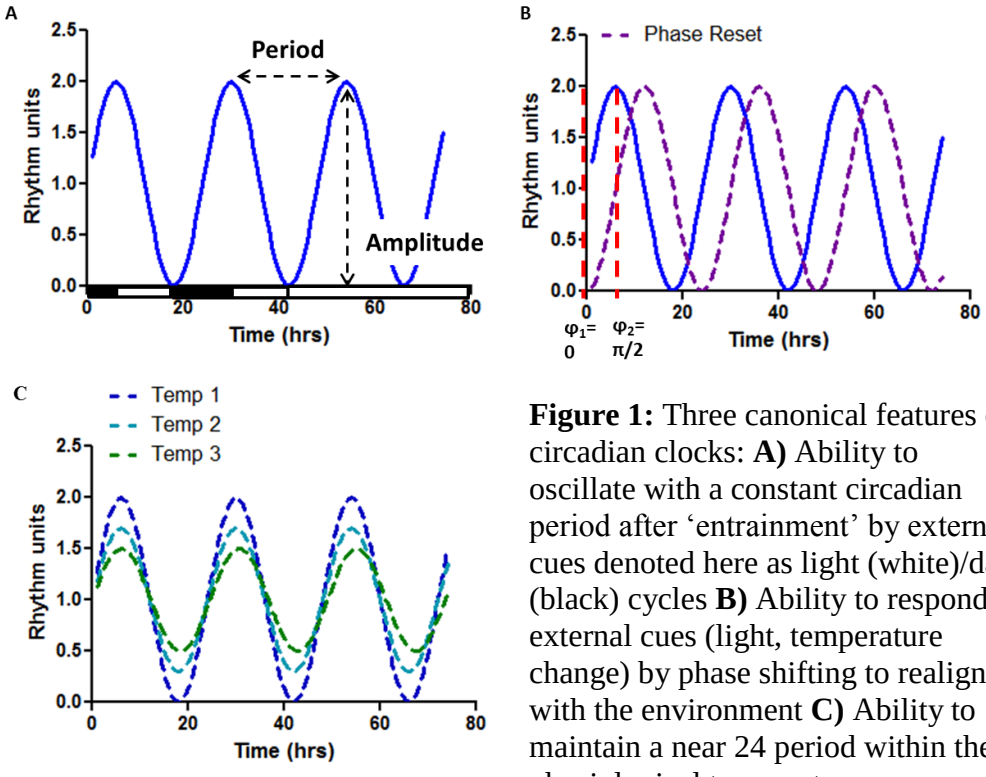


Figure 1: Three canonical features of circadian clocks: **A)** Ability to oscillate with a constant circadian period after 'entrainment' by external cues denoted here as light (white)/dark (black) cycles **B)** Ability to respond to external cues (light, temperature change) by phase shifting to realign with the environment **C)** Ability to maintain a near 24 period within the physiological temperature range or temperature compensation.

humans(2, 6). These circadian clocks system have at least one core oscillator with three canonical features: 1) the ability to autonomously produce rhythms with a near-24 hour period independent of environmental inputs, 2) have components that can sense environment cues to entrain or reset the phase of the clock, and 3) can maintain a near-24 hour period with only slight variation across the organism's physiological temperature range(7). (Fig. 1)

Studies across diverse organisms reveal a strong demand for these clocks to maintain an accurate period, evidenced by the fitness costs and health defects observed to be associated with mutations that abolish rhythmicity or change the clock period (8-10). Similarly, in prokaryotes, fitness studies in the cyanobacteria *S. elongatus* demonstrate that strains with clock periods that do not align with the environmental period and non-rhythmic strains are outcompeted by strains with matching clock periods(11). Concordant with this demand for consistency and accuracy in the period of the circadian clock, single cell studies in cyanobacteria also show that cell-autonomous rhythms in a population of 100,000 to 1,000,000 cells remain precisely in phase with each other for more than 2 weeks after entrainment with a synchronizing dark pulse, displaying extraordinary period precision and robustness to environmental fluctuations and biological noise (12)

General mechanisms of circadian clocks

In eukaryotic organisms, the core oscillators consist of positive and negative clock elements, where for example, positive elements such as WC1/2 in Neurospora, and CLOCK/BMAL1 in higher organisms, drive the transcription of negative elements such as FRQ in Neurospora and PER in mammals(2). These in turn first translocated back in to the nucleus and inhibit the positive elements, closing the negative feedback loop(2). The next cycle is restarted as negative transcript levels fall due to their repression of positive elements and by their active degradation in a phosphorylation-dependent manner(13). The delay in the repressive activity of the negative

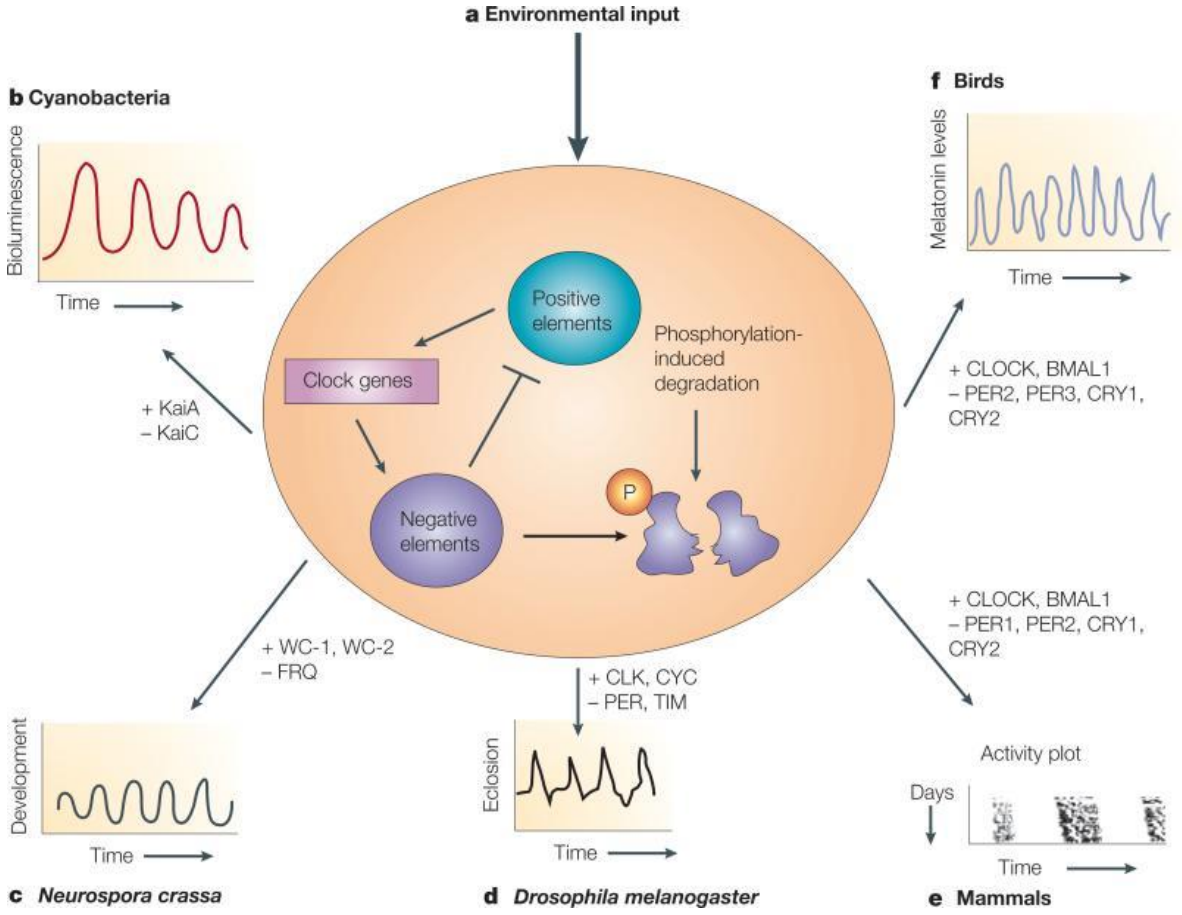


Figure 2: Negative feedback loops are the common mechanism underlying circadian rhythms across various organisms. Adapted from Bell-Perdersen *et al* (2)

element, and their later removal is essential to maintaining the circadian periodicity and depend on various phosphorylation steps which regulate their nuclear translocation and degradation (14). Output from the clock is mediated by positive elements which are basic helix-loop-helix transcription factors that directly drive rhythmic gene expression of clock regulated genes (2). While in eukaryotes, transcriptional oscillation of clock components is essential in driving circadian oscillations, prokaryotes, namely cyanobacteria, have evolved a post-translational core oscillator that is capable of driving rhythms independent of transcription(15). Studies in the model cyanobacteria, S. elongatus identified three genes, KaiA, KaiB and KaiC as the core clock components(16). Together they drive genome-wide transcriptional rhythms that are detectable in over 64% of the transcriptome(17).

Furthermore, the protein products of these three Kai genes can be purified to reconstitute the core oscillator in vitro, where in the presence of ATP, the phosphorylation levels of KaiC oscillate with a circadian period(18) (Fig. 4A). The phosphorylation of KaiC reflects the output signal which produces the genome-wide transcriptional rhythms; known mutations that change the period of KaiC phosphorylation also correspondingly cause the same period change in transcriptional rhythms(19). Additionally, all three canonical features of circadian clocks can be recapitulated in vitro using the purified components, making it the first and only clock system that can be reconstituted in vitro(18). This makes the cyanobacterial clock an ideal system to study the molecular mechanism of how oscillations with a consistent 24-hour period can be generated in a biological system, as well as how specific functional properties, such as robustness of the clock period against various physiological perturbations, can be biochemically achieved.

Biochemistry of the Cyanobacterial Circadian Clock

Previous work characterizing the KaiABC oscillator showed that each KaiC molecule consists of two internally-duplicated RecA/DnaB superfamily ATPase domains which form hexamers in an ATP dependent manner. Phosphorylation occurs at two adjacent sites (Ser431, Thr432) in its C-terminal ATPase domain (CII) (20) (21, 22)(Fig. 3). KaiC, in addition to be being an ATPase, exhibits both auto-kinase and auto-phosphatase activity in its CII domain, auto-catalyzing the circadian dynamics in its phosphorylation state(23-25).

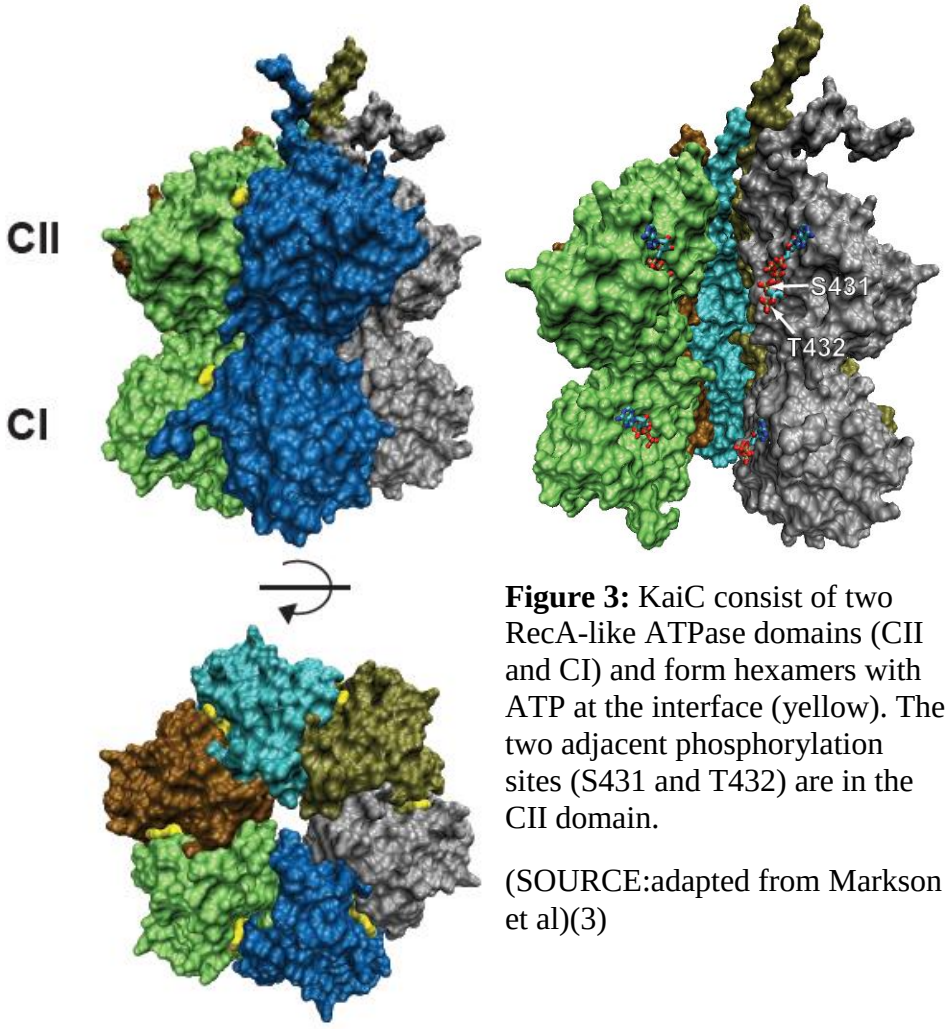


Figure 3: KaiC consist of two RecA-like ATPase domains (CII and CI) and form hexamers with ATP at the interface (yellow). The two adjacent phosphorylation sites (S431 and T432) are in the CII domain.

(SOURCE:adapted from Markson et al)(3)

Underlying these oscillations is the ability of KaiA and KaiB to regulate KaiC's auto-catalytic activities(21). During the day, KaiA will bind KaiC on the A-loop and C-terminal tail regions of

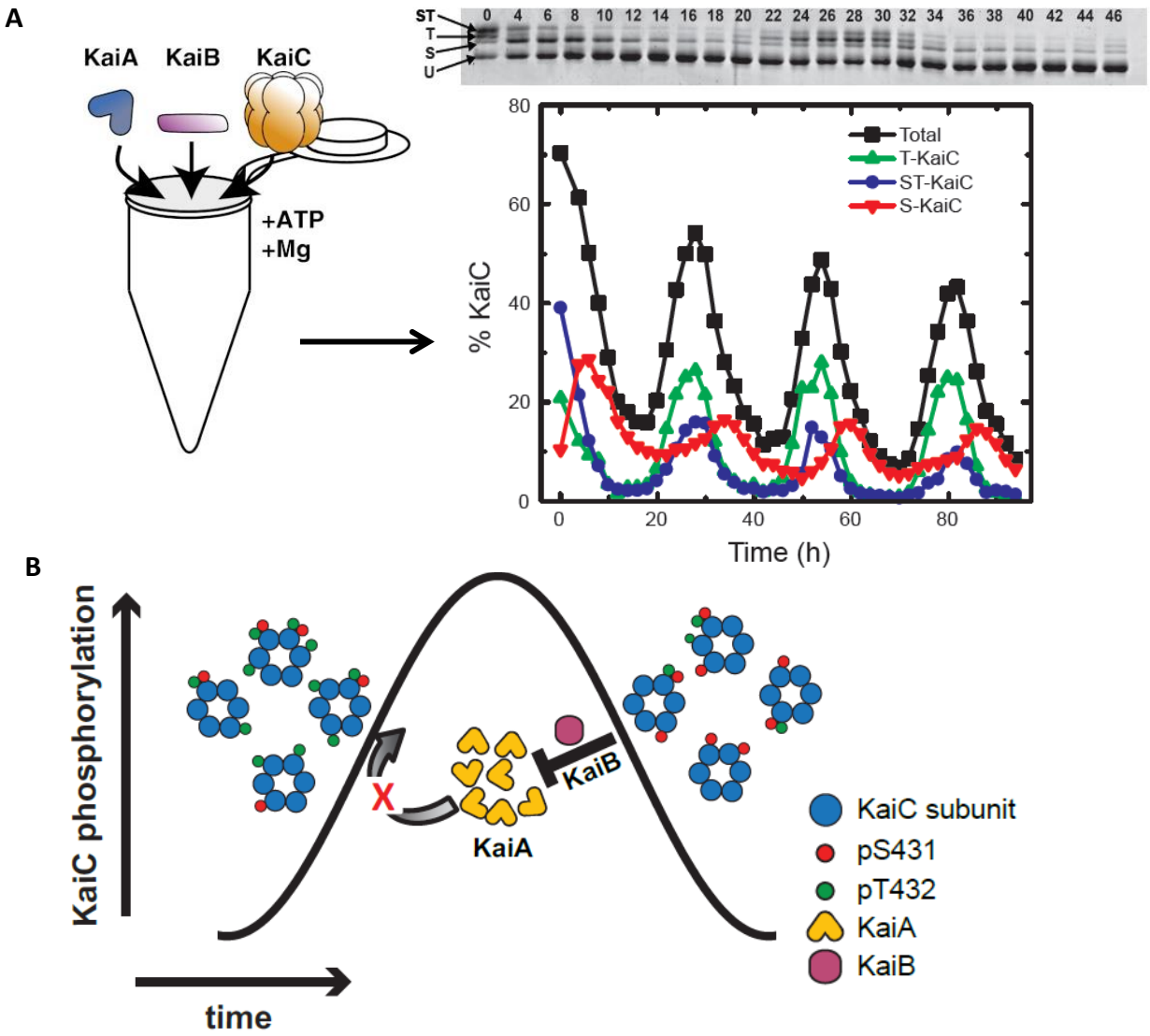


Figure 4: A) In reactions with purified Kai proteins, KaiC phosphorylation and dephosphorylation occur in an ordered manner at the two phosphorylation sites (S431, T432) with T432 reacting first, generating 4 phosphoforms (phospho-T432 = T, phospho-S431 = S, unphosphorylated = U, both phosphorylated= ST) which peak at different times during the cycle. All 4 forms can be resolved on a 10% SDS-PAGE gel (above) and quantified by densiometry (below). **B)** KaiA promotes phosphorylation and KaiB inhibits KaiA in KaiBC complexes, promoting dephosphorylation. Regulation of KaiB binding to KaiC is essential to setting the 24h period and generating oscillations

(SOURCE: adapted from *Markson et al*) (3).

the C-terminal CII domain, and activate KaiC's kinase activity(26). This allows phosphorylation to occur in an ordered manner with threonine phosphorylating to make singly threonine phosphorylated KaiC (T-KaiC) first before the serine is phosphorylated to make the doubly phosphorylated KaiC form (ST-KaiC)(21). After the accumulation of phosphorylated KaiC forms, KaiB –KaiC complexes are formed, and these bind KaiA to sequester it in ternary KaiABC complexes at the CI domain, distant from the kinase-activating KaiA binding site in CII and in a configuration that would prevent KaiA from activating KaiC's kinase activity(27) (Fig. 4B). Since alone and in the presence of KaiB, KaiC auto-phosphatases, this sequestration of KaiA also promotes dephosphorylation, closing the negative arm of the circadian clock feedback loop during the subject night. Similar to phosphorylation, dephosphorylation also occurs in an ordered manner with the threonine residue being dephosphorylated first to form the singly phosphorylated serine form (S-KaiC) before the fully dephosphorylated KaiC (U-KaiC).

One of the most essential and basic questions underlying this oscillator reaction is how are KaiC phosphorylation and dephosphorylation rates regulated so as to produce stable oscillations with a near-24 hour period? All biochemical oscillators consist of at least one negative feedback loop and in the KaiABC oscillator this negative feedback step involves the formation of KaiA inhibiting KaiBC complexes(28). Theoretically, to achieve stable oscillations, the reaction kinetics of an oscillator must be sufficiently non-linear so as to prevent

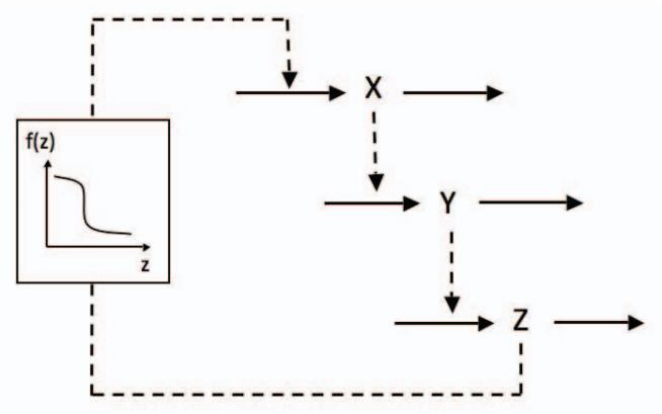


Figure 5 Goodwin oscillator model where negative feedback is sharply activated through a highly cooperative transition, driving stable oscillations. SOURCE: adapted from Gonze et al(1)

equilibrium(28). Sources of non-linearity are varied, for example, in one of the simplest biochemical oscillator models, the Goodwin oscillator, sufficient nonlinearity arises from a highly cooperative or ultrasensitive (Hill Coefficient = 8) negative feedback step to produce stable limit cycle oscillations(1) (Fig. 5).

To investigate how the biochemistry of the Kai proteins can generate oscillations, many mathematical models of the oscillator have been developed (21, 29). These models describe the negative feedback step based largely on the observations that the formation of KaiBC complexes correlate well with the abundance of S-KaiC forms, and that in binding experiments using KaiC phosphomimetics where the two phosphorylation sites are mutated either to an glutamate(E) to mimic phosphorylation or alanine(A) to mimic the unphosphorylated state, KaiB binds to the S and ST-KaiC phosphomimetic only(23, 30). In these mathematical models, KaiC cycle through the four phosphorylation states based on experimentally derived kinetic parameters in an ordered manner, and only S-KaiC can form KaiA inhibitory KaiB-KaiC complexes(21, 29). Non-linearity in the dephosphorylation/phosphorylation rates arises from positive feedback on the formation of KaiA inactivating KaiBC complexes formed from S-KaiC, where KaiA inactivation drives dephosphorylation from ST- KaiC to produce even more KaiB binding S-KaiC forms(21). However, experimental data with phosphomimetics also demonstrate that KaiB binds the ST-KaiC form to a similar extent as S-KaiC form(23), suggesting potentially an alternative or an additional source of non-linearity in the rates of KaiBC complex formation.

Additionally, while these previous Kai oscillator models can generate stable KaiC oscillations, they could not recapitulate some of the experimentally observed robustness features of the

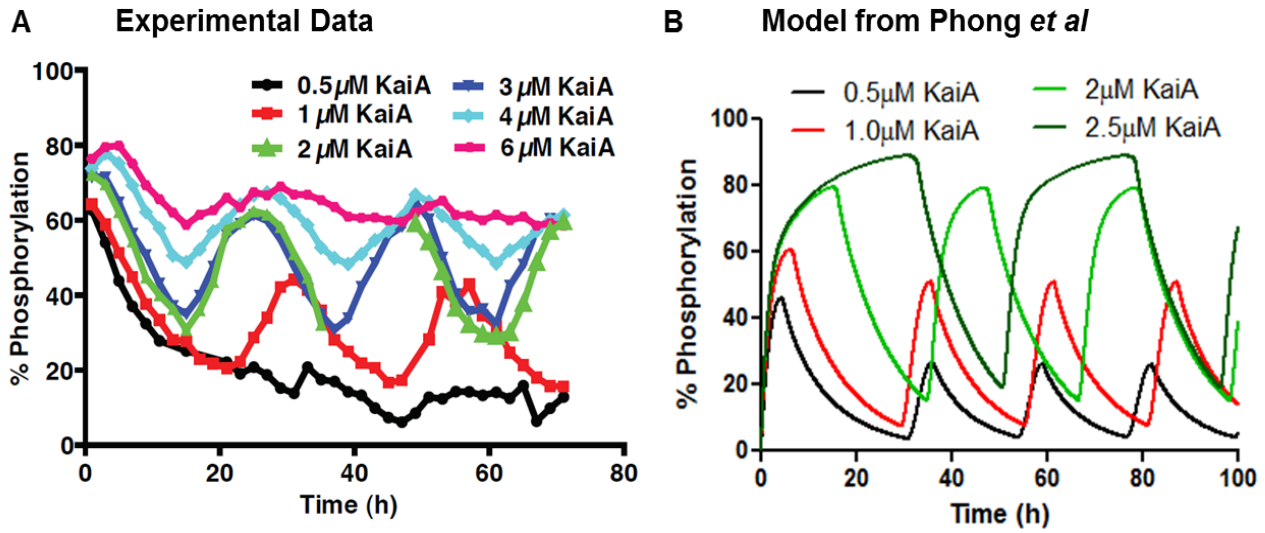


Figure 6: A) KaiC phosphorylation dynamics in clock reactions with varying concentrations of KaiA **B)** A previous clock model of the KaiABC oscillator simulated at varying KaiA concentrations

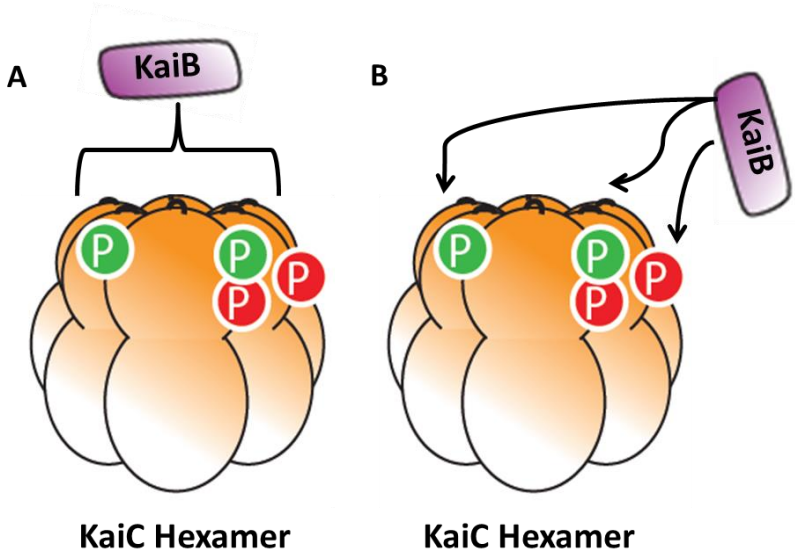


Figure 7: A) Phosphorylation state of all subunits determine a common affinity for the whole hexamer for KaiB **B)** KaiB binds each subunit independently based on their individual phosphorylation states

oscillator. For example, *in vitro* clock reactions, the concentration of KaiA can be increased 4-fold relative to KaiC and KaiB, with very little corresponding change in the period of the oscillations. This feature in particular could not be recapitulated by mathematical models of the KaiABC oscillator before this thesis work(29), which indicated an insufficiency in our biochemical understanding of the oscillator (Figure 6A, B).

From a biochemical perspective, it was also not clear if KaiC phosphomimetics, where all the subunits of a KaiC hexamers are all in the same permanent phosphorylation, accurately represent the pool of KaiC in wildtype KaiABC clock reactions. For example, in an oscillatory reaction, do wildtype KaiC hexamers contain a mixture of differentially phosphorylated subunits, and if so, is the binding affinity to KaiB independently determined by the phosphorylation state of individual subunits, or by the combined phosphorylation state of the hexamer? (Figure 7A, B) Could phosphorylation-dependent regulation of KaiB binding within the context of KaiC hexamers generate non-linearity like ultrasensitivity in the formation of KaiBC complexes and would this better recapitulate experimental data on the period robustness of the clock?

To answer the questions raised in this section, Chapter 2 of this thesis will be focused on determining the role of the two phosphorylation sites in regulating binding of KaiC to its activity modulators, KaiB and KaiA, within KaiC hexamers, and using the biochemical data to build a more accurate mathematical description of the cyanobacterial clock.

Molecular and Structural Basis of Cyanobacterial Circadian Rhythms

Many studies, including one described in chapter 2 of this thesis, focus on understanding how the specific activities of the Kai proteins interplay to generate robust circadian oscillations. Recent studies have revealed high levels of complexity in the molecular basis of the Kai clock oscillations. Outlined below is the current state of understanding on how the structures of the Kai proteins mediate essential biochemical steps.

Phosphorylation and dephosphorylating:

It is thought the periodicity of the clock is largely defined by intrinsic slow catalytic rates of the CII and CI ATPase sites(25). While it is not clear which of the two domains plays a dominant role in setting the period of the clock, slowness in the CI ATPase has been recently traced to sequestration of a lytic water and trans-cis peptide isomerization barrier during catalysis(31). However, it is unknown whether one of the two CI or CII domains plays a more dominant/tunable role in controlling the period of oscillations.

Binding of KaiA to KaiC:

The C-terminal region of KaiC contains the binding site for KaiA. KaiA activation is hypothesized to work by disengaging a hydrogen bonding network that keeps A-loops in a kinase repressive mode which then maintains the two phosphorylation residues (Ser/Thr) away from the γ-phosphate of the ATP substrate (26). Additionally, recent evidence demonstrates that KaiA may also activates KaiC's auto-kinase activity through facilitating nucleotide exchange in the ATP binding site(32). While KaiA could perform both potential mechanisms of action, the molecular basis of its action is unknown. For example, is there a network of finely-tuned interactions that connect the KaiA binding site on the C-terminal tail to the ATP binding site as well as the phosphorylation residues?

