THE UNIVERSITY OF CHICAGO

THE BIOPHYSICAL BASIS OF NUCLEOTIDE SENSITIVITY IN THE CYANOBACTERIAL CIRCADIAN CLOCK

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
AND
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOPHYSICAL SCIENCES

BY
LU HONG

CHICAGO, ILLINOIS
JUNE 2020

Everything to which we concede existence is a posit from the standpoint of a description of the theory-building process, and simultaneously real from the standpoint of the theory that is being built. Nor let us look down on the standpoint of the theory as make-believe; for we can never do better than occupy the standpoint of some theory or other, the best we can muster at the time. (Quine, *Word and Object*)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my advisors Aaron Dinner and Mike Rust for tolerating my irreverence, my labmates for listening to me, Adam Hammond and Pete Dahlberg for their help through the first year of Ph.D., the Biophysical Sciences program for giving me an opportunity to embark on this intellectual journey, and professors Scott Feller and Chad Westphal for introducing me to the world of computation and modeling.

ABSTRACT

Oscillatory dynamics is ubiquitous in biology, yet the complexity of the underlying signaling pathways often makes quantitative understanding of such systems difficult. An ideal model system for understanding biochemical oscillators is the Kai system, a post-translational circadian oscillator found in the cyanobacterium *Synechococcus elongatus*. The Kai oscillator consists of three soluble proteins named KaiA, KaiB, and KaiC, and their periodic interactions, which are driven by ATP hydrolysis in KaiC, generate a near-24-h period oscillation of KaiC phosphorylation states. The Kai oscillator interacts with cellular metabolic conditions, in the form of ATP/ADP concentrations, that can act as external timing cues, such that the period remains robust while the amplitude and phase of the clock adjust. How the clock achieves such robust yet tunable dynamics is not fully understood.

Understanding the interaction between KaiA and KaiC is key to understanding how the clock senses metabolic conditions. KaiA is a nucleotide-exchange factor that regulates the nucleotide-bound state of KaiC and thus the direction of its reversible phosphotransferase activity. In this work, I dissect the KaiA–KaiC interaction at two levels. First, I fit a detailed kinetic model of KaiA–KaiC interactions to experimental kinetic data using Bayesian Markov chain Monte Carlo. This model revealed ultra-sensitivity in KaiC phosphorylation as a result of nucleotide-dependent KaiA binding affinity, which I argue is important for period stability. Second, I use molecular dynamics simulations to probe the molecular mechanism of ADP release from KaiC, which reveals coupling between the nucleotide-binding pockets and the A-loop, an unstructured region in KaiC that binds to KaiA. This provides a hypothesis for how KaiA acts as a nucleotide-exchange factor.

CHAPTER 1

INTRODUCTION

1.1 Overview of the Kai Oscillator

Oscillatory and pulsatile dynamics are ubiquitous in biology, such as cell division control, circadian rhythms in gene expression, and the propagation of action potentials. Although the biological elements underpinning these phenomena are different, it has long been recognized that a few organizing motifs, such as positive and negative feedback loops, are sufficient to characterize the qualitative behavior of such systems [1]. However, the complexity of these signaling networks often makes it challenging to precisely map the organizing motifs onto the underlying biochemical elements. Morever, quantitative differences in spatiotemporal scales and robustness that cannot be easily explained by the network topology suggest that a detailed analysis of their dynamics is necessary to better understand and manipulate such systems.

The Kai system, a simple post-translational oscillator consisting of three soluble proteins named KaiA, KaiB, and KaiC [2], is an ideal model for quantitative analysis of biochemical oscillators. The periodic interactions among the three proteins, driven by ATP hydrolysis, generate a robust, near-24-h period oscillation of the phosphorylation states of KaiC. In vivo, the Kai system is embedded in a transcription-translation feedback loop that controls the circadian rhythms of *Synechococcus Elongatus* PCC 7942 [3–5], a strain of autotrophic freshwater cyanobacteria, where the global, daily promoter activities [6] are controlled by the Kai system via a two-component signaling pathway [7, 8] consisting of two histidine kinases SasA and CikA, which relay the phosphorylation state of KaiC to a response regulator, RpaA.

The fact that species of cyanobacteria exhibit circadian rhythms in such activities as nitrogen fixation, amino acid update, and cell division was known in the 1980s [9–11], but it was not until 1998 when the *kai* genes were identified in *S. elongatus*. In 2005, two

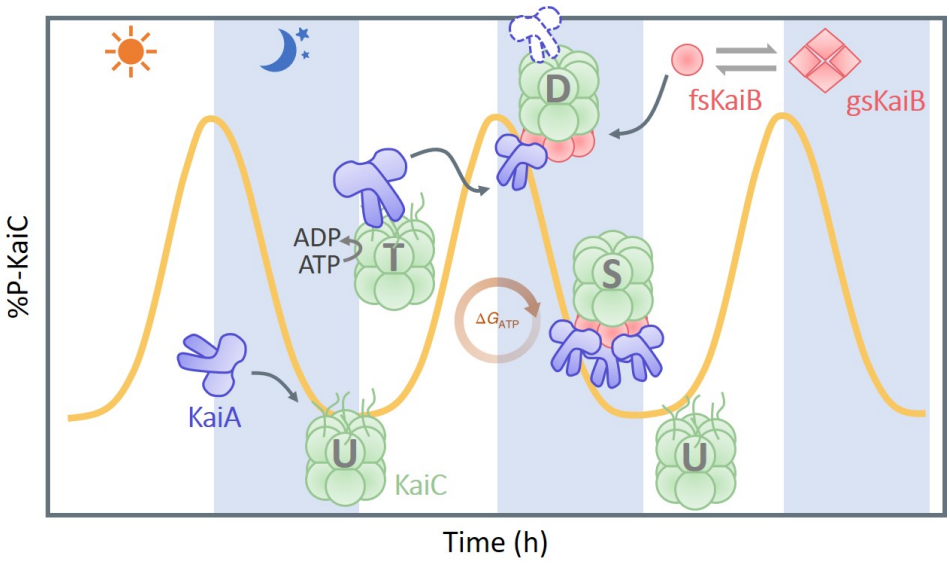


Figure 1.1: Overview of the Kai oscillator. The periodic interactions among KaiA, KaiB, and KaiC, driven by the free energy of ATP hydrolysis, generates a robust, near-24-h period oscillation of the KaiC phosphorylation states. Specifically, the binding of KaiA to KaiC during the day leads to KaiC autophosphorylation via its action as a nucleotide-exchange factor, while the binding of foldswitched KaiB during the night sequesters KaiA, which leads to KaiC autodephosphorylation. Here, U stands for the unphosphorylated state, T stands for the phosphothreonine-432 state, S stands for the phosphoserine-431 state, and D stands for the doubly phosphorylated state.

landmark papers demonstrated that the rhythmic in vivo cycling of KaiC phosphorylation states can act independently of transcriptional feedback [12], and, even more surprisingly, that the KaiC phosphorylation cycle can be reconstituted in vitro using the three purified Kai proteins [13]. These studies prompted significant interest in the fundamental structural, biochemical, and biophysical basis of the Kai oscillator, as well as theoretical considerations into the source of its robust and tunable dynamics, which we briefly outline below. Currently, *S. elongatus* is the prevailing model system in cyanobacterial circaidna biology due to the relative ease of genetic manipulation. Proteins homologous to the clock components in *S. elongatus* has been identified in other cyanobacterial species [14], but whether such orthologous systems can generate self-sustained oscillation in vitro remains unclear.

Details of the clock biochemistry will be discussed in the following chapters when

appropriate. However, it is instructive to take the conceptual viewpoint here and consider the basic requirements for biochemical oscillation and how such requirements are mapped onto the elementary processes of the clock. As summarized by Novák and Tyson [1], biochemical oscillators require a) delayed negative feedback, b) nonlinearity, and c) balance of timescale. In the Kai system, time is encoded in the phosphorylation states of KaiC [12, 15], a hexameric protein [16] where each subunit has two phosphorylation sites at S431 and T432 [17, 18]. At day, the autophosphorylation of KaiC is promoted by KaiA [15], a nucleotide-exchange factor for KaiC [19]. This is then followed at the dusk with the binding of KaiB to KaiC, which sequesters KaiA from its site of action and leads to autodephosphorylation [17, 20]. The opposing effects of KaiA and KaiB on KaiC phosphorylation are the basis of the negative feedback loop in the system (Fig. 1.1).

The (de)phosphorylation of KaiC proceeds in an ordered manner, in that T432 is preferentially (de)phosphorylated before S431 [17, 18]. This "kinetic ordering" has two important consequences. First, it breaks the symmetry between the day and the night phases of the clock; e.g., T432 is preferentially phosphorylated at midday, while S431 is preferentially phosphorylated at midnight, even though the percentage KaiC phosphorylation level is the same at the two time points. Second, the ordered phosphorylation of KaiC provides a mechanism to delay the onset of the the negative feedback. Specifically, the phosphorylation of T432 and S431 appears to have different effects on KaiC conformation, such that only the phosphorylation of the second site, S431, favors KaiB binding [21].

Kinetic ordering, however, is not the only source of delay in the system. KaiB is a metamorphic protein, which means that it can stably fold into two conformations [22]. The conversion from the "ground state" (gs) to the "foldswitched state" (fs) appears to be both kinetically and thermodynamically unfavorable, thus adding an additional layer of delay to the onset of KaiA sequestration. The fsKaiB conformation, however, appears to be stabilized by KaiC binding, such that fsKaiB monomers bind to KaiC subunits in a 1:1 stoichiometry in an all-or-none (i.e., cooperative) manner [23]. The cooperative KaiB–

KaiC interaction is believed to be an important source of nonlinearity in the system [21].

Thirdly, biochemical oscillation also requires a balance of timescales. This is a major unresolved issue concerning the kinetics of the Kai oscillator. Here, balance of timescale does not necessarily imply that all elementary processes in the system must have similar rate constants. Indeed, the Kai oscillator appears to operate with a separation of timescale at physiological conditions, where KaiA (un)binds to KaiC on the timescale of seconds [24, 25], but on the other hand KaiC catalytic activities [26] and KaiB foldswitch/binding operate on the timescale of hours [22]. However, a complete description of the kinetics of each step of the circadian cycle is currently unavailable. It is unclear how these identified slow steps in the system contribute to the 24-h period (e.g., can one write down an analytical expression for the period in terms of elementary rate constants?), nor is it clear what the ranges of kinetic parameters that still give rise to robust oscillations are.

The analysis by Novák and Tyson is primarily rooted in a dynamical systems perspective. From a thermodynamic perspective, however, one should also consider the issue of phase coherence. Under thermal equilibrium, there is no net unidirectional flux in the cycle of KaiC phosphorylation states, and the various phosphorylation states, nucleotide-bound states (of KaiC), and protein complexes eventually reach a steady state where the state populations are dictated by their equilibrium free energy differences. As a result, although on a microscopic level, individual protein complexes can spontaneously pass through the circadian cycle, on a bulk level there is no coherent oscillation. To overcome this problem, the reactions described so far are coupled to the ATPase activity of KaiC, which provides the nonequilibrium free energy to drive the system irreversibly forward in time [27]. The exact relation between free energy dissipation and robustness of the oscillation is an area of active research. A complementary argument is that the slowness of the KaiB foldswitch process also plays a role in this regard by synchronizing the KaiC population in the hyperphosphorylated state before switching to the dephosphorylation phase.

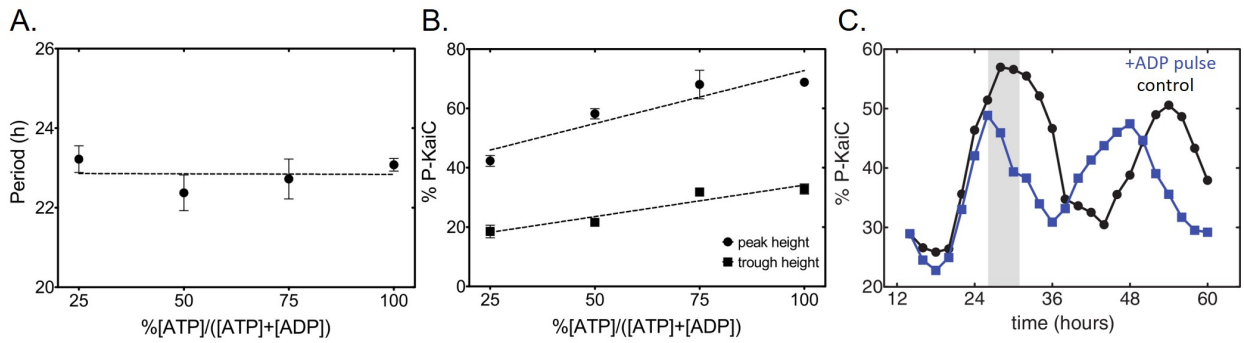


Figure 1.2: The Kai oscillator is robust yet tunable. A) the period of the Kai oscillator is insensitive to changes in a wide range of % ATP conditions. B) both the peak and trough of the KaiC phosphorylation cycle scale with %ATP, such that the size of the cycle becomes larger at higher %ATP. C) the phase response of the Kai oscillator to an ADP pulse; the shaded region represents the addition of 5 mM ADP to the reaction mixture with 5 mM ATP. Panels A) and B) are adapted from [28], and panel C) is adapted from [29].

Lastly, the above account of the Kai oscillator focuses on clock function as it is decoupled from the environment; as such, it is also instructive to consider the clock in the context of environmental conditions. First, as implied in the preceding discussion, the free-running clock under constant conditions has a self-sustained, near-24-h period period [13], which can persist for weeks in vitro [30]. Moreover, the period of the clock appears to be insensitive to environmental inputs such as temperature [31] and metabolic conditions [28] (Fig. 1.2A). For in vitro studies, the cellular metabolic conditions are usually measured and represented in terms of %ATP (i.e., [ATP] as a percentage of total [ATP] and [ADP].) However, this insensitivity does not imply that the clock is decoupled from the environment. Rather, the amplitude and phase of the phosphorylation cycle do respond to such changes (Fig. 1.2B and C), and the clock can be entrained by regular cycling of these signals [32, 33]. Taken together, the robustness of the period allows for accurate timekeeping, while the responsiveness of the amplitude and phase allows the clock to adapt to external timing cues (i.e., the zeitgebers). How to reconcile this remarkable blend of robustness and tunability is another active area of research (see, for example, [28, 31, 34–36]), and a primary driving force behind this thesis.

1.2 Overview of the Thesis

What the above analysis has made clear is that the simplicity of the Kai oscillator allows one to articulate the basic biochemical events in the system and how they contribute to the oscillatory dynamics. However, what is less obvious is a "second-order" description of the clock that accounts for how elementary processes in the clock interact with environmental perturbations to generate robust yet tunable circadian rhythms. Broadly put, therefore, the fundamental goal of this thesis is to investigate how the circadian clock in *S. elongatus* responds to changes in cellular metabolic conditions. Analysis of this question may in turn reveal new insights into the basic biochemistry of the system and the principles of biochemical oscillation.

How the Kai oscillator responds to metabolism is a property of the circuit, and therefore a natural first step is to investigate the kinetics of the reaction network (Chapter 2). Specifically, the observations that the function of KaiB is to sequester KaiA at nighttime and that KaiA promotes phosphorylation via its function as a nucleotide-exchange factor suggest that a detailed understanding of the KaiA–KaiC subsytem is necessary to understand how the direction of phosphotransfer is regulated over the day-night cycle and by changes in metabolic conditions. To this end, we measured the kinetics of KaiC phosphorylation over a wide range of [KaiA] and %ATP conditions, and the data was fitted with a detailed kinetic model using Bayesian parameter estimation. The resulting model revealed unexpected ultrasensitivity in KaiC phosphorylation, which we argued has important implications for the stability of the oscillator period against changes in %ATP condition.

The analysis of the kinetic model is consistent with an emerging understanding that nucleotide exchange in KaiC is a critical process that allows the system to sense the metabolic condition, and how this process is regulated determines the direction of KaiC phosphotransfer reactions. This observation motivated us to examine the molecular mechanism of nucleotide exchange and how it may be accelerated by KaiA (Chapter 3).

We thus performed molecular dynamics simulations to characterize the energetic and structural details of the ADP release pathway in KaiC. These simulations suggest that ADP release requires concerted global motions that destabilize the KaiC subunit interface, and such motions are coupled to conformational changes at the known KaiA binding site, thus providing a molecular hypothesis for how KaiA functions as a nucleotide-exchange factor.

Lastly, we turn our attention to KaiB and the dephosphorylation phase of the clock. It has been argued that the timescale of KaiB binding and the resulting negative feedback is controlled by the timescale of ATP hydrolysis and nucleotide exchange in KaiC, even though KaiB foldswitch itself is comparably slow. To assess how KaiB contributes to the slowness of the clock, it is necessary to determine the kinetics of KaiB foldswitch and how it may be regulated by binding to KaiC. We are currently pursuing molecular dynamics simulations of the KaiB foldswitch pathways and applying recent advance in the theory of chemical kinetics to address these questions (Chaptuer 4). This is an ongoing study, and further details will be made available in future publications.

CHAPTER 2

BAYESIAN MODELING REVEALS METABOLITE-DEPENDENT ULTRASENSITIVITY IN THE CYANOBACTERIAL CIRCADIAN CLOCK

This chapter was published as an article in Molecular Systems Biology (doi:10.15252/msb.20199355) and appears below with minor formatting modifications.

Lu Hong[a], Danylo O. Lavrentovich[b], Archana Chavan[c], Eugene Leypunskiy[a], Eileen Li[d], Charles Matthews[d], Andy LiWang[c,e,f,g,h,i], Michael J. Rust[j,k,l] & Aaron R. Dinner[b,k,m]

[a]Graduate Program in Biophysical Sciences, University of Chicago

[b]Department of Chemistry, University of Chicago

[c]School of Natural Sciences, University of California, Merced

[d]Department of Statistics, University of Chicago

[e]Quantitative and Systems Biology, University of California, Merced

[f]Center for Circadian Biology, University of California, San Diego

[g]Chemistry and Chemical Biology, University of California, Merced

[h]Health Sciences Research Institute, University of California, Merced

[i]Center for Cellular and Biomolecular Machines, University of California, Merced

[j]Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago

[k]Institute for Biophysical Dynamics, University of Chicago, Chicago

[l]Institute for Genomics and Systems Biology, University of Chicago

[m]James Franck Institute, University of Chicago

2.1 Abstract

Mathematical models can enable a predictive understanding of mechanism in cell biology by quantitatively describing complex networks of interactions, but such models are

often poorly constrained by available data. Owing to its relative biochemical simplicity, the core circadian oscillator in *Synechococcus elongatus* has become a prototypical system for studying how collective dynamics emerge from molecular interactions. The oscillator consists of only three proteins, KaiA, KaiB, and KaiC, and near-24-h cycles of KaiC phosphorylation can be reconstituted in vitro. Here, we formulate a molecularly-detailed but mechanistically naive model of the KaiA-KaiC subsystem and fit it directly to experimental data within a Bayesian parameter estimation framework. Analysis of the fits consistently reveals an ultrasensitive response for KaiC phosphorylation as a function of KaiA concentration, which we confirm experimentally. This ultrasensitivity primarily results from the differential affinity of KaiA for competing nucleotide-bound states of KaiC. We argue that the ultrasensitive stimulus-response relation likely plays an important role in metabolic compensation by suppressing premature phosphorylation at nighttime.

2.2 Introduction

Achieving a predictive understanding of biological systems and chemical reaction networks is challenging because complex behavior can emerge from even a small number of interacting components. Classic examples include the propagation of action potentials in neurobiology and chemical oscillators such as the Belousov–Zhabotinsky reaction. The collective dynamics in such systems cannot be easily intuited through qualitative reasoning alone, and thus mathematical modeling plays an important role in summarizing and interpreting existing observations and formulating testable, quantitative hypotheses. However, it can be difficult to rationalize how collective dynamics emerge from specific molecular features. One approach to addressing this issue is to compare mathematical representations of competing molecular mechanisms based on their abilities to fit experimental data.