Binding of KaiB to KaiC:

Biochemical studies demonstrate that the catalytic activities of both ATPase domains of KaiC are essential for oscillations, and both play a role in regulating negative feedback or the formation of KaiB-KaiC complexes in the system. Binding of KaiB to KaiC requires both the correct CII phosphorylation state, as well as catalytic activity in the CI domain(29). NMR structural studies using two separately expressed CI and CII domains suggest that while KaiB

binds the CI domain of KaiC, in an area around the N-terminal B-loops, it requires the stacking of the CII ring in the correct phosphorylation state(33). Structurally, phosphorylation on Ser431 increases rigidity in the CII ring and stacking of the CII to the CI ring increases flexibility in the CI ATPase ring(34). However, it is unknown at the molecular level, how the stacking of the CII domain or the ATPase activity of the CI changes the conformation of the CI domain at the KaiB binding interface to allow for binding. Comprehensive mutagenesis studies of allosteric proteins such as of the PDZ binding domain, have shown that their function (in this case, binding specificity) is controlled from an allosteric site by a network of mutationally sensitive, physically connected, coevolving sectors regions(35). Does a similar network of mutationally sensitive residues mediate allosteric changes between the phosphorylation sites and KaiB binding site of KaiC as well?

Despite these recent advances, many open questions still remain about how the function of the Kai protein is actually encoded in its sequence and structure. For example, while structural and limited mutagenesis data do suggest sites of interaction between KaiC, KaiA and KaiB, without solved structures of the complexes, the binding sites and molecular mechanisms of KaiA kinase activation, KaiB binding and how KaiBC complexes bind and inactivate KaiA, remain largely undetermined. Additionally, while targeted structural studies can reveal how certain biochemical processes are mediated, but it is unclear if we can identify all the structural regions that are involved in the process. It is also uncertain if we have discovered all the biochemical processes mediated by the Kai structures which underlie the function of the cyanobacterial clock. Hence a more comprehensive and direct study of the structural/sequence-function relationships is necessary to fully understand the molecular basis of the cyanobacterial clock function, and how the circadian periodicity of the clock is exactly encoded in the sequence.

Despite the biochemical tractability of the Kai clock system, much remains unknown about the details and principles underlying the molecular architecture of the Kai biochemical oscillator. In Chapter 3, I will be introducing a novel screen technique using high throughput RNA-sequencing to screen a comprehensive library of Kai protein mutants to generate hypotheses for the questions posed above.

High resolution mutagenesis as a tool for biochemical hypothesis generation

Further understanding of the structural basis of function can be investigated by comprehensive mutagenesis screens to identify constraints on the chemical nature of each residue, and revealing their molecular function. On a global perspective, the mutational tolerance of each position may indicate the level of functional constraint on the amino acid identity of the position. While conservation score is often used to identify positions that are critical for function, functions that are unique to a single organism, such as adaptive to its specific environment, or structural features that function only in the context of its own sequence and coevolved binding partners are not detected. Hence, large scale mutagenesis may reveal novel structural

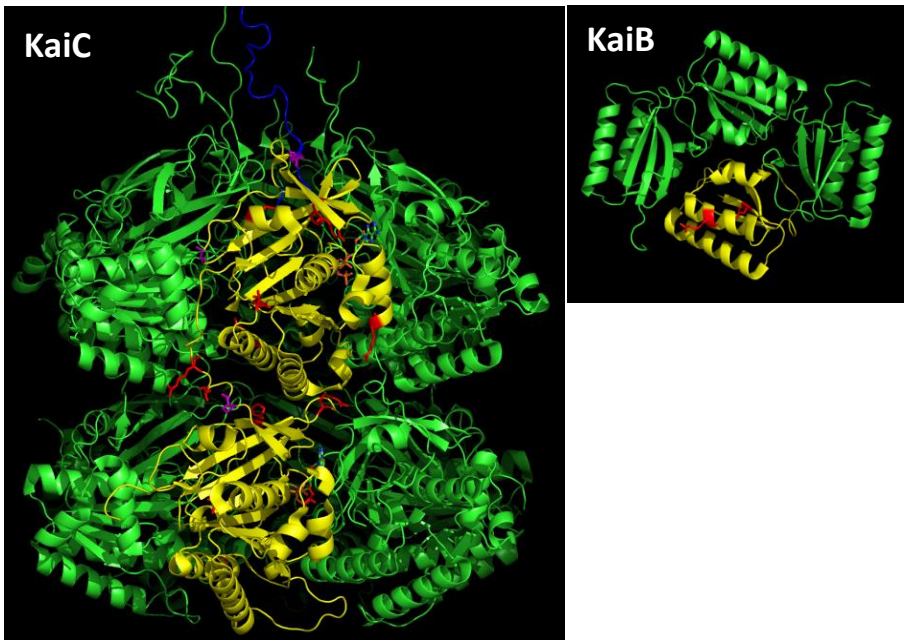


Figure 6: Structures of KaiB and KaiC with the positions of known perturbative mutations (16 shown). Red indicates period-changing mutations, while purple indicates arrhythmic mutants.

elements that are important for the oscillatory output of the Kai oscillatory system, as well as confirm and extend our existing knowledge about how essential biochemical steps in the Kai clock reaction are structurally mediated. For example, high levels of mutational sensitivity at surface exposed positions on KaiC may suggest an interaction site with KaiA or KaiB and lead to hypotheses of how phosphorylation can regulate these binding sites. In terms of determining how period is determined by the structure of the Kai proteins we can examine the spatial pattern of period mutants. So far, non-saturating low-throughput mutagenesis screens have found around 20 oscillation perturbing mutations in the three Kai genes, but these mutations are scattered throughout the three proteins without any clear structural pattern (36)(Fig. 8). A more comprehensive map could lead us to novel hypotheses of how the clock protein structures set the period of oscillations. For instance, a network organization or regional clustering of period mutants may indicate a specific biochemical process that dominantly sets the period of clock and a dispersed distribution may indicate that multiple processes contribute to setting the period of oscillations.

Additionally, the data also hold information about the evolutionary constraints on the Kai proteins for example, general robustness qualities of the sequence against mutations such as higher robustness against less desirable perturbation (i.e. period changes versus arrhythmia).

Beyond the cyanobacterial circadian clock, this high-throughput technique can be used to screen the mutational effects of any auto-regulatory transcriptional feedback system, which include all circadian clocks. This technique would be especially powerful in dissecting the function of multiple phosphorylation sites in eukaryotic clock proteins whose roles are largely unknown, by assessing single mutations or combinatorial mutations in a local sequence region(37).

Chapter 3 of thesis will be focused on developing a method to perform high-throughput saturating mutagenesis are every site on the KaiB and KaiC genes, to generate hypotheses in the areas described above.

Main Objectives

Chapter 2: Mixtures of opposing phosphorylations within hexamers precisely time feedback in the cyanobacterial circadian clock

1. Profile composition of KaiC phosphoform that bind KaiB in wildtype clock reactions and determine how the two KaiC phosphorylation sites regulate KaiB binding in the context of KaiC hexamers. Does KaiB bind KaiC subunits independently, or is KaiB binding affected by the composite phosphorylation state of a KaiC hexamer?
2. Assess phosphorylation-dependent feedback regulation on KaiA activity in the context of KaiC hexamers
3. Develop a hexameric clock mathematical model with terms describing hexamer-level regulation of KaiB and KaiA binding to KaiC. Evaluate the ability of this model to recapitulate period robustness of the oscillation across varying KaiA concentration and explore the functional role of the two phosphorylation sites

Chapter 3: Development of a high-throughput deep sequencing assay for clock mutants

1. Develop workflow for high through-put screening of a saturating single codon mutant library of KaiB and KaiC for circadian clock function including protocols for single-codon mutagenesis, large scale transformation and growth of cyanobacteria, and Illumina sequencing sample preparation.

2. Perform small-scale pilot sequencing to test developed workflow, identity potential weakness, bottleneck or sources of error. Identity and test measures to improve data quality for sequencing the full KaiB-KaiC mutageneic library.
3. Using the pilot sequencing data, develop an analysis pipeline for initial sequencing data processing including quality control, alignment, and variant calling. Explore biological content in the pilot data: spatial distribution of mutational sensitivity and specific phenotypic changes (period and amplitude), compare mutational sensitivity with conservation score/conserved residues to determine level of correlation (sequence-independent constraints) and regions of difference (sequence/organism specific constraints)
4. Perform preliminary analysis of sequencing data for half the KaiB-KaiC mutageneic library, define strategies to complete the sequencing screen

Chapter 2

Mixtures of Opposing Phosphorylations within Hexamers Precisely Time Feedback in the Cyanobacterial Circadian Clock

(Published in Lin *et al*, PNAS, September 6, 2014)

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 (5). 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 (11, 38).

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* (18, 27, 39). 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* (40-42), 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 (24, 43, 44). 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 (22). KaiC autokinase and autophosphatase activities occur at the same active site (45, 46). 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 (19, 47). Roughly speaking, kinase activity predominates during the day, and phosphatase activity predominates during the night (23). Thus, understanding the feedback mechanisms that generate a precise time delay between these modes is crucial to understanding timing in the oscillator (26).

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 (21, 48). 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 (21, 30, 49). 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 (21, 23, 48). 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 (29, 33).

Because of the importance of precisely timing negative feedback via KaiB•KaiC complex formation for generating appropriate rhythms (30), 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 (33, 50). 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. 9A). 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 (21, 23). 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 (21, 23, 30, 39). In contrast, in the 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. 9A).

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.8A). 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. 9B,8B). 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 (46). 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. 9C). This monomerization and rehexamerization procedure does not compromise the ability of the wildtype protein to oscillate (Fig. 11A). 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. 11B-C).

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. 9D). 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. 9E, 9F-G). 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 (51).

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. 9F, 9E).

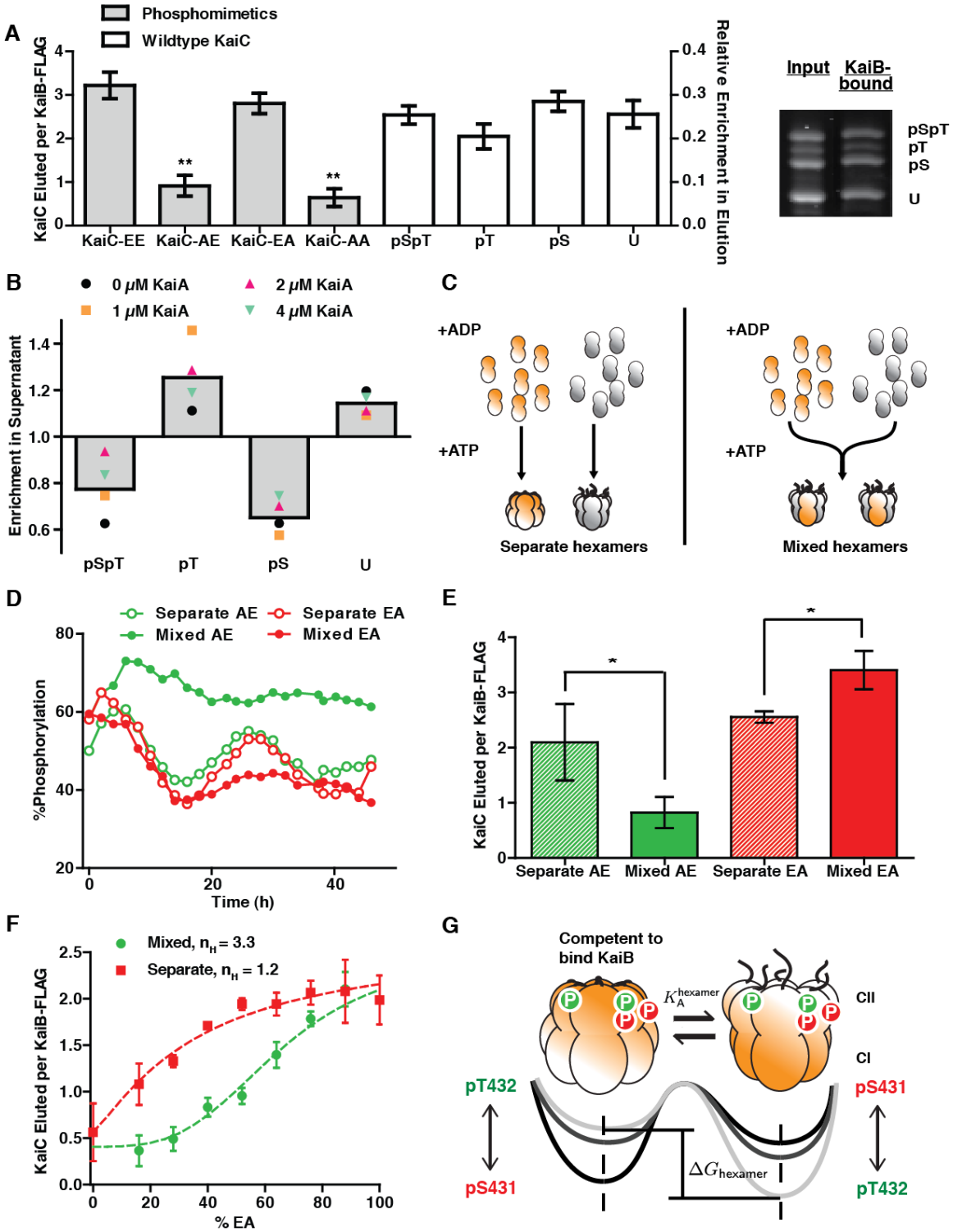


Figure 7: KaiB-KaiC interaction favors KaiC hexamers with appropriate mixtures of phosphorylated subunits. (**A**) *Left:* Phosphorylation site mutants in homogeneous hexamers co-IPed by KaiB-FLAG (*grey bars, right axis*). Error bars represent standard error of three replicates after 24 h incubation. 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-

(Figure 9 continued) 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 $p < 0.01$. (**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 His_6-tagged wildtype KaiC (His_6-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 n_H = 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.

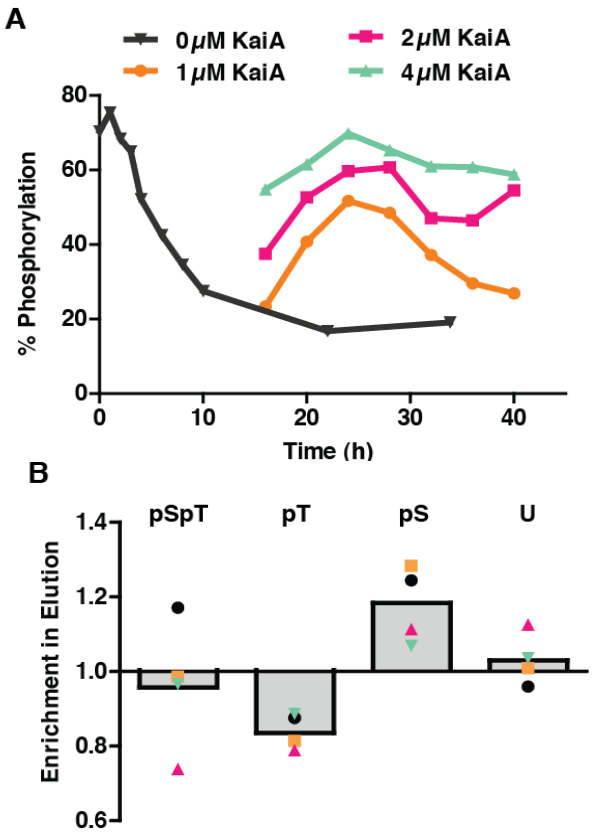


Figure 8: KaiC phosphorylation dynamics in reactions used for KaiB-FLAG co-immunoprecipitation.(**A**)Time course of KaiC phosphorylation in reactions with KaiB-FLAG and 0 μM (black *symbols*), 1 μM (*orange symbols*), 2 μM (*magenta symbols*), and 4 μM (*cyan symbols*) KaiA showing sampled reaction timepoints used for immunoprecipitation. (**B**) Enrichment of each KaiC phosphoform in the material bound to KaiB-FLAG relative to the input in clock reactions. Colored symbols show an average from 7 timepoints taken over 24 h (1-4 μM KaiA) or 9 timepoints taken over 34 h (0 μM KaiA). Bars show averages over all KaiA concentrations and timepoints.

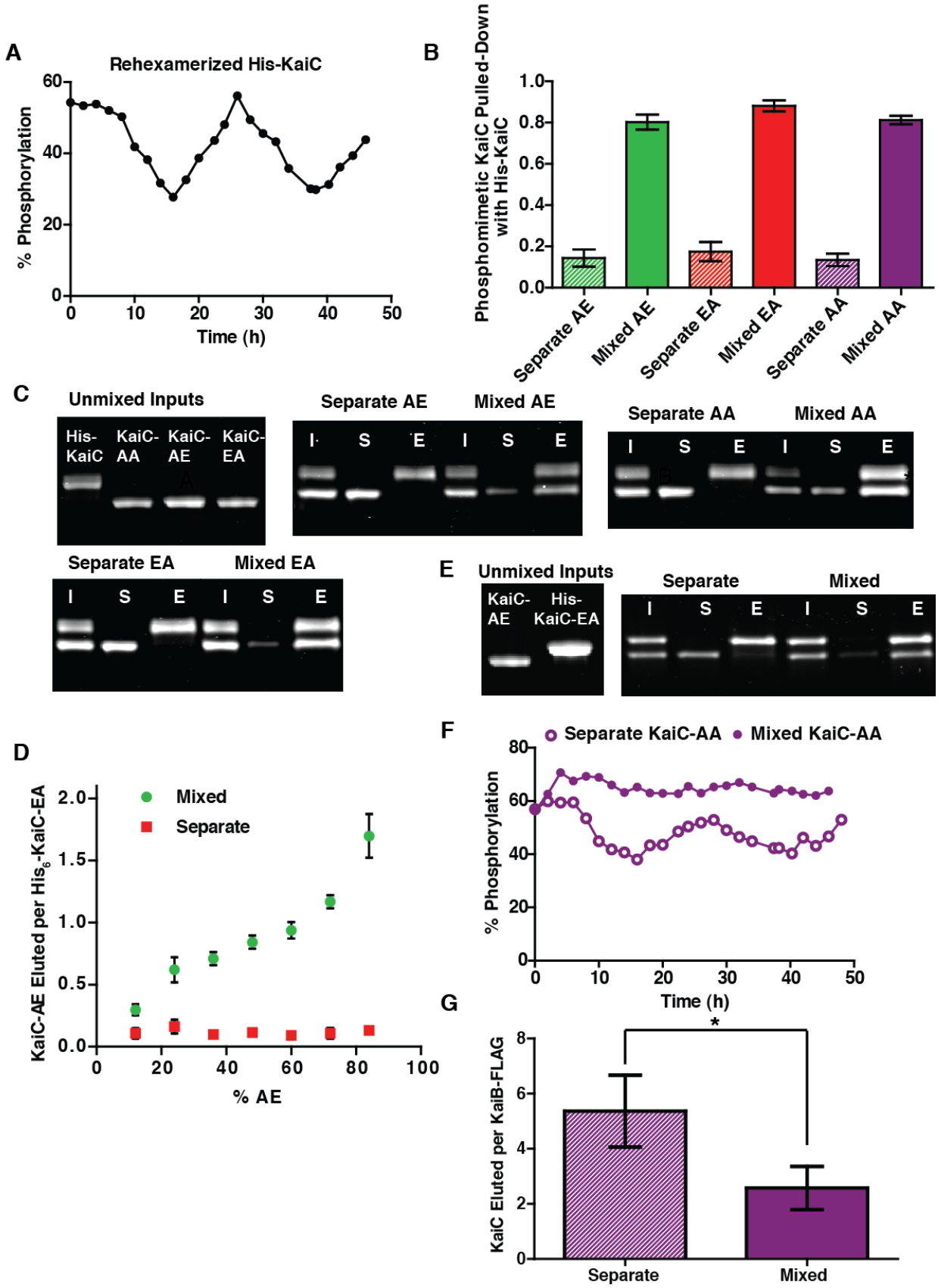


Figure 9: Mixing monomers before re-hexamerization generates more hexameric mixing than mixing hexamers and mixing of KaiC-AA into wildtype hexamers induces hyperphosphorylation

(Figure 11 continued) and inhibits KaiB binding (**A**) Phosphorylation dynamics of re-hexamerized His_6-tagged KaiC in a standard clock reaction. (**B**) Amount of mixing of phosphomimetic KaiC with His_6-KaiC resulting from each preparation averaged over 48 h during clock reactions, shown as a ratio of the densiometry signal in the pulled down material for each phosphomimetic to His_6-KaiC. Error bars indicate the standard deviation over the time course. (**C**) Sample gel images of input (I), supernatant (S) and elution (E) from His_6-tag pull downs of the mixed and separate preparations of KaiC-AA, KaiC-EA and KaiC-AE with His_6-KaiC at t = 0 h of the clock reactions. A sample gel image of each KaiC species alone is also shown. (**D**) Amount of mixing between KaiC-AE and His_6-KaiC-EA at varying input percentages of KaiC-AE from each preparation as an average of three timepoints. Error bars indicate the standard deviation over the time course. (**E**) Sample gel image of input (I), supernatant (S) and elution (E) from His_6-tag pulldowns of the mixed and separate preparations of KaiC-AE with His_6-KaiC-EA. A sample gel image of each KaiC species alone is also shown. (**F**) Phosphorylation dynamics of wildtype His_6-tagged KaiC in clock reactions containing either mixed hexamers of KaiC-AA and wildype KaiC or separate KaiC-AA and wildtype KaiC hexamers at a 1:1 molar ratio. (**G**) Amounts of total KaiC co-IPed by KaiB-FLAG in clock reactions containing either mixed or separate KaiC-AA and wildtype KaiC hexamers. An average over the first dephosphorylation phase (6-15 h) is shown. Error bars indicate standard deviation over the time course. * indicate $p<0.05$ from the Student's *t*-test.

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. 9F). 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 (21, 23). 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 (52).

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 (26, 33, 50, 53). 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 (54).

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. 9G). 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. 9F) 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. 12A-C). 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. 12D).

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. 12D). This mixing-dependent effect indicates that phosphorylation on Ser431 acts allosterically in the KaiC 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 (55) and that phosphomimetic mutation at Ser431 makes the KaiA-binding "A loops" inaccessible to proteolytic cleavage (56).

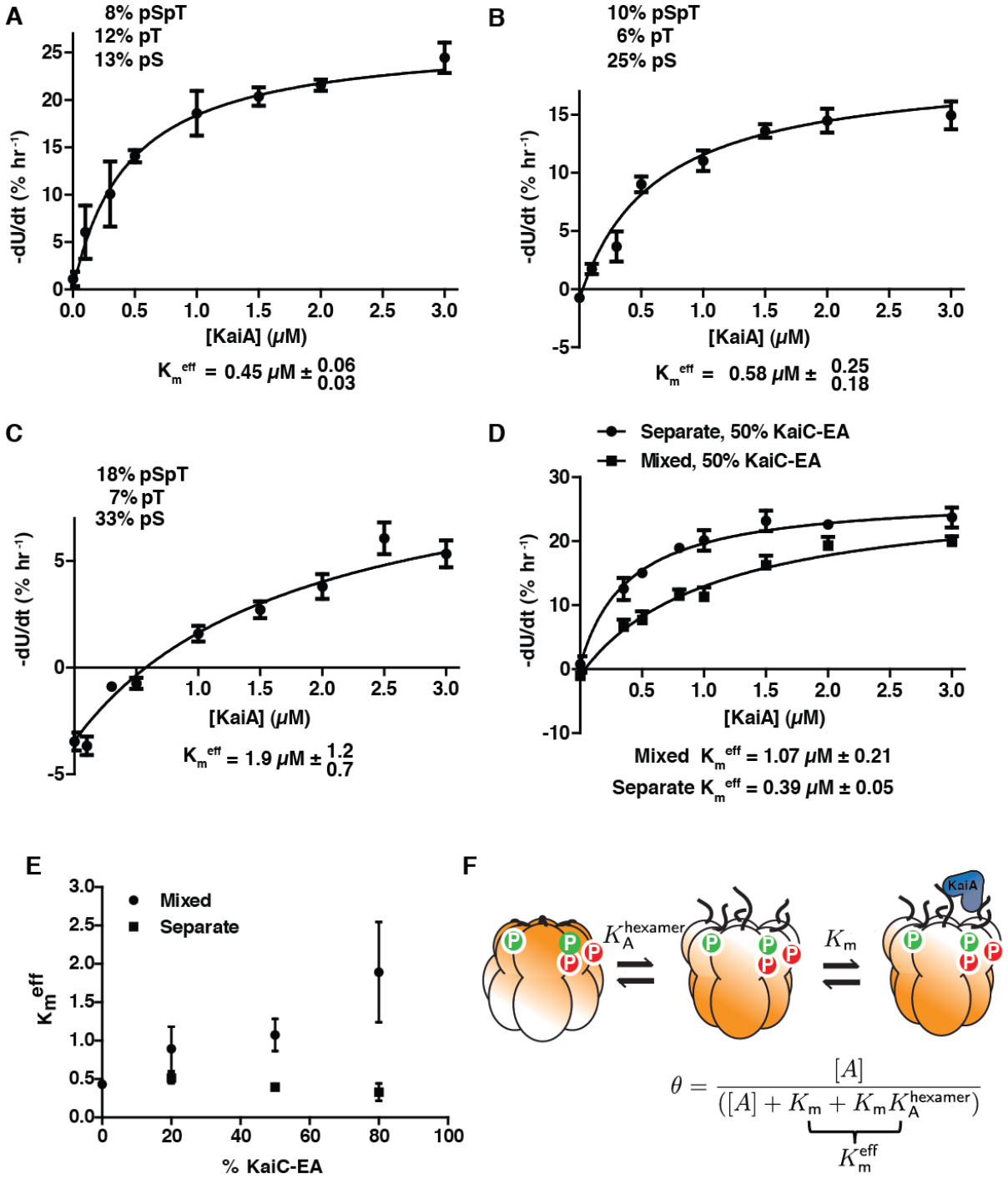


Figure 10: KaiC hexamers with heavy Ser431 phosphorylation are less sensitive to KaiA. (**A**)-(**C**) Rates of KaiA-stimulated KaiC autokinase activity as a function of [KaiA] for various initial phosphorylation states. Fits are to a modified Michaelis-Menten equation with baseline: $V_i = V_{max}[KaiA]/([KaiA] + K_m^{eff}) + V_{dephos,}$ 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 K_A, the allosteric equilibrium constant, and hence also $K_m^{eff} = K_m(1 + K_A)$.

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. 12E)

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 bind KaiB even when the phosphorylation state is held fixed, causing KaiC to resist interaction with KaiB (Fig. 13A). 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. 13B). 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 (42). Similar results held for the KaiC-EA mutant (Fig. 14A).

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 (26). As predicted, this mutation causes severe defects in KaiB interaction (Fig. 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

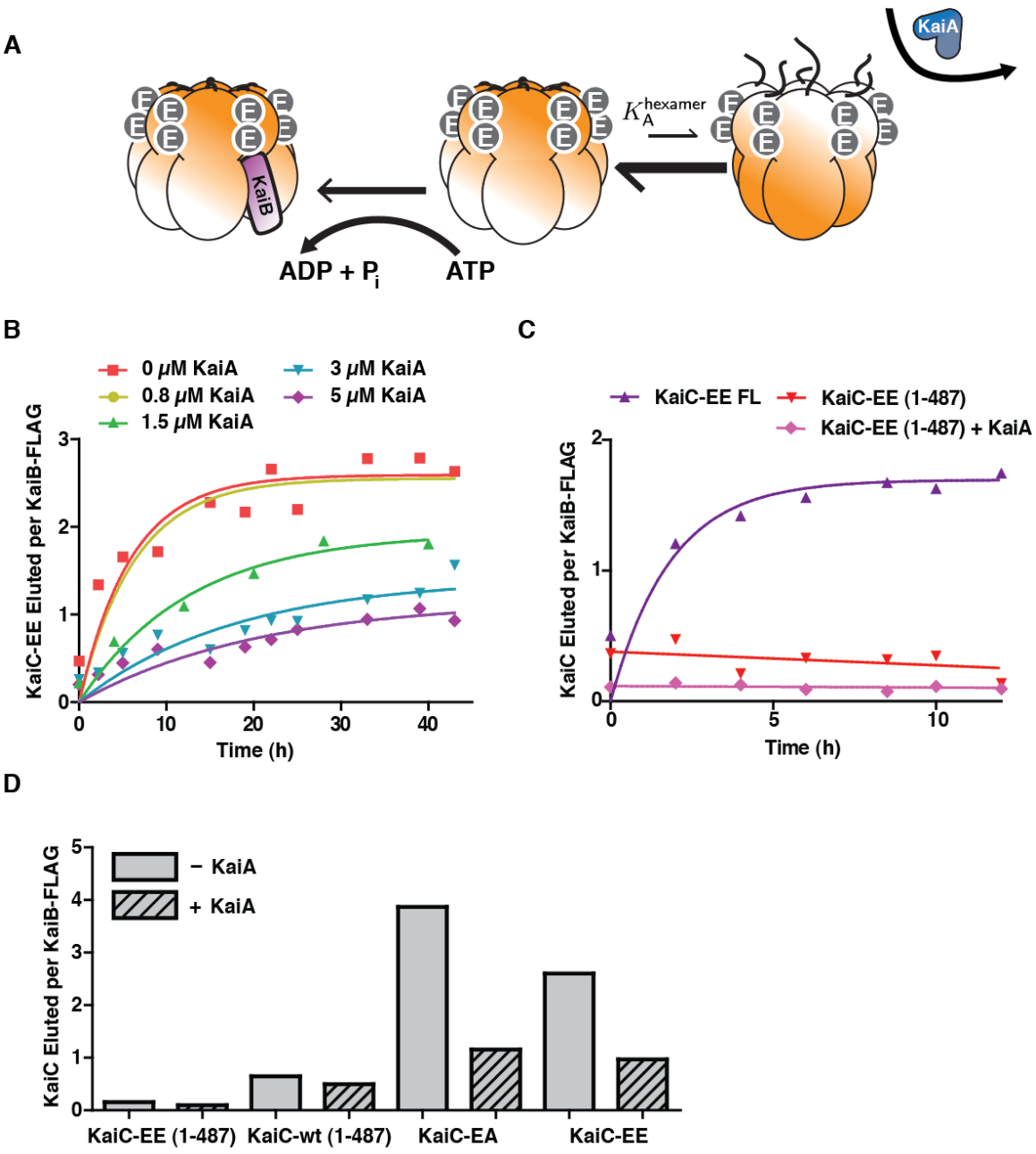


Figure 11: KaiA allosterically stabilizes a KaiC state that KaiB cannot bind. (**A**) Cartoon of allosteric state selection by KaiA and KaiB. Stimulation of the kinase active mode by KaiA shifts the allosteric equilibrium away from KaiB binding. (**B**) Timecourse of KaiC-EE co-IPed by KaiB-FLAG in the 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

(Figure 13 continued) (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.

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. 14B-E).

The mutations made here and the known KaiA binding site on KaiC are distant from proposed KaiB-KaiC interaction sites (33, 57). 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.

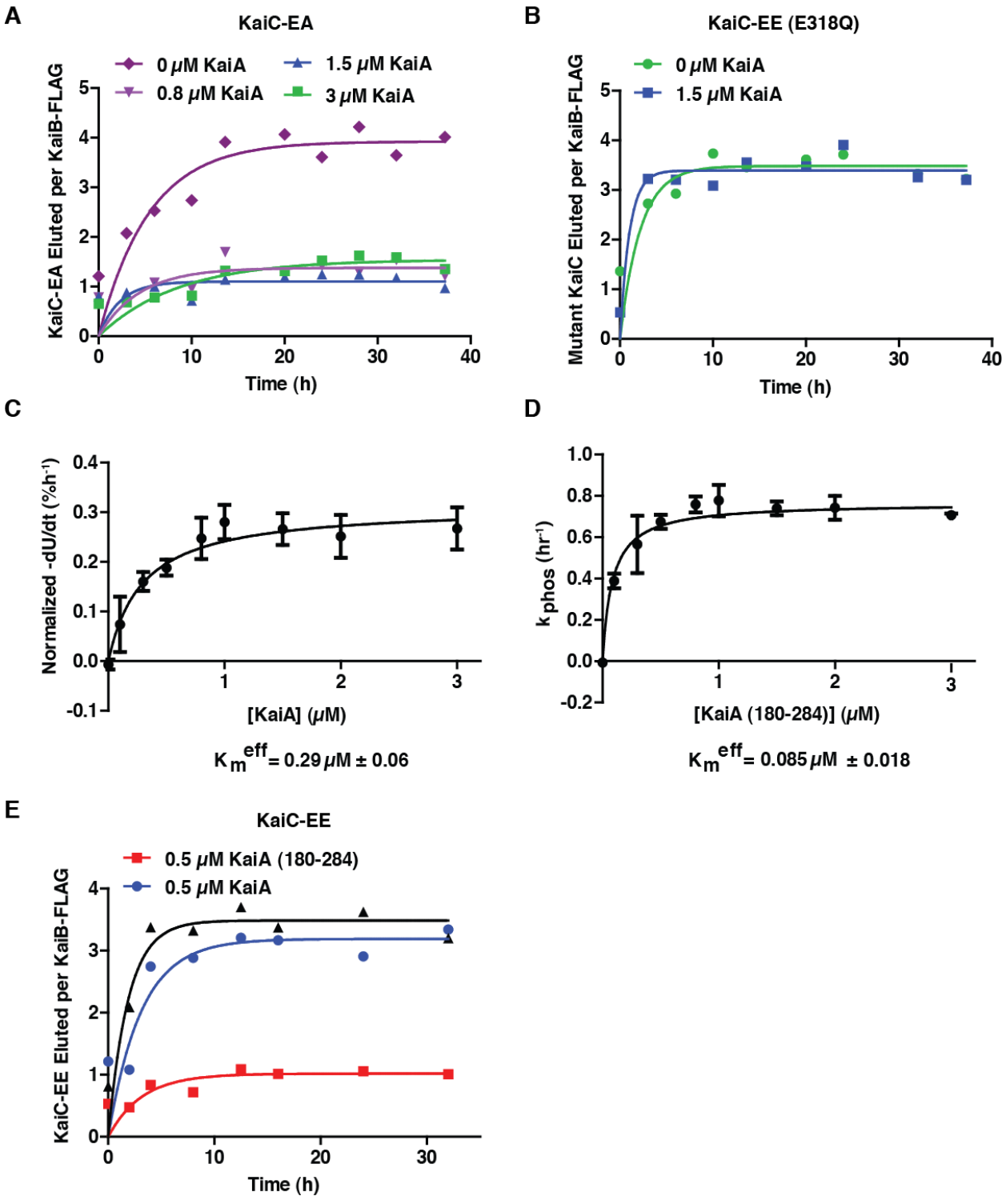


Figure 12: Effect of mutations on KaiA-dependent inhibition of the KaiB-KaiC interaction
(**A**) Time course of KaiC-EA co-precipitated with KaiB-FLAG in the presence of various concentrations of KaiA. Plotted are densitometry ratios of KaiC-EA to KaiB-FLAG in the precipitated material. (**B**) Time course of KaiC-EE E318Q (CII catalytic site mutation) co-precipitated with KaiB-FLAG with (*blue symbols*) and without (*green symbols*) 1.5 μM KaiA.

(Figure 14 continued) (**C-D**) K_m^{eff} for N-terminally truncated hyperactive KaiA (180-284) or for wildtype KaiA acting on nearly unphosphorylated KaiC. (**E**) Time course of KaiC-EE co-precipitated with KaiB-FLAG in the presence of either 0.5 μM KaiA (*blue symbols*), 0.5 μM KaiA (180-284) (*red symbols*), or 0 μM KaiA (*black symbols*).

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 (21). 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. 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 into KaiA•KaiB•KaiC complexes at a stoichiometry of one KaiA dimer to one KaiC subunit, consistent with recent structural work (58). 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 (21, 29). Thus, our allosteric oscillator model shares features with previous treatments of allostery in the KaiABC system (49), 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 (40) (Fig. 15B). 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. 15C).

Remarkably, the range of protein concentrations over which this model can generate circadian rhythms is nearly the same as the experimental system (Fig. 15D). The role of KaiA in stabilizing the allosteric state that cannot bind KaiB helps to enhance the robustness of the period in this model. 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. 15E). 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. 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.

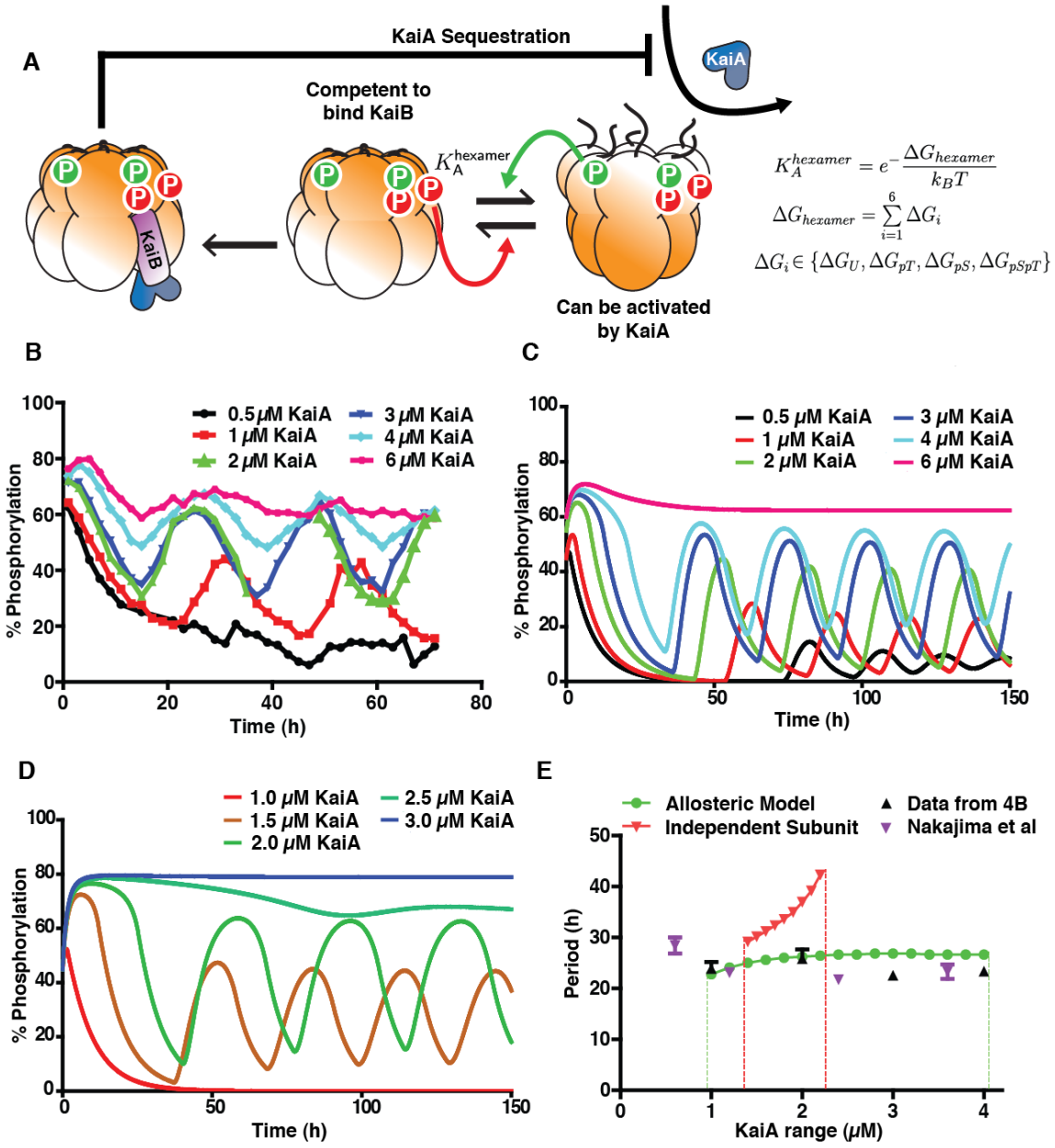


Figure 13: The allosteric model predicts the experimentally observed robustness of the oscillator period to changes in component concentrations. (**A**) Allosteric multisite phosphorylation clock model: KaiC hexamers switch in a concerted fashion between a state competent for KaiB binding and a state 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

(Figure 15 continued) 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.

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. 16A). These findings parallel the enrichment and depletion for pSer431 and pThr432 respectively that we observed experimentally for KaiC hexamers interacting with KaiB (*cf.* Fig 7B). 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 7B). 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. 16B).

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.

16C). 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. 16C). 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. 16E), while the spread in period seen over this KaiA range decreases (Fig. 16E, 15D). Both of these effects are correlated with increased ultrasensitivity of KaiB-KaiC complex formation (Fig. 17A-C). 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. 16D-E).

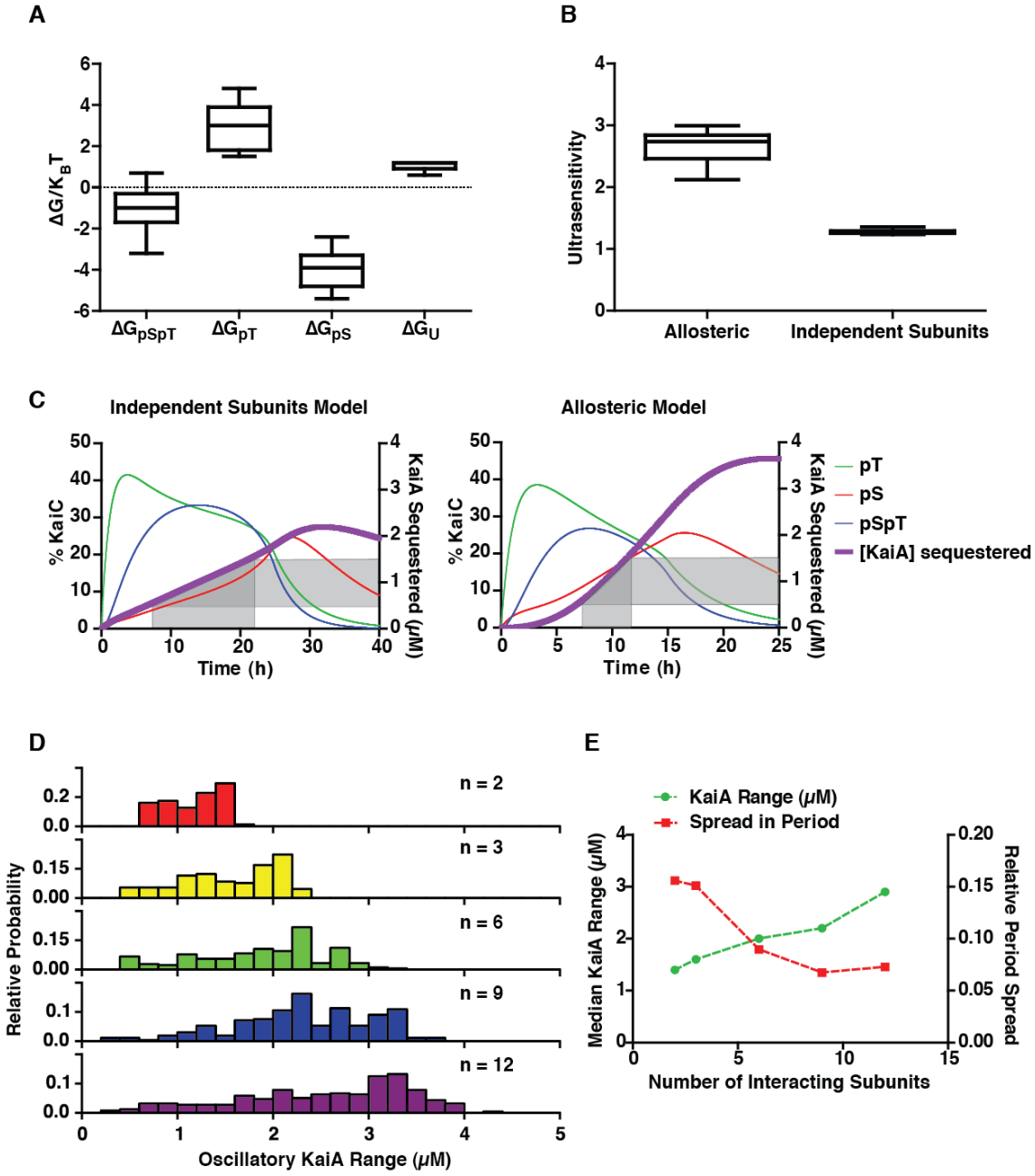


Figure 14: Opposing effects of pSer431 and pThr432 on the allosteric equilibrium produces an ultrasensitive switch in negative feedback necessary for a robust period. (**A**) Box and whisker plot of the free energy (ΔG) distribution associated with each subunit phosphorylation state for parameter sets 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),

(Figure 16 continued) 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.

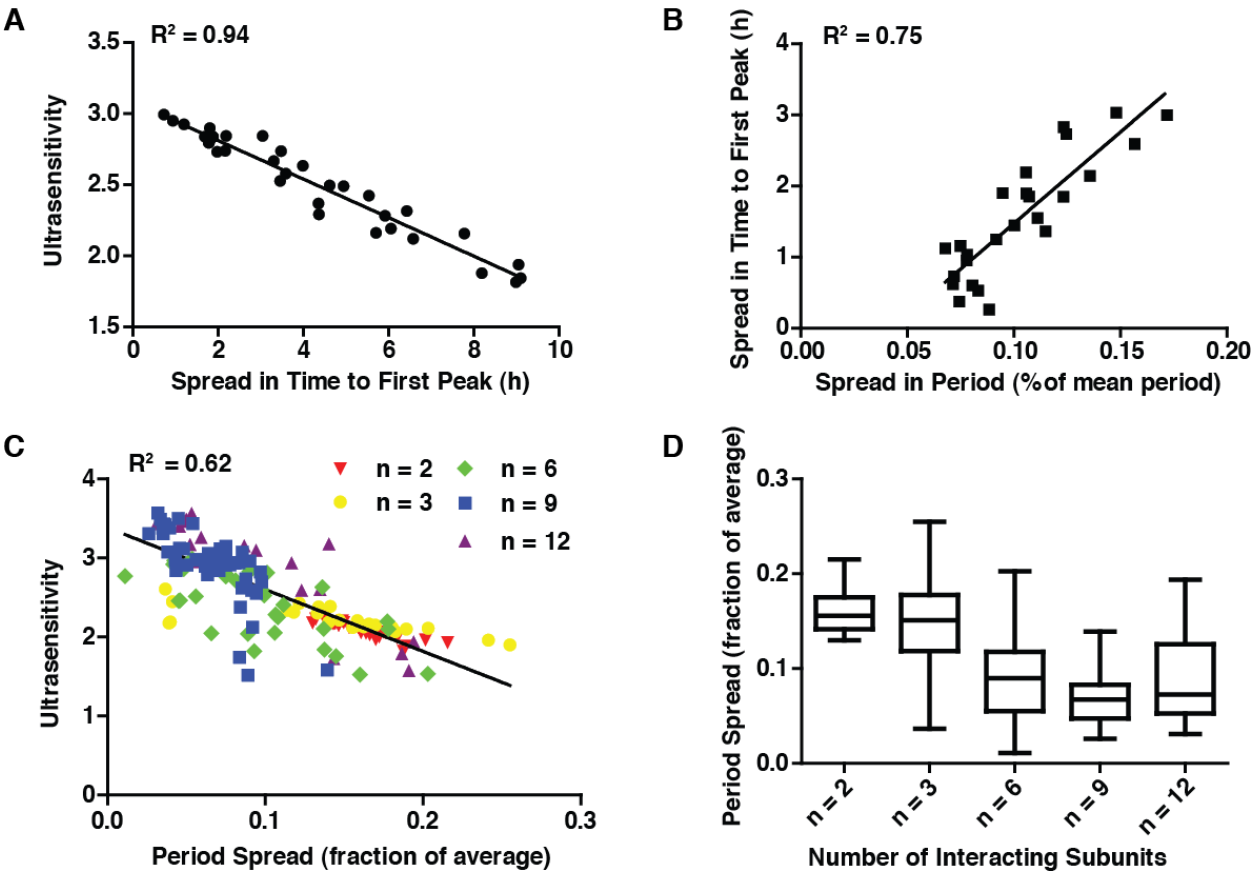


Figure 15: Ultrasensitivity in KaiB-KaiC complex formation gives a smaller spread in time delays before the onset of dephosphorylation allowing for oscillations with a robust period over a larger range of KaiA concentrations; the effect depends on the number of molecules in an allosterically interacting unit.**(A)** Correlation between the sigmoidal kinetics of KaiB-KaiC formation (effective Hill coefficient) and the spread in waiting time to the first phosphorylation peak over the oscillatory KaiA range in clock simulations initiated from 0% KaiC phosphorylation. **(B)** Correlation between the spread in waiting time to the first phosphorylation peak and the standard deviation in oscillatory period over the KaiA concentration range which supports stable oscillations (period spread). Analysis done with ΔG parameter sets that produce

(Figure 17 continued) stable oscillations over a maximal KaiA range (≥ 2.8 μM). **(C)** Correlation between the effective Hill coefficients for KaiB-KaiC formation over time with period spread as defined in (B) for clock models with different numbers of allosterically interacting subunits. **(D)** Distribution of the spread in oscillatory period as defined in (B) for models with different numbers of allosterically interacting subunits. Analysis done with ΔG parameter sets that produce circadian (22 h – 29 h period) oscillations at 1.5 μM KaiA (circadian selection).

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 (59), regulation of transcription through sigma/anti-sigma interactions (60), and microRNA-mRNA buffering (61). 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 (62).

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 (63).

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 (64). 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 (65). 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 (29). 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.

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(21). 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 of this PNAS publication.

Reactions with artificially mixed KaiC hexamers

Monomerization of KaiC was carried out following Nishiwaki *et al.* (46). 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. 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 of this PNAS publication.

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. 10A-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. 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 of this PNAS publication.

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 of the PNAS publication. 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 (21, 29). 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.

Updated Insights and Future Directions

One potential alternative explanation for the cooperative KaiB binding response with respect to changing S-KaiC levels (Fig. 9F) is that this is not necessarily due to the opposing effect of T-KaiC and S-KaiC subunits in KaiC hexamers, but rather due to the oligomeric structure of KaiB. KaiB in previous structural studies have been found to be a tetramer(44). Potentially, as S-KaiC populates the KaiC ring, a single KaiB subunit can bind and increase the local concentration of KaiB subunits due to tetrameric interactions, allowing them to bind to other S-KaiC subunits, mirroring how interacting DNA binding proteins cooperatively bind multiple DNA elements(66). This is model is however is not supported by several recent biochemical findings: a) native mass spectrometry data suggest that KaiB binding occurs with 6:6 ratio to KaiC hexamers with almost no intermediate stoichiometries except a small population of 1:6 KaiB:KaiC complexes, and b) KaiB at *in vitro* reaction concentrations are mostly monomers, and c) KaiB binds KaiC in the CI domain as monomers in a fold-switch conformation that has a distinct secondary structure than KaiB in tetramers(50, 67). All these information suggest that KaiC likely undergoes a symmetric allosteric conformation switch which creates 6 binding sites for 6 KaiB monomers as depicted in my model and where the hexameric structure of KaiC determines complex stoichiometry rather than the number of S-KaiC subunits in a ring or the KaiB oligomeric structure. Both the modeling results (Fig. 16A) and the enrichment pattern of KaiC phosphoforms (Fig. 9B) does indicate that opposing action of the S-KaiC and T-KaiC phosphoforms, where T-KaiC must destabilize the KaiB binding hexameric state of KaiC more than the U-KaiC form, is necessary to recapitulate the oscillatory properties of the oscillator. Thinking theoretically, having equally

opposing U-KaiC and T-KaiC may prevent sufficient dephosphorylation, and limit the amplitude of oscillations and/or prevent oscillations. To confirm this, a similar experiment as in Figure 9F could be repeated for the KaiC-AA (U-KaiC) mimetic mixed with S-KaiC.

What still remains unclear about the mechanism of KaiBC complex formation is how 6 KaiB is able to cooperatively bind KaiC as monomers. This may be a low probability event contributing to the slow kinetics of KaiB binding, an interaction which takes hours to reaches equilibrium. Speculatively, at least 3 KaiB dimers or 6 KaiB monomers must bind a KaiC hexamer at once. Adding to the complexity is the fact that KaiB binding kinetics are not dependent on KaiB concentrations, suggesting that in fact additional rate limiting steps exist involving non-KaiB dependent steps (29). For example, KaiB fold switching may occur in a cooperative manner as induced by symmetric KaiC CI-ATPase dependent transitions in KaiC, and the combined phosphorylation state of KaiC determines whether this induction occurs. In this sense, the free energy assignments described above may not necessarily represent the relative thermodynamic differences between states but could reflect kinetic shifts in the processes involved in KaiB binding or CI/CII communication, for example, in the rates of CI catalysis or CI-CII ring stacking, to increase the lifetime of KaiB in the fold-switch state or the initial KaiB-KaiC complex, and hence the likelihood of forming a stable KaiB-KaiC complex. In this scenario, the CI ATPase reaction, whether or not it is linked to KaiB fold switching, may act as the final irreversible and KaiB-independent step to translate kinetic differences in differentially phosphorylated KaiC hexamers into large differences in the amount of KaiB binding, and complete a potential kinetic proof reading pathway to achieve high levels of specificity in negative feedback. Testing of this hypothesis is currently ongoing in the lab.

Chapter 3

Development of a High-Throughput Deep Sequencing Assay for Clock Mutants

Background and Significance

The circadian clock in cyanobacteria consists of just three gene, KaiA, KaiB, and KaiC, and together these three proteins outputs a signal through a downstream two-component pathway which drives transcriptional rhythms across the cyanobacteria genome with a near 24 hour period. This clock is able to function as a post-translational oscillator, where the three purified proteins *in vitro* can recapitulate all three hallmark features of circadian clocks. Leveraging the clock system's biochemical tractability, previous biochemical studies have revealed a very complex set of biochemical processes and interactions which produce the oscillatory clock output(3). However, given the lack of Kai complex crystal structures, and the dynamic nature of the reaction, it is not yet known about how the structure and sequence of the three proteins mediate the biochemical processes involved in generating the circadian periodicity of the oscillator.

Mutagenesis is a powerful and widely used technique in assessing both the functional role of a specific position or region on a structure. Screening multiple amino acid substitutions at a position additionally gives information about the chemical constraints on that position for function. Applied to the Kai proteins, phenotypic information from a saturating (each amino acid to every other) library of Kai gene mutants would yield biochemical constraints for every sequence position, which when combined with existing structural data, would allow testing and generation of hypotheses about the molecular basis of circadian function, as well as answer the

very fundamental question of how rhythmicity and circadian periodicity is encoded in the Kai gene sequences.

While screening each member of this kind of comprehensive mutant library individually would be infeasible, in this chapter, I will describe a newly developed deep sequencing technique that will allow high-throughput screening of the Kai gene mutants.

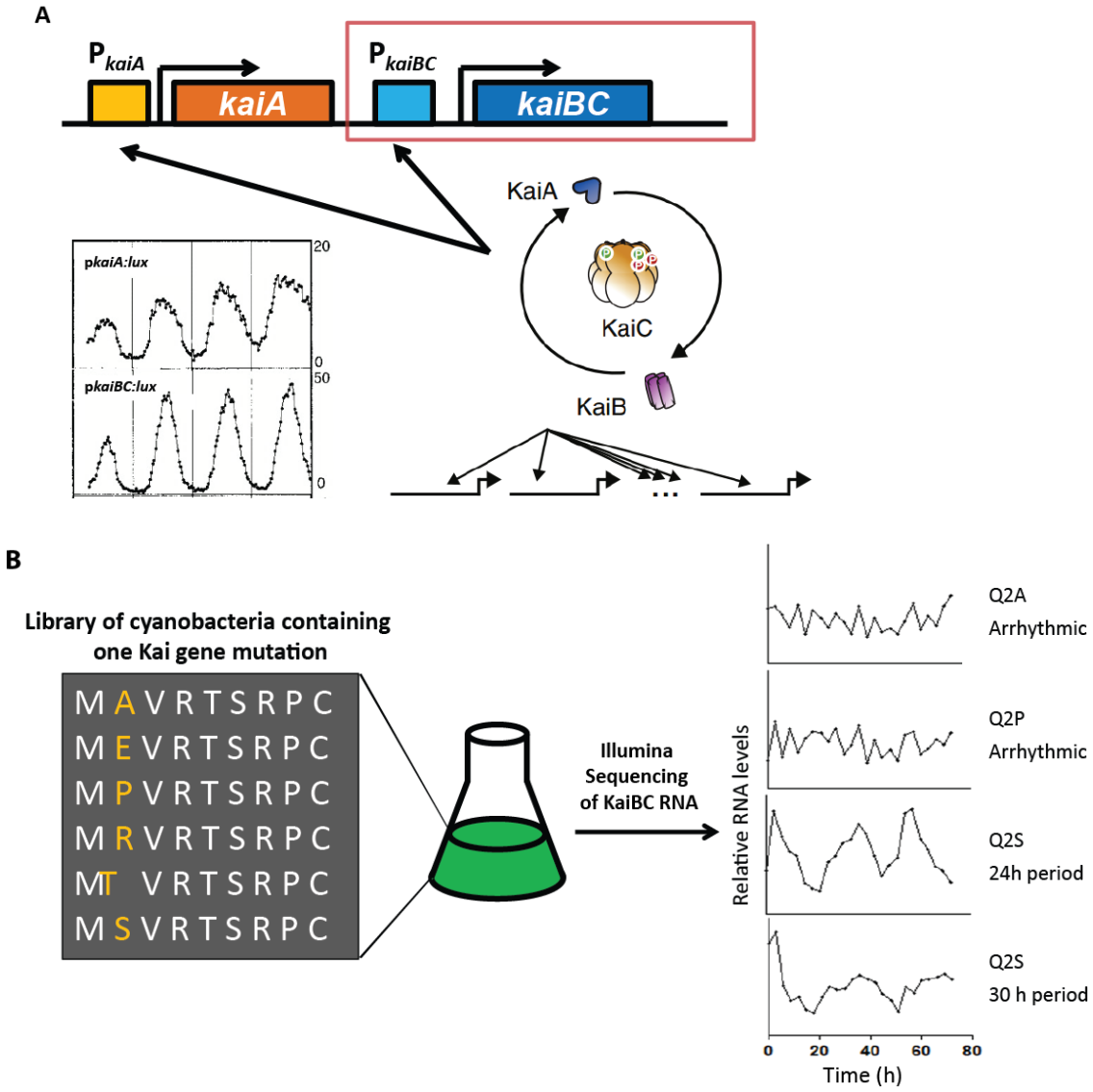


Figure 16: A) Kai gene products function in post-translation clock which drives circadian rhythms in Kai gene expression. Red box indicate the assayed *kaiBC* cotranscript. **B)** Clock phenotype of single codon mutants can be assess by Illumina sequencing of the *kaiBC* gene over time in high throughput manner.