The circadian clock from the cyanobacterium *Synechococcus elongatus* [37] represents

a unique opportunity to use model fitting to learn biochemical mechanisms. This is because the core oscillator can be reconstituted in a test tube from a small number of components. The simplicity of the system makes it possible to both cleanly model the basic biochemical events in the circadian cycle and to collect quantitative data under well-controlled conditions. The core oscillator of *S. elongatus* consists of three proteins, KaiA, KaiB, and KaiC, which self-organize to generate a near-24-h rhythm in KaiC phosphorylation. The basic biochemical events are well-established [2]. KaiC is an ATPase [26] that phosphorylates and dephosphorylates itself by transfer of phosphoryl groups from and to bound nucleotides [38, 39]; KaiA-dependent nucleotide exchange reactions drive the phosphorylation phase of the cycle [19], and KaiB-mediated sequestration of KaiA leads to dephosphorylation. The reconstituted oscillator retains many of the hallmarks of circadian rhythms in living organisms [13, 29, 32, 33].

Yet, questions remain about how the clock couples to environmental conditions while maintaining a robust ∼24 h rhythm [28, 31–33]. The Kai oscillator senses changes in the relative concentrations of ATP to ADP in solution, which allows entrainment to metabolic rhythms [28, 29, 33]. KaiA modulates these dynamics via its function as a nucleotide-exchange factor, but how the system adapts and responds to changes in metabolic conditions is not clear. Models that account for the possible protein complexes, including the interplay of nucleotide-bound and phosphorylation states, can be dauntingly complex [e.g., 21, 40]. How to fit them to data and interpret the results is an area of active research.

Here we use a data-driven Bayesian approach to estimate the parameters of a molecularly-detailed kinetic model of the KaiA-KaiC subsystem, with the goal of learning the features required to capture the behavior of the system during the phosphorylation phase of the clock cycle quantitatively (Fig. 2.1A). The model describes the coupling between KaiA, nucleotides (ATP and ADP) in solution, KaiC phosphorylation, and KaiC nucleotide-bound states. To provide training data for the model, we collected kinetic time

series measuring KaiC phosphorylation kinetics over a wide range of KaiA concentrations ([KaiA]) and %ATP (defined as 100%[ATP]/([ATP] + [ADP])). Although such data do not give us direct access to all relevant states of the KaiA-KaiC subsystem, they place constraints on the underlying molecular interactions. Bayesian statistics [41, 42] have found diverse applications in systems biology [43–50], including circadian biology [51–53]. Here, they provide a unified framework for estimating parameter values, quantifying the importance of specific model elements, and making mechanistic predictions from the model.

By Markov chain Monte Carlo (MCMC) sampling, we obtain an ensemble of parameter sets that fit the data. Even with extensive training data, many microscopic parameters in the model are not tightly constrained. Despite this, we show that this ensemble of fits robustly makes predictions that are borne out in experimental tests [54, 55]. In particular, the model reveals a previously unappreciated ultrasensitive dependence of phosphorylation on the concentration of KaiA, with strong nonlinearity at low [KaiA], conditions that likely apply near the nighttime to daytime transition point, when a large fraction of KaiA molecules are inhibited. Importantly, we find that the threshold KaiA concentration varies with the %ATP in solution. This ultrasensitive response primarily arises from a differential affinity of KaiA for different nucleotide-bound states of KaiC. This mechanism is analogous to substrate competition [56], where kinetic competition of multiple enzyme substrates leads to ultrasensitivity.

Lastly, we consider the implications of these results for the full oscillator, in which KaiC rhythmically switches between phosphorylation and dephosphorylation. A well-known mechanism for the inhibition of KaiA is its sequestration into KaiBC complexes [20] that form when KaiC is sufficiently phosphorylated [17]; this mechanism serves as a delayed negative feedback loop in the oscillator. The KaiB-independent thresholding phenomenon we describe here is strongest when KaiC phosphorylation is low and when KaiC is ADP-bound. This suggests there are at least two mechanisms that work together

during the cycle to prevent KaiA from acting at inappropriate times, and that the relative strength of the two mechanisms varies with the nucleotide pool.

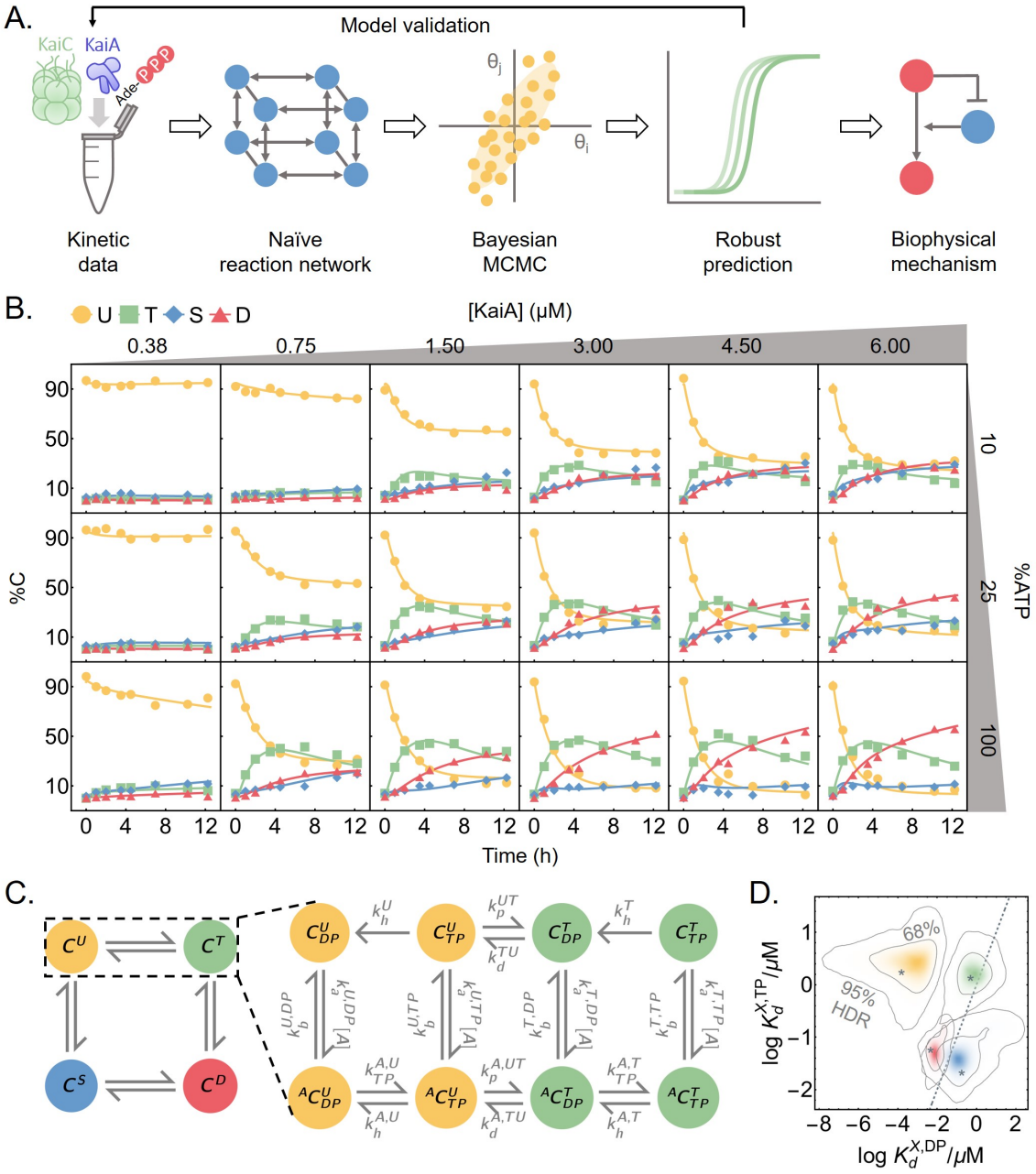


Figure 2.1: Phosphorylation data are fit by a mechanistically naive kinetic model. A. An outline of the data-driven Bayesian model fitting approach employed in this work. B. To constrain the model, measurements of KaiC phosphorylation kinetics were collected at six [KaiA] and three %ATP conditions. The curves represent the best fit model prediction. C. A schematic of the mass-action kinetics model. The model elaborates on the auto-phosphorylation reactions of KaiC by explicitly keeping track of the time evolution of the KaiC phosphoforms and nucleotide-bound states; conversions among these states are

(Figure 2.1, continued) mediated by phosphotransfer, nucleotide exchange, ATP hydrolysis, and KaiA (un)binding. Note that the KaiA binding reactions are second-order, but KaiA concentration ([A]) is written as part of the effective first-order rate constant. See the main text for a discussion of the state and rate constant nomenclature and Fig. S2.5A for a schematic of the full model. D. The posterior distributions for log KaiA dissociation constants (base 10). The horizontal axis represents the affinity for ADP-bound KaiC, and the vertical axis represents the affinity for ATP-bound KaiC; X $\in$ {U, T, S, D} and corresponds to the four colors of the KaiC phosphoforms, as in panel C. The asterisks represent the best fit, and the contour lines represent the 95% and 68% highest posterior density regions (HDR). The dashed line represents the $K_d^{\mathrm{X,TP}} = K_d^{\mathrm{X,DP}}$ line, so that densities above the line indicate higher affinity for the ADP-bound states and densities below the line indicate higher affinity for the ATP-bound states.

Incorporation of the ultrasensitive response to KaiA into a mathematical model of the full oscillator suggests that this effect both stabilizes the period against changes in the nucleotide pool and allows oscillations to persist even when KaiB binds KaiA relatively weakly. Consistent with this prediction, we find that a substantial amount of KaiA is not bound by KaiB even when the clock is dephosphorylating. These results shed new light on metabolic compensation, a property that allows robust 24-h oscillation in spite of changes in %ATP conditions [57]. Taken together, our results show how the Bayesian framework combined with extensive training data can be used to discover unanticipated mechanisms and direct experimental investigations.

2.3 Results

2.3.1 A molecular model of KaiA-KaiC dynamics

To probe the response of KaiC phosphorylation to a wide range of metabolic conditions, we made kinetic measurements of KaiC phosphorylation at three %ATP conditions and six [KaiA] conditions while holding the KaiC concentration constant (Fig. 2.1B). KaiC is a homohexamer and each subunit has two duplicated ATPase domains, termed CI and CII [16, 26, 58]. The CII domain can autophosphorylate via a bidirectional phosphotransferase mechanism [38, 39] with two phosphorylation sites. Each KaiC subunit thus has four

phosphoforms: the unphosphorylated (U), phosphoserine-431 (S), phosphothreonine-432 (T), and doubly phosphorylated (D) states, each of which peaks at a distinct time in the oscillation.[17, 18, 59].

Our strategy is to fit these data with a model of the KaiC catalytic cycle in CII with a minimum of simplifying assumptions. To this end, we formulate a model based on mass-action kinetics. We explicitly keep track of three properties of the CII domain of each KaiC subunit: its phosphorylation status (right superscripts in Fig. 2.1C), nucleotide-bound state (right subscript), and whether or not KaiA is bound (left superscript). We do not consider CI or the hexameric nature of KaiC explicitly (see Appendix for further discussion). There are thus 16 possible KaiC states, 8 of which are shown in Fig. 2.1C, along with the phosphotransfer, nucleotide exchange, KaiA (un)binding, and hydrolysis reactions that connect the states (see Fig. S2.5A for the full model structure). We also considered the possibility that nucleotides might interact directly with KaiA, which could allow KaiA's activity to directly depend on nucleotides in solution. However, we did not detect any direct interaction between KaiA and ATP or ADP using NMR spectroscopy (Fig. S2.6), so we do not allow for this scenario in the model.

In the rest of this section, we further elaborate our model by stepping through the four classes of reactions that we include; additional details can be found in Materials and Methods. In the next section, we describe in qualitative terms the fitting procedure and analyze the extent to which the model was constrained by the training dataset. For readers who are primarily interested in the biochemical conclusions of the model, these results can be skipped without loss of continuity.

Phosphotransfer The KaiC CII domain has reversible phosphotransferase activity [38, 39]; it can transfer a γ-phosphate group from a bound ATP to a phosphorylation site, but

unlike a typical phosphatase, it regenerates ATP from ADP during dephosphorylation, i.e.,

$$\mathrm{C}^{\mathrm{X}}_{\mathrm{TP}} \underset{k_{\mathrm{d}}}{\overset{k_{\mathrm{p}}}{\rightleftharpoons}} \mathrm{C}^{\mathrm{Y}}_{\mathrm{DP}} \tag{2.1}$$

where (X, Y) $\in$ {(U, T), (U, S), (T,D), (S, D)}. This mechanism implies that the nucleotide-bound state of KaiC has a significant impact on the net direction of its phosphotransferase activity: an ATP-bound KaiC presumably cannot dephosphorylate, and an ADP-bound KaiC cannot phosphorylate.

Nucleotide exchange KaiA binding to the CII domain [60, 61] stimulates KaiC autophosphorylation [15, 24, 62]. Recent work has shown that KaiA can bind to KaiC and act as a nucleotide-exchange factor [19] by facilitating conformational changes at the subunit interface that promote solvent exposure of the nucleotide-binding pocket [63]. It is currently unclear whether this nucleotide exchange activity is responsible for all of KaiA's effect on KaiC or whether it alters the KaiC catalytic cycle in other ways (see Appendix for further analysis of this issue). The reversible binding of KaiA

$$\mathrm{C} \underset{k_{\mathrm{b}}}{\overset{k_{\mathrm{a}}[\mathrm{A}]}{\rightleftharpoons}} {}^{\mathrm{A}}\mathrm{C} \tag{2.2}$$

contributes two classes of rate constants, k_{a} and k_{b}.

Because the CII domain of KaiC releases its bound nucleotide very slowly in the absence of KaiA [19], we ignore the possibility of KaiA-independent nucleotide exchange in the model. Under the assumptions that i) the apo state is in a quasi-steady state, ii) the ADP and ATP on-rates are identical, and iii) ATP release is slow, nucleotide exchange can be modeled as a one-step reaction:

$$\mathrm{ATP} + {}^{\mathrm{A}}\mathrm{C}_{\mathrm{DP}} \xrightarrow{k^{\mathrm{A}}_{\mathrm{TP}}} {}^{\mathrm{A}}\mathrm{C}_{\mathrm{TP}} + \mathrm{ADP} \tag{2.3}$$

where

$$k_{\mathrm{TP}}^{\mathrm{A}} = k_{\mathrm{r}}^{\mathrm{DP}} \frac{[\mathrm{ATP}]}{[\mathrm{ATP}] + [\mathrm{ADP}]} \tag{2.4}$$

and $k_{\mathrm{r}}^{\mathrm{DP}}$ is the ADP dissociation rate constant. Nucleotide exchange thus contributes one class of rate constant, $k_{\mathrm{r}}^{\mathrm{DP}}$. See Materials and Methods for the derivation of (2.4).

ATP hydrolysis Finally, we allow for irreversible ATP hydrolysis in the CII domain

$$\mathrm{C}_{\mathrm{TP}} \xrightarrow{k_{\mathrm{h}}} \mathrm{C}_{\mathrm{DP}} + \mathrm{P}_i \tag{2.5}$$

which contributes one class of rate constants, k_{h}. This assumption is important because this pathway is required in the model for sustained dephosphorylation by allowing ADP-bound KaiC to accumulate in the absence of KaiA. Because each KaiC molecule consumes relatively little ATP on the timescale of the simulations [26], we assume the solution ATP and ADP concentrations are constant.

State-dependent rates Given the six classes of rate constants, k_{p}, k_{d}, k_{a}, k_{b}, $k_{\mathrm{r}}^{\mathrm{DP}}$, and k_{h}, we make the model maximally general, or mechanistically naive, by allowing each rate constant to potentially depend on the specific molecular state involved in the reaction. For example, the KaiA dissociation rate constant is allowed to vary depending on the nucleotide-bound state and phosphoform background of KaiC, and thus the dissociation rate constants for the ADP-bound U ($k_{\mathrm{b}}^{\mathrm{U,DP}}$) and ATP-bound T phosphoforms ($k_{\mathrm{b}}^{\mathrm{T,TP}}$) are two independent model parameters. In this way, the parameter fitting and model comparison procedures automatically test specific biochemical hypotheses about the functions of KaiA and KaiC. For example, allowing the KaiA off-rates to depend on the nucleotide-bound states is equivalent to the hypothesis that KaiA has different dwell times for ATP- versus ADP-bound states of KaiC. In fact, because each reaction has an independent rate constant, except for thermodynamic constraints of detailed balance, the fitting procedure effectively allows for simultaneous testing of all possible two-way

interactions of the three categories of KaiC properties, without a priori preference for any particular mechanism.

2.3.2 The data constrain the parameters to widely varying degrees

We estimate the model parameters through a Bayesian framework. In this framework, we maximize the posterior probability, which is proportional to the product of the prior distribution and the likelihood function. Here, we interpret the prior as representing subjective beliefs on the model parameters before experimental inputs, while the likelihood function quantifies the goodness of fit. Bayesian parameter estimation reduces to least-squares fitting under the assumption of normally distributed residuals and uniform priors. In practice, we find that direct numerical optimization of the posterior usually results in fits that are trapped in low probability local maxima (Fig. S2.7B). Thus we instead draw parameters from the prior distribution and then use a heuristic combination of Markov chain Monte Carlo (MCMC) sampling and optimization (Powell's algorithm) to explore the parameter space. The MCMC method that we use [64, 65] efficiently searches the parameter space by simulating an ensemble of parameter sets in parallel; the spread of the ensemble reflects the geometry of the posterior distribution and is used to guide the directions of Monte Carlo moves. See Materials and Methods for a more mathematical treatment of the fitting procedure and comparison of different numerical optimization and sampling methods.

We use this approach to fit the phosphorylation data (Fig. 2.1B) together with previously published data on dephosphorylation [29], ATP hydrolysis rate [26], and the KaiA dwell time for each KaiC phosphoform [24, 25] (see Materials and Methods). Overall, the model achieves excellent agreement with the training data (Fig. 2.1B and Fig. S2.8A–C). In the following analyses, we refer to model predictions using the best fit parameter values, and quantify the uncertainties using the posterior distribution (see Appendix for further discussion of the convergence of the simulations).

We find that certain parameters, such as the hydrolysis rates in the U and T phosphoforms and the KaiA off-rates from the U phosphoform, are tightly constrained, while many others, mainly involving S and D phosphoforms, are less constrained, in the sense that their posterior distributions span multiple orders of magnitude, exhibit multimodality, or cannot be reproduced over multiple independent runs (Fig. S2.5B). Some parameters are highly correlated and certain combinations of the parameters are much better constrained than the individual parameters. For example, the posterior distributions for the KaiA binding affinities (Fig. 2.1D) appear better constrained than the on/off rates (Fig. S2.9B).

Taken together, these results are consistent with the notion that collective fits of multiparameter models are generally "sloppy," meaning that the sensitivities of different combinations of parameters can range over orders of magnitude with no obvious gaps in the spectrum [54, 55]. As we will see, we can nonetheless make useful predictions using the ensemble of model parameters, because the model behavior is constrained along the stiffest directions of the posterior distribution. By contrast, direct parameter measurements need to be both complete and precise to achieve similar predictive validity [55]. We further characterize the structure of the parameter space in Appendix and Fig. S2.9.

2.3.3 KaiC (de)phosphorylation goes through transient kinetic intermediates

In the model, we can breakdown the kinetics of KaiC phosphorylation reactions and interpret the underlying molecular events. Here we consider the phosphorylation kinetics at the standard reaction condition (3.5 µM KaiC, 1.5 µM KaiA, 100% ATP; Fig. 2.2A and B, solid curves); we examine the effect of varying [KaiA] and %ATP in the following sections.

At the beginning of the phosphorylation reaction, KaiC molecules are predominantly in the ADP-bound U state (C^{U}_{DP}), the end product of the dephosphorylation pathway

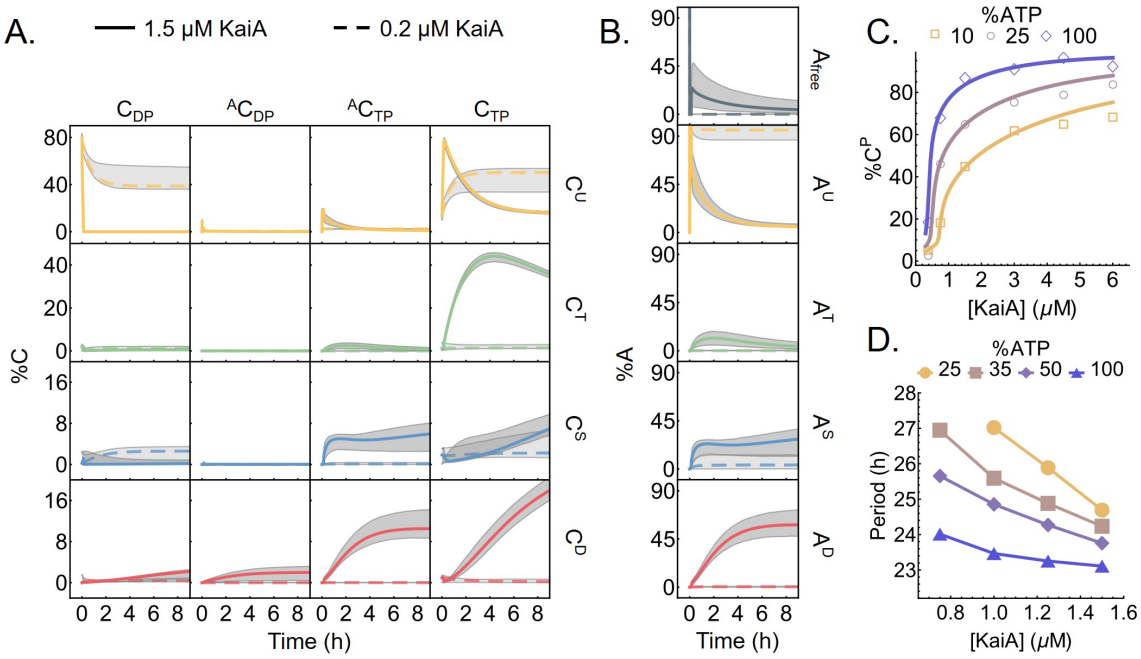


Figure 2.2: The model captures the kinetics of KaiC phosphorylation. A. The time evolution of all 16 KaiC states in a phosphorylation reaction with 100% ATP and either 1.5 μM (solid curves) or 0.2 μM (dashed curves) KaiA. The gray regions represent the 95% posterior interval. Refer to Fig. S2.5A for the KaiC state nomenclature. B. The corresponding KaiA kinetics, broken down according to the phosphoform of the bound KaiC. C. KaiA concentration can tune the sensitivity of the KaiC phosphorylation level to %ATP. The points represent the measured total percentage phosphorylation levels at $t = 12.25$ h (see Fig. 2.1B for the full kinetics), and the curves represent the model prediction at the same time point. D. KaiA concentration can tune the sensitivity of the clock period to %ATP. The period of the full KaiABC oscillator is calculated from fluorescence polarization measurements (see Fig. S2.10 A and B for further analysis).

in the absence of KaiA (Fig. 2.2A). With the addition of abundant KaiA, the C^U_{DP} state becomes rapidly depleted within the first 10 minutes of the reaction and enters the C^U_{TP} state. Consistent with the kinetic ordering observed in the full oscillator, the C^U_{TP} population is primarily converted into the T phosphoform over the S phosphoform. The mechanism underlying the preference for the T phosphoform is not well constrained by the data, but it appears to be the result of more than just a difference in the relative U→T and U→S phosphorylation rates; a sensitivity analysis shows that the ordering of phosphorylation is also dependent on KaiA (un)binding kinetics (see Appendix and Fig. S2.11). The ADP- and KaiA-bound T phosphoform states are unstable kinetic

intermediates, and the population accumulates at the C^{T}_{TP} bottleneck for the first 4 hours. As phosphorylation reaches completion, the T phosphoform is converted first into $^{A}C^{D}_{TP}$ through the unstable ADP-bound intermediates, and then to the C^{D}_{TP} state; the populations of the $^{A}C^{D}_{TP}$ and C^{D}_{TP} states are comparable at steady state. We note here, however, that previous measurements indicate that approximately 30% of CII nucleotide-binding pockets should be ADP-bound in the presence of KaiA at steady state [19], which suggests that the stability of the ADP-bound form is systematically underestimated by the model fit.

During the phosphorylation reaction, the amount of free KaiA is initially transiently depleted due to association with the ADP-bound U phosphoform (Fig. 2.2B). Afterwards, KaiA primarily associates with the ATP-bound S and D phosphoforms as they appear, but does not bind to the T phosphoform strongly. Therefore, even in the absence of KaiB, not all KaiA is free during the phosphorylation phase, and the amount of free KaiA is predicted to depend on both the affinities of the nucleotide-bound states and the mixture of KaiC phosphorylation states (Fig. 2.1D).

We were surprised to see that the model fit predicts that the KaiA binding affinity for the ATP-bound T phosphoform is lower than those for the S and D phosphoforms. This is apparently in contradiction with experimental results that show that S-phosphomimetic mutants reduce A-loop exposure and weaken KaiA binding, while T-phosphomimetic mutants have opposite effects [66, 67]. Our model predicts that the nucleotide bound to CII has large impact on the affinity of KaiA for KaiC. One possibility is that phosphomimetic mutations may change the average nucleotide-bound state by altering the catalytic cycle, and that these nucleotide changes have a large effect on the experimental measurements. Future tests of this hypothesis through experiments that directly control for the bound nucleotide and models that represent both nucleotide- and phosphoform-dependent KaiA binding affinity will likely be able to separate the direct effect of phosphorylation from nucleotide effects.

The dephosphorylation pathway is simpler because KaiA is not involved. In the model, KaiC by itself has no nucleotide exchange activity in CII, and thus phosphorylated KaiC molecules enter a cycle of dephosphorylation by the transfer of phosphoryl groups from the phosphorylation sites back to bound ADP molecules, followed by ATP hydrolysis and removal of inorganic phosphate, until the protein reaches the C^{U}_{DP} state (Fig. S2.8D). The ADP-bound forms of the T, S, and D phosphoforms are only transiently populated, suggesting that the dephosphorylation bottleneck is ATP hydrolysis, which makes bound ADP available as a cofactor for dephosphorylation, rather than the phosphotransfer itself. The kinetic preference for the D→S dephosphorylation pathway is the direct result of faster dephosphorylation via the D→S reaction compared to the D→T reaction (Fig. S2.5B; compare the posterior distribution of k_d^{DS} with that of k_d^{DT}). During this process, KaiC can occasionally autophosphorylate, but it is driven irreversibly towards the dephosphorylated state by ATP hydrolysis. We note here that the independence of the dephosphorylation reaction from solution ADP [29] is a built-in feature of the model, since solution %ATP only affects the nucleotide exchange rate, which is assumed in the model to be zero in the absence of KaiA.

2.3.4 KaiA concentration tunes clock sensitivity to %ATP

Our data and the model allow us to ask in detail how [KaiA] modulates the effect of %ATP on KaiC phosphorylation (Fig. 2.2C). Consistent with previous measurements [28, 29], these results indicate that the near-steady-state ($t = 12.25$ h) total phosphorylation level of KaiC ($\%C^{P} = \%C^{T} + \%C^{S} + \%C^{D}$) is lower in the presence of ADP. Since we simultaneously vary %ATP and [KaiA], the data reveal that this inhibitory effect can be tuned by [KaiA]. In particular, the system is most insensitive to %ATP at either very low or very high [KaiA], while the %ATP sensitivity is the highest around [KaiA] = 0.75 µM.

Some %ATP sensitivity remains even at saturating KaiA concentrations (Fig. 2.2C).

This effect can be interpreted qualitatively in terms of the structure of the model. When there is more KaiA in solution, more KaiC goes through intermediate states that are in complex with KaiA, by Le Châtelier's principle. This shifts a larger fraction of the KaiC population to states that allow for exchange of bound ADP for ATP, which promotes phosphorylation. On the other hand, as the %ATP decreases, the ATP to ADP exchange rate decreases according to (2.4). When nucleotide exchange becomes less efficient, more KaiC stays in ADP-bound states, which are prone to dephosphorylation. In summary, [KaiA] and %ATP both act on the phosphorylation kinetics via the nucleotide exchange step. Solution %ATP directly regulates the exchange rate constant and sets its upper bound, while [KaiA] controls the population of exchange-competent KaiC and thus the effective exchange rate. Therefore, the effects of KaiA and increasing solution %ATP are not equivalent; because the effective exchange rate cannot exceed the limit set by %ATP, even a saturating amount of KaiA cannot fully compensate for low %ATP. In this sense, KaiA acts as an input regulator—high [KaiA] blunts the intrinsic sensitivity of KaiC to ADP in solution.

Given that the metabolic sensitivity of the KaiA-KaiC subsystem can be tuned by KaiA concentration, we asked whether metabolic sensitivity of the full oscillator period may also be tuned by KaiA. Such a scenario might allow the [KaiA]/[KaiC] ratio to be an important parameter in vivo for adjusting the sensitivity of the clock to the daily metabolic rhythm of the cell. To address this question, we characterized the dependence of the period of the in vitro KaiABC oscillator on [KaiA] and %ATP using a fluorescence polarization assay [30, 33] (Fig. 2.2D and Fig. S2.10A and B). Consistent with the hypothesis, we found that low KaiA concentration enhances the period sensitivity to %ATP compared to the standard condition (1.5 µM KaiA). These results suggest that the KaiA activity, and how it is controlled, plays a critical role in determining how responsive the oscillator is to metabolic changes.

2.3.5 KaiC phosphorylation exhibits ultrasensitive dependence on KaiA levels

In addition to inferring kinetics of states not easily accessible to experiments, the model allows us to interpolate between the training data points and study the relation between KaiC phosphorylation, [KaiA], and %ATP at a much finer resolution. This analysis shows an ultrasensitive dependence of the steady-state $\%C^P$ on KaiA concentration (Fig. 2.3A and D left). Specifically, we see a threshold-hyperbolic stimulus-response relation [68, 69], where KaiC phosphorylation is highly suppressed near the sub-micromolar [KaiA] regime, but then follows a right-shifted hyperbolic stimulus-response function once [KaiA] exceeds a threshold. Importantly, the threshold depends on %ATP. The model makes similar predictions for the steady-state T, S, and D phoshoforms as well (Fig. S2.12A). However, because of the T→D and S→D phosphotransfer reactions, the stimulus-response relations of T and S are not monotonic functions of [KaiA] because high [KaiA] and high %ATP conditions stabilize the D phosphoform at the expense of the T and S phosphoforms.

Previous studies of KaiA-KaiC interactions examined the response of KaiC at relatively high KaiA concentrations ($\geq$ 1.2 µM), comparable to the total amount of KaiA used in an oscillating reaction. [70] investigated the steady-state stimulus-response relation, but did not consider the effect of %ATP or fully characterize the low [KaiA] regime. Previous reports of initial phosphorylation rates suggest that they exhibit a hyperbolic dependence on [KaiA] [17, 21], similar to simple Michaelis-Menten enzyme-substrate systems. However, this does not imply that the steady-state stimulus-response relation is hyperbolic as well.

To assess the robustness of the model prediction of ultrasensitivity across the ensemble, we use two metrics proposed by [71] to quantify the shape of the predicted stimulus-response curves for $\%C^P$ at any fixed %ATP: we use EC10 to measure the extent to which the curve acts as a threshold and EC90 – EC10 to measure the extent to which the curve acts as a switch. Here, ECx is the KaiA concentration required to

reach x% of the steady-state phosphorylation level at saturation. Fig. 2.3E shows the distribution of these quantities in the ensemble at 25% ATP. Overall these statistics are tightly constrained by the training data set, and are clearly distinct from those from hyperbolic stimulus-response relations (Fig. 2.3E, dashed gray line).

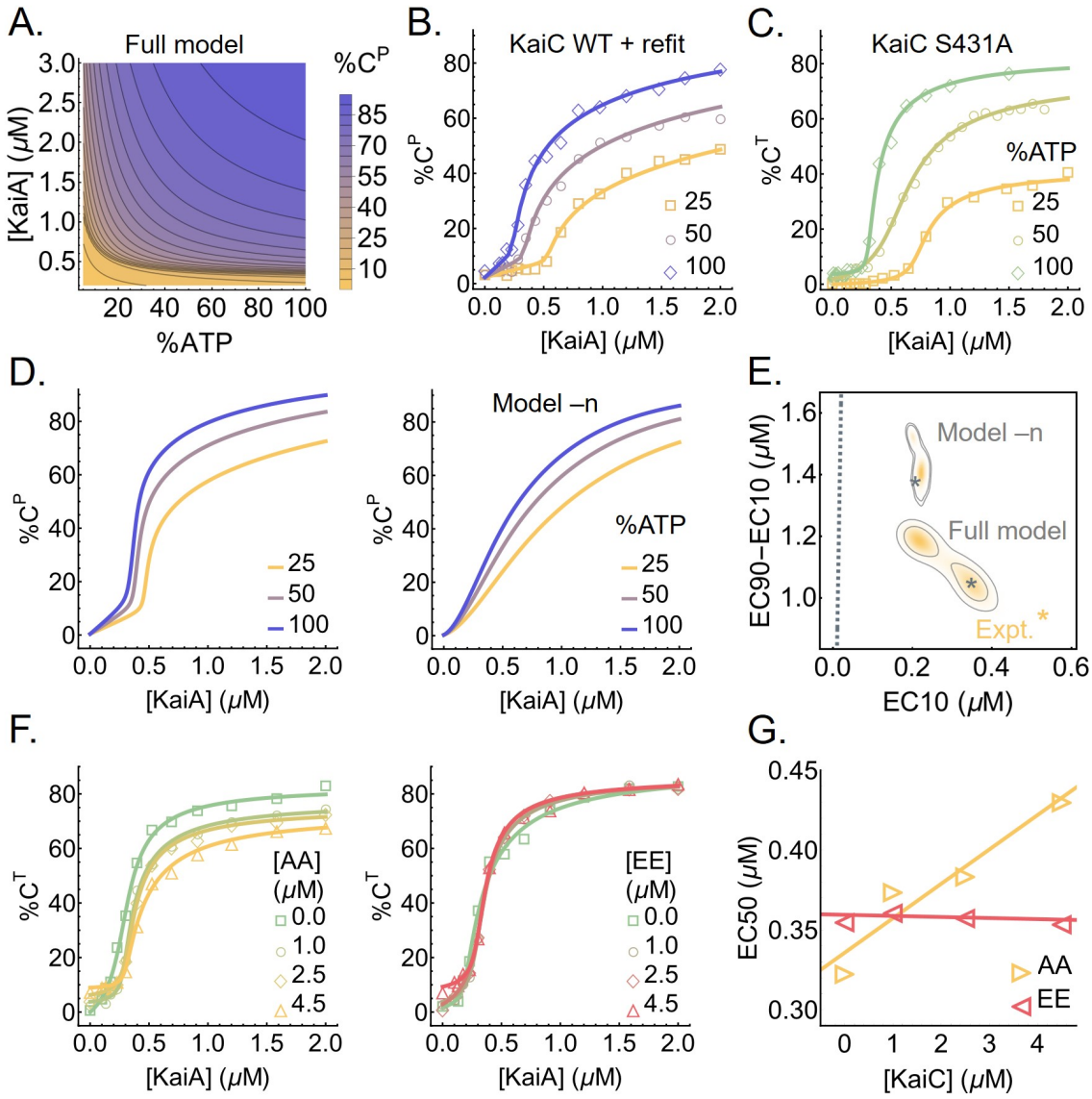


Figure 2.3: Substrate competition explains KaiC phosphorylation ultrasensitivity. A. The predicted stimulus-response relation of the total steady-state KaiC phosphorylation level as a function of %ATP and [KaiA]. B. Experimentally-determined stimulus-response function of KaiC at three %ATP conditions; the curves are based on refitting the best fit of the full model to the steady-state measurements. C. Similar to B. but for KaiC S431A, which has only one phosphorylation site; the curves are based on independent fits to a simple phenomenological substrate competition model.

(Figure 2.3, continued) D. Cross sections of the stimulus-response relation at three %ATP, computed using the full model (left) and model –n (right). E. Posterior distributions for the shape measures of the stimulus-response functions at 25% ATP predicted by the full model and model –n. The contours represent the 68% and 95% HDRs, and the gray stars represent the model best fits. The shapes of the stimulus-response functions are quantified using two metrics: EC10, which quantifies threshold-like behavior, and EC90 – EC10, which quantifies switch-like behavior. The shape measures of the experimentally-determined stimulus-response function at 25% ATP is shown as the yellow star. The dashed line represents $(\mathrm{EC10}, \mathrm{EC90} - \mathrm{EC10}) = (K/9, 80K/9)$, which characterizes the shape of a hyperbolic stimulus-response function $[\mathrm{A}]/(K + [\mathrm{A}])$ that has no switching or thresholding. F. The stimulus-response functions of KaiC S431A at 100% ATP in the presence of KaiC S431A/T432A (AA; left) and S431E/T432E (EE; right) phosphomimetic mutants to probe the effect of kinetic competition on KaiC phosphorylation. G. The relations between EC50 (the midpoint of a stimulus-response function) and KaiC AA/EE concentrations, quantified using the curves shown in F.

We then sought to experimentally determine the shape of the stimulus-response function. We measured KaiC phosphorylation at $t = 24$ h at various concentrations of [KaiA] at three %ATP conditions (Fig. 2.3B and Fig. S2.12B). Consistent with model prediction, the experimentally-derived stimulus-response relations are ultrasensitive with a %ATP-dependent phosphorylation threshold, and the stimulus-response relation of the S phosphoform at 100% ATP is non-monotonic. We then quantified the shape of the stimulus-response curve for $\%\mathrm{C}^{\mathrm{P}}$ at 25% ATP using the same two metrics defined above (Fig. 2.3E, yellow star). At 25% ATP, the shape of the experimentally-derived stimulus-response curve is close to that of the model prediction, but the model fit is systematically less threshold-like (i.e., smaller EC10) and less switch-like (i.e., larger EC90 – EC10). This inconsistency is likely due to a combination of training data under-determining the shape of the curve at the sub-micromolar range (compare Fig. 2.2C with Fig. 2.3B) and the fitting method under-estimating uncertainties (see Appendix).

Lastly, the saturating phosphorylation levels in the steady-state measurements appear systematically lower than those implied by the training data set (compare Fig. 2.3B with D left). This may be a result of batch-to-batch variations in protein and nucleotide quantification. This difference can be corrected by refitting the full model to the steady-

state measurement (Fig. 2.3B and Fig. S2.12C). The refit results suggest that errors in protein and nucleotide concentrations primarily affect the kinetic properties of the S phosphoform in the model (Fig. S2.12D), but the refitting does not change the qualitative conclusions.