Using transcriptional feedback for Kai mutant screening

Aside from regulating the transcriptional oscillations of loci across the cyanobacterial genome, the Kai post-translational oscillator, in a cell-autonomous manner, also drives the rhythmic expression of the Kai gene transcripts, forming a transcriptional feedback loop(36, 68). This transcriptional feedback loop or circadian oscillation in Kai gene expression is dependent on the function of the post translational oscillator, where known mutations that abolish or change the period of oscillations in the phosphorylation level of KaiC will also cause the same phenotype at the level of gene expression from the Kai gene promoter(18). In this way, in each cyanobacterial cell, the dynamics of the Kai gene transcripts acts a reporter for the ability of the Kai protein products to function properly in the post-translational clock. This then allows us, to generate a library of cyanobacteria with mutations in the Kai genes, and use high-throughput Illumina sequencing of the Kai transcript RNA over time to identity mutants and track the amount of that mutant transcript over time, simultaneously genotyping and phenotyping a library of mutants in one sequencing run (Fig. 18A). This technique is not only high throughput, but allows the evaluation of all Kai mutants in a internally controlled experiment and culture, limiting artificial phenotypic effects caused by variable growth or experimental conditions (Fig. 18B). For the development and testing of this technique, mutagenesis and screening will be limited to the KaiB-KaiC co-transcript, as this displays higher amplitude transcriptional rhythms than KaiA, and is better characterized experimentally.

Materials and Methods

Workflow overview

This high throughput mutations screen workflow will consist of three main steps: mutant cyanobacterial library construction, clock function assessment time course and sampling, and the custom preparation of timepoint samples for Illumina sequencing (Fig.19). These will be explained in detail below.

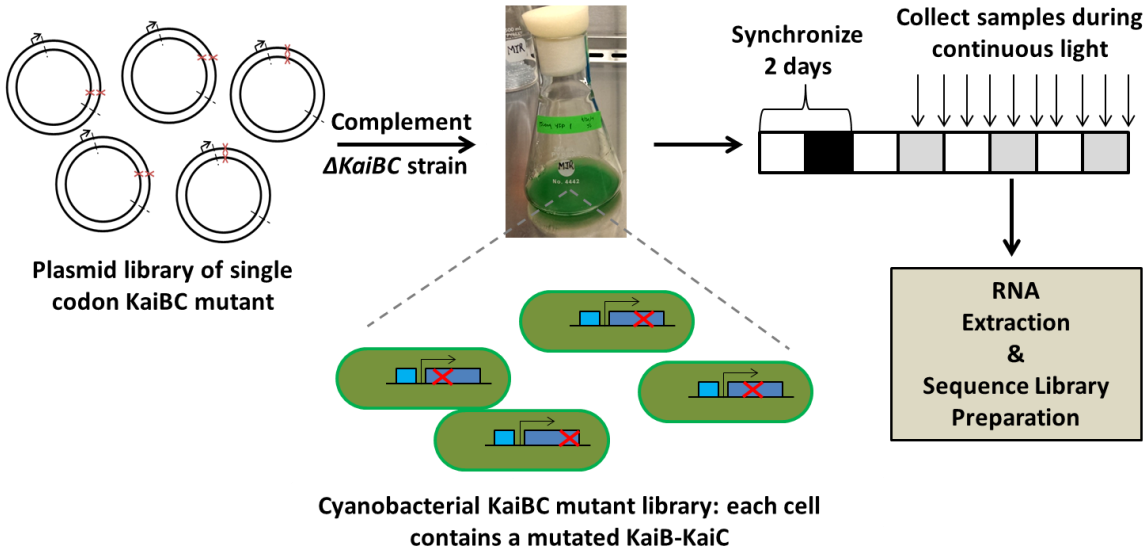


Figure 17: Overall screening workflow – 1. Mutagenesis of *kaiBC* and subcloning into a genomic integration vector pAM2134; 2.Complement a Δ*kaiBC* strain with the plasmid library and collect transformants into a single culture; 3. Entrainment of the library culture by 2X 12h light- 12h dark synchronizations cycles; 4. Prepare Illumina sequencing sample from total cellular RNA.

Mutagenesis: To generate a library of all single amino acid mutations in the KaiB and KaiC genes, we used PFunkel mutagenesis, which allows single mutations at different sequence position to be made in one reaction(69) (Fig. 20). Briefly, the method uses a uracil containing single-stranded plasmid template derived from M13K07 helper phage proliferated in *e.coli* K12 CJ326 (*ung-1*$^{-}$*/dut-1*$^{-}$). The complementary strand is synthesized from a mutagenesis primer by a polymerase with uracil read through ability (*pfu* Turbo Cx), and is then circularized with *Taq* ligase. The second mutant strand is synthesized from a second non-mutagenic forward primer, displaying the wildtype strand which is then degraded by uracil DNA glycosylase and exonuclease III. To mutate each amino acid to the 20 others, we used one degenerate 'NNR' (N = A/T/C/G, R = C/G) substitution mutagenesis primer for each codon position (except terminal STOPs in KaiC and KaiB). The 'NNR' codon substitutions allow us to access all amino acids while excluding two out of the three STOP codons. To make multiple single codon substitutions at in one reactions but minimize the rate of double mutations, the primer to template ratio was maintained at

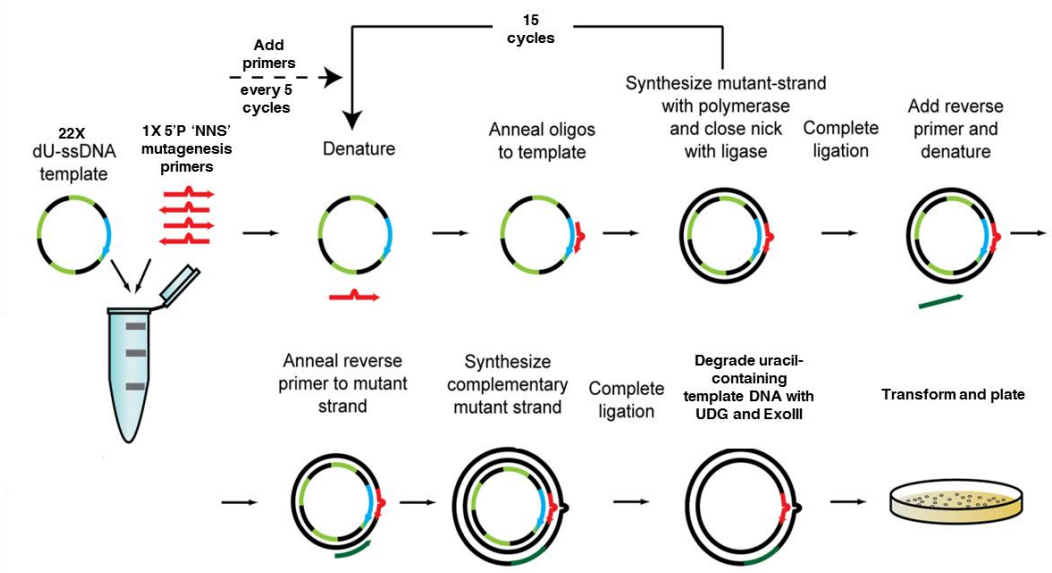


Figure 18: Pfunkel mutagenesis process to make a saturating single codon mutant plasmid library where every codon in the KaiB and KaiC genes is mutated to 'NNS' using a mutagenic forward primer for each codon position and one common reverse primer.

<1:22, limiting the probability of two mutagenic primers binding the same template molecule to less than 4.5%. Sanger sequencing of 48 colonies yielded 8 wildtype and no double mutants. The mutagenesis was done with the F1 origin containing pGEMT-EASY cloning vector phagemid containing the wildtype KaiC and KaiB gene and the region 400bp upsteam of the KaiB gene serving as the promoter. For the pilot library, mutagenesis was done with 39 oligos, one for each of the amino acid positions of KaiC. For the full library, 68 – 71 oligos was added to 9 different Pfunkel reactions to cover all amino acids in KaiB and KaiC. The mutagenized pKaiBC-KaiB-KaiC section is then subcloned into the multi-cloning site of a *S. elongatus* genomic integration vector pAM2314. The flanking regions around the multi-cloning site are homologous to a site in the cyanobacterial genome where the insertion of genetic elements is known to be non-perturbative, a so-called 'neutral site' (NS1)(70). Upon screening 12 colonies transformed the ligation reaction, no colony contained an empty pAM2314 vector, indicating a high ligation efficiency. In the construction of the pilot and the full KaiB-KaiC mutagenesis libraries, at least 12 colonies from each Pfunkel reaction, and subsequently after subcloning into the pAM2314 vector, were screened to test for mutagenesis efficiency. Sanger sequencing of colonies revealed that around 15% of the pAM2314 vectors contained the wildtype KaiB-KaiC gene, a double mutation rate of <3% resulting likely from $pfuC_x$ polymerase errors, and a low level of non-library mutants around 1-3% in the form of deletions, likely due to mutagenesis primer manufacturing errors.

The remainder of this chapter will be divided into two sections, the first describing the specific methods, results and conclusions from a pilot mutagenesis experiment targeting the first 39 amino acids of KaiC, and the next section will describe the approach to performing the experiment at a larger scale for all of KaiC and KaiB, and progress thus far on this.

Sequencing pilot methods:

Cyanobacterial Transformation: To generate a library of cyanobacteria, *S. elongatus* cells lacking the native *kaiBC* gene region are transformed by homologous recombination with the mutant plasmid library (Fig. 19). This is done as described in Clerico *et al* with 500mL of cells at OD = 0.7 and 7μg of plasmid DNA for the pilot library(70). Briefly, cells were washed in 0.1mM NaCl before resuspension in 10mL of BG-11 with the plasmid DNA and incubated with shaking overnight in the dark. Cells are then plated on selective BG-11 media in two large square (260mmX 260mm) plates to select for transformants. For the pilot library, there are >20 colonies per variant (given 1231 codon variants) for a total of around 25,000 colonies. As a check of library diversity, 12 individual cyanobacteria colonies were picked and inoculated into 5mL selective BG11 cultures. Genomic DNA was extracted from these cultures, and sequenced through the mutated KaiC region; 11 and out of 12 colonies sequenced contained a single codon mutation, one was wildtype, and none showed a second mutation in the 800bp sequenced region. Cyanobacterial library colonies were then collected by scraping and re-suspension into selective liquid Bg11 media. This pilot library culture was then grown in selective BL11 media in 3X500mL cultures until dense (OD = 0.7 – 0.9), before being inoculated into a 4L selective Bg11 media in a large 6L beveled spinner flask at OD = 0.1 – 0.2 for the clock function time course.

Time course conditions: To assess clock function of the variants, timepoints are collected over a period in constant light/temperature conditions where a wildtype clock would promote rhythmic circadian transcription. To synchronize or "entrain" the circadian clocks of the all cells in the culture, the culture was subjected to two days of 12-hour light and 12-hour dark, causing all functional clocks to align their clocks to the external environment. Due to the large volume of

the spinner flask, aeration was too poor to allow for sufficiently rapid growth to replace the sampled cells so that the culture volume and density remain relatively constant and changes in light penetration limited during the timecourse. To increase carbon availability, sodium bicarbonate was added at a final concentration of 5mM at the beginning of the entrainment period. Addition of sodium bicarbonate though has metabolic effects, have not been reported to perturb clock function or Kai gene rhythms(71). Additionally, there were no identifiable changes to the morphology of the cells after the sodium bicarbonate addition. After entrainment, 2X 90mL culture samples were taken starting two hours after the last entrainment dark pulse, every 3 hours for the next 72 hours, collecting a total of 25 timepoints (Fig 17). The samples were collected on rapid-flow 0.45μM SFCA filter paper through vacuum filtration, flash frozen in liquid nitrogen and immediately stored in -80°C.

Illumina library construction: Given the small area of mutation, we decided on using paired

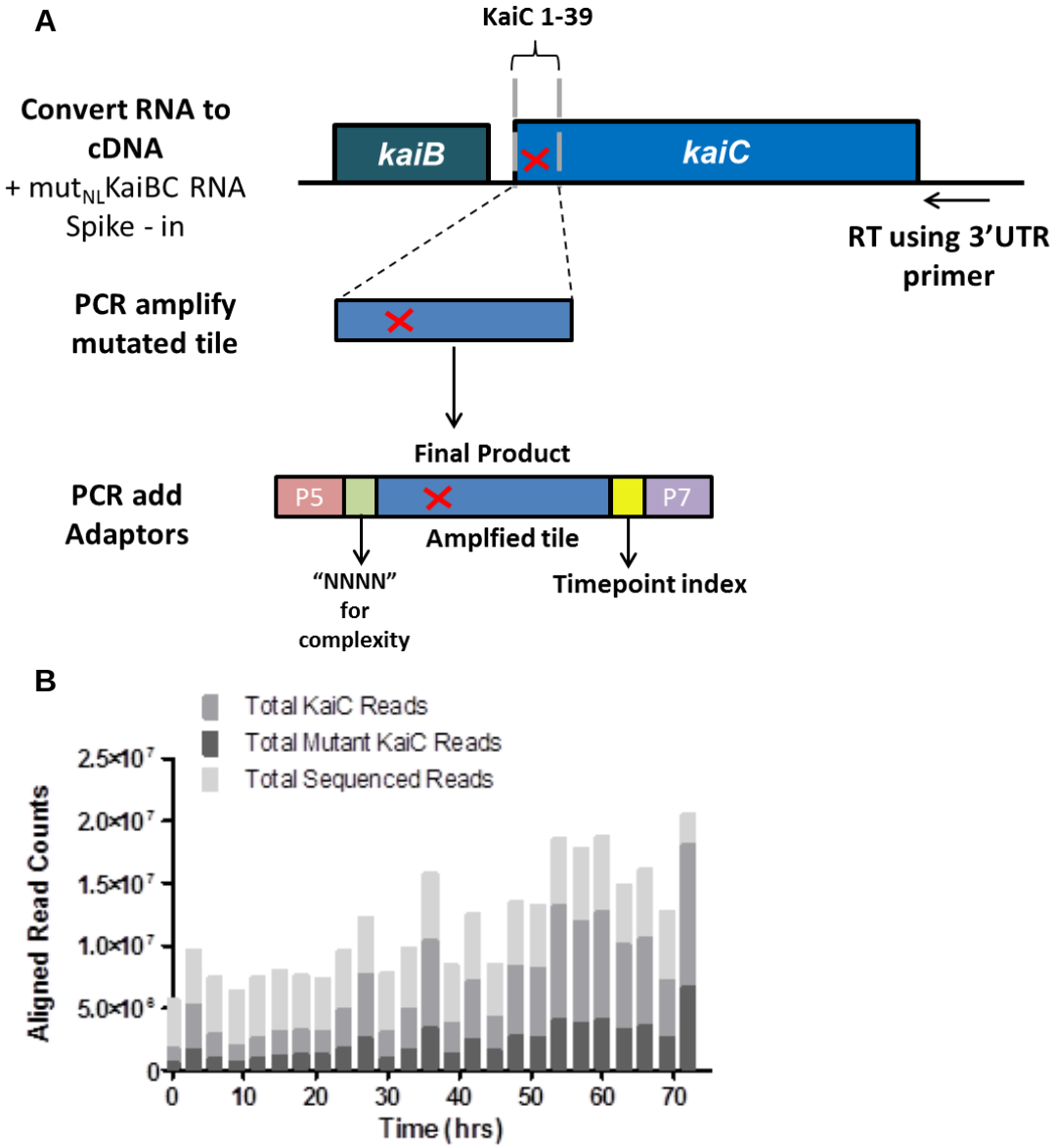


Figure 19: Pilot sequencing A) Sample preparation- KaiBC RNA is converted to cDNA from total RNA, the mutated region is amplified, and sequencing adaptors and timepoint indices are added by PCR. A non-library mutant KaiBC RNA (mut$_{NL}$KaiBC) is added to the reverse transcription master mix to normalize for procedural variability between timepoint. **B) Sequencing ouptut -** Number of total sequenced reads by index (timepoint) which align to KaiC (mutant or wildtype) or library 'NNS' mutant chromosomes by bwa-aln algorithm (Burrows-Wheeler). A filler library added to increase sequence complexity represent the vast majority of the 41% of total reads that are non-KaiC. A large portion of KaiC reads are wildtype (~76%) due mutations being outside the sequencing range (~50%) and incomplete mutagenesis (~15%)

100bp sequencing to span the mutational area in the two non-overlapping reads (Fig, 21A, 23). To prepare the sequencing sample, first total RNA was extracted and purified via phenol-chloroform extraction using the protocol developed in Vijayan et al(72). DNA was degraded by DNAse digestion, and the sample is re-precipitated and re-purified. All total RNA samples had 260/280 of >1.95 and 230/280 ratios of >2.0.

The total RNA is then used for reverse transcription(RT) with superscript III (Invitrogen) and a reverse 3'UTR primer. To account for variation between timepoint samples due to handling error, an *in vitro* transcribed *kaiBC* RNA containing an orthogonal three nucleotide mutation was added to the RT reaction master mix. The mutagenized section of KaiB-KaiC co-transcript cDNA is then specifically amplified to form the sequencing fragment. A second and last PCR add the P7/P5 adaptors, 'NNNN' degenerate bases before the beginning of the Read 1 region, and a timepoint index for each of the 25 timepoints. Multiplex paired-end 100bp Illumina sequencing in one lane of an 8 lane Hiseq2500 flow cell using the high-output run mode was performed. Despite the addition of the 4X random nucleotides at the start of read 1, library sequence complexity was not sufficient for cluster identification so a filler library was added to the sample. Overall library quality and output slightly exceeded manufacturer standards; the average quality score was 35.4, with 91% of base calls with a quality score ≥30 and a total output of 298,407,480 reads (290,407,584 reads with library indices) (Fig. 21B).

Sequencing Pilot Results

In the section below, the analysis procedure, results and conclusions from the sequencing pilot examining the mutational effects of codon substitutions for the first 39 amino acids of KaiC will be discussed.

Analysis procedure: Demultiplexed (separated by timepoint index) sequencing data in FASTQ format was downloaded to a local desktop for analysis. To determine reads counts for each variant at every timepoint, a pipeline using the following step was established and applied:

1. Alignment of reads from each timepoint to a reference genome where each non-wildtype variant from the 'NNR' single codon library (1230 variants), the wildtype sequence, and the RNA spike-in variant is a chromosome using the Burrows-Wheeler *aln* function, with a maximum edit length of 0 (-n 0), and trimming of 4 bp (-B 4) from the 5'end for Read 1 to remove the degenerate bases.
2. Alignment file from the bwa alignment is filtered with SAM tools, removing all non-mapped reads, PCR/optical duplicates, or reads that fail platform/vendor quality checks (-F 1540). The same alignment is also processed without removing PCR/optical duplicates (-F 516) in case this removes reads of the same variant, however, this did not change the number of mapped reads for each timepoint. Output data was in the BAM format.
3. Mapped counts to each variant "chromosome" is determine using the BED tools coverage function, which returns counts of the alignments in the BAM file that overlaps with the variant chromosomes for each variant chromosomes, with the output in BED format.

Further analyses of the mapped mutant counts were performed in Matlab. About 170M reads aligned to this reference genome across all the timepoints. The majority of the non-aligned reads corresponded to the filler sequence. Among the aligned sequences, around 76% are wildtype KaiC reads, which is expected given that ~50% of the sequenced DNA have a mutation in a section sequenced by the reverse primer (second read), and due to a ~15% wildtype rate from incomplete mutagenesis (Fig. 21B)

To investigate the rate of non-library mutant KaiC reads in the sequencing data and rate of double mutants, alignment was performed to the wildtype KaiC genome with the maximum edit distance increased to 3, and this yielded 11% more aligned reads (~6% of total reads). This indicated a low-level presence of non-library mutants, which may be due to sequencing error, mutagenesis primer synthesis error or double mutants arising from mutagenesis. Only the last possibility would compromise our ability to correctly genotype KaiC mutants using these short Illumina sequencing reads. We further characterized these non-library mutants by applying a custom Matlab script for alignment and variant calling to the Read 1 data of the first timepoint index. This showed that out of all reads had that the same first 5'-20bp sequence as wildype (priming regions for PCR), around 2.5% also had insertions or deletions that led to long stretches of misalignment. These indel likely resulted from mutagenesis primer synthesis errors/manufacturing defects. Those reads that had mutations in more than one codon (ex. mutation in 2 codons) as percent of all mapped KaiC reads (reads matching the first 20bp of the wildtype KaiC sequence) was slightly lower than the expected % from sequencing error (3.5% at Q = 34.6 for read 1, timepoint 1) at around 2%. By sampling, most of these read errors fell within positions with very low quality Phred scores (<20), suggesting that the variant calling accuracy/mutant genotyping could be increased by excluding reads where identified mutations is at a low Phred score position. While this analysis was conducted only for 1/50 of all the sequencing data, the average Phred quality score and Phred quality score distribution is very similar for the rest of the timepoint indices and for the second read. Hence, this analysis suggests that rate of double mutation is very low in the library, the sequenced mutations are the only mutation in the KaiC/KaiB transcript and the mutant genotyping is accurate for the vast majority of library members.

Normalization and filtering procedure: Several normalization steps are taken to remove noise and to accurately extract mutant dynamics from the data and to allow for comparison between mutants. First, the read counts over time for each mutant was normalized by the spike-in counts to remove timepoint preparatory variability, and then normalized by their total read counts, so that each variant is equally represented, removing differences in library representation (Fig. 22A-B). Given results from competition experiments between *S.elongatus* Kai point mutant strains, growth rates differences between mutants should not significantly change mutant distributions within the time frame of experiment, so differences in total read counts arise likely from

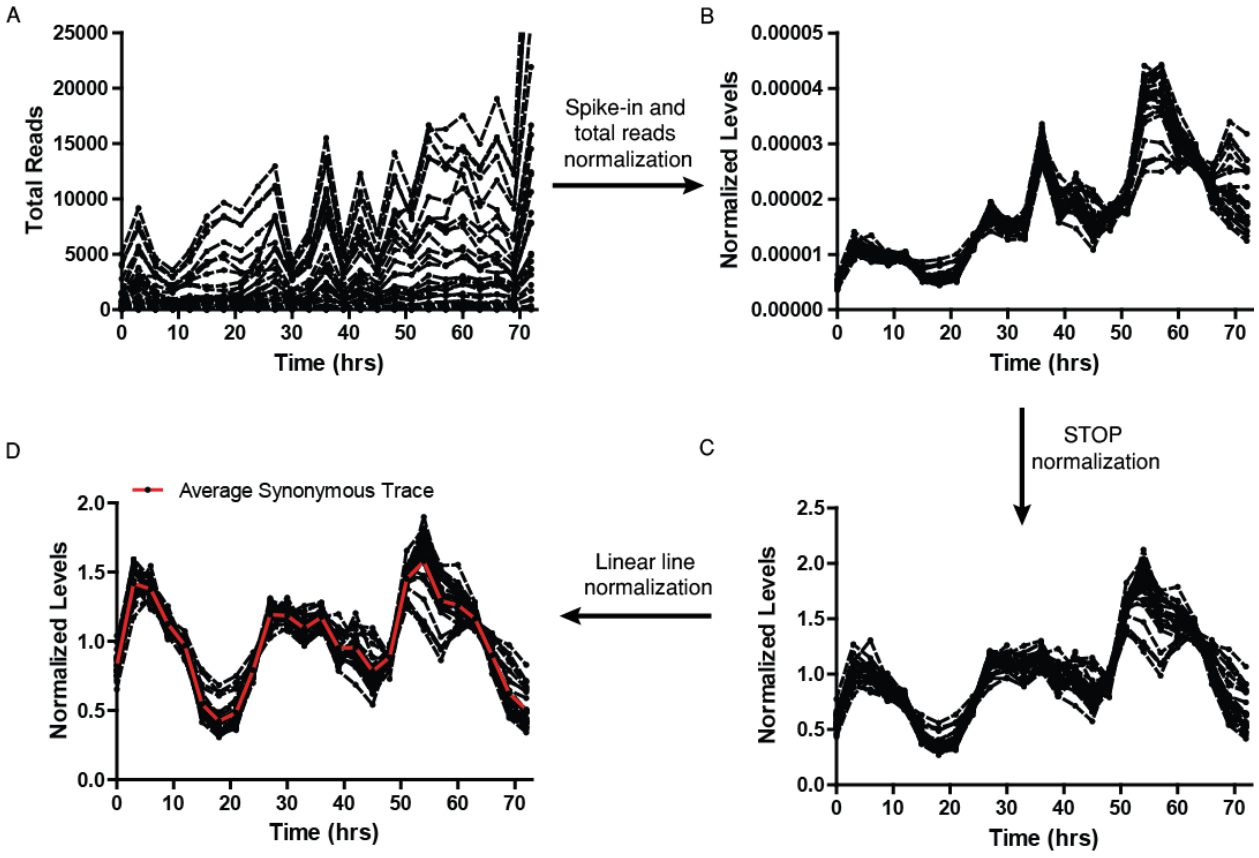


Figure 20: Normalization procedure shown for all synonymous mutations used for all 28 synonymous mutations (same amino acid as wildtype KaiC). A)-B) The raw data is first normalized by the total count of the mutant to equalize representation, and by the spike-in read counts to remove timepoint procedural variation. C) This is then normalized by an average STOP codon trace (using STOP mutants after residue 6). D) Last step is normalization by a linear growth line to de-trend the data and allow for sine fitting and parameterization.

mutagenesis or KaiC expression levels(11, 73)

Then two steps were taken to remove the contribution of random and systematic noise to the data. First, all mutants with low mean read counts (μ_{reads}) where Poisson error would make a large contribution to the timepoint dynamics were removed using this upper bound threshold: $\sqrt{\mu_{reads}}/\mu_{reads} < 20\%$ (average read count > 25) Secondly, aside from the oscillatory behavior, it is apparent that there is also a systematic noise signature across all the samples. Since all STOP codons after the second methionine in position 6 appear largely arrhythmic aside from this systematic noise signature, we normalize all mutant traces by an average of these STOP traces to remove the systematic noise (Fig. 22C). Additionally, there is also a distinct upward slope in the number of reads over time. This may be a slow response in gene expression to the constant light conditions after the light-dark pulses during the entrainment period. To remove differences between the rates of increase in expression in different mutants, the upward trend is determine by fitting the mutant traces to a linear line (y = a + bt), and the mutant traces are then detrended by normalizing by that linear line (Fig. 22D).

Sequencing Error: Sequencing error is potentially a large contributor of noise given the abundance of wildtype reads((76% of all KaiC reads) arising from both from actual wildtype KaiC RNA, and from unmutated sections of mutant fragments owning to the mutation being on the other read (Fig. 23). To empirically measure the rate of non-library single

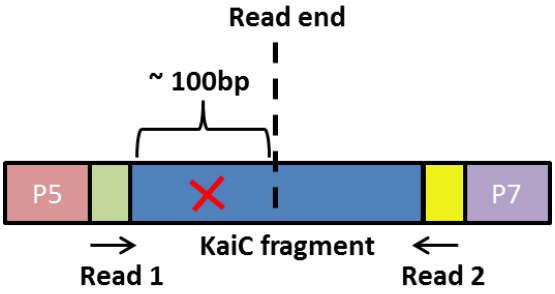


Figure 21: Schematic of DNA submitted for Illumina sequencing. P5 and P7 refer to sequencing flow cell annealing regions. KaiC fragment is sequenced by two non-overlapping reads. There is one mutation per KaiC fragment, so only about 50% of the reads from one direction will be a non-wildtype sequence

nucleotide mutants, the counts of all reads that mapped to a non-wildtype A or a T in the third codon position, which is not included in the 'NNR' mutant library, was determined for each timepoint. Any read with this mutation likely arose from sequencing error. The average rate of these mutations as a fraction of wildtype reads over all timepoints is $3.23X10^{-4}$, which is comparable to what is expected given an overall Phred quality score of around 35.4 (theoretical = ~$2X10^{-4}$) . This is taken as the estimated average sequencing error rate per 100bp read. In comparison, given the lower average total read count of any mutant in the library, estimated at ~$1.9\ X10^{-4}$ assuming equal distribution among mutants, the contribution of erroneous single nucleotide calls to 1nt mutants read counts can be significant for those that are highly represented in the library. Fortunately, on average single nucleotides mutants comprised only about 5-10% of all possible mutants at a position (except histadines = 25%), so the vast majority of mutants phenotypes are not compromised by sequencing errors.

Read count threshold were set to remove mutants whose read counts would be largely dominated by reads that result from erroneously sequenced wildtype DNA. The expected read count contribution due to sequencing error for any 1nt mutant is calculated as the average error rate × wildtype counts, and for any 2nt mutants is calculated as (average error rate)^2 × wildtype counts. For 1nt mutants, those with a total read counts less than 1.6X the expected error read contribution are removed from the analysis, for 2nt mutants those with a total read counts less than 10X the expected error read contribution are removed. This very high tolerance threshold of 60% for potential sequencing error contribution was set for 1nt mutant to maximize use of the sequencing data, especially given uncertainty in the error rate estimation at the mutant positions. Additionally, above this read count there did not seem to be a common phenotype for 1nt mutant traces at lower read counts as compared to mutants with a higher total read count, which suggests

that the read count dynamics of the included 1nt mutants are not dominated by an average 'wildtype read' phenotype (wildtype reads over time should show the average dynamics of reads from actual wildtype KaiC molecules ~ 15%, plus all the mutations on the other section of sequenced the gene fragment ~50%). This 1nt read count threshold also allows 28 out of the 39 possible synonymous (different codons coding for the wildtype amino acid) mutations to be retained then for phenotypic analysis. An improved strategy to prevent sequencing error from limiting and compromising data on 1nt mutants will be addressed in the conclusions section.