2.3.6 A substrate competition mechanism underlies ultrasensitivity in KaiC phosphorylation

What is the mechanism of ultrasensitivity in KaiC phosphorylation? Given that each KaiC subunit has two phosphorylation sites, a plausible explanation is multisite phosphorylation, whereby the concentration of the maximally phosphorylated state exhibits an ultrasensitive dependence on the kinase concentration [71] (or in this case, the nucleotide-exchange factor concentration), even if each consecutive phosphorylation step follows mass-action kinetics. To examine this possibility, we measured the stimulus-response relation of the KaiC S431A mutant, which has only one phosphorylation site, and the results show ultrasensitivity comparable to that of the WT protein (Fig. 2.3C). Furthermore, because KaiC is its own phosphatase, it violates the assumption of distributivity (i.e., at most one modification takes place before the dissociation of the enzyme and substrate) [71]. Multisite phosphorylation thus cannot explain the observed ultrasensitivity.

In the ensemble of parameter sets, the KaiA dissociation constant of the ADP-bound (but not ATP-bound) U phosphoform (C^{U}_{DP}) is in or below the nanomolar range, much smaller than that of any other state of KaiC (Fig. 2.1D). This is consistent with recent single molecule observations suggesting that the unphosphorylated form of KaiC can bind very tightly to KaiA [25] and native mass spectrometry measurements suggesting that KaiA binding to KaiC is enhanced by ATP hydrolysis, which would be needed to produce ADP-bound KaiC [72]. Here, we argue that the key to understanding the origin of ultrasensitivity in the model lies in the differential binding affinity of KaiA to the ADP-

and ATP-bound states of KaiC. We note here that since the model does not consider the hexameric structure of KaiC, we cannot rule out possible hexameric cooperative effects that may contribute to ultrasensitivity.

In the model, differential KaiA binding affinity leads to the following dynamics: KaiA promotes phosphorylation by catalyzing the exchange of the bound ADP for ATP, but this process is in a kinetic competition with ATP hydrolysis, which returns KaiC to the ADP-bound state. At the beginning of the phosphorylation reaction, almost all the KaiA is bound to C_{DP}^{U} (Fig. 2.2A and B) due to its high abundance and high affinity for KaiA (Fig. 2.1D). When [KaiA] is low, the competition between nucleotide exchange and hydrolysis in the U phosphoform reaches a steady-state where $[C_{DP}^{U}]$ stays above [KaiA] (Fig. 2.2A, dashed curves). Therefore, KaiA stays trapped by C_{DP}^{U} and the phosphorylation products (mostly T) cannot undergo nucleotide exchange. In the absence of KaiA, the autophosphatase activity of KaiC dominates, and the phosphorylation products revert back to the U phosphoform.

When [KaiA] is high, however, the competition between nucleotide exchange and hydrolysis in the U phosphoform pushes C_{DP}^{U} below [KaiA] (Fig. 2.2A, solid curves), which frees KaiA to catalyze the nucleotide exchange reactions of the phosphorylation products. Once the flux of phosphorylation, KaiA binding, and nucleotide exchange outweighs that of hydrolysis, dephosphorylation, and KaiA unbinding, the phosphorylation products stay phosphorylated at steady state. Furthermore, the formation of phosphorylated products positively feeds back to deplete C_{DP}^{U}, further removing a KaiC state that traps KaiA and leading to rapid saturation of phosphorylation past the [KaiA] threshold. The [KaiA] threshold for phosphorylation depends on %ATP (Fig. 2.3A), because when %ATP is low, more KaiA is needed to counteract the reduced ADP-to-ATP exchange rate.

This mechanism is a form of substrate competition [56, 73], a previously identified general scheme where the kinetic competition of multiple substrates for enzyme binding

leads to ultrasensitivity. Here, KaiA plays the role of the enzyme, while the ADP-bound U phosphoform and the T phosphoform (as well as the S and D phosphoforms to a lesser extent due to phosphorylation ordering) are the substrates that compete for KaiA binding. However, the fact that the phosphorylated and unphosphorylated forms of KaiC can interconvert through phosphotransfer reactions distinguishes the Kai system from a typical substrate competition scheme, where the substrates cannot interconvert.

The model suggests that the U phosphoform plays a special role in generating ultrasensitivity due to the significant difference in the affinity of KaiA for its ATP- vs. ADP-bound states (Fig. 2.1D). This observation leads to two testable predictions. First, the amount of KaiA required to activate phosphorylation should be higher when more U phosphoform is present. We tested this prediction experimentally by measuring the stimulus-response relation of KaiC S431A in the presence of KaiC S431A/T432A (AA), which mimics the U phosphoform, or KaiC S431E/T432E (EE), which mimics the D phosphoform. The KaiC AA and EE mutants act as competitors for the KaiA-KaiC interaction (Fig. 2.3F). Consistent with the hypothesis, the EC50 (i.e., the midpoint of the ultrasensitive switch) is positively correlated with the concentration of KaiC AA, while varying the concentration of KaiC EE has little effect (Fig. 2.3G). Interestingly, the presence of KaiC-AA in an oscillatory reaction also appears to reduce the amplitude of the oscillation, but the effect is most pronounced at low %ATP (Fig. S2.10C).

Second, the substrate competition mechanism suggests that the model should exhibit weaker nonlinearity if KaiA has the same affinity to ATP- vs. ADP-bound states of a given KaiC phosphoform. To computationally test this prediction, we constructed simplified models where KaiA on/off rates are set to be independent of the nucleotide-bound state (model –n) or phosphorylation state (model –p) and fit the new models to the experimental data ab initio. Consistent with the prediction, decoupling KaiA on/off rates from the nucleotide-bound states results in a significant loss of ultrasensitivity (Fig. 2.3D right and Fig. S2.13A). Model –p by contrast behaves similarly to the full model (Fig.

Table 2.1: Effects of differential KaiA (un)binding kinetics

Model	Log likelihood			Bayes factor[a]
	Phosphorylation	Dephosphorylation	Hydrolysis	
Full model	422.9	346.8	–0.8	1
–n[b]	249.2	275.4	–0.3	10.4
–p[c]	392.4	303.2	–2.4	2.2
–n,–p	204.4	266.4	–2.6	19.7

[a] We define the Bayes factor as the ratio of the marginal likelihood function of the full model over that of the simplified models. We adopt the convention that a Bayes factor larger than 3.2 is substantial evidence against the model [74].
[b] –n: on/off rates decoupled from nucleotide-bound state.
[c] –p: on/off rates decoupled from phosphorylation state.

S2.13C); consistent with the substrate competition mechanism, the ADP-bound states of KaiC in model –p have higher affinity to KaiA than the ATP-bound states, regardless of the phosphorylation state (Fig. S2.13B). We quantify the effects of such model reductions by computing the Bayes factor, which is a metric for systematic model comparison that favors goodness of fit but penalizes model complexity and parameter fine tuning [42]; it is similar to the Bayesian information criterion [75], but makes no asymptotic assumptions. The analysis shows that the loss of ultrasensitivity in model –n degrades the fit quality significantly, while model –p is only marginally worse than the full model (Table 2.1). Interestingly, a model where the KaiA on/off rates are completely independent of the state of KaiC (model –n,–p; Fig. S2.13D and E) is much worse than either model –n or model –p (Table 2.1). We conclude that the nucleotide-bound state of KaiC plays a key role in regulating its interaction with KaiA and thus in determining phosphorylation kinetics.

2.3.7 Substrate competition may underlie metabolic compensation

Finally, we consider the implications of the ultrasensitivity for the full oscillator. For the sake of clarity, we make a distinction in this section among three subpopulations of KaiA: the sequestered KaiA, which refers to inactive KaiA in a KaiABC complex; the active

KaiA, which refers to (free or bound) KaiA not sequestered by KaiB; and the free KaiA, which is not associated with either KaiB or KaiC.

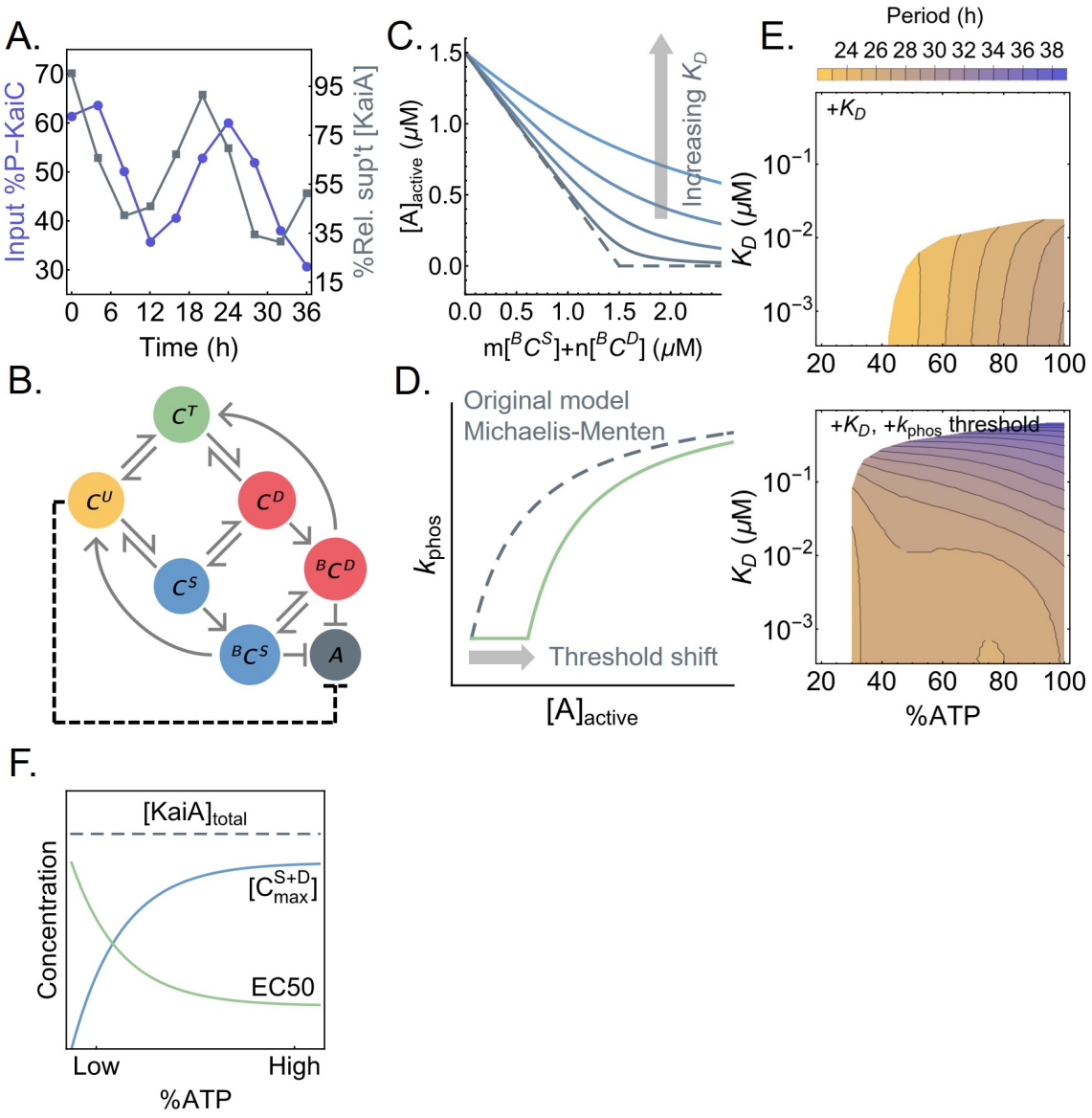


Figure 2.4: Ultrasensitivity contributes to metabolic compensation. A. The time series of the total input KaiC phosphorylation level (purple, left scale) and residual KaiA concentration not precipitated with KaiB-FLAG (gray, right scale). B. A schematic of the oscillator model by [28]. Here, $^{B}C^{S}$ and $^{B}C^{D}$ represent the KaiB-bound S and D phosphoforms, respectively, which can sequester KaiA. The dashed line represents the effect of introducing ultrasensitivity to the model. C. A cartoon representation of introducing a KaiA sequestration affinity, K_D, into the Phong model. The original model has an effectively infinite sequestration affinity (dashed curve). D. A cartoon representation of introducing a KaiA threshold to the Michaelis-Menten-type phosphorylation rate constant in the Phong model.

(Figure 2.4, continued) E. The period of the oscillator model as a function of %ATP and K_D, a measure of KaiA sequestration affinity, without (top) or with (bottom) a phosphorylation threshold. All model simulations were done with 3.5 µM of KaiC and 1.5 µM of KaiA. White regions indicate unstable or no oscillation. F. The extent to which KaiA can be sequestered by KaiB depends on the maximal S and D phosphoform concentration, $[C^{S+D}_{max}]$, achieved over the phosphorylation cycle. The scaling of the EC50 of the phosphorylation stimulus-response function, which is a measure of the capacity of the U phosphoform to suppress KaiA activity, compensates for the scaling of $[C^{S+D}_{max}]$ with %ATP.

We first consider the current understanding of how KaiA activity is regulated over the circadian cycle. It is well-established that the regulation of KaiA during nighttime by KaiB plays an essential role in producing the negative feedback loop. At dusk, the buildup of KaiC D and S phosphoforms triggers the binding of KaiB to CI [17, 20, 21, 28, 76–79] and subsequently the sequestration of KaiA by CI-bound KaiB [20, 24, 80]. In the absence of active KaiA, the CII domain autodephosphorylates, and the KaiABC ternary complex disassembles [78] at dawn as KaiC reaches its unphosphorylated state [12], freeing KaiA and readying the clock for the next cycle.

This understanding of the negative feedback loop implies that the sequestration of KaiA by KaiB is a source of nonlinearity in the system that is critical for generating oscillation. Indeed, in many models of the Kai oscillator, the complete sequestration of KaiA during dephosphorylation is either a built-in or required feature for stable oscillation [e.g., 28, 40, 81, 82]. However, our observation that phosphorylation is suppressed nonlinearly at low [KaiA] suggests that complete sequestration of KaiA by KaiB is not necessary to prevent phosphorylation at night. Indeed, there is mounting evidence that KaiB sequestration by itself is insufficient to completely inactivate KaiA during dephosphorylation. Specifically, measurements using native mass spectrometry, co-immunoprecipitation (co-IP), and native PAGE suggest that there is a significant amount of $KaiA_2C_6$ complex [24, 83] and free KaiA [80] throughout the entire phosphorylation cycle.

To confirm that KaiA is not fully sequestered by KaiBC complexes, we used immunoprecipitation of FLAG-tagged KaiB to monitor the amount of uncomplexed KaiA in supernatant, which we interpret to be a measure of active KaiA concentration (Fig. 2.4A and Fig. S2.10D). The experiment shows that there is indeed a substantial amount of active KaiA in solution in the first half of the dephosphorylation stage of the oscillation. Taken together, these results suggest that either the binding of KaiA to KaiBC has lower affinity than previously assumed, or that the sequestration kinetics are slow compared to the length of the dephosphorylation stage. In either case, some KaiA appears to remain free of KaiABC complexes at all times during the oscillation.

Given these results, we consider the role ultrasensitivity may play in regulating KaiA's ability to stimulate KaiC phosphorylation at nighttime. In particular, the fact that the phosphorylation threshold scales with %ATP suggests that ultrasensitivity may also lead to insensitivity of the period of the Kai oscillator to %ATP [28], a phenomenon termed "metabolic compensation" [57]. As a proof of principle, we examine this possibility using a simple model of the Kai oscillator proposed by [28], which we refer to as the Phong model. The Phong model explicitly keeps track of the monomer phosphorylation cycle and uses KaiB binding to the S phosphoform to generate negative feedback (Fig. 2.4B). In the Phong model, the KaiA sequestration affinity is effectively infinite. In light of the co-IP experiment, we modify the model by assuming that the KaiA sequestration reaction is in a quasi-equilibrium with a dissociation constant for KaiA binding to the KaiBC complex, K_D (Fig. 2.4C; see Appendix for mathematical details). When K_D is small (i.e., $< 10^{-3}$ µM), the modified model exhibits the same robust oscillations as the original model over a large range of %ATP, but the range of %ATP that allows for stable oscillation shrinks as K_D increases (Fig. 2.4E top), and the model is unstable when K_D is in the micromolar range regardless of %ATP.

In the original Phong model, the dependence of KaiC phosphorylation on KaiA is described by a Michaelis-Menten-like function with no ultrasensitivity. In this scenario, a

small increase in active KaiA leads to a proportional increase in phosphorylation, making the dephosphorylation phase of the clock strongly dependent on the strength of KaiB-mediated KaiA sequestration. To test if ultrasensitivity can increase the robustness of oscillations in the model, we introduce a phenomenological patch to the model in the form of an ultrasensitive KaiA threshold to the phosphorylation rate function, which varies as a function of %ATP and U phosphoform concentration (Fig. 2.4D; see Appendix for mathematical details). Given that the ultrasensitivity is a result of substrate competition, this modification effectively introduces an inhibitory interaction between the U phosphoform and KaiA (Fig. 2.4B, dashed arrow). This modification amounts to the assumption that the EC50 measured at steady state (Fig. 2.3A) in the absence of KaiB corresponds to the active KaiA concentration required to re-enter the phosphorylation phase at the trough of the circadian oscillation. Remarkably, the resulting model can generate stable oscillations over a larger range of both %ATP and K_D conditions, and the period of oscillation is made much less dependent on %ATP. (Fig. 2.4E bottom). This observation suggests that ultrasensitivity in KaiC phosphorylation plays a role in clock stability that complements the function of KaiB-dependent KaiA sequestration.

Why does ultrasensitivity in KaiC phosphorylation allow for metabolic compensation? The binding of KaiB to KaiC, and thus the sequestration and inactivation of KaiA, depends on S431 phosphorylation of KaiC (i.e., the S and D phosphoforms). At low %ATP, the maximal S and D concentrations, $[\mathrm{C}^{\mathrm{S+D}}_{\mathrm{max}}]$, are lower (Fig. 2.4F). Thus the maximal amount of KaiA sequestered by the KaiBC complex is smaller. This is problematic for the stability of the clock at low %ATP, since the active KaiA can promote premature KaiC U→T phosphorylation of some molecules, which can lead to phase decoherence that manifests as decay of the oscillation (Fig. S2.10F). The ultrasensitive stimulus-response that we report here implies that a finite amount of KaiA must be liberated from KaiB before there is a noticeable impact on KaiC phosphorylation. In other words, the inhibitory effect of ultrasensitivity is a synchronization mechanism at

night. Importantly, the EC50 of the stimulus-response function scales with %ATP, such that the capacity of phosphorylation suppression by C^U is enhanced at low %ATP, which compensates for weaker KaiB-mediated KaiA sequestration (Fig. 2.4F). The compensatory relation between these two mechanisms as %ATP varies is likely key to period stability. This relation likely also contributes to the scaling of the phosphorylation limit cycle size with %ATP. At higher %ATP, the EC50 is smaller and thus more KaiA needs to be sequestered to trigger dephosphorylation, which implies that higher concentrations of the S and D phosphoforms need to accumulate to enable KaiB binding. Since KaiC phosphorylation is ordered, this means that the T phosphoform concentration scales with %ATP as well.

2.4 Discussion

In this work we undertook a data-driven kinetic modeling approach to understand the mechanism and metabolic sensitivity of the KaiA-KaiC subsystem, part of the *S. elongatus* circadian oscillator. We constructed a detailed yet mechanistically naive kinetic model, which was fit to extensive experimental measurements of KaiC phosphorylation kinetics within a Bayesian parameter estimation framework. Approaches that are similar in spirit have been pursued in eukaryotic systems [e.g., 84–88]. However, owing to the greater complexity of eukaryotic clocks, these studies combined direct experimental measurements, cost function optimization, and hand tuning of selected parameters to account for unknown or unconstrained biochemical processes. Because the Kai system can be studied as a well-defined mixture of purified components, the participating molecular species are known, and all the parameters in the model can be treated in a consistent manner to enable objective comparison of mechanisms underlying collective oscillations.