Coverage and Mutant Distribution: To assess the diversity of our library, the level of mutagenesis saturation across the mutated region, we determined the percentage of all possible mutants (32 possible codons substitution with the 'NNR' codon) found by sequencing, and what

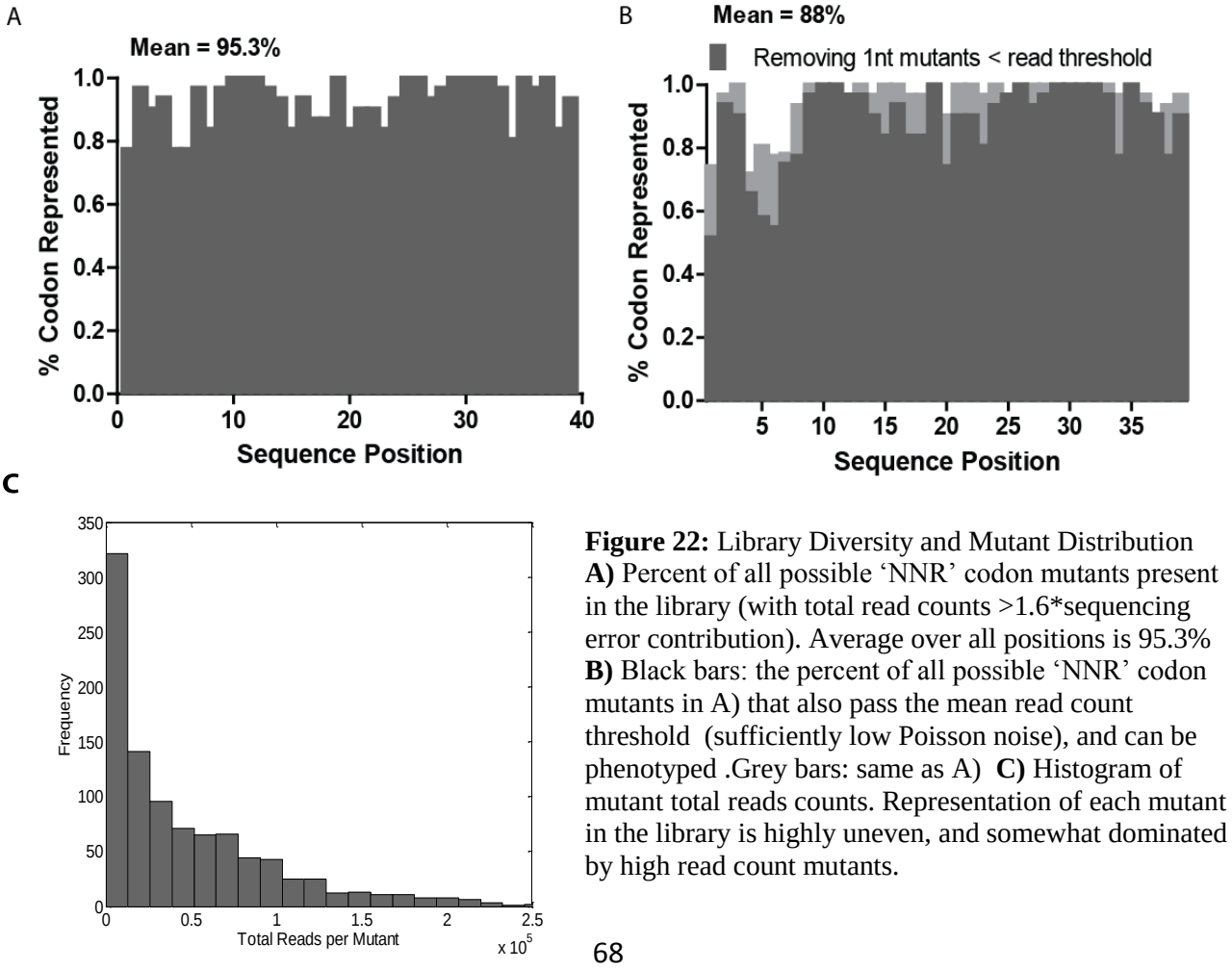


Figure 22: Library Diversity and Mutant Distribution **A)** Percent of all possible 'NNR' codon mutants present in the library (with total read counts >1.6*sequencing error contribution). Average over all positions is 95.3% **B)** Black bars: the percent of all possible 'NNR' codon mutants in A) that also pass the mean read count threshold (sufficiently low Poisson noise), and can be phenotyped .Grey bars: same as A) **C)** Histogram of mutant total reads counts. Representation of each mutant in the library is highly uneven, and somewhat dominated by high read count mutants.

the distribution of mutants was in terms of total read counts. We first examined the percentage of all possible mutants at each position that can be found in the library at total read count significantly above the expected read count contribution from sequencing errors (>1.6*sequencing error contribution), and this was near saturation, with an average of 95.3% across all positions, and was largely limited by the high sequencing error burden for 1nt mutants (Fig. 24A). The average percentage of all possible codon substitution at each position with sufficient read counts where we could extract reliable information about their dynamics (meets the read count thresholds for Poisson and sequencing error noise described above) was around 88% (Fig. 24B). Hence, overall the library coverage was very high with almost all codons/amino acid substitutions covered at all positions, indicating sufficient depth of coverage in sequencing. Across the sequence, loss of coverage occurs unevenly, where mutants at some positions, especially in the N-terminal region, are especially under-represented. This is likely due to the effect of these mutations leading to low average KaiC transcription or poor mutational efficiency in those regions. Upon examining a histogram of mutant total read counts (read counts summed over all 25 timepoints), it is apparent from trailing tail of the distribution the sequencing output is somewhat dominated by mutants with the largest total read counts, where 99% of all KaiC mutant reads corresponds to reads from 80% of the library mutants (Fig 22C). If this broad read count distribution and regional evenness in mutant representation hold true for mutants across the whole *kaiBC* gene region given our library generation protocol, a high depth of coverage (average reads/variant) may be required to to gain information on the more lowly represented sequence positions. Relative to this pilot where the average read per variant was around 40,000 reads, the amount of sequencing feasible for each mutant in the full library will likely be ~100 fold less (see Conclusions section).

Synonymous mutations: Out of the 28 synonymous mutants retained, most show clear oscillatory behavior with a near-24 hour period (Fig. 25). The three that showed the lowest amplitude had also the lowest number of counts, corresponding to less than 14th percentile in total read counts. These non-wildtype codons may decrease translational levels of the KaiC which may decrease oscillator amplitude which is observed when the ratio of KaiC to KaiA is lowered *in vitro*. The average period is slightly longer than 24 hours at 25.5 hours. Overall, the circadian oscillatory behavior of the synonymous mutations indicates that this assay is sufficiently sensitive to derive information on period and amplitude

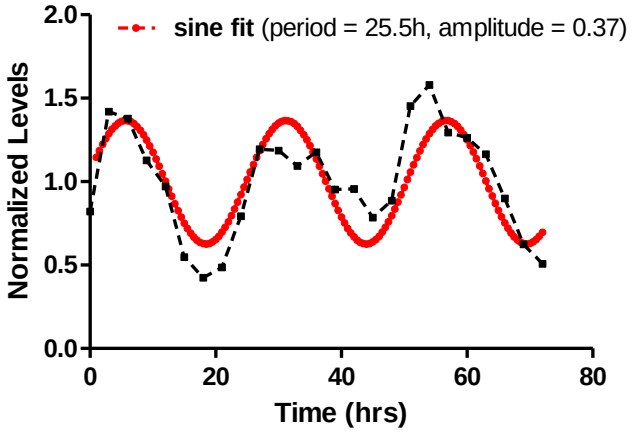
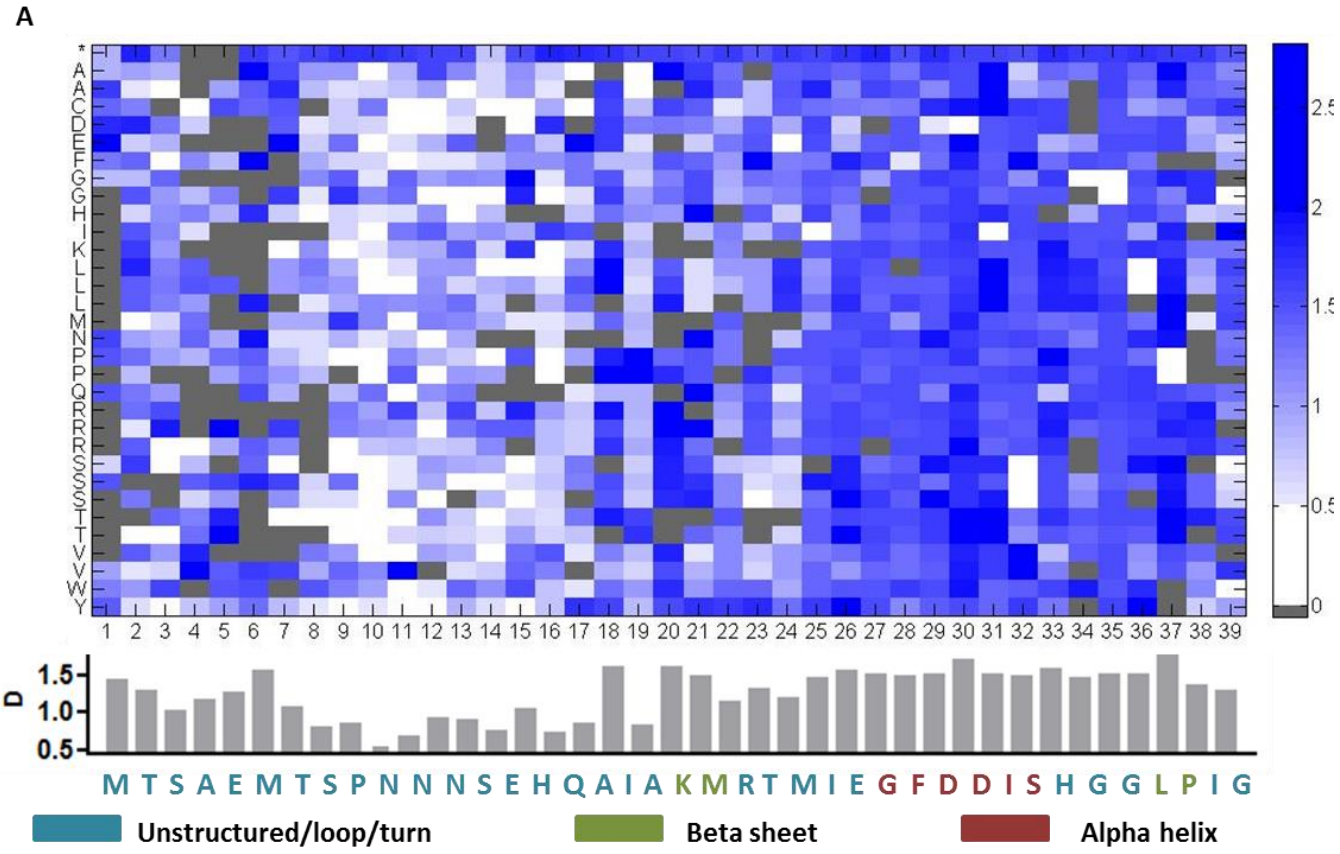


Figure 23: Average normalized trace of all synonymous mutants, and the best fit sinusoidal curve.

Mutational analysis:

In the following section, metrics of characterizing non-synonymous mutant phenotypes and the biological significance of patterns in the mutant data will be discussed.

Aggregate positional mutational effects: To assess the level of perturbation caused by a specific



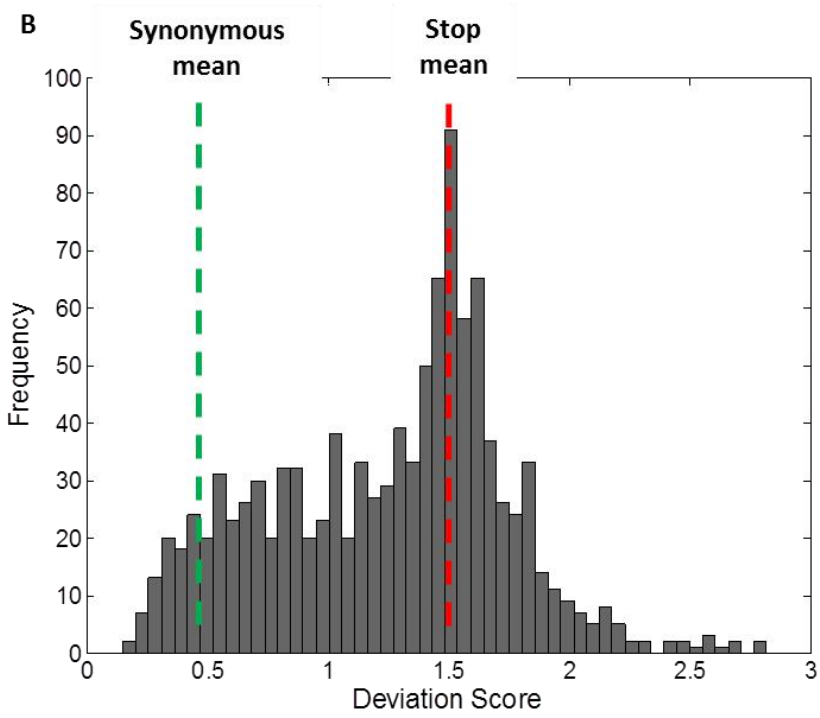


Figure 24: A) Matrix displaying deviation scores for each codon at each sequence position **B)** Histogram of mutant deviations scores, the green dotted lines indicates the median deviation score for synonymous mutations and red dotted line indicates the median deviation score for STOP codons after position 6

mutation, all normalized reads over time are compared with a standard synonymous time trace using a deviation score, which similar to a root mean squared deviation. This standard synonymous time trace used here is the average of all the 10 synonymous time traces. The deviation score is calculated as:

$D = \sqrt{\sum_{t=0}^{72}(syn_t - mut_t)^2}$ where syn_t the average synonymous normalized levels at a specific timepoint t, and mut_t is the mutant normalized levels at a specific timepoint t. Below is a bar graph showing the median deviation score at that position. (Fig. 26A, B). This serves as a measure of how mutational sensitive a position is, where a higher average deviation score indicates that on average substitutions at this position causes larger differences from the synonymous phenotype, and hence is more mutationally sensitive.

Mutant oscillatory parameters: To further characterize the effects of mutation, normalized reads over time was fit to sine curve ($y = Asin(\omega t + \theta) + 1$) using non-linear regression, where A is amplitude, and ω is the period, and 95% confidence intervals of these parameters were also determined using the *nlparci* with the Jacobian option (Matlab). The amplitude parameter was also used to categorize mutants as oscillating or not. For those mutants that did not have an amplitude parameter that was significantly above zero, where the best fit amplitude parameter was less than 0.1 or where the best fit amplitude was not significantly non-zero relative to the 95% confidence interval (amplitude Z-score < 0.85), it is phenotypically classified as not oscillating. These cutoffs are set so that using higher cutoff resulted in mostly oscillating traces to be removed by visual inspection. For each position whether the codons substitutions caused the loss of rhythmicity or not is shown for all substitutions across the first 39 amino acids. The %

of all represented codon substitutions that caused arrhythmicity at each position correlates well with the median deviation score. (Fig 25A, B).

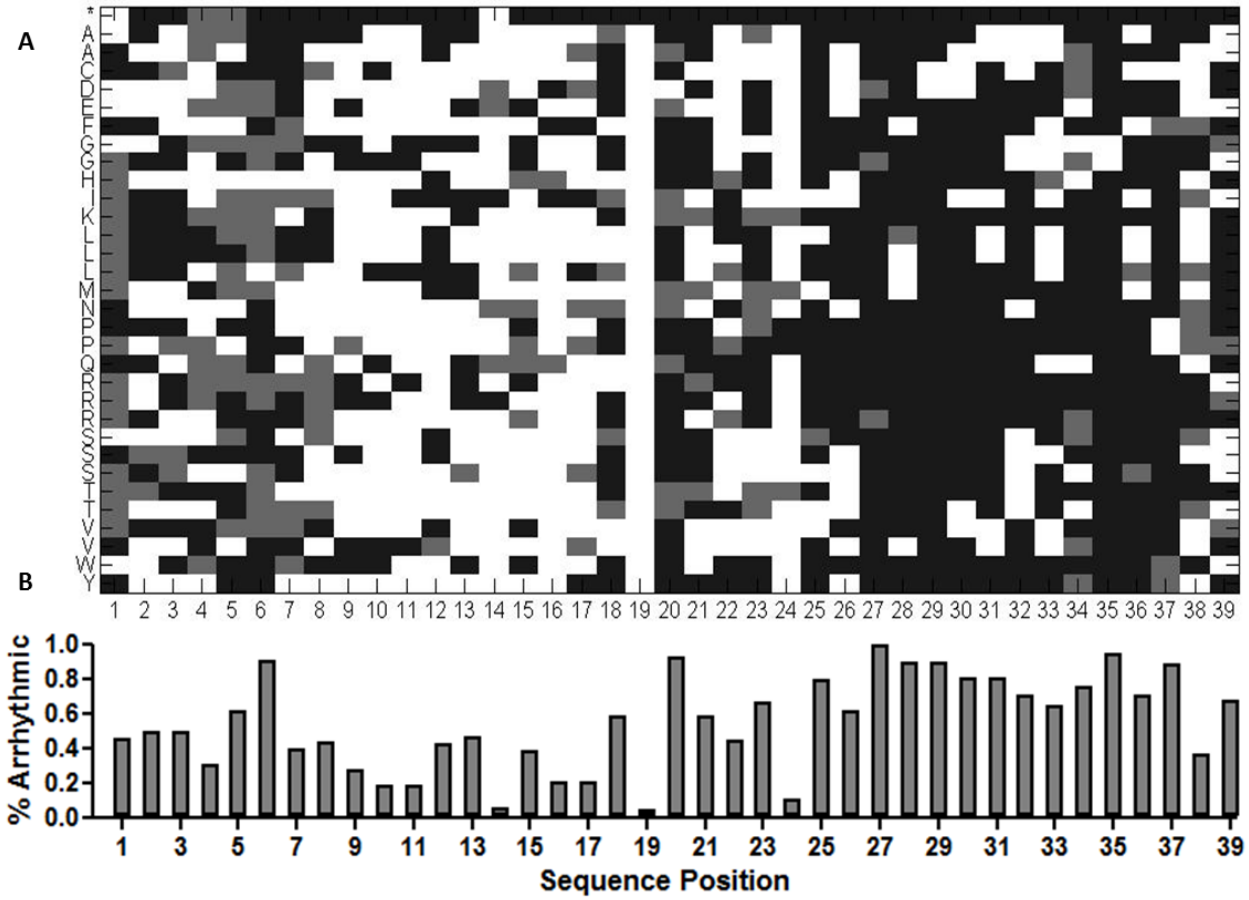


Figure 25: A) Matrix displaying whether a codon substitution results in rhythmicity (> 0.1 amplitude, colored white), or in non-rhythmicity(black). Mutants not in the library are colored grey. B) % of codons out of those meeting the mean count threshold that are arrhythmic at each position.

Global mutational sensitivity patterns: Globally it looks like there are stronger mutational effects or overall less mutation tolerance after around position 18, with the highest mutational tolerance or lowest average D score from position 8 to 17. Surprising, positions 1 – 6 seems relatively intolerant to mutation as compared to region 8-17, with a high % of arrhythmicity, although it is an unstructured regions not resolved on crystal structures of KaiC. This region also has the lowest % percent coverage, as well as the lowest average total reads of mutations, suggesting that this region may play a role in translation of KaiC and in maintaining proper KaiC expression

levels and stoichiometric ratios with the other two Kai molecules. The sensitivity of position 6 to mutation, where every substitution resulted in arrhythmicity other than a leucine 'CUC' codon. Even this substitution showed very low amplitude oscillations, suggesting a strong functional constraint for a methionine at this position. The unstructured nature of this region does not support for a structural role for this methionine, but rather this site may be an alternative (or the dominant) translational start site given that the actual start site of KaiC translation have not been reported previously. Unfortunately, confidence in this observation is limited due to the particularly low mutational coverage (~55%) of the position owning to insufficient mutant read counts and the 1nt mutant error burden.

Additionally, it is also apparent that almost all non-synonymous mutations between 25 and 37 cause relatively high levels of perturbation. This region corresponds to a turn –alpha-helix-turn regions which may play a precise functional or structural role. For example, glycines 34 and 35 involved an alpha helix to beta sheet turn structure in this region are extremely sensitive to mutation where G35 is arrhythmic for all non-synonymous substitutions but comparatively a later glycine (G39) at a beta hairpin is relatively more permissive to substitutions.

Amplitude vs period perturbations: To determine whether a mutation has caused any period perturbation, period differences ($\Delta\theta$) between all the rhythmic non-synonymous mutants (θ_{mut}) and the synonymous mutants (θ_{syn})were calculated using the period parameter from sine fitting where $\Delta\theta = \theta_{mut} - \theta_{syn}$ and where θ_{syn} = 25.6hrs as determined by fitting to the average synonymous trace. To discover statistically significant period mutants, period differences $\Delta\theta$ are normalized by the 95% confidence interval and represented as a Z-score, those with a Z-scores > 0.85 are designated as long period mutants, and those with Z-score <-0.85 are designated as short

period mutants (Fig. 29A-G) The period difference Z-scores are represented by colors for each rhythmic mutant with the grey standing for arrhythmic mutants (Fig. 28A,B).

In examining mutational effects on oscillator period, the vast majority of mutation that do not abolish oscillations also do not cause significant period changes. Out of the 1085 mutants passing read count thresholds, there are only about 5 potential long period mutants (3 different amino acids substitutions), and 6 short mutants (5 different amino acid substitutions) (Fig. 29 A-G). Interesting, all identified period mutants fall within positions that overall have low mutational tolerance, where the majority of the codon substitutions result in arrhythmicity. These mutationally sensitive positions may be functionally critical in mediating a specific biochemical/kinetic step where most non-synonymous substitutions result in perturbations that lead to either arrhythmicity or period changes. Hence, control of period and amplitude (rhythmicity) may then be biochemically controlled by the same structurally elements, whether this proves true for the rest of the sequence will depend on analysis of the fully library of KaiC mutants.

In contrast to the high mutational robustness in the oscillator period, there 592 codon

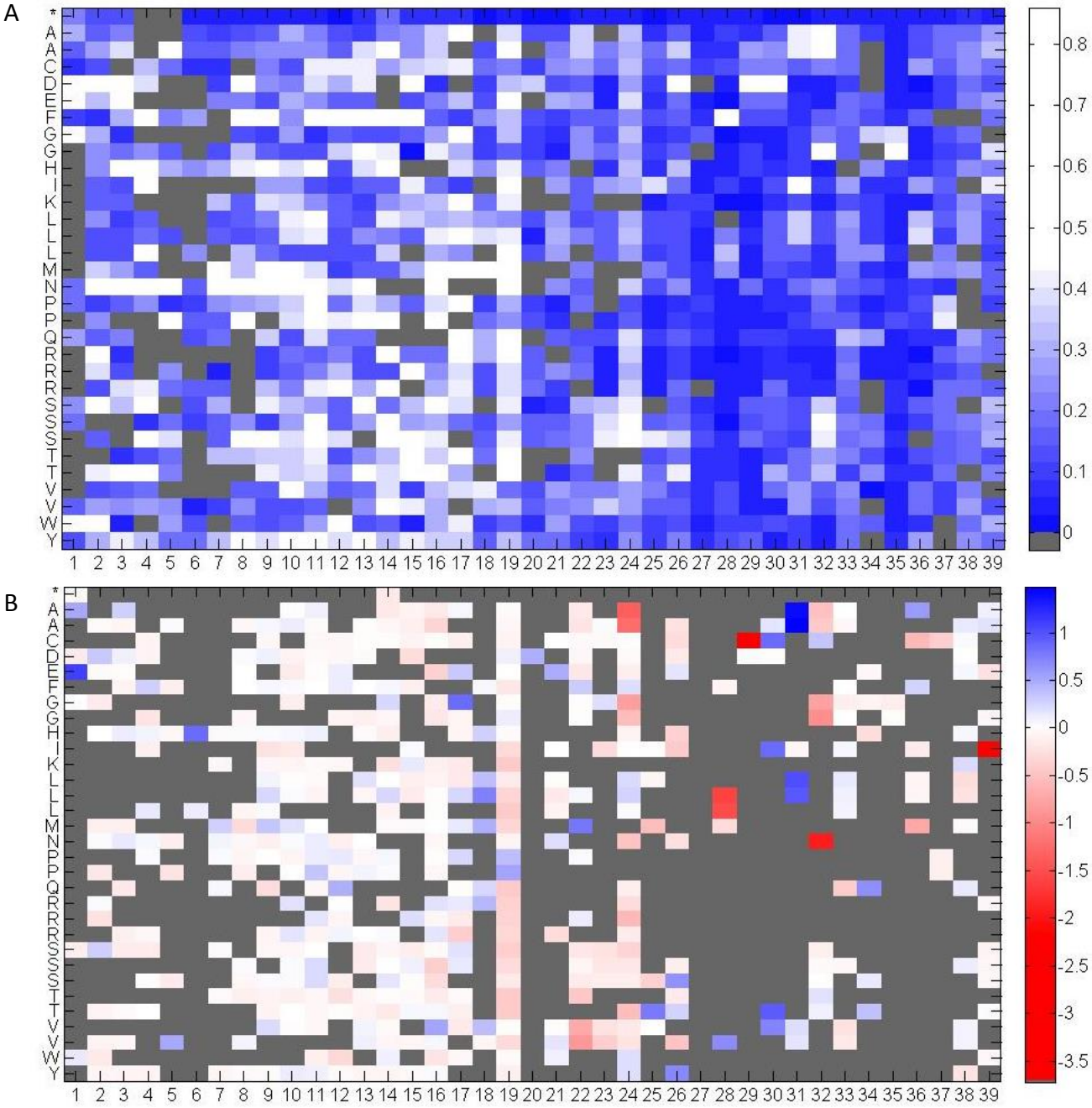


Figure 26: A) Matrix displaying the best fit amplitude parameter at each position for every codon substitution. Going from blue to white, amplitude increases, with grey being codons not present in the library. Color saturates at 0.45, which is only slightly higher than the average synonymous amplitude. **B)** Matrix of the period difference form the average synonymous period as a Z-score. Grey are arrhythmic mutants or mutants missing in the library. Blue indicates a longer period and red indicates a shorter period. (by visual inspection, 32N and 29C display a high frequency "noise" signature are likely not short period mutants)

substitutions that result in arrhythmicity, or over half of the codon substitutions considered in the library result in substantial loss of amplitude so that oscillations are no longer detectable. Looking at the relationship between amplitude and period deviation (Z-score of the absolute difference from the average synonymous period) for all mutants, as noted above, period is relatively robust to single codon substation relative to amplitude, where most mutants that are clearly oscillating with an amplitude near or higher than the average synonymous trace also had a period that was near the synonymous period (Fig. 30A). Concordantly, the amplitudes of the period mutants satisfying the criteria described above anti-correlated with the magnitude of their period difference from the synonymous average (wildtype period), suggesting a strong structural linkage between the two properties (Fig. 30D). Additionally, it would also appear from their amplitude distribution that the long period mutants tend to have higher amplitudes than the short period mutants (Fig. 30B). Additionally, short period mutants also had best fit period parameters that were closer to the synonymous period (<4 h difference) than the long period mutants (>5h difference). This suggests an underlying propensity of the Kai clock system to oscillate with a longer period more stably with higher amplitude upon mutational or biochemical perturbation.

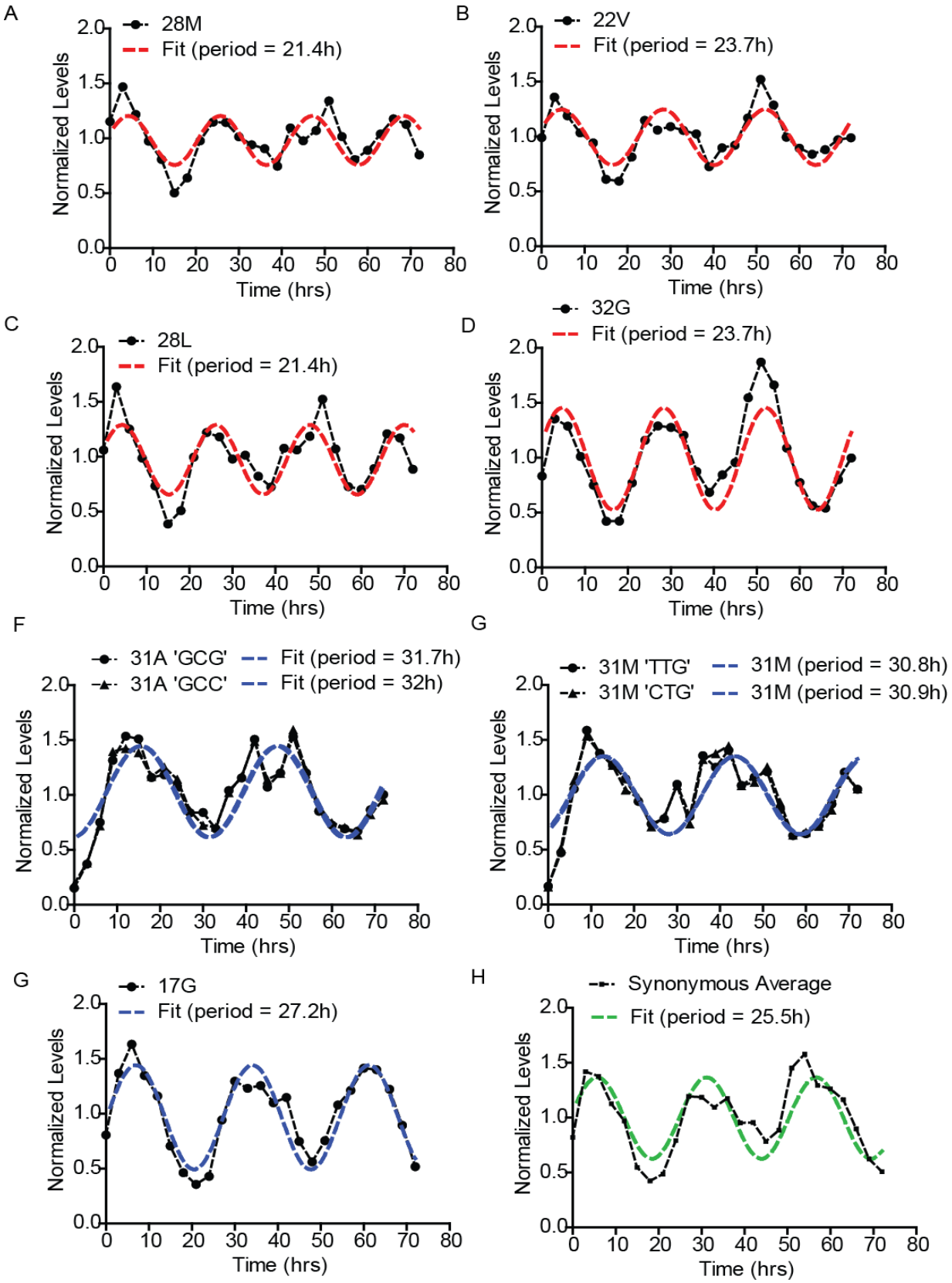


Figure 27 A-D) Examples of mutants with a shorter than the average synonymous period. **F-G)** Examples of mutants with a longer than average synonymous period. All designated period mutant have a period difference Z-score >0.85

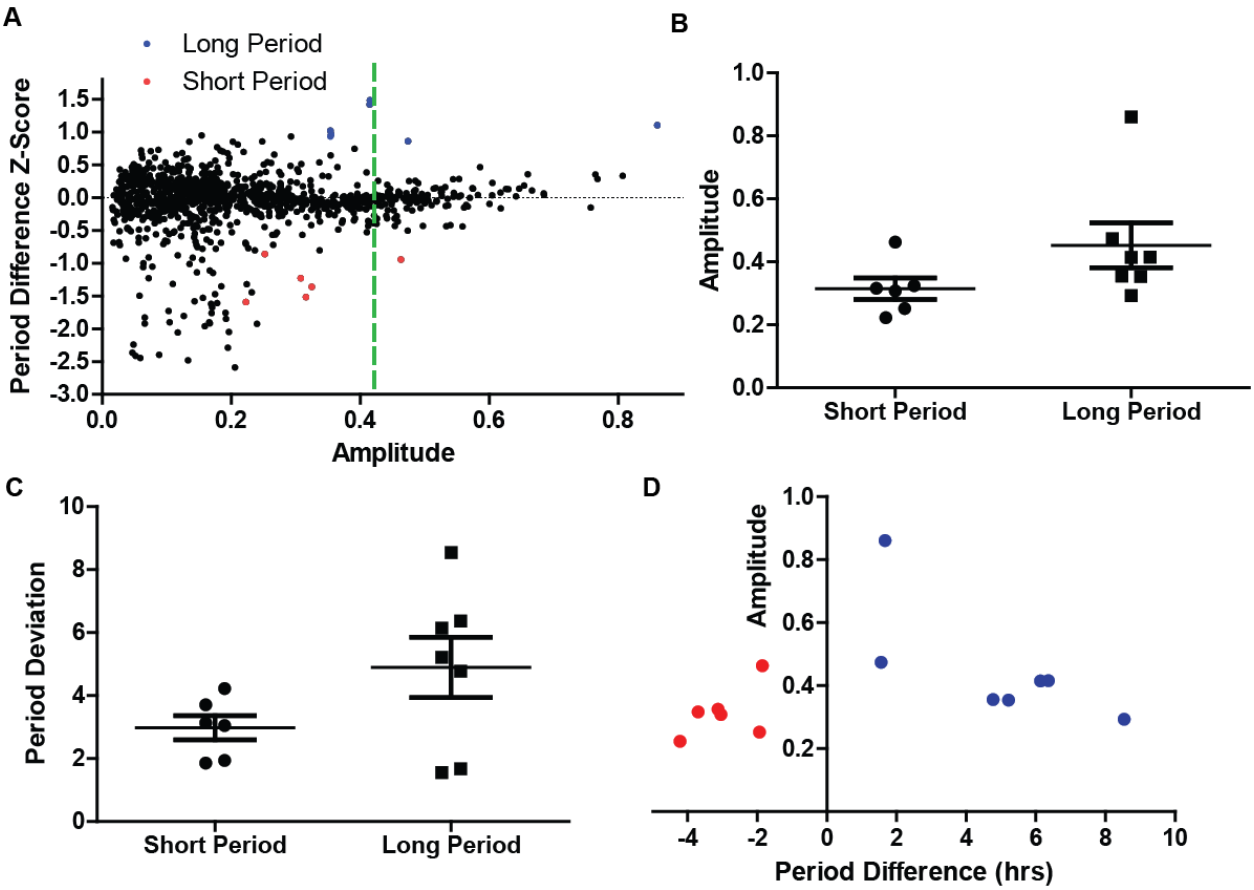


Figure 28: A) Period deviation Z-score versus amplitude comparison B) Distribution of amplitudes for short and long period mutants C) Distribution of period deviation (in hours) from the synonymous average for long and short period mutants D) Amplitude versus period difference (from the synonymous average) for all period mutants (Red = short period, blue = long period)

Structural basis of mutational sensitivity: To explain the mutational sensitivity of residues, we found a distinct and statistically significant anti-correlation between the percent solvent accessibility of a residue and percentage of substitutions by mutagenesis that result in arrhythmicity for a position (Fig. 27A). This indicated that residues that served a more structural role were more mutational sensitive. In support of this, we examined the structural role of all positions where substitutions would result in either a high percentage of arrythimicity (>75%), or any long or short period phenotypes. We found each of these residues were either buried in the hydrophobic core, were involved in turn structures (G35, P37), or a polar interactions with either the main or side chains or both. The only except was D30, which was highly solvent exposed,

did not form any polar bonds, but was exceptionally sensitive to mutation, where even conservative mutations such as to small polar residues like a serine or threonine or to a similarly charged glutamate would result in either a long period phenotype or arrhythmicity. It also does not fit the trend shown in Figure 31A. While H15 also appear an outlier, it is located at the beginning of the resolved section in the crystal and potentially in non-native position. Given the proximity of D30 to the CI-CII interface, it is possible that it may play a role in intra-domain communication or allostery, and in stabilizing conformations of KaiC that is not represented by known crystal structures, including the one shown in Figure 31B.

Comparing mutational sensitivity with conservation scores: The level of conservation of a sequence position across homologues generally correlates with function, reflecting constraints on the chemistry of the amino acid side chain to support structure, enzymatic activity and nor interaction with binding partners. Comparing the mutational tolerance at each position to the level of conservation in KaiC sequences of related organisms that have all three Kai genes (presumed to have a functional clock) allows us to determine whether the level of sequence

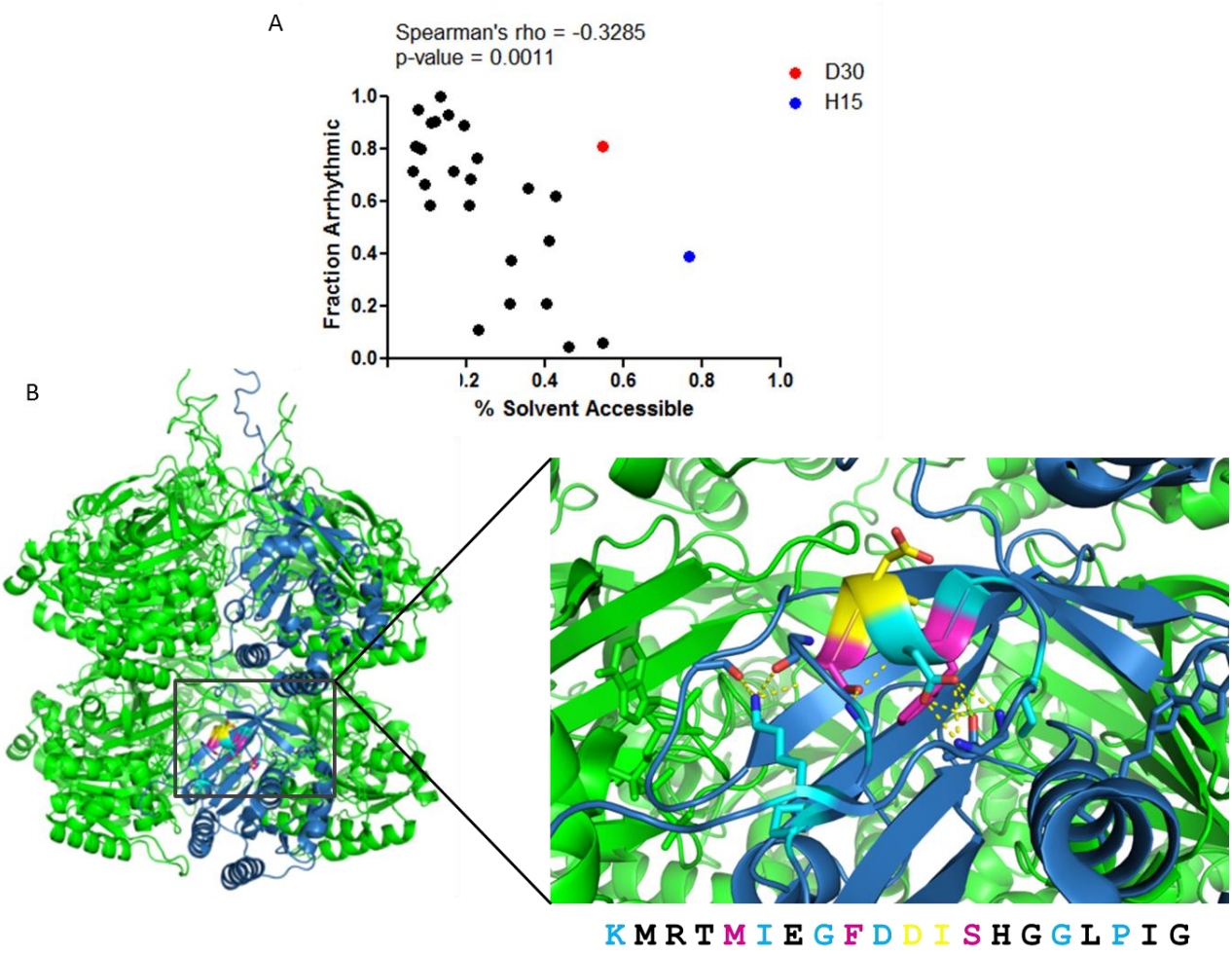


Figure 29: A) Correlation between solvent accessibility and the fraction of substitutions resulting in an arrhythmic phenotype for each resolvable residue (14 – 39) on the KaiC crystal structure (PDB ID: 3dvl). Solvent accessibility was calculated using the POPS* algorithm as the total solvent accessible surface area over the total surface area of the isolated residue(4) **B)** Structure of KaiC showing the most mutational sensitive positions, where cyan indicates >75% of substitutions resulting in an arrhythmic phenotype, magenta indicate positions where substitutions can result in a shorter period, and yellow indicate positions where substitutions can result in a longer period.

conservation at each position reflects function constraints that are also present in *S. elongatus* KaiC. Positions where the relative sequence conservation and mutational sensitivity diverge may indicate selective pressures in *S. elongatus* or in the context of the *S. elongtaus* KaiC sequence that are different from that of homologous clocks KaiC genes, suggesting non-canonical forms of function and/or regulation. To calculate conservations scores at each position, multiple sequence alignment using MUSCLE of all KaiC protein sequences from cyanobacterial species that also

have KaiA and KaiB were performed using 96 unique KaiC sequences. Then conservation scores for each position in KaiC were calculated using Consurf, which employs the empirical Bayesian paradigm to calculate a rate of evolution based on a phylogenetic tree generated from the MSA using Rate4Site and the JTT evolutionary substitution model(74). The conservation score is 1/rate of evolution, and Consurf also calculates 50% confidence intervals (Fig. 32A).

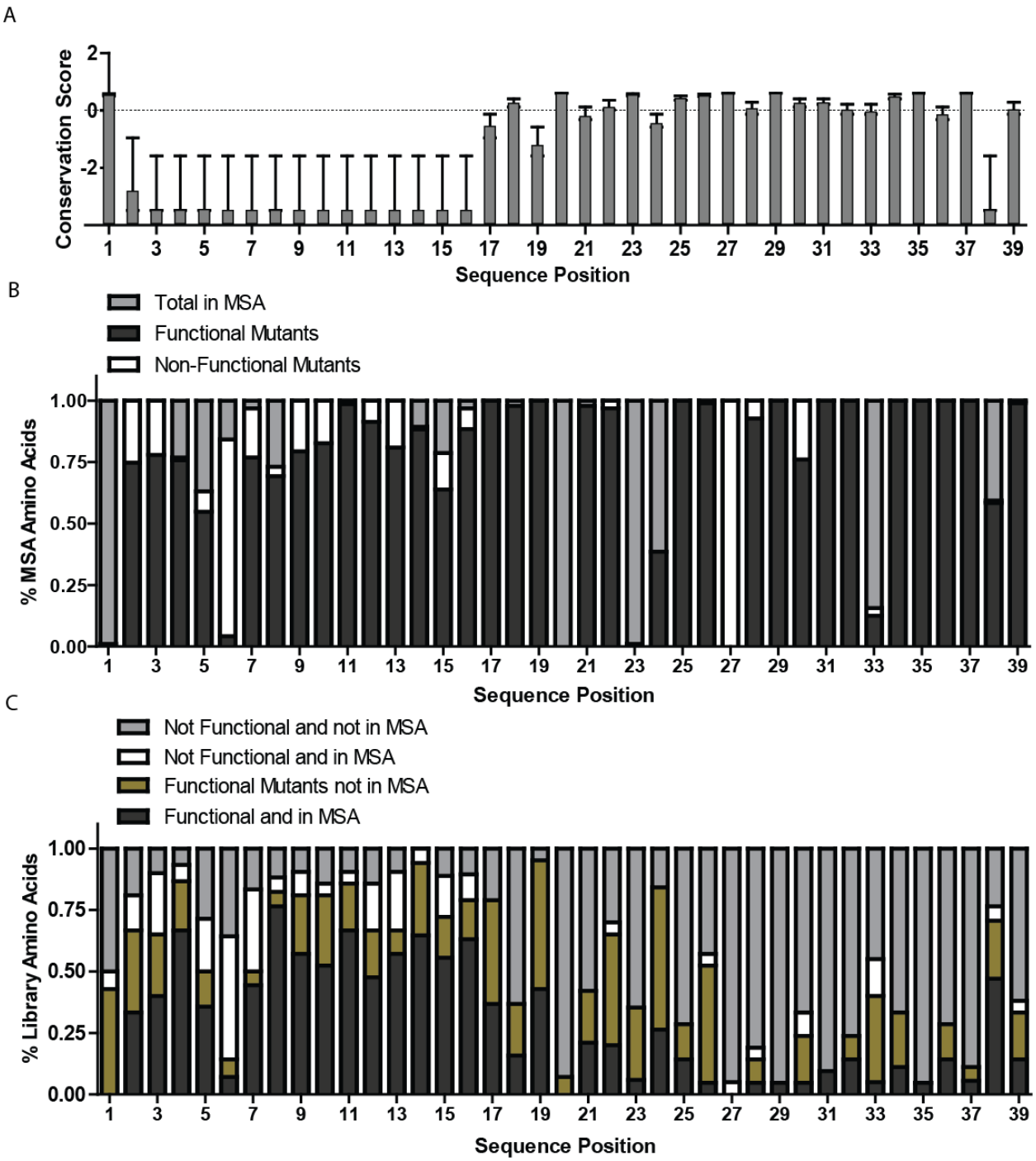


Figure 30: A) Relative conservation score based on evolutionary rates calculated by the Rate4Site method using the JTT substitution matrix (Consurf) and a multiple sequence alignment (MSA) of cyanobacterial KaiC from species with both KaiA and KaiB. Higher scores indicate a lower rate of evolution relative to the rest of the KaiC sequence. **B)** Percentage of all amino acids substitutions observed in KaiC MSA at each position that is represented in sequencing data by at least one mutant/codon (Total in MSA), this is split into amino acids that result in a functional clock phenotype as indicated by least one codon (Functional Mutants) and the % that disrupt clock function as indicated by at least one codon (Non-Functional Mutants). The remaining %

(Figure 32 continued) (light grey) is MSA amino acids not in the library. Amino acid substitutions are weighted by their MSA frequency. **C)** Percentage of sequenced amino acids substitutions (represented by at least one codon) that results in a functional clock phenotype (Functional Mutants), and a subset of this can also be found in the KaiC MSA (Functional and in MSA). Stacked is the segment of amino acid substitutions that disrupt clock function and are in the MSA (Non-Functional and in MSA). The remaining % (light grey) corresponds to non-functional substitutions that not in the MSA.

Comparing the relative conservation scores to the median deviation score (Fig. 26A vs Fig. 32A), both looks well correlated except for the 1-6 regions, where KaiC sequences are relatively mutational sensitive but homologous KaiC sequences are highly variable. This variability may indicate either low levels of functional constraints at these positions or just variable functional constraints among homologous KaiC genes at those positions. At least for S. elongatus KaiC, there seem to be a particular set of functional constraints on the sequence in that region which are most constrained at Met6.

Additionally, we compared how well the natural range of substitutions in clock KaiC genes matched the functional set of KaiC substitutions, and whether then evolutionary selection for specific residues was independent of sequence context. Indeed, we found that most of the residues found in the multiple sequence alignment (MSA) at each position matched the mutants in the library that were largely non-perturbative (oscillating with a period difference Z-score of <0.85) or still had clock function (Fig. 32B). This was especially true for the relatively conserved residues from 17 – 39 (excluding 38), where only a few amino acids were found at each position in natural KaiC sequences, the large majority of which were in the library and were functional. The only exception in this region is for G27, which is found in 100% of all homologous KaiC sequences used in the MSA, and synonymous mutant in the library was found to be arrhythmic. Upon closer examination, while the time trace for the mutant was quite noisy, but did appear to have some circadian rhythmicity. It is possible that a non-optimal G codon at this position for

KaiC could have led to change in protein levels (total mRNA counts were around the library average), leading to a loss of amplitude. In the most conserved regions of the sequence, where only 1 or 2 residues are naturally found (7 positions), at three of these positions (29, 31, and 35) only the residues conserved in the MSA were found functional by mutagenesis with 100% coverage of all 32 possible codon at each of these positions, a fourth position (27) as mentioned above, was functional for none of the non-wildtype substitutions. There are also a couple of positions (20 and 23), where only one amino acid was found naturally at these positions, however, synonymous substitutions at these positions were not found in the library/did not pass read count thresholds. Overall, this evidence indicates that at the most conserved positions of the KaiC sequence, a precise amino acid identity is required independent of sequence context. Additionally, for almost every position, the percentage of functional substitution out of all those observed in nature (Fig. 32A, dark grey) is greater than the percentage of functional substitutions out of those not observed in nature (Fig. 32B, brown relative to light grey). This indicates that natural sequences for the most part have similar functional constraints as the *S. elongatus* sequences at the first 39 positions, and mostly have not evolved very different sequence contexts.

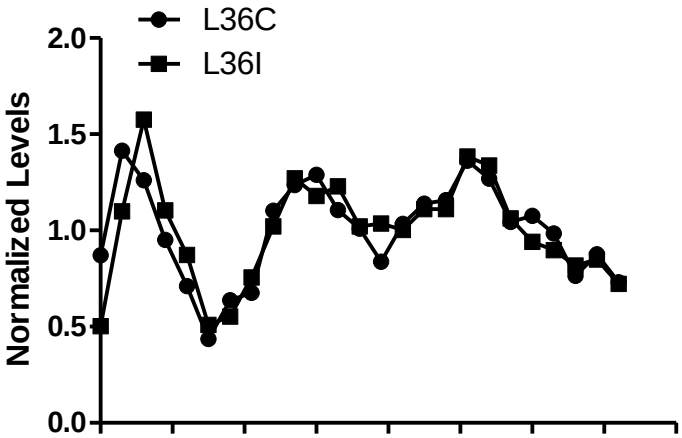


Figure 31: Example of two potentially dampening oscillator mutants that are classified as rhythmic with a nearly wildtype period

For all mutants that produced a functional clock phenotype in the mutagenesis library, many additionally were not found in other cyanobacterial species.

Especially at positions with intermediate levels of conservation from 17 to 24, only a small percentage of all the possible functional residue substitutions are observed in other cyanobacteria clock KaiC sequences (Fig. 32C). This is especially the case for residue 19, where no substitutions prevented circadian rhythmicity except for the stop codon. The expanded number of functional residues at each position than what is observed in naturally occurring clock KaiC sequences indicates potentially either an evolutionary selection for a more precise phenotype that is used in the analysis (i.e. specific range of amplitudes), or that only a limited subset of all permissible residues have been explored evolutionarily. As evidence for the former explanation, dampening oscillations are difficult to confidently characterize due the limited time course/number of cycles and by the criterions used here are sometimes characterized as functional, though studies demonstrate they have lower competitive fitness than cyanobacteria with a stable wildtype Kai oscillator (73) (Fig. 33).

Codons Effects: Using the sequencing data we also examined whether synonymous codon pairs (two degenerate codons which code for the same amino acid) resulted in similar phenotypes based on the three metrics developed above 1) deviation score 2) amplitude 3) period difference (from wildtype) Z-score. The distribution of the absolute difference for the above three metrics are then compared between all possible synonymous pairs in the library versus the same number of randomly drawn non-synonymous pairs from the library to demonstrate whether if indeed amino acid differences generate the most phenotypic differences so that the non-synonymous pairs should have a wider or right-shifted distribution (Fig. 34A-C). This is true for all three metrics, especially for the period parameter, where all Z-score differences greater than 1 were found only between non-synonymous pairs. However, through this analysis, it is apparent that some synonymous codons pairs fall within the overlapping tails portions of the distributions

indicating that codons substitutions can produce significant phenotypic differences without corresponding amino acid changes in this region of KaiC.

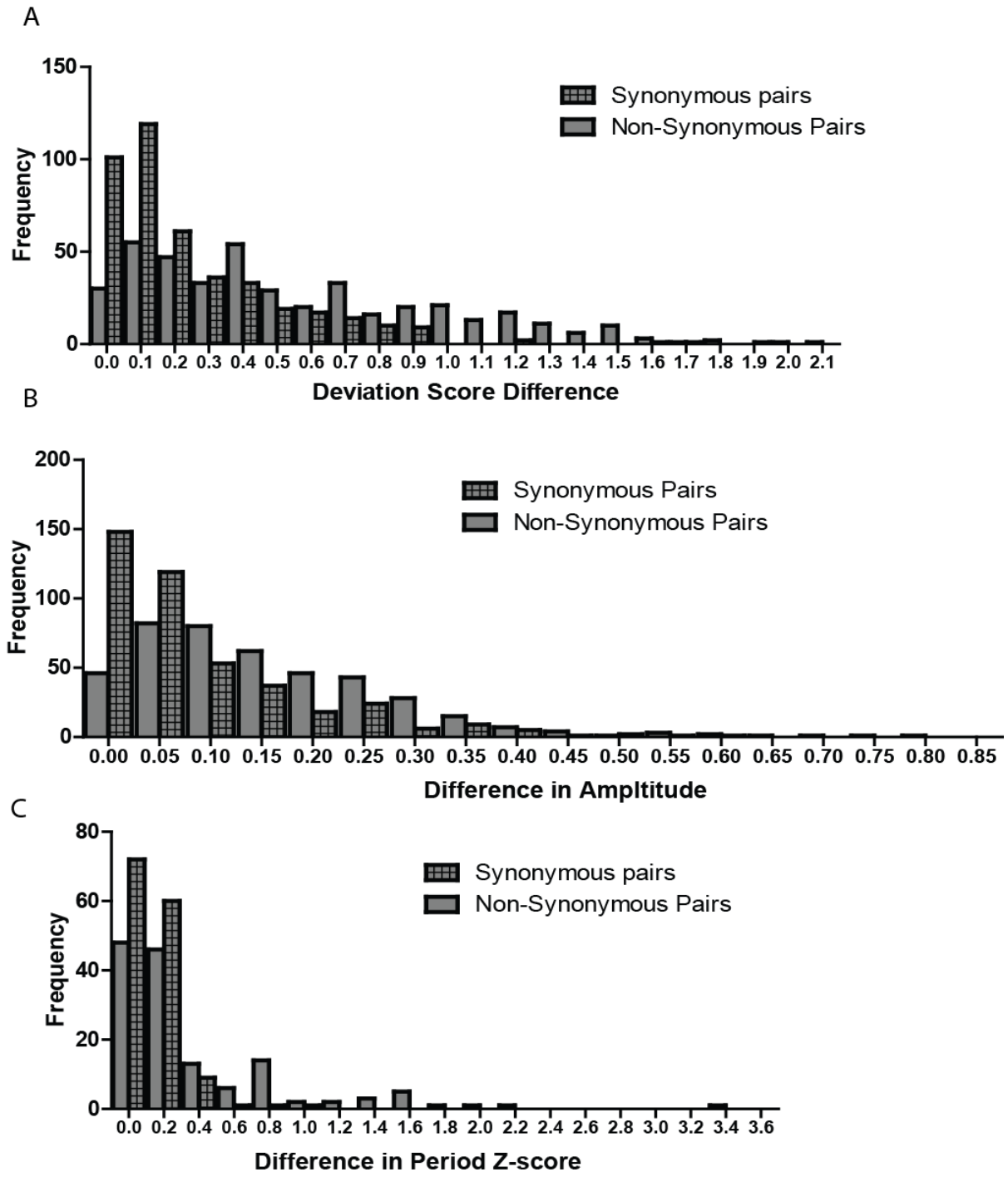


Figure 32: A) Distribution of differences in deviation score between synonymous codon pairs as compared to the distribution of deviation score differences between the same number of randomly chosen non-synonymous codon pairs **B)** Distribution of differences in the amplitude (from sine curve fitting) between synonymous codon pairs as compared to the distribution of amplitude differences between the same number of randomly chosen non-synonymous codon pairs **C)**Distribution of differences in the period difference Z-score between synonymous codon pairs as compared to the distribution of Z-score differences between the same number of randomly chosen non-synonymous codon pairs

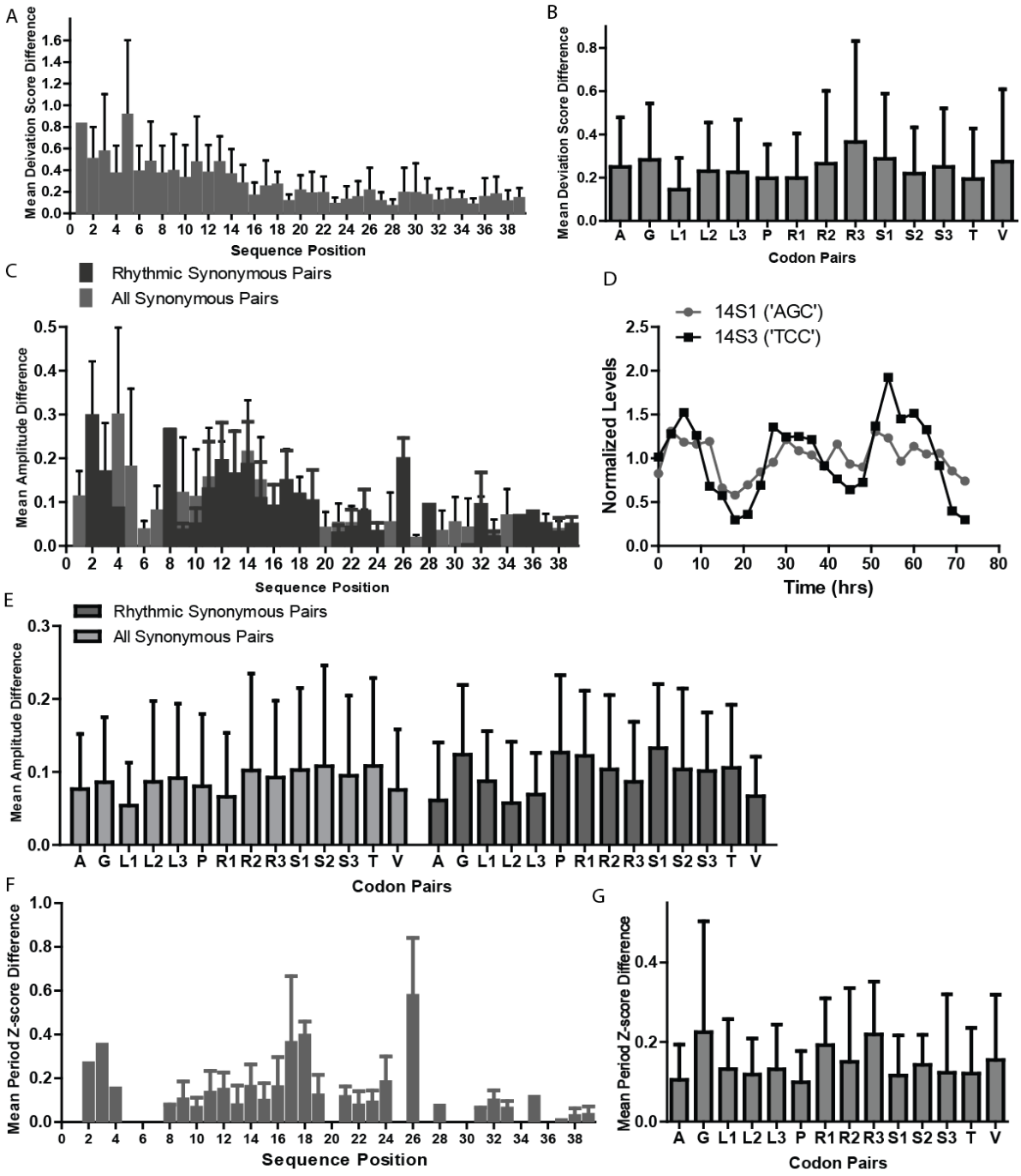


Figure 33: A) Mean deviation score difference between all synonymous codon pairs at each sequence position, error bars indicate standard deviation **B)** Mean deviation score difference for each synonymous pair across KaiC positions 1-39, error bars indicate standard deviation **C)** Mean amplitude difference between all synonymous codon pairs at each sequence position, error bars indicate standard deviation **D)** Example of a relatively large amplitude difference between two synonymous codons for Serine 14 **E)** Mean amplitude difference for each synonymous pair across KaiC positions 1-39 , error bars indicate standard deviation **F)** Mean period Z-score difference between all synonymous codon pairs at each sequence position, error bars indicate standard deviation **G**) Mean period Z-score difference for each synonymous pair across KaiC positions 1-39, error bars indicate standard deviation