This data-driven approach is to be contrasted with the more common hypothesis-driven approach, whereby a model is built to examine how features of the oscillator arise from proposed mechanisms. This hypothesis-driven approach has been employed

extensively in the study of the cyanobacterial clock. These studies have revealed insights into specific aspects of the oscillator function, such as entrainment [33, 83], synchronization [81, 82, 89], irreversibility [90], and robustness against variations in temperature [31, 34, 35, 91], ATP/ADP concentration [28, 36], protein copy numbers [21, 83, 92], and environmental noise in general [93–95]. This hypothesis-driven approach is pedagogically powerful but relies on the ingenuity of the modeler to identify a priori which mechanisms should be included. These approaches give little indication of the range of the parameter space consistent with a proposed mechanism, which makes it difficult to quantify the uncertainties of model predictions and validate them experimentally.

Not all parameters in our model were fully constrained by the data, as expected given the complexity of the model [55]. Nevertheless, the ensemble of parameter sets still led to consistent predictions. In particular, the model revealed unexpected ultrasensitivity in KaiC phosphorylation as a function of KaiA, which we confirmed experimentally. The source of ultrasensitivity in the model is a substrate competition mechanism that arises from the differential affinity of ADP- and ATP-bound KaiC for KaiA. Previous studies have considered the importance of the differential affinity of KaiA for KaiC states but have focused on phosphorylation [25, 40]. We note here that the ultrasensitivity in KaiC phosphorylation that we discovered and the ultrasensitivity in KaiB-dependent KaiA sequestration that arises from opposing S and T phosphorylations within hexamers [21] perform fundamentally different roles. Specifically, the cooperative binding of KaiB to KaiC during the dephosphorylation stage provides the nonlinear, delayed negative feedback loop essential for oscillation [1], whereas the mutual inhibition between KaiA and ADP-bound KaiC, which effectively provides positive feedback on KaiA's activity, generates a KaiC phosphorylation threshold that contributes to period stability as %ATP changes. The presence of nonlinearities and delayed feedback at multiple steps in a molecular oscillator allows the system to achieve greater robustness [88, 96–98].

We hypothesized that ultrasensitivity in KaiC phosphorylation plays a role in

stabilizing the oscillator at low %ATP conditions by suppressing premature phosphorylation during the dephosphorylation stage and thus promoting phase coherence. Currently, the Kai oscillator model most robust against yet tunable by metabolic conditions appears to be that of [36, 40]. In the Paijmans model, metabolic compensation is achieved both at the hexamer and ensemble level. At the hexamer level, the onset of dephosphorylation is primarily controlled by the antagonistic effects of the T and S phosphoforms. Since fewer subunits in the T phosphoform accumulate at low %ATP, fewer subunits in the S phosphoform are needed to trigger dephosphorylation; therefore, the reduced amplitude of oscillation counteracts the slower phosphorylation rate at low %ATP. At the ensemble level, low %ATP limits the fraction of hexamers that are able to trigger dephosphorylation before the onset of KaiB-mediated delayed inhibition; this makes the dephosphorylation phase shorter, which compensates for the longer phosphorylation phase. It is worth noting that, unlike the data, the Paijmans model is not oscillatory when %ATP reaches below 50%, partly due to phase decoherence during dephosphorylation, an issue that can potentially be addressed with ultrasensitivity in KaiC phosphorylation. In our model the coupling between KaiA binding affinity and KaiC nucleotide-bound states is critical in generating ultrasensitivity, a feature that is missing in the Paijmans model. It remains an open question whether a hexameric model with no such coupling can nevertheless produce ultrasensitivity in KaiC phosphorylation (see Appendix for further comparison between this work and the Paijmans model). The ultimate goal of this line of research is to obtain a molecularly detailed model of the complete Kai oscillator that is constrained by direct fitting to empirical measurements.

In *S. elongatus*, the Kai oscillator is embedded in a transcription-translation feedback loop [3–5]. However, with the exception of peroxiredoxin oxidation cycles [99, 100], cell-autonomous circadian rhythms in eukaryotes are thought to be generated by interlocked transcription-translation feedback loops [1]; the cooperative autoregulation of transcription is a key source of nonlinearity and robustness in the circuit [e.g., 85, 101–104]. Our results

raise the possibility that post-translational modifications and protein-protein interactions may also contribute to robustness generally in circadian clocks by introducing ultra-sensitivity, even if these processes do not generate self-sustaining rhythms that can be decoupled from transcription. Overall, it is clear that post-translational steps such as (de)phosphorylation [105–108], protein degradation [105, 109], and complex formation [88] play an important role in eukaryotic circadian oscillators, but to our knowledge there is currently no complete experimental characterization of the stimulus-response relations of these processes.

2.5 Materials and Methods

2.5.1 Computational methods

Treatment of nucleotide exchange Here we derive (2.4) in Results. The nucleotide exchange process can be modeled as a two-step reaction that includes an apo intermediate state of KaiC, i.e.,

$$ {}^{\mathrm{A}}\mathrm{C}_{\mathrm{TP}} \underset{k_{\mathrm{on}}^{\mathrm{TP}}}{\overset{k_{\mathrm{r}}^{\mathrm{TP}}}{\rightleftharpoons}} {}^{\mathrm{A}}\mathrm{C} \underset{k_{\mathrm{r}}^{\mathrm{DP}}}{\overset{k_{\mathrm{on}}^{\mathrm{DP}}}{\rightleftharpoons}} {}^{\mathrm{A}}\mathrm{C}_{\mathrm{DP}} \tag{2.6} $$

where we have omitted the free ATP and ADP from the chemical equation. Here, $k_{\mathrm{r}}^{\mathrm{TP}}$ and $k_{\mathrm{r}}^{\mathrm{DP}}$ are the dissociation rate constants and $k_{\mathrm{on}}^{\mathrm{TP}}$ and $k_{\mathrm{on}}^{\mathrm{DP}}$ are the binding rate constants for ATP and ADP, respectively. Since KaiC requires nucleotides for hexamerization [16, 76, 110], the apo state of KaiC is presumably both thermodynamically and kinetically unstable in the presence of saturating amount of nucleotide (5 mM in our experiments). Therefore, under the assumption that the KaiC apo state is in a quasi-steady state throughout the reactions, we can eliminate the apo state and model nucleotide exchange as a one-step reaction

$$ {}^{\mathrm{A}}\mathrm{C}_{\mathrm{TP}} \underset{k_{\mathrm{TP}}^{\mathrm{A}}}{\overset{k_{\mathrm{DP}}^{\mathrm{A}}}{\rightleftharpoons}} {}^{\mathrm{A}}\mathrm{C}_{\mathrm{DP}} \tag{2.7} $$

where

$$k_{\mathrm{TP}}^{\mathrm{A}} = k_{\mathrm{r}}^{\mathrm{DP}} \frac{[\mathrm{ATP}]}{[\mathrm{ATP}] + K_{\mathrm{on}}[\mathrm{ADP}]} \tag{2.8}$$

$$k_{\mathrm{DP}}^{\mathrm{A}} = k_{\mathrm{r}}^{\mathrm{TP}} \left(1 - \frac{[\mathrm{ATP}]}{[\mathrm{ATP}] + K_{\mathrm{on}}[\mathrm{ADP}]}\right) \tag{2.9}$$

and $K_{\mathrm{on}} = k_{\mathrm{on}}^{\mathrm{DP}}/k_{\mathrm{on}}^{\mathrm{TP}}$ is a ratio of the two nucleotide binding rate constants.

We make two further simplifying assumptions. First, we assume the on rates are completely diffusion controlled and are thus the same for ATP and ADP, which allows us to set $K_{\mathrm{on}} = 1$. Second, based on fit results (Fig. S2.8F) showing that the posterior for $k_{\mathrm{r}}^{\mathrm{TP}}$ has a long tail to negative infinity in log space, we follow the approach proposed by [111] and set $k_{\mathrm{r}}^{\mathrm{TP}} = 0$; i.e., the dwell time of ATP-bound states are sufficiently long that a bound ATP cannot be released without first giving up its γ-phosphate group. This assumption implies that the only ways for KaiC to enter an ADP-bound state are through hydrolysis and phosphorylation, and solution ADP has no effect on the system except to slow down the ADP to ATP exchange process. With these two assumptions, we eliminate (2.9), and (2.8) reduces to (2.4).

Model parameterization In Results, we introduced a model parameterization scheme in which rate constants for phosphotransfer, nucleotide exchange, KaiA (un)binding, and ATP hydrolysis reactions depend on the participating molecular species. Although we use this independent-rate scheme to interpret the model, including computing the sensitivity ODEs, during the fitting itself we represent state-dependent effects by modifying each of the six basic rate constants (k_{p}, k_{d}, k_{a}, k_{b}, $k_{\mathrm{r}}^{\mathrm{DP}}$, and k_{h}) by multiplicative Δk factors. For example, the KaiA dissociation rate $k_{\mathrm{b}}^{\mathrm{T,TP}} = k_{\mathrm{b}} \Delta k_{\mathrm{b}}^{\mathrm{T,TP}}$ is represented by the product between a base rate k_{b} and a modifier $\Delta k_{\mathrm{b}}^{\mathrm{T,TP}}$ (compare Fig. S2.5 and Fig. S2.14). The multiplicative-factor scheme introduces 38 Δk parameters. Because of the requirement for detailed balance (see below), only 34 of these parameters are free; these free parameters

Table 2.2: Full model parameters and their priors

Category	State-dependent effect	Parameters	Prior	Unit
Basic	N/A	$k_{\mathrm{h}}, k_{\mathrm{p}}, k_{\mathrm{d}}, k_{\mathrm{a}}, k_{\mathrm{b}}, k_{\mathrm{r}}^{\mathrm{DP}}$	$10^{\mathcal{N}(\mu,3)\ b}$	s^{-1} [d]
Nucleotide exchange	KaiA & phos.[a]	$\Delta k_{\mathrm{TP}}^{\mathrm{A,T}}, \Delta k_{\mathrm{TP}}^{\mathrm{A,S}}, \Delta k_{\mathrm{TP}}^{\mathrm{A,D}}$		
Hydrolysis	phos.	$\Delta k_{\mathrm{h}}^{\mathrm{T}}, \Delta k_{\mathrm{h}}^{\mathrm{S}}, \Delta k_{\mathrm{h}}^{\mathrm{D}}$		
	KaiA & phos.	$\Delta k_{\mathrm{h}}^{\mathrm{A,U}}, \Delta k_{\mathrm{h}}^{\mathrm{A,T}}, \Delta k_{\mathrm{h}}^{\mathrm{A,S}}, \Delta k_{\mathrm{h}}^{\mathrm{A,D}}$		
KaiA on	nuc.[a] & phos.	$\Delta k_{\mathrm{a}}^{\mathrm{U,DP}}, \Delta k_{\mathrm{a}}^{\mathrm{D,DP}}, \Delta k_{\mathrm{a}}^{\mathrm{D,TP}}$		
KaiA off	nuc. & phos.	$\Delta k_{\mathrm{b}}^{\mathrm{U,DP}}, \Delta k_{\mathrm{b}}^{\mathrm{T,DP}}, \Delta k_{\mathrm{b}}^{\mathrm{S,DP}}, \Delta k_{\mathrm{b}}^{\mathrm{D,DP}}, \Delta k_{\mathrm{b}}^{\mathrm{T,TP}}, \Delta k_{\mathrm{b}}^{\mathrm{S,TP}}, \Delta k_{\mathrm{b}}^{\mathrm{D,TP}}$		
(De)phosphorylation	phos.	$\Delta k_{\mathrm{p}}^{\mathrm{US}}, \Delta k_{\mathrm{d}}^{\mathrm{SU}}, \Delta k_{\mathrm{p}}^{\mathrm{TD}}, \Delta k_{\mathrm{d}}^{\mathrm{DT}}, \Delta k_{\mathrm{p}}^{\mathrm{SD}}, \Delta k_{\mathrm{d}}^{\mathrm{DS}}$	$10^{\mathrm{Laplace}(\mu,1)\ b}$	N/A
	KaiA & phos.	$\Delta k_{\mathrm{p}}^{\mathrm{A,UT}}, \Delta k_{\mathrm{d}}^{\mathrm{A,TU}}, \Delta k_{\mathrm{p}}^{\mathrm{A,TD}}, \Delta k_{\mathrm{d}}^{\mathrm{A,DT}}, \Delta k_{\mathrm{p}}^{\mathrm{A,SD}}, \Delta k_{\mathrm{d}}^{\mathrm{A,DS}}, \Delta k_{\mathrm{p}}^{\mathrm{A,US}}, \Delta k_{\mathrm{d}}^{\mathrm{A,SU}}$		
Global error	N/A	σ^2	Inv-Gamma (1, 0.01)	$\mu\mathrm{M}^2$
Initial conditions	N/A	$[\mathrm{C}_{\mathrm{TP}}^{\mathrm{U}}]_0, [\mathrm{C}_{\mathrm{DP}}^{\mathrm{U}}]_0, [\mathrm{C}_{\mathrm{TP}}^{\mathrm{T}}]_0, [\mathrm{C}_{\mathrm{DP}}^{\mathrm{T}}]_0, [\mathrm{C}_{\mathrm{TP}}^{\mathrm{S}}]_0, [\mathrm{C}_{\mathrm{DP}}^{\mathrm{S}}]_0, [\mathrm{C}_{\mathrm{TP}}^{\mathrm{D}}]_0, [\mathrm{C}_{\mathrm{DP}}^{\mathrm{D}}]_0$	Dirichlet(a)[c]	$\mu\mathrm{M}$

[a] phos., phosphoform; nuc., nucleotide-bound state

[b] The mean of the priors, μ, is zero unless specified by Table 2.5.

[c] $a = (20, 100, 1, 1, 1, 1, 1, 1)$; points drawn from the distribution are scaled by the total KaiC concentration. The support of the Dirichlet distribution implies that only seven of the eight initial conditions are free fitting parameters.

[d] or $\mathrm{s}^{-1}{\cdot}\mu\mathrm{M}^{-1}$ for the second-order rate constant k_{a}.

Table 2.3: Detailed balance conditions

Cycle	Detailed balance condition
$\{C^{S}_{TP}, C^{D}_{DP}, {}^{A}C^{D}_{DP}, {}^{A}C^{S}_{TP}\}$	$\delta k_a^{S,TP} = \Delta k_b^{S,TP} \frac{\Delta k_d^{A,DS}}{\Delta k_p^{A,SD}} \frac{\Delta k_a^{D,DP}}{\Delta k_b^{D,DP}}$
$\{C^{T}_{TP}, C^{D}_{DP}, {}^{A}C^{D}_{DP}, {}^{A}C^{T}_{TP}\}$	$\delta k_a^{T,TP} = \Delta k_b^{T,TP} \frac{\Delta k_d^{A,DT}}{\Delta k_p^{A,TD}} \frac{\Delta k_a^{D,DP}}{\Delta k_b^{D,DP}}$
$\{C^{U}_{TP}, C^{T}_{DP}, {}^{A}C^{T}_{DP}, {}^{A}C^{U}_{TP}\}$	$\delta k_a^{T,DP} = \Delta k_b^{T,DP} \frac{\Delta k_p^{A,UT}}{\Delta k_d^{A,TU}}$
$\{C^{U}_{TP}, C^{S}_{DP}, {}^{A}C^{S}_{DP}, {}^{A}C^{U}_{TP}\}$	$\delta k_a^{S,DP} = \Delta k_b^{S,DP} \frac{\Delta k_p^{A,US}}{\Delta k_d^{A,SU}}$

are listed on Table 2.2. The advantage of the multiplicative parameterization scheme is that it facilitates ℓ^1 regularization, discussed below.

Detailed balance All elementary reactions, except ATP hydrolysis and nucleotide exchange, are assumed to occur in equilibrium, and thus the free energy change over each reversible cycle must be equal to zero. In practice, this means that the product of all rate constants in the forward direction of each cycle listed on Table 2.3 must be equal to that in the reverse direction (see also Fig. S2.14A). This introduces an additional algebraic constraint for each such cycle, which is used to eliminate one free Δk parameter. In total, one can eliminate four parameters.

Fitting data set To constrain the model parameters, we collected experimental measurements that characterized different aspects of the KaiA-KaiC subsystem, which are summarized in Table 2.4.

The dephosphorylation reaction data taken from [17] constrains the dephosphorylation rates and the ATP hydrolysis rates of KaiC in the absence of KaiA, because the model structure dictates that dephosphorylation requires alternating phosphotransfer and hydrolysis reactions. There is currently no direct measurement of CII hydrolysis rate

Table 2.4: Fitting data set

Measurement (source)	T (°C)	[KaiA][a] (µM)	ATP	Time points	Phosphoform
Phosphorylation (this work)	30	0.375, 0.75, 1.50, 3.00, 4.50, 6.00	10, 25, 100%[b]	8	U, T, D[c]
Dephosphorylation [29]	30	1.4	5 mM	21	
ADP production [26]	30	1.2	1 mM	1	N/A
KaiA on/off rates [24]	25	Variable	1 mM	N/A	Likely U
KaiA dwell time [25]	25–28	1.0	N/A	N/A	T, S, D[d]

[a] We report here on the KaiA monomer concentration. However, since KaiA functions as a dimer, all KaiA concentration is divided by two in the models.
[b] %ATP, defined as 100%[ATP]/([ATP] + [ADP]); total [ATP] + [ADP] is held constant at 5 mM.
[c] The conservation of mass constraint implies that one of the four phosphoforms is not a free state variable. We have chosen the S phophoform to be the constrained state variable.
[d] Phosphomimetic mutants

in the presence of KaiA. However, the maximum ADP production rate of KaiC in the presence of 1.2 µM of KaiA was reported to be 29.8 ± 5.1 $KaiC^{-1}\cdot day^{-1}$ [26], which we take as an upper bound on the average CII hydrolysis rate in phosphorylation reactions with [KaiA] = 0.375, 0.75, and 1.50 µM.

Because the phosphorylation reactions were measured in the presence of varying %ATP and higher [ADP] inhibits phosphorylation by slowing down nucleotide exchange (see equation 2.4), such measurements provide indirect constraints on the nucleotide exchange rate. Similarly, because the reactions were measured in the presence of varying [KaiA] conditions, they directly constrain the phosphorylation rates of KaiC with and without KaiA, as well as the KaiA binding affinity, i.e., the ratio of KaiA on/off rates. Although there are direct experimental measurements of KaiA binding and dissociation [24, 25], these results cannot be directly mapped onto model rate constants. This is primarily because the KaiC nucleotide-bound fractions are not measured in these experiments, or, in the case of phosphomimetic mutants, it is unclear if the mutations

affect nucleotide binding affinities. As a consequence, the experimental constraints on KaiA on/off rates enter through the priors rather than the likelihood function, in contrast to the other data (see below).

Initial conditions For each phosphorylation reaction in Table 2.4, we solve the ODE model with the corresponding [KaiA] and %ATP condition. The predicted phosphoform composition, as well as the ATP hydrolysis rate when appropriate, is compared to the experimental measurements in the Bayesian parameter estimation formalism described below. However, since the experimental data do not resolve the initial conditions for all 16 KaiC states in the model, we have chosen to directly estimate the initial concentrations by treating them as free parameters in the fitting procedure. We take $t = 0$ to be the time point at which KaiA is mixed with KaiC, and thus all eight KaiA-bound KaiC states have zero concentration at the onset of the experiment. Because total KaiC concentration is conserved, this introduces seven additional parameters (see Table 2.2).