In terms of amplitude, the synonymous pairs with the highest differences fall largely within the regions of the sequences where most mutations result in a functional clock phenotype. This is especially apparent, when examining synonymous pairs where both mutants are rhythmic (Fig. 35C). For these synonymous codons, as demonstrated by an example in Figure 33D, one codon of the synonymous pair will result in high amplitude rhythms but the other will produce rhythms with a lower amplitude. This pattern is largely mirrored by the differences in deviation score between synonymous pairs (Fig.. 33A), and suggests that codons differences may modulate amplitude, likely through changing the protein levels of KaiC. Previous studies have also demonstrated cases where codon usage in the Kai genes can affect oscillations by

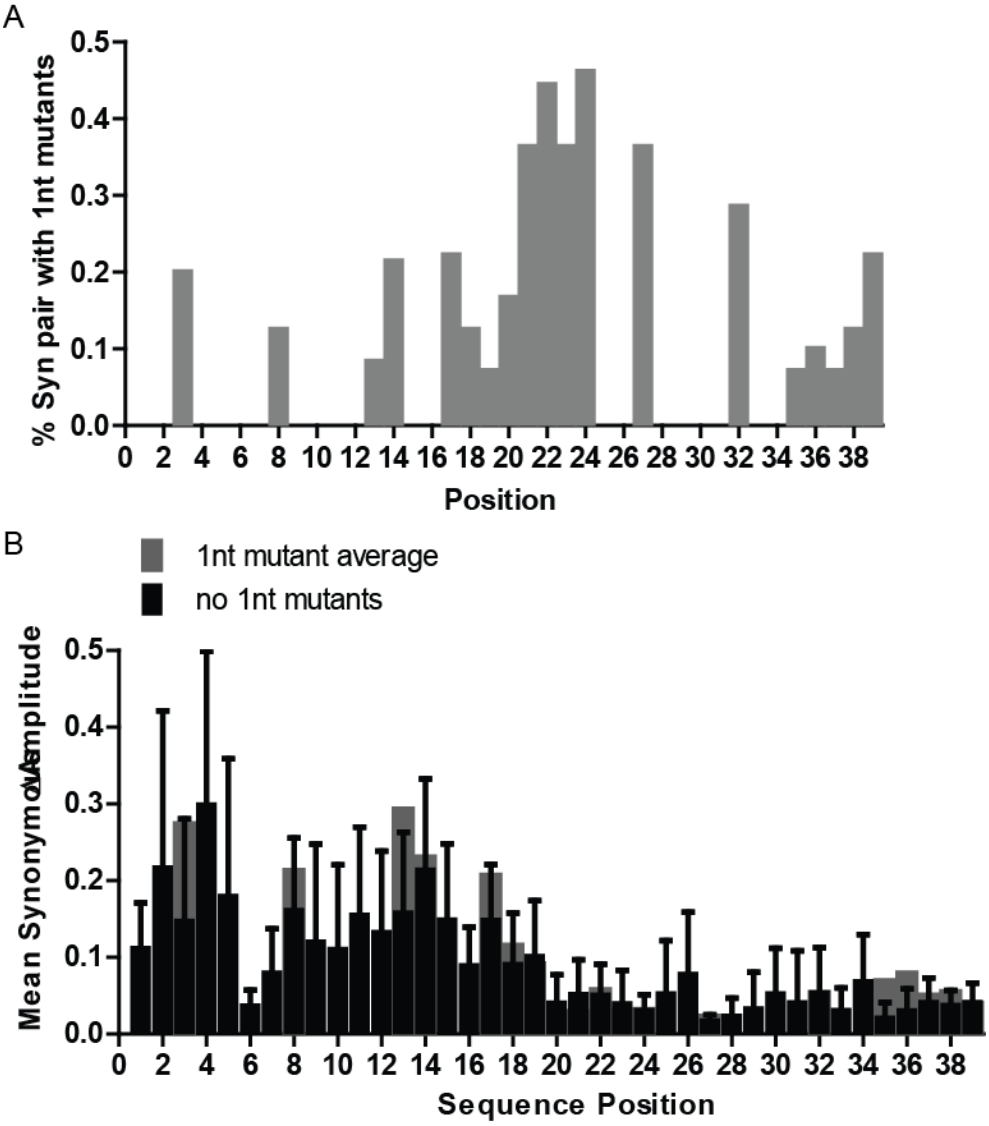


Figure 34: A) Percent of synonymous pairs of which at least one is a 1 nucleotide mutant. **B)** Amplitude difference between synonymous pairs excluding pairs with 1nt mutants, and the average amplitude difference of all pairs including 1nt mutants

broading the functional range of the oscillator(75).

Given the low read count threshold set for 1nt mutants, we also considered the possibility that synonymous pairs involving 1nt mutants with higher levels of noise/error may be artificially increasing the amplitude difference at positions that are enriched for these pairs involving 1nt mutants. However, the % of pairwise synonymous pairs involving 1nt mutants is lower at the part of the sequence where the average phenotypic difference between synonymous codon pairs are the highest, and removing all pairs involving 1nt mutants does not significantly the average synonymous amplitude differences across all positions (Fig. 36A,B).

Additionally, there does not appear to be synonymous codon pairs that consistently produce phenotypic differences across the first 39 KaiC positions, though one of the leucine pairs (L1) seemed most phenotypically similar upon substitution across this portion of KaiC (Fig. 35B). There were also no synonymous pairs with significant period Z-score differences (>1).

Sequencing Pilot Conclusions

- There is a clear correlation in the various measures of mutational tolerance (median deviation score from wildtype, % arrhythmic) at each position with both structure and conservation score. Largest discrepancy can be observed in the N-terminal region from 2-6, where the sequence is varied across cyanobacteria species, but the area is relative sensitive to mutation.
- Most naturally occurring clock KaiC residues that are covered in the library are functional in KaiC, especially in conserved KaiC regions, where almost all naturally occurring residues are functional. This suggests sequence-independent evolutionary constraints for function at these positions. The main exception is Met6, where almost all

naturally occurring KaiC residues substitutions result in arrhythmicity, suggesting a highly constrained functional role for this position. One hypothesis is that Met6 may serve as an important alternative start site. Alternatively, given the low number of mutant reads mapping to that region (Fig. 24B) , it could also play a role in RNA stability

- Most codon substitutions in the KaiC 1- 39 cause amplitude changes or arrhythmicity, but relatively very few mutations change the period of oscillations and these can be generally found in conserved regions that are most sensitive to mutation. Speculatively, it is possible that different amino acid substitutions at these positions may causes varying degrees of perturbation to essential activitie(s) or kinetic parameter(s) of KaiC, and period mutations result from smaller perturbations to these parameters than mutations that result in arrhythmicity. The observed anti-correlation between amplitude and the magnitude of period difference from the wildtype period also suggests that period and amplitude are structurally linked properties where higher period changes (larger biochemical perturbations) will decrease amplitude until arrhythimicity results.
- Longer period mutant tend to have large amplitudes than short period mutants
- Synonymous codons substitutions may modulate the amplitude of oscillations

Further improvements for full library sequencing

cDNA bottleneck: Insufficient mutant cDNA copies at the start of the sequencing library preparation process can contribute random Poisson noise to mutant RNA quantifications at each timepoint. We measured the total KaiC cDNA yield at each time point by qPCR with a primer set targeting the N-terminal region of KaiC to determine whether if the cDNA copies were sufficiently high enough in the pilot to allow for accurate quantification. This was estimated based on an absolute standard curve of the *kaiBC* gene fragment. Accounting for the library

distribution, when divided among the small number of mutants in pilot library, cDNA copies were sufficiently high as to be not a significant bottleneck except for the lowest 10% of mutants (Fig. 37A). To see if cDNA copies would be a constraint for the full library, we extrapolated using the pilot mutant distribution by assuming a 10-fold reduction in cDNA copies per mutant, and found that the amount of mutants with sufficient cDNA copies may drop to an average of 67%, with some region in the N-terminal region where only 25% of the codons are represented (Fig. 37B). The final result would rather depend on the evenness of the library and at this level

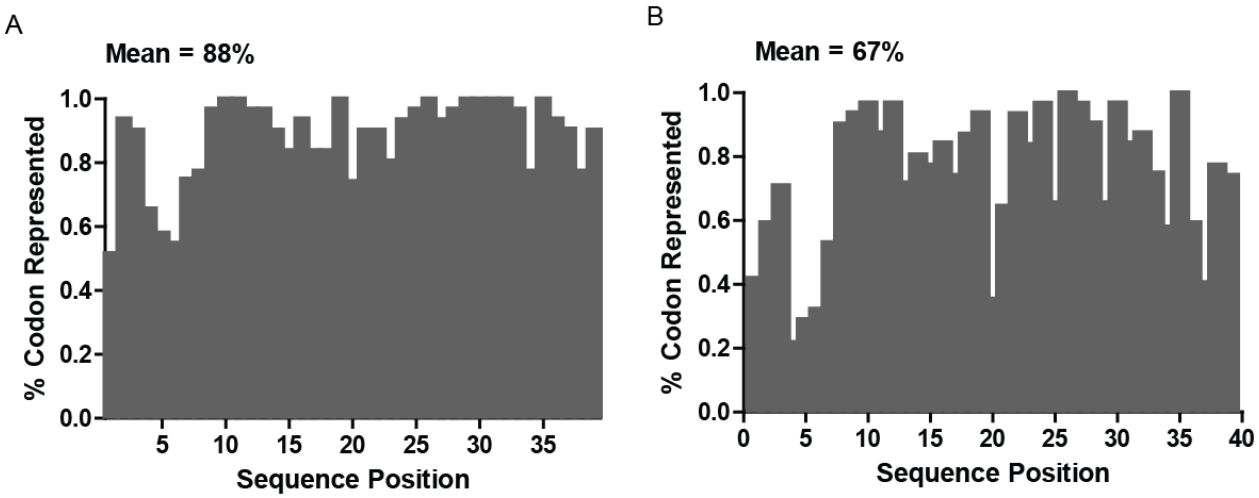


Figure 35: Mutant cDNA bottleneck may limit library coverage **A)** Applying a cDNA filter where all mutants whose mean cDNA copy number coefficient of variation ($1/\mu_{cDNA}$) is >20% of their coefficient of variation over time are removed. **B)** Same as A, but given a 1/10 reduction in the number of copies for each mutant

of coverage much mutant data can still be recovered, optimization of the reverse transcription to improve cDNA yield would increase robustness of the technique.

Sampling window: Throughout time course, it is apparent that the expression or mRNA levels of KaiC trends upwards independent of the constant total RNA concentration used in the reverse transcription reaction (Fig. 38). Since we are largely interested in examining oscillatory, free running clock function, the optimal sampling window would after the first day where amplitudes of oscillations are higher. This would increase the signal to noise ratio of the data set, by

decreasing noise resulting from lower cDNA copies and increase the confidence around our fitting parameters (amplitude, period, phase) in phenotypic assessment.

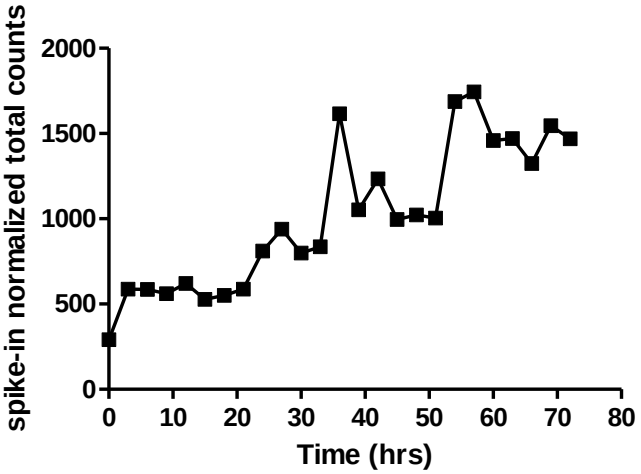


Figure 36: Spike in normalized total read counts for all mutants over time

Full Library Sequencing Potential Pitfalls: It is clear given the unevenness in the distribution of mutants that increasing the efficiency of sequencing by reducing wildtype reads would allow for greater depth of coverage and phenotypic information on more mutants, without prohibitive costs. Additionally, to phenotype 1nt mutants given the sequencing error burden, overlapping paired-end reads is necessary to ensure accurate 1nt mutant read calls. Hence, the best sequencing approach would be then to cover as much of the *kaiBC* gene as possible in one read, and the longest read possible on the HiSeq sequencer (higher sequencing output than the MiSeq which is capable of slightly longer reads) is 250bp. However even at this read length, covering the *kaiBC* gene will still require at least 9 different reads, limiting the chance of sequencing a mutation to 1/9. Hence, to increase read efficiency, an enrichment strategy for mutant regions is necessary.

Testing Strategies for Sequencing the Fully Mutagenesis Library

Library preparation strategy

To enrich for the region containing the single codon mutations, we examined this following approach (Fig. 39):

1. Mutate in tiles of ~230bp on plasmid containing a unique 3'UTR tag that marks the region of mutation. There will be 9 areas of mutations or tiles
2. Use a tag specific reverse transcription primer to specifically reverse transcribe all KaiC/KaiB RNA molecules with a specific tag, essentially enriching for KaiC/KaiB mutants in one specific region
3. PCR amplify the mutated tile region on the tag selected cDNA
4. Add on P5/P7 adaptors with the appropriate timing barcode index and 'NNNNN' at the start of read 1 (the number degenerate bases is increased relative to the pilot to decrease use of filler libraries and increase read efficiency)

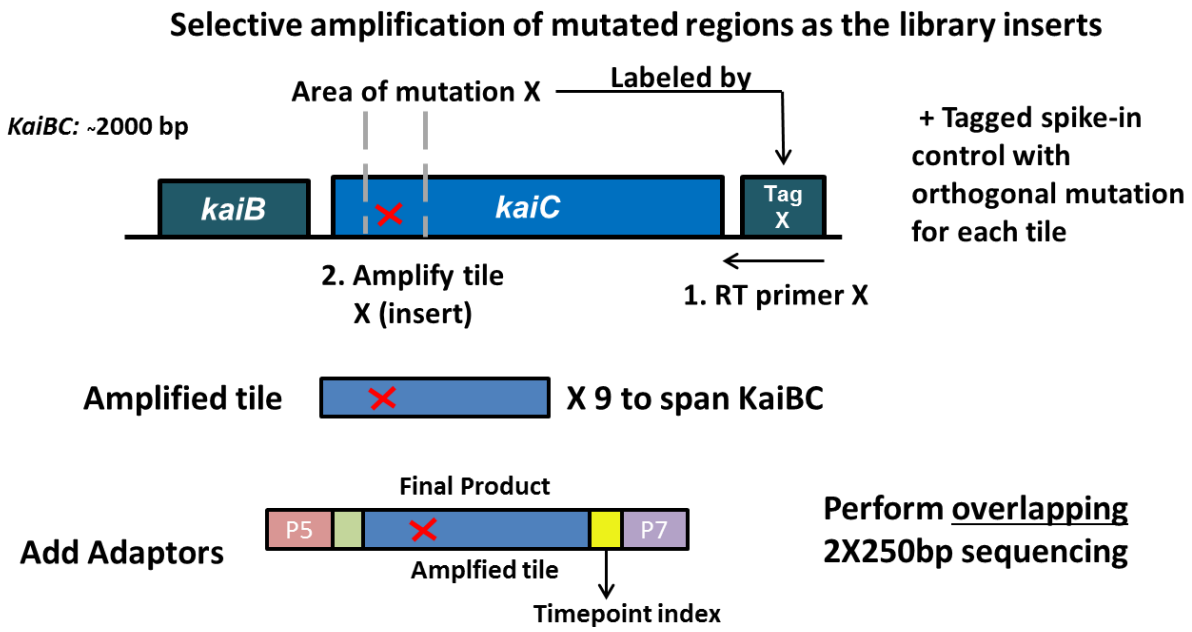


Figure 37: Schematic of the fully KaiB-KaiC mutagenesis library sequencing strategy

Step 1 of this strategy has been completed, where 9 different Pfunkel mutagenesis reactions were performed to mutate 9 different regions of KaiB and KaiC. Nine different pAM2314 plasmid libraries with mutations in these 9 separate regions or tiles were transformed separately into *S. elongatus* following the same protocol and cell to DNA ratio as for the pilot library. Estimated

library coverage by colony count was around 8-10X (150 – 200,000 total colonies), and was generally even among all 9 different transformations. After collecting all the transformants, each mutant tile library was then grown until OD 0.7 – 0.9 and mixed together evenly and redistributed to 4X100mL cultures in selective BG-11 media. This was then propagated to16X100ml cultures and grown for 3 weeks while maintained at OD 0.3 to allow for sufficient segregation of the wildtype genome copies. About 20mL from each of these separate cultures were then combined into one culture which was then used to inoculate a 4 liter culture at OD 0.2 for the time course experiment.

Full library mutant time course:

We decided to avoid using sodium bicarbonate as a carbon source as we did in the pilot because since it was consumed and could fluctuate over time, contributing a potential source of systematic variation to expression levels in cells. Instead, to enable constant aeration of the media, we used a 10G aquarium pump to bubble air consistently into the large 4 liter library culture (in a continuously spinning 6L spinner flask). The culture was then subjected to 2 days of 12 hours light and 12 hours dark in the Percivial growth chamber. For the time course sampling was done in duplicate, with 90mL of culture for each sample, starting 14 hours after the last dark pulse, and was done for next 99 hours every 3 hours using the same protocol as in the pilot. Only 25 time points between 32 and 104 hours were then used for generating the sequencing library. RNA extractions of the timepoint samples were done using the same protocol as the pilot library.

Enrichment by 3'UTR selection:

Unfortunately, initial testing with the earlier library timepoints indicated that selectivity of a specific 3'UTR tag on the KaiC mRNA was very low in reverse transcription, where the relative

enrichment of the selected tag versus others was low in the same total RNA timepoint sample, mostly around 2X -3X the non-selected tag, and can vary depending on the tag (Fig. 40A, B). Given the number of tags (9 tags) in the same RNA sample, this low level of enrichment would result in the selected tag to consist of only 20-30% of the total KaiC cDNA. Enrichment ratios were calculated using ΔCt values from qPCR with a specific primer for the two tag sequences compared and by assuming an equal PCR efficiency of 1.95. PCR efficiency between tags was found to be similar using dilutions of a library timepoint sample (ranging from 1.91-1.98), but their actual variation between tags could contribute error to the exact calculations of enrichment – however, this does not change the conclusion that selectivity was insufficient to achieve majority level of the selected KaiC cDNA in the total KaiC pool.

Given the data from the pilot sampling and sequencing test, maximizing both tag selection specificity (tag purity) and cDNA yield from the reverse transcription reactions are necessary to prevent bottlenecking at the cDNA and sequencing (mutant depth of coverage) steps for the lower 30 percentile of mutants represented in the library. To improve reverse transcription by these criteria the following options were explored.

Strategies to Improve Tag Selectivity: First, to address the specificity issue, we first hypothesized

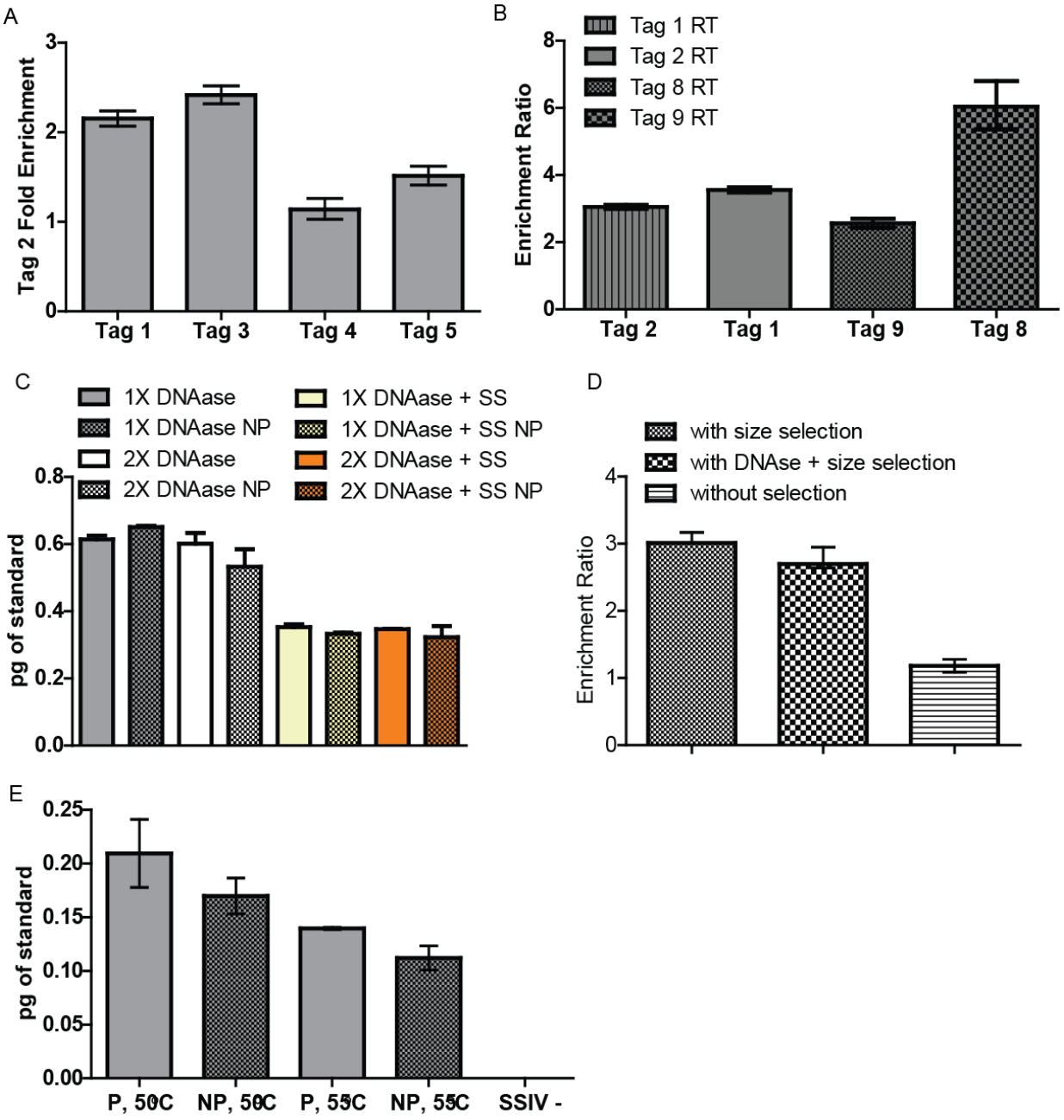


Figure 38: Tag Selectivity Limited by Primer Independent Reverse Transcription **A)** Relative enrichment of Tag 2 cDNA compared to Tag 1, 3, 4, and 5 cDNA in a reverse transcription reaction using a Tag 2 primer. **B)** Relative enrichment of the tag selected by reverse transcription (RT) primer compared to the tag indicated on the x-axis **C)** Amount of Tag2 cDNA in a reverse transcription reaction with a Tag 2 primer or without (NP) using total RNA that was size selected with Ampure beads (SS) or treated with 1X or 2X concentration of DNAase.**D)** Enrichment of Tag 2 in a RT reaction with the Tag 2 primer using a RNA timepoint sample with DNAse treatment, size selection or both. **E)** Amount of Tag 2 cDNA in a RT reaction at different temperatures, with or without the Tag 2 RT primer. Enrichment determined by qPCR using ΔCt values.

that small DNA oligo contamination due to incomplete DNAse digestion of the genomic DNA may be acting as unintended primers in reverse transcribing some regions of the *kaiBC* transcript. This was supported by the fact that in the absence of any added reverse transcription primer, KaiC cDNA yield was comparable to a when a specific primer was added. To remove possible DNA contamination, digestion with 2X the concentration of DNAse was done, as well as size selection using Agencourt Ampure RNAse-free beads (RNAClean XP) at a ratio of 0.65X bead: liquid volume, removing all RNA<200bp. However, neither procedure or when combined reduced the amount of primer independent KaiC reverse transcription, and while tag-selection slightly increased with these two steps, they also reduced overall KaiC cDNA yield by half, and hence did not quantitatively increase the cDNA yield of a selected KaiC mRNA (Fig. 40C,D).

To increase the specificity of priming or reduce RNA secondary structures which may either inhibit tag primer binding or act as a primer-independent start site for reverse transcriptase, a higher temperature reverse transcriptase, superscript IV was used and reverse transcription performed at 50 and 55°C instead of 45°C. However, this also did not decrease primer independent reverse transcription (Fig. 40E).

To increase the yield of the primer-dependent reverse transcription in case the tag-specific primers were being sequestered by non-specific interactions with the overwhelming amount of other RNAs from the cyanobacteria cell, we performed mRNA enrichment using the Ambion MICROBExpress kit per the manufacturer's instructions. However, the enriched mRNA did not

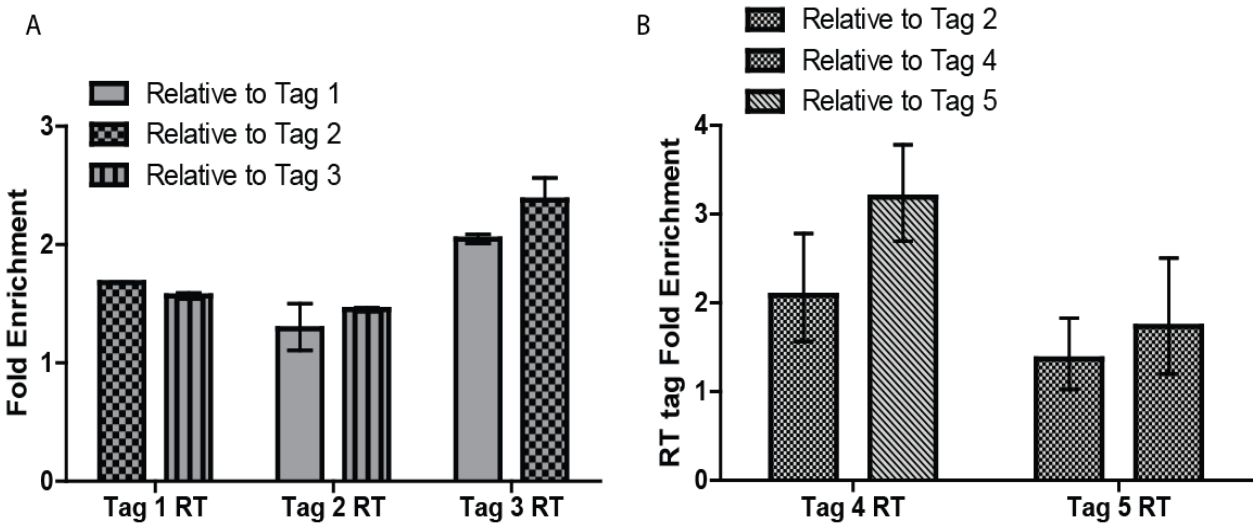


Figure 39: Tag selectivity in RT reactions with mRNA enriched and IVT RNA A) Enrichment of reverse transcription (RT) selected tag relative to indicated tag in cDNA pool using mRNA-enriched RNA sample. **B)** Enrichment of the RT selected tag relative to indicated tag in cDNA pool using in-vitro transcribed mixture of all 9 tagged KaiB-KaiC RNAs

demonstrate tag selection specificity in reverse transcription nor higher KaiC reverse transcription efficiency (cDNA yield/ug of total RNA input) (Fig. 41A).

Lastly, a similar non-specificity was observed for the in vitro transcribed spike-in RNA, which was pure and added at magnitudes lower amounts (50pg) than the total RNA to the reverse transcription reaction, suggesting that this was something that was intrinsic to the tagged RNA, such as very stable RNA secondary structure or a product of the exact reverse transcription reaction conditions (Fig. 39B).