For the dephosphorylation reaction, we do not estimate the initial conditions. To mimic the way the experiment was done, the dephosphorylation reaction is simulated in two stages. In the first stage, we assume that 3.4 µM of dephosphorylated KaiC is phosphorylated in the presence of 1.3 µM KaiA and 100% ATP for 20 h. Since the protein is initially dephosphorylated, we assume that $[\mathrm{C}^{\mathrm{U}}_{\mathrm{TP}}]_0 = 3.4$ µM while the concentrations of all other KaiC states are set to zero. In the second stage, we simulate the autodephosphorylation reaction after KaiA pull-down, which corresponds to eliminating all free KaiA as well as KaiA-bound KaiC states from the simulation. The amount of KaiC lost in the pull-down experiment was not reported in the original experiment [29]. We therefore make the assumption that the amount of KaiC lost in the pull-down experiment in the simulation is exactly equal to that in the experiment for the purpose of computing the likelihood function.

Table 2.5: Priors incorporating KaiA on/off constraints

Parameter	Prior mean (μ)	Experimental measurements	Source
k_{a}	$\log k_{\mathrm{a,exp}}$	$k_{\mathrm{a,exp}} = 0.0279\ \mathrm{s}^{-1} \cdot \mu\mathrm{M}^{-1}$	[24]
k_{b}	$\log k_{\mathrm{b,exp}}$	$k_{\mathrm{b,exp}} = 0.0663\ \mathrm{s}^{-1}$	
$\Delta k_{\mathrm{b}}^{\mathrm{T,DP}}$, $\Delta k_{\mathrm{b}}^{\mathrm{T,TP}}$	$-\log \tau_{\mathrm{b,exp}}^{T} k_{\mathrm{b,exp}}$	$\tau_{\mathrm{b,exp}}^{\mathrm{T}} = 1.0\ \mathrm{s}$	[25]
$\Delta k_{\mathrm{b}}^{\mathrm{S,DP}}$, $\Delta k_{\mathrm{b}}^{\mathrm{S,TP}}$	$-\log \tau_{\mathrm{b,exp}}^{S} k_{\mathrm{b,exp}}$	$\tau_{\mathrm{b,exp}}^{\mathrm{S}} = 0.43\ \mathrm{s}$	
$\Delta k_{\mathrm{b}}^{\mathrm{D,DP}}$, $\Delta k_{\mathrm{b}}^{\mathrm{D,TP}}$	$-\log \tau_{\mathrm{b,exp}}^{D} k_{\mathrm{b,exp}}$	$\tau_{\mathrm{b,exp}}^{\mathrm{D}} = 0.26\ \mathrm{s}$	

Bayesian parameter estimation We directly fit numerically integrated ODEs to experimental data in the Bayesian parameter estimation framework [41]. The best fit model parameters, $\hat{\theta}$, are obtained from the maximum a posteriori estimator:

$$\hat{\theta} = \arg\max_{\theta} p(\theta|\mathcal{D}), \tag{2.10}$$

where $p(\theta|\mathcal{D})$ is the posterior distribution of the parameters θ, conditioned on the training data set $\mathcal{D}$. Using Bayes' theorem, the posterior distribution can be written as

$$p(\theta|\mathcal{D}) = \frac{\mathcal{L}(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}. \tag{2.11}$$

Here, $p(\theta)$ is the prior distribution, which represents subjective belief in the model parameters θ prior to experimental input; $\mathcal{L}(\mathcal{D}|\theta)$ is the likelihood function, which represents a probabilistic model of the experimental data set $\mathcal{D}$ given a particular model $\mathcal{M}$ (implicit in the formulas) that depends on the parameters θ; $p(\mathcal{D})$ is the evidence, which is analogous to the partition function in statistical mechanics. Note that the evidence $p(\mathcal{D})$ does not depend on the parameter choice, and is thus an irrelevant constant for the purpose of parameter estimation. The specific choices for the functional forms of the likelihood function and priors are discussed further below.

Model priors The priors for all model parameters used in Bayesian parameter estimation are given in Table 2.2. Here, the choice of the prior distributions is primarily motivated by the need for regularization (see below). In addition, as discussed above, the experimental measurements on KaiA binding kinetics are incorporated into the priors rather than the likelihood function (Table 2.5). Note that all the rate constants and their multiplicative factors are estimated in the log space (base 10). This ensures that all rate constants are positive.

ℓ^1 regularization As model complexity grows, the parameters become less constrained by the data. To address this problem, we impose sparsity on the state-dependent effects (i.e., the Δk factors) through ℓ^1 regularization [112]. This is accomplished in the Bayesian parameter estimation framework by using a Laplace prior centered at zero in the log parameter space (or one in the real space). Intuitively, the Laplace prior imposes sparsity by forcing the (marginalized) posterior distribution for each $\log \Delta k$ to peak at zero unless there is experimental evidence in the fitting data set to suggest otherwise. Since the Δks are multiplicative factors modifying the six basic rate constants, $\log \Delta k = 0$ implies that the state-dependent rate is identical to that of the base rate. This method is directly analogous to the use of the lasso estimator in the context of linear least squares models [112]. To see this, consider the Laplace distribution

$$p(\theta; b) = \frac{1}{2b} e^{-\|\theta\|_1 / b} \tag{2.12}$$

where $\|\theta\|_1$ is the ℓ^1-norm of θ. Then from (2.11) the negative log-posterior distribution becomes

$$-\ln p(\theta|\mathcal{D}) = -\ln \mathcal{L}(\mathcal{D}|\theta) + \lambda \|\theta\|_1 + \text{const.} \tag{2.13}$$

where $\lambda = 1/b$. In a linear model $Y = X\beta + \epsilon$ where Y is the response vector, X is the design matrix, β is the parameter vector, and ϵ is the error vector, the negative log-

likelihood function reduces to the sum of squares $\|Y - X\beta\|_2^2/N$, where N is the number of dependent variables. Thus, maximizing the posterior is equivalent to minimizing the sum of squares with an ℓ^1 penalty, which is the lasso estimator.

Likelihood function To determine the functional form of the likelihood function, we consider a kinetic experiment where measurements on some observables y are made at a set of time points $\{t_i\}$ with uncertainties $\{\sigma_i\}$. If we assume that the experimental errors σ_i are independent and normally distributed, then the likelihood function is given by

$$\mathcal{L}(\mathcal{D}|\theta) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} e^{-[y_{\text{exp}}(t_i) - y_{\text{model}}(t_i;\theta)]^2/2\sigma_i^2}. \tag{2.14}$$

In other words, the likelihood function gives the probability for observing a given data set provided that the model prediction is true. In practice, all posterior evaluations are done in the log space (base e) for numerical stability. Thus, taken together, (2.11) can be rewritten as,

$$\ln p(\theta|\mathcal{D}) = -\sum_i \frac{\left[y_{\text{exp}}(t_i) - y_{\text{model}}(t_i;\theta)\right]^2}{2\sigma_i^2} + \ln p(\theta) + \text{constant}. \tag{2.15}$$

For the sake of simplicity, we assume that there is a single global error, σ, for all (de)phosphorylation measurements, which is then estimated during fitting as a hyperparameter (see Table 2.2).

The choice of the Gaussian likelihood function applies to all (de)phosphorylation data sets, but not the hydrolysis constraint, which only provides an upper bound on the average hydrolysis rate per day [26]. Therefore, for the hydrolysis data a "half harmonic" is used

as the log-likelihood function:

$$\ln \mathcal{L}(\mathcal{D}|\theta) = \begin{cases} -\dfrac{\Big([\mathrm{ADP}]_{\mathrm{model}}(\theta) - [\mathrm{ADP}]_{\mathrm{exp}}\Big)^2}{2\sigma_{\mathrm{h}}^2}, & [\mathrm{ADP}]_{\mathrm{exp}} \leq [\mathrm{ADP}]_{\mathrm{model}} \\ 0, & 0 \leq [\mathrm{ADP}]_{\mathrm{model}} < [\mathrm{ADP}]_{\mathrm{exp}} \end{cases} \tag{2.16}$$

The total amount of ADP produced by the model during a phosphorylation reaction over 12 h, $[\mathrm{ADP}]_{\mathrm{model}}$, is given by the sum of all P_i production over time plus all ADP-bound KaiC states at $t = 12$ h.

The log-likelihood values from the appropriate phosphorylation, dephosphorylation, and hydrolysis reactions are added together to determine the log-likelihood of the data set for each given model parameter choice. Since the phosphorylation data set is much larger than the dephosphorylation data set, the fitting procedure tends to favor fitting the phosphorylation data set at the expense of fitting the dephosphorylation data set. To overcome this problem, the log-likelihood function for the dephosphorylation reaction is multiplied by a factor of 4 to increase the weight of the dephosphorylation data points.

Model fitting procedure To determine the posterior mode and the uncertainties associated with the estimate, we employ a heuristic combination of ensemble MCMC sampling and numerical optimization methods (Fig. S2.7A). This fitting procedure can be divided into four steps that are analogous to those in a genetic algorithm:

1. Initialization. An ensemble MCMC method evolves a set of random walkers (i.e., parameter sets) simultaneously; we thus begin by drawing 224 walkers from the prior distribution $p(\theta)$ and use these walkers for simulated annealing [113, 114]. In annealing, instead of sampling $p(\theta|\mathcal{D}) \propto \mathcal{L}(\mathcal{D}|\theta)p(\theta)$ (in the log space), a flattened distribution $\mathcal{L}(\mathcal{D}|\theta)^{\beta}p(\theta)$ is sampled with an annealing schedule of $\beta = 0.3$, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0. Note that instead of letting $\beta \to \infty$, the simulation ends with the target distribution at $\beta = 1.0$. Each temperature is sampled over 20,000 steps.

2. Selection. The fitnesses of the walkers are determined by their log-posterior values (equation 2.15). 10 walkers from the best 300 parameter sets sampled in the previous step are chosen and subjected to numerical optimization to find the nearby local maximums, which are then used to seed a sampling run in the next step. In the spirit of elitist selection, the best walker is always included for the next generation.

3. Recombination and mutation. The initial walkers for the sampling run are generated by adding a Gaussian noise $\mathcal{N}(0, \sigma I)$ to the 10 optimized walkers (here, I is the identity matrix and $\sigma = 0.001$), and the number of initial walkers centered around each optimized walker θ_j is given by

$$224 \frac{p(\theta_j|\mathcal{D})^\beta}{\sum_k p(\theta_k|\mathcal{D})^\beta}. \tag{2.17}$$

That is, the proportion of the initial walkers generated from each optimized walker is weighted by its posterior value with a temperature factor of $\beta = 0.6$; the temperature factor is chosen to allocate more walkers to optimized walkers with lower posterior values. The sampling run consists of 50,000 steps. Note that the purpose of the Gaussian noise is to ensure that the proposal distribution is valid for any pair of walkers for the ensemble MCMC method (see below), rather than to control the mutation strength, as is done in evolution strategy [115].

4. Termination. The best walkers from the sampling run are compared to the optimized walkers. If the best walkers do not escape to new local maximums with higher posterior values, then the procedure is terminated after an additional 50,000 sampling steps. If, however, new local maximums are discovered during sampling, the algorithm loops back to the selection step. This process is repeated until no better local maximum is discovered at the end of sampling. Unless otherwise specified, only the last 30,000 sampling steps (downsampled every 100 steps) are used for post-analysis.

In general, the number of walkers in ensemble MCMC needs to be larger than the number of free parameters; here the number 224 is chosen to optimize parallel performance on a local computer cluster (8 nodes $\times$ 28 CPU cores/node).

We found that this procedure outperformed either ensemble MCMC or numerical optimization by itself (compare Fig. S2.7A and B). For the full model we ran this procedure three times to assess the reproducibility of the fit (see Appendix for further discussion).

Markov chain Monte Carlo One major challenge in efficient MCMC sampling in systems biology is that the target distributions are often poorly scaled. In the context of ODE kinetic modeling, this means that different reaction rates and their associated uncertainties can be separated by several orders of magnitude. This is almost certainly true for the KaiA-KaiC subsystem because, among other things, the experimentally measured rates of KaiA binding and dissociation are much faster than the ATP hydrolysis rate of KaiC. Without a priori knowledge of the natural time scales, conventional MCMC schemes are inefficient in such sampling problems, because only very small displacements are accepted at appreciable rates. In this work we employ an ensemble MCMC method developed by [64]. The advantage of the Goodman-Weare algorithm is that it is affine invariant, which means that it performs equally well for isotropic and poorly scaled measures, providing that the two can be related by a linear transformation of the coordinate system. This appears to be the case for the present problem since the Goodman-Weare algorithm vastly outperforms a standard Metropolis-Hastings scheme with a (preconditioned) Gaussian proposal distribution (Fig. S2.7B).

In brief, the Goodman-Weare algorithm evolves an ensemble of walkers, rather than a single one. At each step, individual walker positions are updated sequentially. For a given walker θ_k at step τ, a walker θ_j is drawn randomly from the rest of the ensemble and a

new position, η, on the line connecting θ_k and θ_j is proposed by a "stretch move"

$$\eta = \theta_j + z\left[\theta_k(\tau) - \theta_j\right] \tag{2.18}$$

where z is a random number drawn from the distribution

$$Z \sim g(z;\alpha) = \begin{cases} 1/\sqrt{z}, & z \in [1/\alpha, \alpha] \\ 0, & \text{otherwise} \end{cases} \tag{2.19}$$

The "stretch factor" α is a tunable parameter that controls the step size. In an N-dimensional parameter space, the new walker η is accepted with the probability

$$q = \min\left(1, z^{N-1}\frac{p(\eta|\mathcal{D})}{p(\theta_k(\tau)|\mathcal{D})}\right) \tag{2.20}$$

which guarantees that the scheme obeys detailed balance and thus converges to the target distribution $p(\theta|\mathcal{D})$ as $\tau \to \infty$. Note that no derivative of the posterior distribution is required to draw from the proposal distribution. In this work we use $\alpha = 1.1$, which gives an average acceptance rate of 47% in steps 3 and 4 of the fitting procedure.

Numerical optimization The numerical optimization method used in this work is a modified version of Powell's method [116, 117]. Briefly, given an initial guess and direction set, which is usually the Cartesian coordinate set, Powell's method performs a line search to minimize the objective function, here $-\ln p(\theta|\mathcal{D})$, sequentially along each vector in the direction set. The direction set is then updated by replacing the direction of largest decrease in the objective function in the current iteration with the displacement vector from the estimated minimum at the beginning to that at the end of the line minimizations, provided that certain technical conditions are met to avoid the build-up of linear dependence in the direction set. This process is repeated until a convergence threshold is met. Note that unlike the original method, the modified Powell's method does

not guarantee that the vectors in the direction set are mutually conjugate.

Similar to the Goodman-Weare algorithm, Powell's method is derivative-free. For the current system Powell's method converges faster than the Nelder-Mead method [118], another commonly used derivative-free method, although the Nelder-Mead method appears less prone to becoming trapped in local metastable states (Fig. S2.7C).

Software implementation The fitting procedure is implemented in an in-house Python script that interfaces with several existing Python modules: numerical integration of the model ODEs is done using the odespy package [119] with the BDF method; the Goodman-Weare algorithm is implemented in emcee (version 2.2.1) [65]; Powell's method and the Nelder-Mead method are implemented in scipy (version 1.2.1) [120]. The derivative evaluation step in ODE integration is accelerated using numba [121], and the script is parallelized using mpi4py 2.0.0 [122–124].

The most computationally expensive step in the fitting procedure is the MCMC sampling, because each move requires multiple ODE evaluations to compute the posterior function. With 224 walkers and 8 nodes (each with 28 Intel E5-2680v4 2.4GHz cores), the speed of MCMC sampling is 46,000 steps/hour.

Model comparison In the preceding sections all definitions of probability distributions implicitly assume that there is a model $\mathcal{M}$ with a well-defined functional form, whose parameters θ need to be determined. For the sake of model comparison, we make this assumption explicit and rewrite (2.11) as

$$p(\theta|\mathcal{D},\mathcal{M}) = \frac{\mathcal{L}(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})}. \tag{2.21}$$

To compare two models $\mathcal{M}_i$ and $\mathcal{M}_j$, we need to compare the posterior probabilities for each model, usually in the form of their ratios

$$\frac{p(\mathcal{M}_i|\mathcal{D})}{p(\mathcal{M}_j|\mathcal{D})} = \frac{p(\mathcal{D}|\mathcal{M}_i)}{p(\mathcal{D}|\mathcal{M}_j)} \frac{p(\mathcal{M}_i)}{p(\mathcal{M}_j)}. \tag{2.22}$$

Assuming that we have no prior preference for any model, the ratio becomes the Bayes factor

$$B_{ij} = \frac{p(\mathcal{D}|\mathcal{M}_i)}{p(\mathcal{D}|\mathcal{M}_j)}, \tag{2.23}$$

which we adopt as the metric for model comparison.

The primary difficulty in computing the Bayes factor is thus estimating the evidence, or the marginal likelihood function, for each $\mathcal{M}_i$. There are several methods for computing the evidence [74]. Here we derive a formula compatible with the ensemble MCMC scheme that is directly analogous to free energy perturbation [125]. For the sake of simplicity, we drop the model index i from this point on. First, note that for a given model $\mathcal{M}$,

$$p(\mathcal{D}|\mathcal{M}) = \int \mathcal{L}(\mathcal{D}|\theta, \mathcal{M}) p(\theta|\mathcal{M}) \, d\theta = \langle \mathcal{L}(\mathcal{D}|\theta, \mathcal{M}) \rangle_{p(\theta|\mathcal{M})} \tag{2.24}$$

where the first equality follows from the law of total probability and the second equality assumes that the prior $p(\theta|\mathcal{M})$ is normalized (as a probability density function of θ). Equation (2.24) suggests that the marginal likelihood function can be computed by estimating the average of the likelihood function $\mathcal{L}(\mathcal{D}|\theta, \mathcal{M})$ against the prior. Using MCMC to estimate this integral is inefficient since there is very little overlap between the likelihood function and the prior for the models of interest. Instead, we define

$$q_\lambda(\theta) = \mathcal{L}(\mathcal{D}|\theta, \mathcal{M})^\lambda p(\theta|\mathcal{M}) \quad \text{and} \quad Z_\lambda = \int q_\lambda(\theta) \, d\theta$$

for $0 \leq \lambda \leq 1$ and then note that (2.24) can be recast as

$$p(\mathcal{D}|\mathcal{M}) = \frac{Z_1}{Z_0} = \left(\frac{Z_{\lambda_0}}{Z_{\lambda_1}} \frac{Z_{\lambda_1}}{Z_{\lambda_2}} \cdots \frac{Z_{\lambda_{N-1}}}{Z_{\lambda_N}} \right)^{-1} \tag{2.25}$$

for $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_N = 1$, and the λs are chosen to allow for sufficient overlap between successive $q_\lambda(\theta)$s. Each fraction in (2.25) is given by

$$\frac{Z_{\lambda_{n-1}}}{Z_{\lambda_n}} = \frac{\int \mathcal{L}(\mathcal{D}|\theta,\mathcal{M})^{\lambda_{n-1}-\lambda_n} q_{\lambda_n}(\theta)\, d\theta}{\int q_{\lambda_n}(\theta)\, d\theta} = \left\langle \mathcal{L}(\mathcal{D}|\theta,\mathcal{M})^{\lambda_{n-1}-\lambda_n} \right\rangle_{q_{\lambda_n}} \tag{2.26}$$

Therefore,

$$p(\mathcal{D}|\mathcal{M}) = \prod_{n=1}^{N} \left\langle \mathcal{L}(\mathcal{D}|\theta,\mathcal{M})^{\lambda_{n-1}-\lambda_n} \right\rangle_{q_{\lambda_n}}^{-1} = \prod_{n=1}^{N} \left\langle e^{\ln q_{\lambda_{n-1}}(\theta) - \ln q_{\lambda_n}(\theta)} \right\rangle_{q_{\lambda_n}}^{-1} \tag{2.27}$$

where the averages $\langle \cdot \rangle_{q_{\lambda_n}}$ can be approximated with MCMC. Equation (2.27) is a version of the free energy perturbation formula. Note that (2.27) requires that the likelihood function $\mathcal{L}(\mathcal{D}|\theta,\mathcal{M})$ be properly normalized (as a probability density function of $\mathcal{D}$), but does not require the prior $p(\theta|\mathcal{M})$ to be normalized, as any missing normalization constant cancels out in each term of the product.

For each simplified model in Table 2.1 and Table S2.6, the ensemble of walkers from the last time step of the model fitting procedure is used to initialize an MCMC sampling run with $\lambda = 1.00$. The lambda value is reduced by 0.01 at each subsequent stage until λ reaches 0.01. Each stage is sampled for 2,000 time steps using the Goodman-Weare ensemble sampler [64]. Only the last 1,000 time steps from each stage is used to compute the ensemble average in (2.27). The Bayes factors are then computed as the ratios of the evidence for the full model to each simplified model.

Refitting The steady-state KaiC phosphorylation measurements (Fig. 2.3B and Fig. S2.12B) are fit to the full model using Powell's method, starting from the best fit based on

the training data set. The priors on the kinetic parameters (Table 2.2) are centered on the best fit values, so that the refit model can be interpreted as the "minimal" perturbation to the best fit that enables agreement with the steady-state measurement.

Curve fitting The experimentally determined stimulus-response relations for KaiC S431A (Fig. 2.3C and F) are fit to the simple inhibitor ultrasensitivity scheme described in Box 5 of [56]. Using their notation, the amount of phosphorylated protein substrate ($\%XP$) as a function of kinase concentration ($[K]$) is given by

$$\%XP([K]) = P_{\max}\frac{K_1[I] + K_1K_2 - K_1[K] + 2K_2[K] - K_1\sqrt{[I]^2 + 2(K_2 - [K])[I] + (K_2 + [K])^2}}{2K_1[I] - 2(K_1 - K_2)(K_1 + [K])} + b \quad (2.28)$$

Here, $P_{\max}$, $[I]$, K_1, K_2, and b are free model parameters. Unlike the Hill function, EC50 is not an explicit parameter of (2.28) and thus needs to be determined numerically. Note that (2.28) can be reduced to a right-shifted hyperbolic function as $K_2 \to 0$:

$$\%XP([K]) = P_{\max}\frac{[I] - [K] - |[I] - [K]|}{2[I] - 2(K_1 + [K])} + b, \quad (2.29)$$

which is equivalent to a threshold-hyperbolic stimulus-response function,

$$\%XP([K]) = \frac{P_{\max}([K] - [I])}{K_1 + ([K] - [I])}H([K] - [I]) + b, \quad (2.30)$$

where H is the unit step function.

Stimulus-response relations are fit using the NonlinearModelFit function in Mathematica 12.0. We stress here that the curve fits are purely phenomenological and are thus not intended to be interpreted in terms of the biochemical assumptions underlying the model.

Numerical method for the Phong model All numerical simulations of the modified Phong model (Fig. 2.4E and Fig. S2.10E and F) were carried out in Mathematica 12.0 using the NDSolve function. To determine the period of the model for a given parameter set, the model was simulated for 200 h and the first 100 h was discarded to eliminate transient responses. The NDSolve function returns ODE solutions as InterpolatingFunction objects, which were converted to time series sampled at every 0.1 h for subsequent analyses. The troughs in the time series of the total KaiC phosphorylation level (i.e. $[\mathrm{C^P}] = [\mathrm{C^T}] + [\mathrm{C^S}] + [\mathrm{C^D}] + [^{\mathrm{B}}\mathrm{C^S}] + [^{\mathrm{B}}\mathrm{C^D}]$) were determined using the FindPeaks function, and the average of the trough-to-trough time was taken to be the period of the model. The model was considered non-oscillatory if one of the following conditions was met:

1. The standard deviation of the $[\mathrm{C^P}]$ time series is less than 0.1 µM (weak oscillation).
2. The standard deviation of the trough amplitude is larger than 0.001 µM (unstable limit cycle).
3. The number of troughs is less than 3 (abnormally long period).

2.5.2 Experimental methods

Protein expression and purification KaiA, KaiB, KaiB-FLAG, and KaiC were expressed and purified as previously described [28] with two modifications to the protocol: anion exchange chromatography was performed using HiTrap Q columns (GE Healthcare), and KaiC was purified using Ni-NTA affinity chromatography followed by size-exclusion chromotography, omitting the anion exchange step. The expression, purification, and 6-iodoacetofluorescein (6-IAF) labeling of KaiB K25C mutant as a fluorescence reporter in the plate reader assay is described in [33]. All mutants of KaiC were constructed using QuikChange II XL Site-Directed Mutagenesis Kit (Agilent). For the KaiC-AA and -EE mutants, the His-tags were not cleaved during purification; this ensures that these mutant

proteins have shifted mobility in SDS-PAGE, allowing their bands to separate from those of KaiC S431A.

The U-[^{15}N] labeled N-terminal (residues 1–135) and C-terminal (residues 181–284) domains of KaiA were expressed in BL21(DE3) *E. coli* (Novagen) in minimal (M9) media supplemented with ^{15}N-enriched NH_4Cl. For the expression of the C-terminal domain, M9 media enriched with ^{15}N-NH_4Cl was prepared using 98% deuterated water (D_2O). The proteins were purified by Ni-NTA affinity chromatography followed by size-exclusion chromatography using Superdex 75 1660 prep grade column, as described previously [20, 66, 67]. N-terminal KaiA eluted as ∼15 kDa monomer [126], while C-terminal KaiA eluted as a ∼23 kDa homodimer [127].

GFP was expressed as a N-terminal 6xHis-tag fusion from the pET28a plasmid in the BL21 (DE3) strain of *E. coli*. Harvested cells were sonicated for lysis and clarified lysate was loaded onto a HisTrap FF column (GE Healthcare). The His tag was cleaved by overnight incubation at 4 °C with SUMO protease (Invitrogen), after which the sample was loaded again onto a HisTrap FF column to recover the cleaved products. The cleaved proteins were further purified on a 5 mL HiTrap Q HP column (GE Healthcare) and then a Superdex 200 10/300 GL size-exclusion column. The eluted fractions were concentrated in a sample buffer (20 mM HEPES [pH 7.4], 150 mM KCl, 2.5 mM $MgCl_2$, 2 mM DTT), aliquoted, and snap frozen in liquid nitrogen for storage in –80 °C.

In vitro clock reactions All in vitro clock reactions were done in the standard reaction buffer (20 mM Tris-HCl [pH 8], 150 mM NaCl, 5 mM $MgCl_2$, 0.5 mM EDTA , 10% glycerol, 50 µg/mL Kanamycin). KaiC concentration was 3.5 µM in all experiments unless otherwise specified; KaiB concentration was 3.5 µM for all oscillatory reactions, and 6-IAF-labeled KaiB K25C concentration was 0.2 µM for plate reader assays. KaiA concentration and %ATP were determined by each individual experiment, while the total nucleotide concentration (i.e., [ATP] + [ADP]) was held constant at 5 mM. Phosphoryla-

tion kinetics was resolved using SDS-PAGE on 10% acrylamide gels (37.5:1 acrylamide:bis-acrylamide) run for 4.5 h at 30 mA constant current at 12 °C; the gels were stained in SimplyBlue SafeStain (Invitrogen) and then imaged using Bio-Rad ChemiDoc Imager. The oscillatory reactions (Fig. 2.2D and Fig. S2.10C) were also monitored using the plate reader assay described in [33].

NMR spectroscopy A Bruker 600 MHz AVANCE III spectrometer equipped with a TCI cryoprobe was used for all of the NMR experiments of the N- and C-terminal domains of KaiA (Fig. S2.6). Chemical shifts were referenced to internal 2,2-dimethyl-2-silapentane-5-sulfonate (10 µM). Data were processed using NMRPipe and visualized using NMRDraw [128]. NMR samples were prepared with 100 µM monomer concentration of protein in 20 mM Tris-HCl [pH 8], 150 mM NaCl, 5 mM $MgCl_2$, and 5% D_2O buffer. All experiments were performed at 30 °C. Samples were incubated with 1mM ATP or ADP, when needed, for 30 minutes before spectral measurement.

Immunoprecipitation Immunoprecipitation of KaiB-FLAG and associated protein complexes in a clock reaction (Fig. 2.4A) was done as previously described [28] using monoclonal anti-FLAG M2 antibody (Sigma-Aldrich, product number F1804), and elution with 3xFLAG peptide. The supernatant was analyzed by SDS-PAGE on 4–20% Criterion TGX Stain-Free Precast Gels (BioRad) and stained with SYPRO Ruby (BioRad). 1.5 µM GFP was added to the reaction mixture at the beginning of the time course and served as an internal standard to correct for changes in protein concentration due to handling. The relative supernatant KaiA concentration was determined as a ratio of KaiA band intensity in each lane to the GFP band intensity and is normalized as percentage of the largest ratio in the time course (Fig. S2.10D).

2.6 Appendix

2.6.1 Additional biochemistry of the KaiC (de)phosphorylation reactions

In this work we construct a general model of the KaiA-KaiC subsystem based on a set of assumptions of basic clock biochemistry; that is, KaiC is an ATPase and a reversible phosphotransferase with two phosphorylation sites at S431 and T432, while KaiA is a nucleotide-exchange factor that promotes the exchange of bound ADP for ATP in CII nucleotide-binding pockets, and may directly stimulate hydrolysis. Through model fitting, we demonstrate that this set of assumptions is sufficient to explain the (de)phosphorylation kinetics of KaiC and its dependence on %ATP and [KaiA]. Our results, however, do not imply that the model includes all possible biochemical mechanisms; in this section and the next we briefly discuss aspects of KaiC that we do not consider in the model.

First, the current model does not account for the CI domain. The CI domain of KaiC is required for the hexamerization of *S. elongatus* KaiC [110, 129] and its ATPase activity is necessary to allow the binding of KaiB needed to transition into the dephosphorylation phase of the oscillation [28, 77]. However, the ATPase activity of the CI domain has no significant effect on the kinetics of the CII (de)phosphorylation reactions [28], and in the current study we are not concerned with KaiB-dependent processes. Therefore the current model does not keep track of the CI hydrolysis state or any allosteric coupling between the CI and CII domains.

Second, the model does not explicitly consider the function of Mg^{2+}. The presence of Mg^{2+} is required for the assembly of the KaiC hexamer [76, 110], and computational analyses indicate that release of Mg^{2+} independent of the bound nucleotide is highly energetically unfavorable [63]. Given these results, we assume that Mg^{2+} and nucleotide cannot act independently of each other, and the model implicitly assumes that each bound nucleotide is always in complex with a Mg^{2+} ion. A recent study, however, shows that

depletion of Mg^{2+} promotes KaiC autophosphorylation, especially in buffers that lack EDTA, [130]. Moreover, some structures of KaiC have two Mg^{2+} ions modeled in the nucleotide-binding pocket, which has been interpreted to mean that KaiC kinase activity relies on a two-metal-ion phosphotransfer mechanism [131]. Currently, the functions of Mg^{2+} have not been characterized kinetically or mechanistically at a level necessary to constrain its role in a molecularly detailed model.

2.6.2 The hexameric structure of KaiC

The current model does not consider any hexameric effects. There is evidence to suggest that intersubunit interaction regulates KaiC autokinase activity [132] as well as the ultrasensitive dependence of KaiBC complex formation on the KaiC hexamer phosphoform composition [21]. However, explicitly accounting for the hexameric nature of KaiC, as in [133] or [21], would lead to a significant increase in the number of model parameters, which likely cannot be constrained by available data and makes interpretation of the model difficult. Therefore, we only keep track of the states of KaiC subunits considered independently, and the rate constants should be considered averages over hexameric background configurations, weighted by their nonequilibrium state populations.

This leads to two simplifications concerning the KaiA-KaiC interactions. The first issue relates to the stoichiometry of KaiAC complexes. During phosphorylation, KaiA dimers bind to KaiC hexamers likely with either a 1:1 or 2:1 stoichiometry [72, 134]. Because the model does not consider the hexameric structure of KaiC, this stoichiometry is not enforced explicitly, and each KaiC monomer can bind independently to KaiA.

The second issue relates to the effect of the phosphorylation state of the entire hexamer on KaiA (un)binding kinetics. KaiA can inhibit KaiB binding to KaiC [21], even though the KaiA and KaiB binding sites are on opposing sides of KaiC. This has been interpreted as a result of cooperative allosteric transition between a kinase mode of KaiC, stabilized by KaiA binding and the T phosphoform, and a phosphatase mode of KaiC,

stabilized by KaiB binding and the S phosphoform [21]. A possible implication of this proposed mechanism is that KaiC hexamers in the phosphatase mode may have uniformly diminished affinity for KaiA at the CII interface, regardless of the phosphorylation state of the subunits. Given that the current model is trained using primarily the phosphorylation data set, the predicted KaiA dwell time (Fig. S2.8B) and dissociation constant (Fig. 2.1D) likely reflect the property of KaiC subunits in the kinase mode, whereas experiments with phosphomimetic mutants mimicking the S and D phosphorylation states presumably probe the system kinetics in the phosphatase mode. Indeed, in the KaiC-EE titration experiment (Fig. 2.3F right), the presence of KaiC-EE has virtually no effect on the KaiC S431A stimulus-response curve, which may be partly due to the fact that the KaiC-EE hexamers are in the phosphatase mode and may therefore have additionally reduced affinity for KaiA, a condition not considered in our model.

2.6.3 Correlation structure in the posterior distribution

As discussed in the literature [55], often ratios of parameters are better constrained than the parameters themselves. The parameter pairs that have a correlation coefficient larger than 0.9 in log space are shown in Fig. S2.9A–C. Such correlations typically reflect that thermodynamic, rather than kinetic, properties of the system are constrained. These include the free energy of phosphotransfer between the S and D phosphoforms (Fig. S2.9A) and the free energy for KaiA binding to the ATP-bound states of KaiC (Fig. S2.9B; compare with Fig. 2.1D). Interestingly, there is a linear relation among $k_{\mathrm{TP}}^{\mathrm{A,T}}$, $k_{\mathrm{d}}^{\mathrm{A,TU}}$, and $k_{\mathrm{b}}^{\mathrm{T,\,DP}}$ (Fig. S2.9C); this implies that the data constrain the flux out of the ${}^{\mathrm{A}}\mathrm{C}_{\mathrm{DP}}^{\mathrm{T}}$ state, but the exact pathway is underdetermined.

More generally, we characterize the extent to which the model parameters, or linear combinations thereof, are constrained by the data using the principal components of the posterior distribution; that is, the eigenvectors of the covariance matrix from MCMC sampling. The eigenvalues of the covariance matrix span multiple orders of magnitude

with no obvious gap (Fig. S2.9D left), except for the stiffest direction (Fig. S2.9G), which is almost entirely aligned with σ^2, the global error hyperparameter. In addition, the directions of the principal components are in general not aligned with the directions of the bare coordinates (Fig. S2.9D right), and there is no obvious interpretation for the directions of most of the principal components (Fig. S2.9G). These features of the ensemble indicate that the model is "sloppy" [55], and many model parameters are poorly constrained by the data. Nevertheless, as we demonstrate in Results, the model can still be used to make consistent predictions because the variabilities in the ensemble of parameter sets obtained from MCMC sampling align with the softest degrees of freedom of the posterior distribution.

2.6.4 Convergence of the model fit

We assessed the quality of the fit in three ways. First, we repeated the full procedure three times to assess reproducibility of the fitting procedure (Fig. S2.5B). The three independent runs yielded marginalized posterior distributions that are remarkably consistent and tightly constrained for some parameters but diverge over several orders of magnitude for others. The best fits from the three runs have log posterior values of 720, 714, and 705, respectively. Unless otherwise specified, in this work we base our analyses on the run that produced the best fit with the highest posterior value. The ruggedness of the posterior distribution demonstrates that given the model and training data set, the parameter estimation problem is far from the asymptotic (i.e., large sample) regime. Moreover, in our fitting procedure the vast majority of the walkers from the annealing step are discarded for the sake of improving the MCMC sampling efficiency (the acceptance rate is $\leq 7\%$ without pruning). Given the presence of multiple local maxima, this choice likely resulted in an underestimation of the uncertainties in parameter values.

Second, we compared the model predictions with a test data set that probed the phosphorylation reaction at two non-standard KaiC concentrations (Fig. S2.8E). The fit

quality on the test data set is somewhat worse compared to the training set (compare with Fig. 2.1B). In particular, the model overestimates the D phosphoform concentration at 1.75 µM KaiC and underestimates the T and D phosphoform concentrations at 7 µM KaiC. This result suggests some degree of overfitting. This, however, is not a significant issue because we base our conclusions on the ensemble of walkers rather than the behavior of the best fit.

Lastly, we assessed the convergence of the MCMC simulations using the integrated autocorrelation times for the 48 principal components of the posterior distribution (Fig. S2.9E). The autocorrelation times for the largest and smallest principal components are 8,500 and 3,300 steps, respectively, which gives rough estimates of the times between independent samples for the slowest and fastest degrees of freedom. These estimates are far shorter than the length of the final MCMC runs in the fitting procedure (100,000 steps). We also checked the autocorrelation time for the KaiA binding affinities (Fig. S2.9F), which is within the bound given by the principal components.

2.6.5 KaiA function cannot be solely explained in terms of nucleotide exchange

Because of the generality of the model, the function of KaiA is not restricted to that of a nucleotide-exchange factor. In particular, the phosphotransfer and ATP hydrolysis rates are allowed to depend on KaiA binding. There is some experimental evidence to support such effects—KaiA binding inhibits dephosphorylation [135] and the addition of KaiA increases the ATPase activity of KaiC [26, 136]. However, the biochemical mechanisms underlying these effects are not clear from the experiments; for example, does KaiA increase KaiC ATPase activity by reconfiguring the KaiC active site, or does KaiA binding indirectly promote ATP hydrolysis by shifting the KaiC population towards phosphoforms that have high ATPase activity?

To test whether KaiA binding has a direct effect on KaiC catalytic activities, we

construct simplified models where hydrolysis and/or phosphotransfer are/is independent of KaiA binding and compare the resulting models to the full model using the Bayes factor (Table S2.6). We find that decoupling either phosphotransfer (model –P) or hydrolysis (model –H) from KaiA binding decreases the evidence for the simplified models, but the effects are weak, especially in comparison to a model where both classes of reactions are decoupled from KaiA binding (model –P,–H). These results indicate that the function of KaiA cannot be solely explained by nucleotide exchange, but we cannot conclusively distinguish between models –P and –H.

2.6.6 The experimental data admit two S phosphorylation pathways

We analyzed a random selection of 500 walkers to understand the implications of their variations for the mechanisms of ordered phosphorylation. To simplify the analysis, we converted each walker to two single-site models in which either S431 or T432 was available for phosphorylation but not both. We asked how important each reaction rate constant is to the overall T and S phosphoform concentrations in the single-site models using the relative first-order sensitivities computed at the standard reaction condition (i.e., 100% ATP with 1.5 µM KaiA). We focus on the initial phosphorylation rates because the steady-state rates are determined by balances of many contributing processes, making them harder to interpret. The parameter sensitivities at $t = 1$ h are used as proxies for the sensitivities of the initial phosphorylation rates.

Because each single-site model has 18 parameters, there are 36 sensitivities for the two phosphoforms. To characterize this high-dimensional space, we used spectral clustering (Fig. S2.11A). Overall, the parameter sensitivities are much more constrained by the data for the T phosphoform than the S phosphoform, which is unsurprising given the relatively low concentrations of the S phosphoform under all experimental conditions in the training data. The clustering furthermore indicates that there are at least two plausible kinetic ordering mechanisms, which differ primarily in terms of the phosphorylation pathways

taken by the S phosphoform (Fig. S2.11A and B). In both clusters, the U $\rightarrow$ T transition is most sensitive to $k_{\mathrm{p}}^{\mathrm{UT}}$ (i.e., the phosphorylation rate in the absence of KaiA).