Tag-specific cDNA amplification PCR Step: Since the 3'UTR tags could not be specifically reversed transcribed, to select for a specific tag and its associated region of mutation for sequencing, an additional tag specific PCR step was introduced after reverse transcription (Fig. 42). While the reverse primer anneals to a specific primer, the forward primer would anneal 5' to

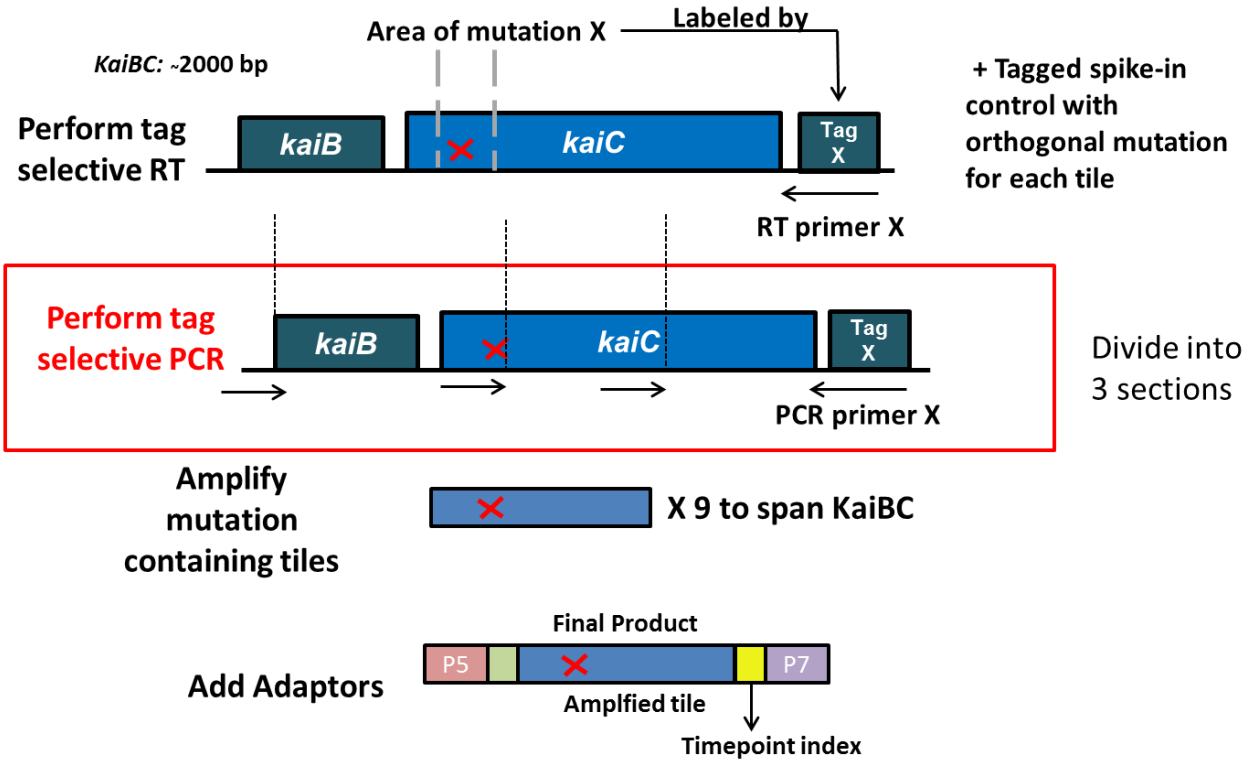


Figure 40: Schematic of the library preparation process with an additional tag selection PCR step

the region of mutation. There will be three different forward primers used to generate 3 differently sized fragments (~700, 1340, 1980bp) where longer fragments covered mutational regions further from the 3'UTR tag. In testing this additional step, it became apparent that a potential cDNA bottleneck may be generated due to limits on the volume of the reverse transcription reaction (cDNA template) that could be added to this PCR. Especially problematic was amplifying the PCR product which covers the mutational regions furthest from the tag, where the % RT reaction tolerated by the PCR reaction without significantly decreasing yield or increasing non-specificity was only 5% as opposed to 10% for the more C-terminal mutational regions/shorter PCR fragments. By qPCR against an absolute standard, the average copies of each variant was only around 200 – 500 given a 10% reverse transcription reaction volume in the downstream PCR, depending on the timepoint and the tag (assuming 30% of total quantified cDNA by qPCR is the selected tag or around a 3X enrichment). This is aroud the limit where lowly represented library member (at least 10 fold less than average) would be subjected to high levels of Poisson noise or be entirely absent at most timepoints. To circumvent this limitation, various methods were attempted to increase cDNA copies and reverse transcription efficiency and yield.

Increasing cDNA yield: To increase KaiC cDNA yield, increasing the amount of superscript III by 1.5X was tested, as well as increasing the input total RNA amount to 2X and 3X the

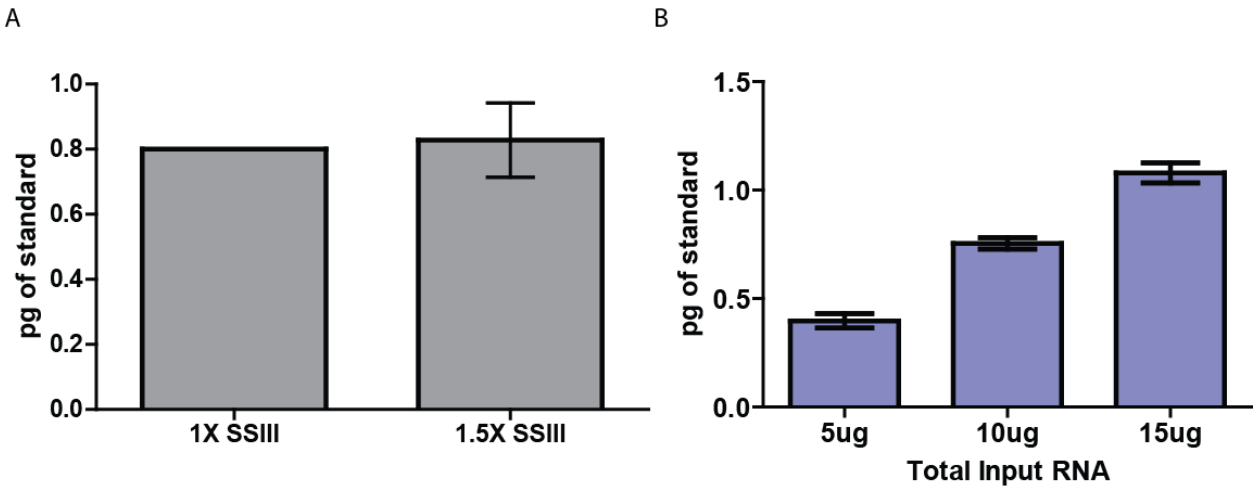


Figure 41: A) KaiC cDNA yield using 1X and 1.5X of the Superscript III for reverse transcription **B)** KaiC cDNA using varying input total RNA amount in a 20µL reverse transcription reaction. Amounts determined by qPCR using an absolute DNA standard.

recommended amount for the reverse transcriptase (Fig. 43A, B). While cDNA yield was increased, subsequent tag-selection PCRs with the higher yielding RT reactions resulted in much higher levels of non-specificity as compared to the same PCRs with an equivalent input volume of 1X RT reaction (at 1X reverse transcriptase and total RNA input). Hence, it is not clear if increasing superscript and total input RNA concentrations in the RT reaction allows more cDNA copies per variant to be amplified given the low amplification specificity at the next PCR step.

Sequencing of the partial KaiB-KaiC mutagenesis library time course

Given uncertainty surrounding the new sequencing format (2X250bp sequencing), and possible sources of bottlenecking, noise, or non-specificity introduced by the possible tag selection methods, two sequencing approaches will be tested which if both successful would produce phenotypic information for 56% of all library mutants. First, due the potential for mutant cDNA

bottlenecking especially during the tag enrichment PCR step of mutational regions distant from the 3'UTR tag, a sequencing library covering two distant tiles was constructed for 1 lane of overlapping 2X250bp sequencing on a Rapid-mode HiSeq flowcell using tag-specific RT only with no subsequent tag enrichment PCR step. For the second lane, a sequencing library covering mutants in the three mutational regions closest to the 3'UTR tag was constructed with the additional tag enrichment PCR step after reverse transcription. This was done using one common forward primer and specific tag primers corresponding to the three different associated tags to produce around a 850bp fragment. Both libraries were constructed based on the procedures outlined in Figure 37 (for KaiB and N-terminal region of KaiC in tile 8 and 9) and Figure 42 (for N –terminal region of KaiC in tile 1, 2, and 3).

Future Directions

Full mutagenesis library analysis plan

Initial alignment and variant calling will be done with the custom Matlab script described previously. This computational procedure is able to replicate the variant counts (within a <2% difference except for variants with very low counts <20, then difference goes up to 10-15%) made by the bwa alignment procedure used for the pilot data. Additionally, it allows for the elimination of reads with very low Phred quality scores at the variant bases.

First, beyond the normalization steps used in the pilot data, some preliminary quality control steps needs to be first completed before further analysis based on positive controls mutants such as synonymous mutations. One concern is bottlenecking at the cDNA amplification step, which may reduce the number of mutants sequenced, but also introduce random Poisson noise in the levels of variants present at very low copies (i.e. <10). To identify if random noise is significant,

and to set a read count threshold to eliminate mutant traces where noise makes a dominant contribution, for synonymous mutation, correlation between noise and read counts will be made by visual inspection. In addition to synonymous mutation, other known mutants/previously characterized mutants will be used as positive controls to determine the sensitivity and accuracy of the assay for the two library preparation procedures. These include a number of period and arrhythmic mutants determined by a previous mutagenesis screen which fall within the regions covered by this partial library (see Table 1, Ishiura *et al*). Additionally, the phosphorylation sites T431 and S432 also fall within the sequenced region where substitutions should result in arrhythmicity or dampened oscillations. These positive controls would also allow us to determine 1) sufficient read count thresholds where noise from low cDNA copy numbers does not inhibit accurate mutant phenotyping/parameterization, 2) if read count independent sources of noise are sufficiently removed by normalization procedures with STOP codons and the spike-in control to accurately phenotype mutants.

Table 1: List of clock mutants from Ishiura *et al* covered by the partial mutagenesis library

Mutant	Phenotype
KaiB L11F	21h period
KaiC R74W	22h period
KaiC R321Q	21hr period
KaiC T409A	27hr period
KaiC G421R	44hr period
KaiC Y442H	60hr period
KaiC G460A	Arrhythmic
KaiC T495A	Arrhythmic

The sequencing strategy which yields the highest library mutant coverage given a similar mean read count cutoff and/or the least amount of noise by visual inspection of positive controls will then be used to complete the rest of the library. If data is of sufficient quality (as determined by

visual inspection of positive controls), phenotypic characterization of rhythmicity, amplitude and period will be done as for the pilot data set for both the partial library. The following questions will be explored using the full library sequencing dataset.

***Spatial distribution of mutations*:** Deviation score from the synonymous 'wildtype' trace as well as % of residues characterized as arrhythmic will be used as measures of mutational tolerance at each position. This will be mapped to the crystal structures of KaiC and KaiB, to determine hotspots locations of low mutational tolerance which would indicate high levels of constraint on the amino acid identity of the position, and a specific role in function. These hotspot positions/regions will be further explored in the following ways:

- Chemical similarities in tolerated residue substitution, do they indicate a specific chemical constraint (size, hydrophobicity, charge)?
- Do these hotspots consist of physically interacting networks?
- Do their physical interactions with other residues and the chemical profile of tolerated substitutions indicate a structural role for these residues (forms the protein core, tertiary structures)?

Additionally, further parameterization of mutant traces by sine curve fitting to the determine period, phase and amplitude of mutant oscillations will be used to address the following questions:

- Spatial pattern of period changing mutations, do they overlap with hotspots for mutational intolerance? Are they also concentrated in a specific structural location (period modulating hotspot)?

- What is the spatial pattern of positions enriched for amplitude changes (increasing and decreasing amplitude) relative to average synonymous amplitude? While lower amplitude is expected in functional hotspots due to arrhythmicity, are there regions where mutations would cause increased amplitudes?

Use spatial distribution of mutation to make biochemical inferences:

Validate and extend existing biochemical hypotheses:

- ***KaiA activation:*** disruption of A-loop interactions by removing hydrogen bonding partners would lead to arrhythmicity or other severe oscillatory perturbations.

- ***KaiB binding:*** B-loop region in KaiB including F121 is hypothesized by structural models to be essential for KaiB binding mutations in this region should abolish arrhythmicity and potentially define a binding interface on KaiB for KaiC. Mutational tolerance at positions essential for maintaining the two secondary structure states of KaiB (G88, D90) may also reveal the phenotypic consequence of modulating the fold-switch equilibrium and sequence constraints set by this regulatory mechanism.

- ***CI/CII stacking:*** define inter-domain stacking interactions or residues necessary for allosteric communication (i.e. network of interacting residues) between the CI/CII to connect the two active sites/and the KaiB binding interface.

- ***CI/CII ATPase sites:*** Does the kinetics of one of the ATPase sites play a dominant and tunable role in setting the period where you would observe a period modulating hotspot?

- **Define regions that are important to protein folding and structure**: Are there residues that are functionally constrained to be hydrophobic? Do they have a hydrogen bonding

donor/ acceptor? Do they form the protein core or are involved in secondary/tertiary structures?

- **Are there regions of high mutational sensitivity on the protein surface or unstructured regions?** Do they present an essential binding surface? Are they possible involved in translation regulation? To investigate whether a codon substitution has effect on the average expression level of the KaiB/KaiC RNA, DNA sequencing of the library can be performed and these reads can be used to determine relative RNA/DNA ratios

Additionally some system-level relationships can be draw about how the circadian function is encoded by examining the following questions:

- Given the complexity of KaiC's enzymatic function, are there multiple structural positions that control period or are period mutant concentrated in region or a network of physically connected residues? Are period mutations mostly found in relatively mutational sensitive regions? Does the spatial pattern of period mutants inform us about a potential rate limiting step in the oscillatory kinetics?

- Are the parameters defining the circadian output controlled modularly where specific structural regions control period but not amplitude? Or rather are parameters linked in mutational space where period change is always concordant with amplitude changes?

- Are period changing mutations relatively rare compared to amplitude changing /arrhythmic mutant? And among all the period mutants, are there more short or long period mutants? Do short period mutants on average have lower amplitudes relative to long period mutants?

- Do the above phenotypic relationships inform us about the evolutionary constraints faced by the system (i.e. long more advantageous than short period)? Can this also constrain mathematical models of the system (i.e.amplitude and period must change concordantly)?

Evolutionary insights: using comparisons between mutational tolerance and the phenotypes of mutational substitutions we can further explore these following questions:

- **Mutational sensitivity and conservation:** How well does the mutational tolerance at each site match with relative evolutionary conservation scores?
  - Are there regions of low conservation but high mutational sensitivity? Are these physically interacting or is there evidence of coevolution of these positions in the multiple sequence alignment? Do these indicate another unique S. elongatus adaption/function?
  - Are there period mutants that can be found in other homologous Kai clock sequences? (none from the pilot library)
  - What are the structural roles for the relatively conserved positions where most, if not all functional mutations are the conserved residues in other clock KaiC sequences? Are these residues interacting or do they independently support a highly conserved/finely tuned structural/functional role?

- **Mutational robustness:** Is there a correlation between number of nucleotide changes and the severity of the clock phenotype (ex. lower % arrhymicity with non-synonymous 1nt mutants vs 2nt vs 3nt). What phenotypes are less likely with 1nt mutations vs 2 or 3nt mutations? (i.e. are period mutations more likely to result from a >1nt mutation than

expected by chance?). While we expect from the genetic code that less perturbative amino acid substitutions will result from 1nt mutations, what phenotype is selected against may reveal evolutionary pressures on the clock. This is difficult to assess accurately in the pilot library given the 1 nucleotide sequencing error burden.

Technical Improvements: Some suggested solutions for potential problems are listed in the table below:

Table 2: Potential sequencing data problems and proposed technical remediation

Problem	Indications	Potential Remediation
cDNA bottleneck	Random noise associated with lower read counts	Adjust $MgCl_2$ concentration in RT buffer to increase % RT reaction in downstream PCR
non-linear PCR amplification	High levels of random (or systematic) noise independent of read count*	Apply SAFE-seq method of using degenerate sequences on primers to get cDNA counts
Low % Mutant Reads (Poor Tag Selectivity)	Low library coverage or most variants have insufficient read counts or no reads for some timepoints	Optimize RT buffer conditions to increase selectivity.
		Do single-ended reads to can get more unique reads for the same cost (tradeoff: sequencing error may interfere with 1nt mutant phenotyping
		Separate cultures for each 250bp read region (tile) during the sampling timecourse
High Levels of Non-Library Mutants	Levels of double mutants > $10^{-(Qavg/10)}$	Decrease PCR cycles used or use Q5 polymerase**

* Systematic noise should be removed by spike-in controls. Non-linear PCR amplification may also be associated with lower read counts, where mutants with lower average read count may be non-linearly amplified at specific timepoints leading to noise or misrepresentation of the mutant phenotype.

** Not expected given polymerase error rates and number of cycles used (<35 cycles in total)

REFERENCES

1. Gonze D & Abou-Jaoudé W (2013) The Goodwin Model: Behind the Hill Function. *PLoS ONE* 8(8):e69573.

2. Bell-Pedersen D*, et al.* (2005) Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nat Rev Genet* 6(7):544-556.

3. Markson JS & O'Shea EK (2009) The Molecular Clockwork of a Protein-based Circadian Oscillator. *FEBS letters* 583(24):3938-3947.

4. Cavallo L, Kleinjung J, & Fraternali F (2003) POPS: a fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Research* 31(13):3364-3366.

5. Winfree AT (2000) *The Geometry of Biological Time* (Springer-Verlag, New York) 2nd Ed p 777.

6. Young MW & Kay SA (2001) Time zones: a comparative genetics of circadian clocks. *Nat Rev Genet* 2(9):702-715.

7. Eckardt NA (2005) Temperature Entrainment of the Arabidopsis Circadian Clock. *The Plant Cell* 17(3):645-647.

8. Beaver LM*, et al.* (2002) Loss of circadian clock function decreases reproductive fitness in males of Drosophila melanogaster. *Proceedings of the National Academy of Sciences* 99(4):2134-2139.

9. Turek FW*, et al.* (2005) Obesity and Metabolic Syndrome in Circadian Clock Mutant Mice. *Science (New York, N.Y.)* 308(5724):1043-1045.

10. Green RM, Tingay S, Wang Z-Y, & Tobin EM (2002) Circadian Rhythms Confer a Higher Level of Fitness to Arabidopsis Plants. *Plant Physiology* 129(2):576-584.

11. Ouyang Y, Andersson CR, Kondo T, Golden SS, & Johnson CH (1998) Resonating circadian clocks enhance fitness in cyanobacteria. *Proc Natl Acad Sci U S A* 95(15):8660-8664.

12. Mihalcescu I, Hsing W, & Leibler S (2004) Resilient circadian oscillator revealed in individual cyanobacteria. *Nature* 430(6995):81-85.

13. Panda S, Hogenesch JB, & Kay SA (2002) Circadian rhythms from flies to human. *Nature* 417(6886):329-335.

14. Partch CL, Green CB, & Takahashi JS (2014) Molecular architecture of the mammalian circadian clock. *Trends in Cell Biology* 24(2):90-99.

15. Tomita J, Nakajima M, Kondo T, & Iwasaki H (2005) No transcription-translation feedback in circadian rhythm of KaiC phosphorylation. *Science* 307(5707):251-254.

16. Ito H*, et al.* (2009) Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in Synechococcus elongatus. *Proc Natl Acad Sci U S A* 106(33):14168-14173.

17. Vijayan V, Zuzow R, & O'Shea EK (2009) Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proc Natl Acad Sci U S A* 106(52):22564-22568.

18. Nakajima M*, et al.* (2005) Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science* 308(5720):414-415.

19. Iwasaki H, Nishiwaki T, Kitayama Y, Nakajima M, & Kondo T (2002) KaiA-stimulated KaiC phosphorylation in circadian timing loops in cyanobacteria. *Proc Natl Acad Sci U S A* 99(24):15788-15793.

20. Hayashi F*, et al.* (2004) Roles of two ATPase-motif-containing domains in cyanobacterial circadian clock protein KaiC. *The Journal of biological chemistry* 279(50):52331-52337.

21. Rust MJ, Markson JS, Lane WS, Fisher DS, & O'Shea EK (2007) Ordered phosphorylation governs oscillation of a three-protein circadian clock. *Science* 318(5851):809-812.

22. Xu Y*, et al.* (2004) Identification of key phosphorylation sites in the circadian clock protein KaiC by crystallographic and mutagenetic analyses. *Proc Natl Acad Sci U S A* 101(38):13933-13938.

23. Nishiwaki T*, et al.* (2007) A sequential program of dual phosphorylation of KaiC as a basis for circadian rhythm in cyanobacteria. *EMBO J* 26(17):4029-4037.

24. Pattanayek R*, et al.* (2004) Visualizing a circadian clock protein: crystal structure of KaiC and functional insights. *Mol Cell* 15(3):375-388.

25. Terauchi K*, et al.* (2007) ATPase activity of KaiC determines the basic timing for circadian clock of cyanobacteria. *Proc Natl Acad Sci U S A* 104(41):16377-16381.

26. Kim YI, Dong G, Carruthers CW, Jr., Golden SS, & LiWang A (2008) The day/night switch in KaiC, a central oscillator component of the circadian clock of cyanobacteria. *Proc Natl Acad Sci U S A* 105(35):12825-12830.

27. Kageyama H, *et al.* (2006) Cyanobacterial circadian pacemaker: Kai protein complex dynamics in the KaiC phosphorylation cycle in vitro. *Mol Cell* 23(2):161-171.

28. Novak B & Tyson JJ (2008) Design principles of biochemical oscillators. *Nat Rev Mol Cell Biol* 9(12):981-991.

29. Phong C, Markson JS, Wilhoite CM, & Rust MJ (2013) Robust and tunable circadian rhythms from differentially sensitive catalytic domains. *Proc Natl Acad Sci U S A* 110(3):1124-1129.

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

31. Abe J, *et al.* (2015) Atomic-scale origins of slowness in the cyanobacterial circadian clock. *Science* 349(6245):312-316.

32. Nishiwaki-Ohkawa T, Kitayama Y, Ochiai E, & Kondo T (2014) Exchange of ADP with ATP in the CII ATPase domain promotes autophosphorylation of cyanobacterial clock protein KaiC. *P Natl Acad Sci USA* 111(12):4455-4460.

33. Chang YG, Tseng R, Kuo NW, & Liwang A (2012) Rhythmic ring-ring stacking drives the circadian oscillator clockwise. *Proc Natl Acad Sci U S A*.

34. Chang Y-G, Tseng R, Kuo N-W, & LiWang A (2012) Rhythmic ring–ring stacking drives the circadian oscillator clockwise. *P Natl Acad Sci USA* 109(42):16847-16851.

35. McLaughlin Jr RN, Poelwijk FJ, Raman A, Gosal WS, & Ranganathan R (2012) The spatial architecture of protein function and adaptation. *Nature* 491(7422):138-142.

36. Ishiura M, *et al.* (1998) Expression of a Gene Cluster kaiABC as a Circadian Feedback Process in Cyanobacteria. *Science* 281(5382):1519-1523.

37. Querfurth C, *et al.* (Circadian Conformational Change of the Neurospora Clock Protein FREQUENCY Triggered by Clustered Hyperphosphorylation of a Basic Domain. *Molecular cell* 43(5):713-722.

38. Scheer FA, Hilton MF, Mantzoros CS, & Shea SA (2009) Adverse metabolic and cardiovascular consequences of circadian misalignment. *Proc Natl Acad Sci U S A* 106(11):4453-4458.

39. Rust MJ, Golden SS, & O'Shea EK (2011) Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator. *Science* 331(6014):220-223.

40. Nakajima M, Ito H, & Kondo T (2010) In vitro regulation of circadian phosphorylation rhythm of cyanobacterial clock protein KaiC by KaiA and KaiB. *FEBS letters* 584(5):898-902.

41. Hosokawa N, Kushige H, & Iwasaki H (2013) Attenuation of the posttranslational oscillator via transcription-translation feedback enhances circadian-phase shifts in Synechococcus. *Proc Natl Acad Sci U S A* 110(35):14486-14491.

42. Kitayama Y, Nishiwaki T, Terauchi K, & Kondo T (2008) Dual KaiC-based oscillations constitute the circadian system of cyanobacteria. *Genes & development* 22(11):1513-1521.

43. Ye S, Vakonakis I, Ioerger TR, LiWang AC, & Sacchettini JC (2004) Crystal structure of circadian clock protein KaiA from Synechococcus elongatus. *The Journal of biological chemistry* 279(19):20511-20518.

44. Hitomi K, Oyama T, Han S, Arvai AS, & Getzoff ED (2005) Tetrameric architecture of the circadian clock protein KaiB. A novel interface for intermolecular interactions and its impact on the circadian rhythm. *J Biol Chem* 280(19):19127-19135.

45. Pattanayek R, Sidiqi SK, & Egli M (2012) Crystal structure of the redox-active cofactor dibromothymoquinone bound to circadian clock protein KaiA and structural basis for dibromothymoquinone's ability to prevent stimulation of KaiC phosphorylation by KaiA. *Biochemistry* 51(41):8050-8052.

46. Nishiwaki T & Kondo T (2012) Circadian autodephosphorylation of cyanobacterial clock protein KaiC occurs via formation of ATP as intermediate. *The Journal of biological chemistry* 287(22):18030-18035.

47. Vakonakis I & LiWang AC (2004) 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(30):10925-10930.

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

49. van Zon JS, Lubensky DK, Altena PRH, & ten Wolde PR (2007) An allosteric model of circadian KaiC phosphorylation. *P Natl Acad Sci USA* 104(18):7420-7425.

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

51. Kitayama Y, Nishiwaki-Ohkawa T, Sugisawa Y, & Kondo T (2013) KaiC intersubunit communication facilitates robustness of circadian rhythms in cyanobacteria. *Nat Commun* 5(2897).

52. Monod J, Wyman J, & Changeux JP (1965) On Nature of Allosteric Transitions - a Plausible Model. *J Mol Biol* 12(1):88-&.

53. Egli M, *et al.* (2013) Loop-Loop Interactions Regulate KaiA-Stimulated KaiC Phosphorylation in the Cyanobacterial KaiABC Circadian Clock. *Biochemistry* 52(7):1208-1220.

54. Snijder J, *et al.* (2014) Insight into cyanobacterial circadian timing from structural details of the KaiB-KaiC interaction. *Proc Natl Acad Sci U S A* 111(4):1379-1384.

55. Ma L & Ranganathan R (2012) Quantifying the rhythm of KaiB-C interaction for in vitro cyanobacterial circadian clock. *PLoS One* 7(8):e42581.

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

57. Villarreal SA, *et al.* (2013) 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(18):3311-3324.

58. Villarreal SA, *et al.* (2013) 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(18):3311-3324.

59. Buchler NE & Louis M (2008) Molecular Titration and Ultrasensitivity in Regulatory Networks. *J Mol Biol* 384(5):1106-1119.

60. Brown KL & Hughes KT (1995) The Role of Anti-Sigma Factors in Gene-Regulation. *Mol Microbiol* 16(3):397-404.

61. Mukherji S, *et al.* (2011) MicroRNAs can generate thresholds in target gene expression. *Nat Genet* 43(9):854-U860.

62. Kim JK & Forger DB (2012) A mechanism for robust circadian timekeeping via stoichiometric balance. *Mol Syst Biol* 8.

63. Tsai TYC, *et al.* (2008) Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science* 321(5885):126-129.

64. Gribun A, *et al.* (2005) 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(16):16185-16196.

65. Tang CT, *et al.* (2009) Setting the pace of the Neurospora circadian clock by multiple independent FRQ phosphorylation events. *P Natl Acad Sci USA* 106(26):10722-10727.

66. Spitz F & Furlong EEM (2012) Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* 13(9):613-626.

67. Chang Y-G*, et al.* (2015) A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science* 349(6245):324-328.

68. Chabot JR, Pedraza JM, Luitel P, & van Oudenaarden A (2007) Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock. *Nature* 450(7173):1249-1252.

69. Firnberg E & Ostermeier M (2012) PFunkel: Efficient, Expansive, User-Defined Mutagenesis. *PLoS ONE* 7(12):e52031.

70. Clerico EM, Ditty JL, & Golden SS (2007) Specialized Techniques for Site-Directed Mutagenesis in Cyanobacteria. *Circadian Rhythms: Methods and Protocols*, ed Rosato E (Humana Press, Totowa, NJ), pp 155-171.

71. McEwen JT, Machado IMP, Connor MR, & Atsumi S (2013) Engineering Synechococcus elongatus PCC 7942 for Continuous Growth under Diurnal Conditions. *Applied and Environmental Microbiology* 79(5):1668-1675.

72. Vijayan V & O'Shea EK (2013) Sequence Determinants of Circadian Gene Expression Phase in Cyanobacteria. *Journal of bacteriology* 195(4):665-671.

73. Woelfle MA, Ouyang Y, Phanvijhitsiri K, & Johnson CH (The Adaptive Value of Circadian Clocks. *Current Biology* 14(16):1481-1486.

74. Glaser F*, et al.* (2003) ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics* 19(1):163-164.

75. Xu Y*, et al.* (2013) Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. *Nature* advance online publication.