In the first cluster (319 parameter sets), the U $\rightarrow$ S transition is most sensitive to $k_{\mathrm{p}}^{\mathrm{A,US}}$ in the presence of KaiA. This is primarily because $k_{\mathrm{p}}^{\mathrm{US}}$ is very small in this cluster relative to $k_{\mathrm{p}}^{\mathrm{A,US}}$ (Fig. S2.11C); however, since, in this cluster, the U $\rightarrow$ S transition is dominated by the KaiA-bound states, the S phosphoform has negative sensitivity to the KaiA dissociation rate constant $k_{\mathrm{b}}^{\mathrm{U,TP}}$. This suggests that in the first cluster, KaiA (un)binding to the U phosphoform is important in determining the initial phosphorylation rate of S. The best fit belongs to this first cluster.

The phosphorylation pathway suggested by the second cluster (181 parameter sets) is more complex. In this cluster, the U$\rightarrow$S transition is mostly independent of KaiA, similar to the U$\rightarrow$T reaction. However, the S phosphoform is limited by the dephosphorylation reaction $k_{\mathrm{d}}^{\mathrm{SU}}$, which is much faster than the corresponding phosphorylation rate (Fig. S2.11C). In addition, the S phosphoform is sensitive to the rate constant for KaiA binding, $k_{\mathrm{a}}^{\mathrm{S,DP}}$, which is important for facilitating nucleotide exchange for the ADP-bound S phosphoform, but tends to be slower in the second cluster (Fig. S2.11C). Therefore, faster dephosphorylation and slower KaiA binding is important for determining the initial S phosphorylation rate in the second cluster.

A comparison of Fig. S2.11C with Fig. S2.5B (blue distributions) shows that the two clusters correspond to the bimodal posterior distributions for the rate constants $k_{\mathrm{p}}^{\mathrm{US}}$, $k_{\mathrm{p}}^{\mathrm{A,US}}$, $k_{\mathrm{d}}^{\mathrm{SU}}$, and $k_{\mathrm{a}}^{\mathrm{S,DP}}$. The two clusters, however, do not cleanly separate along the two modes of $k_{\mathrm{a}}^{\mathrm{S,TP}}$ and $k_{\mathrm{b}}^{\mathrm{S,TP}}$; the kinetic significance of this bimodal distribution is unclear.

As discussed above, the posterior distribution is fairly rugged and thus the fitting procedure is not fully reproducible over independent runs. As a result, there are likely multiple potential kinetic ordering mechanisms that remain unexplored through this analysis. Regardless, the analysis suggests that kinetic ordering is likely a result of an interplay between (de)phosphorylation and KaiA (un)binding kinetics, rather than purely

the product of equilibrium free energies of phosphotransfer.

2.6.7 Comparison to the Paijmans model

Among all the computational work on the Kai oscillator, the model most similar to the current work in terms of the treatment of the KaiA-KaiC subsystem is that by [40], although the latter is a full oscillator model including KaiB, the CI domain, and the allosteric transition between the active (i.e., phopshorylation phase) and inactive (dephosphorylation phase) KaiC conformations. The Paijmans model and the full model in this work are both molecularly detailed, and describe how phosphotransfer, ATP hydrolysis, KaiA (un)binding, and nucleotide exchange reactions control the phosphoform and nucleotide-bound states of KaiC. However, there are some significant differences between these two models, which we examine below.

The Paijmans model is more general than this work in two ways. First, the Paijmans model explicitly considers the hexameric nature of KaiC. There is no intersubunit coordination of phosphotransfer in the Paijmans model, but it explicitly considers the binding of one KaiA dimer to a KaiC hexamer, which is assumed to uniformly accelerate nucleotide exchange in all six subunits. In this work, however, we do not consider the hexameric states of KaiC, and each KaiC monomer is allowed to bind to a KaiA dimer independently. In this way the affinities and kinetics of KaiA binding in this work may not be directly comparable to those in the Paijmans model. Second, the Paijmans model allows for the exchange of bound ATP for ADP, such that KaiA accelerates the exchange rates of both ATP and ADP while leaving the binding affinity unchanged. In our model, however, we assume that there is no exchange of bound ATP for ADP, effectively assuming that the affinity of KaiC for ATP is infinite (i.e., $K^{\mathrm{CII}}_{\mathrm{ATP/ADP}} = 0$ in the Paijmans model terminology). The treatments of nucleotide exchange in both models are otherwise similar, in that both assume that the ATP/ADP on-rates are identical, that the apo state is in a quasi-steady state, and that there is no KaiA-independent nucleotide

exchange.

The current work goes beyond the Paijmans model in the following ways. First, we determine the rate constants under the framework of Bayesian parameter estimation, which enables more rigorous uncertainty quantification, while the parameters in the Paijmans model were hand-tuned to reproduce selected experimental observations. Second, for simplicity the Paijmans model does not consider any possible coupling between ATP hydrolysis and KaiC phosphorylation states, between nucleotide exchange and KaiC phosphorylation states, between ATP hydrolysis and KaiA binding, or between KaiC nucleotide-bound states and KaiA binding. Although many of the state-dependent effects are not fully constrained by data in this work, as we describe in Results, the ultrasensitivity in KaiC phosphorylation depends critically on the coupling between KaiC nucleotide-bound states and KaiA binding. It is an open question whether a model that lacks such effects but explicitly accounts for the hexameric nature of KaiC can generate ultrasensitivity.

The difference in the treatment of KaiA binding affinity implies that some detailed balance conditions are incompatible between the two models. In the Paijmans model, the binding affinity of KaiA to KaiC hexamers during the phosphorylation phase depends on the phosphoform composition of the subunits, and each subunit i in the phosphoform X_i other than U contributes an additive factor of $\delta g_{\mathrm{bind}}^{\mathrm{CII{\cdot}KaiA}}(\mathrm{X}_i)$ to the changes in KaiA binding free energy, $\Delta G_{\mathrm{bind}}^{\mathrm{CII{\cdot}KaiA}}$. Due to detailed balance, the fact that KaiA binds to different KaiC phosphoforms with differential affinities implies that KaiA binding changes the free energy of phosphotransfer [see eq. 8 in [40]]. This condition is also present in our model, but is complicated by the nucleotide-bound states of KaiC. Using the multiplicative-factor parametrization scheme (see Materials and Methods), the detailed

balance conditions in this work can be related to those in the Paijmans model by

$$\delta g_{\mathrm{bind}}^{\mathrm{CII{\cdot}KaiA}}(\mathrm{T}) - \delta g_{\mathrm{bind}}^{\mathrm{CII{\cdot}KaiA}}(\mathrm{U}) = -kT \ln \frac{\delta k_{\mathrm{a}}^{\mathrm{T,DP}}}{\Delta k_{\mathrm{b}}^{\mathrm{T,DP}}} \tag{2.31}$$

$$\delta g_{\mathrm{bind}}^{\mathrm{CII{\cdot}KaiA}}(\mathrm{S}) - \delta g_{\mathrm{bind}}^{\mathrm{CII{\cdot}KaiA}}(\mathrm{U}) = -kT \ln \frac{\delta k_{\mathrm{a}}^{\mathrm{S,DP}}}{\Delta k_{\mathrm{b}}^{\mathrm{S,DP}}} \tag{2.32}$$

$$\delta g_{\mathrm{bind}}^{\mathrm{CII{\cdot}KaiA}}(\mathrm{D}) - \delta g_{\mathrm{bind}}^{\mathrm{CII{\cdot}KaiA}}(\mathrm{T}) = -kT \ln \frac{\Delta k_{\mathrm{a}}^{\mathrm{D,DP}}}{\Delta k_{\mathrm{b}}^{\mathrm{D,DP}}} \frac{\Delta k_{\mathrm{b}}^{\mathrm{T,TP}}}{\delta k_{\mathrm{a}}^{\mathrm{T,TP}}} \tag{2.33}$$

$$\delta g_{\mathrm{bind}}^{\mathrm{CII{\cdot}KaiA}}(\mathrm{D}) - \delta g_{\mathrm{bind}}^{\mathrm{CII{\cdot}KaiA}}(\mathrm{S}) = -kT \ln \frac{\Delta k_{\mathrm{a}}^{\mathrm{D,DP}}}{\Delta k_{\mathrm{b}}^{\mathrm{D,DP}}} \frac{\Delta k_{\mathrm{b}}^{\mathrm{S,TP}}}{\delta k_{\mathrm{a}}^{\mathrm{S,TP}}} \tag{2.34}$$

In general, this set of equations are inconsistent. That is, one cannot express the $\delta g_{\mathrm{bind}}^{\mathrm{CII{\cdot}KaiA}}(\mathrm{X})$s in the Paijmans model in terms of the Δks in our model. The only condition under which these equations can be made consistent is when

$$\frac{\delta k_{\mathrm{a}}^{\mathrm{T,DP}}}{\Delta k_{\mathrm{b}}^{\mathrm{T,DP}}} = \frac{\delta k_{\mathrm{a}}^{\mathrm{T,TP}}}{\Delta k_{\mathrm{b}}^{\mathrm{T,TP}}} \quad \text{and} \quad \frac{\delta k_{\mathrm{a}}^{\mathrm{S,DP}}}{\Delta k_{\mathrm{b}}^{\mathrm{S,DP}}} = \frac{\delta k_{\mathrm{a}}^{\mathrm{S,TP}}}{\Delta k_{\mathrm{b}}^{\mathrm{S,TP}}}$$

that is, when the nucleotide-bound states have no effect on KaiA binding affinities to the T and S phosphoforms.

2.6.8 Phenomenological modifications to the Phong model

In the model of [28], KaiA sequestration is determined by the equation

$$[\mathrm{A}]_{\mathrm{active}} = \max(0, [\mathrm{A}]_{\mathrm{total}} - m[{}^{\mathrm{B}}\mathrm{C}^{\mathrm{S}}] - n[{}^{\mathrm{B}}\mathrm{C}^{\mathrm{D}}]). \tag{2.35}$$

Here, $[\mathrm{A}]_{\mathrm{total}}$ is the total KaiA concentration; $[{}^{\mathrm{B}}\mathrm{C}^{\mathrm{S}}]$ and $[{}^{\mathrm{B}}\mathrm{C}^{\mathrm{D}}]$ are the concentrations of the S and D phosphoforms in complex with KaiB, respectively; m and n are model parameters describing the binding stoichiometries between KaiA and the KaiBC complex. In this way, the KaiA binding affinity to the inhibitory complex is effectively infinite, with no KaiA dissociation until $[{}^{\mathrm{B}}\mathrm{C}^{\mathrm{S}}]$ and $[{}^{\mathrm{B}}\mathrm{C}^{\mathrm{D}}]$ drop below a threshold. To make the

representation of KaiA sequestration more realistic, we introduce a KaiA dissociation constant K_D (Fig. 2.4C). Similar to the treatment in [88], we assume that the KaiA sequestration reaction is in a quasi-equilibrium and replace (2.35) with

$$[\mathrm{A}]_{\mathrm{active}} = \frac{1}{2}\left([\mathrm{A}]_{\mathrm{total}} - \mathrm{SD} - K_D + \sqrt{4[\mathrm{A}]_{\mathrm{total}}K_D + ([\mathrm{A}]_{\mathrm{total}} - \mathrm{SD} - K_D)^2}\right) \tag{2.36}$$

where we have defined $\mathrm{SD} = m[\mathrm{B}\mathrm{C}^{\mathrm{S}}] + n[\mathrm{B}\mathrm{C}^{\mathrm{D}}]$.

To introduce ultrasensitivity to the Phong model, we first note that the four phosphorylation rate constants for the U→T, U→S, T→D, and S→D transitions are given by Michaelis-Menten kinetics with ADP serving as a competitive inhibitor,

$$k_{\mathrm{phos}} = \frac{k^{\mathrm{A}}_{\mathrm{phos}}[\mathrm{A}]_{\mathrm{active}}}{K_{1/2} + [\mathrm{A}]_{\mathrm{active}}} \frac{1}{1 + K_I[\mathrm{ADP}]/[\mathrm{ATP}]} \tag{2.37}$$

where $k^{\mathrm{A}}_{\mathrm{phos}}$ varies with the specific phosphorylation reaction. To introduce ultrasensitivity, we add a threshold term T (Fig. 2.4D),

$$k_{\mathrm{phos}} = \frac{k^{\mathrm{A}}_{\mathrm{phos}}([\mathrm{A}]_{\mathrm{active}} - T)}{K_{1/2} + ([\mathrm{A}]_{\mathrm{active}} - T)} \frac{1}{1 + K_I[\mathrm{ADP}]/[\mathrm{ATP}]} H([\mathrm{A}]_{\mathrm{active}} - T) \tag{2.38}$$

where

$$T = (0.3 + 1.5e^{-5\%\mathrm{ATP}/100\%})(1 + 0.08[\mathrm{C}^{\mathrm{U}}]/\mu\mathrm{M}) \tag{2.39}$$

and H is the Heaviside function. The first part of the expression for the phosphorylation threshold, $0.3 + 1.5e^{-5\%\mathrm{ATP}/100\%}$, describes how the threshold changes as a function of %ATP; the constants are determined by approximating the [KaiA] threshold in Fig. 2.3A as an exponential function. The second part of the expression, $1 + 0.08[\mathrm{C}^{\mathrm{U}}]/\mu\mathrm{M}$, describes how the threshold changes as a function of C^{U} concentration. This formula is defined by

$$\frac{a/\mu\mathrm{M} + b([\mathrm{C}^{\mathrm{U}}]/\mu\mathrm{M} - 3.5)}{a/\mu\mathrm{M} + b(-3.5)}$$

where parameters a and b are determined by taking a linear fit, $a + b$[AA], of the data from Fig. 2.3G (yellow line); that is, $1 + 0.08[\mathrm{C^U}]/\mu\mathrm{M}$ gives the fractional changes to the phosphorylation threshold as a result of any U phosphoform (in the form of AA phosphomimetic mutant) additional to the 3.5 µM KaiC S431A in the experiment.

We use the original Phong model parameters in all analyses of the model with one exception. In the final modified model with both a K_D and a phosphorylation threshold (Fig. 2.4E bottom), the period is systematically longer than 24 h due to a slow down in phosphorylation. To fix this problem, we change k_{ds}^A and k_{ds}^0, the two rate constants controlling the D→S transition, to 0.94 k_{ds}^A and 1.1 k_{ds}^0.

To assess whether the ultrasensitivity we introduced into the model with (2.38) is consistent with our measurements, we simulated the KaiC-AA titration experiment using the modified Phong model (Fig. S2.10E) and compared the predicted EC50 with the experimental results (Fig. 2.3G). To perform this simulation, we set all phosphorylation rates to the S and D states to zero, and represent KaiC-AA concentration as a constant offset to the $[\mathrm{C^U}]$ term in (2.39). A comparison of the model prediction with the experimental result shows that EC50 scales with [AA] in the model similarly to the experiment, but the EC50 in the model is consistently larger than the experiment by about 0.15 µM. This result suggests that the ultrasensitivity we introduced into the model by hand is reasonable, but to achieve precise agreement with experiment would require fine tuning the Phong model parameters, which is outside the scope of the current work.

2.7 Supplementary Figures

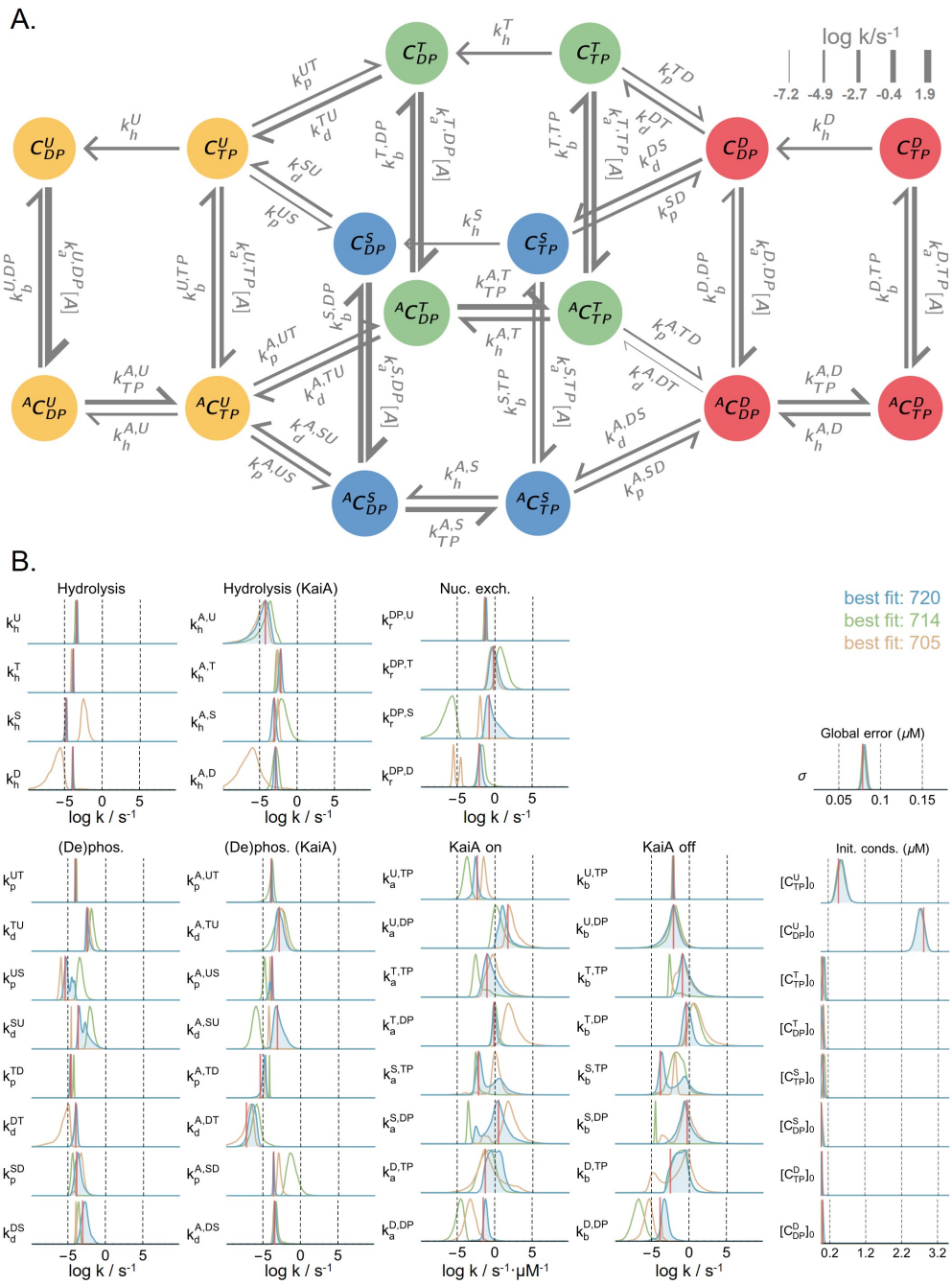


Figure S2.5: Overview of the model. A. A schematic of the full mass-action kinetic model. Here, each arrow represents a reaction, and the associated rate constant is represented using the notation introduced in the main text. The thickness of the arrows is proportional to the best fit rate on a log scale (base 10) at 100% ATP and 1.5 μM KaiA. B. The posterior distributions for all rate constants, initial conditions, and the global error hyperparameter. The three distributions represent the results from three independent runs; the log posterior values for the best fits from the three runs are listed. The red lines represent the best fit from the best run (i.e., the blue distributions). See Materials and Methods and Fig. S2.14 for further details on the model parameterization method.

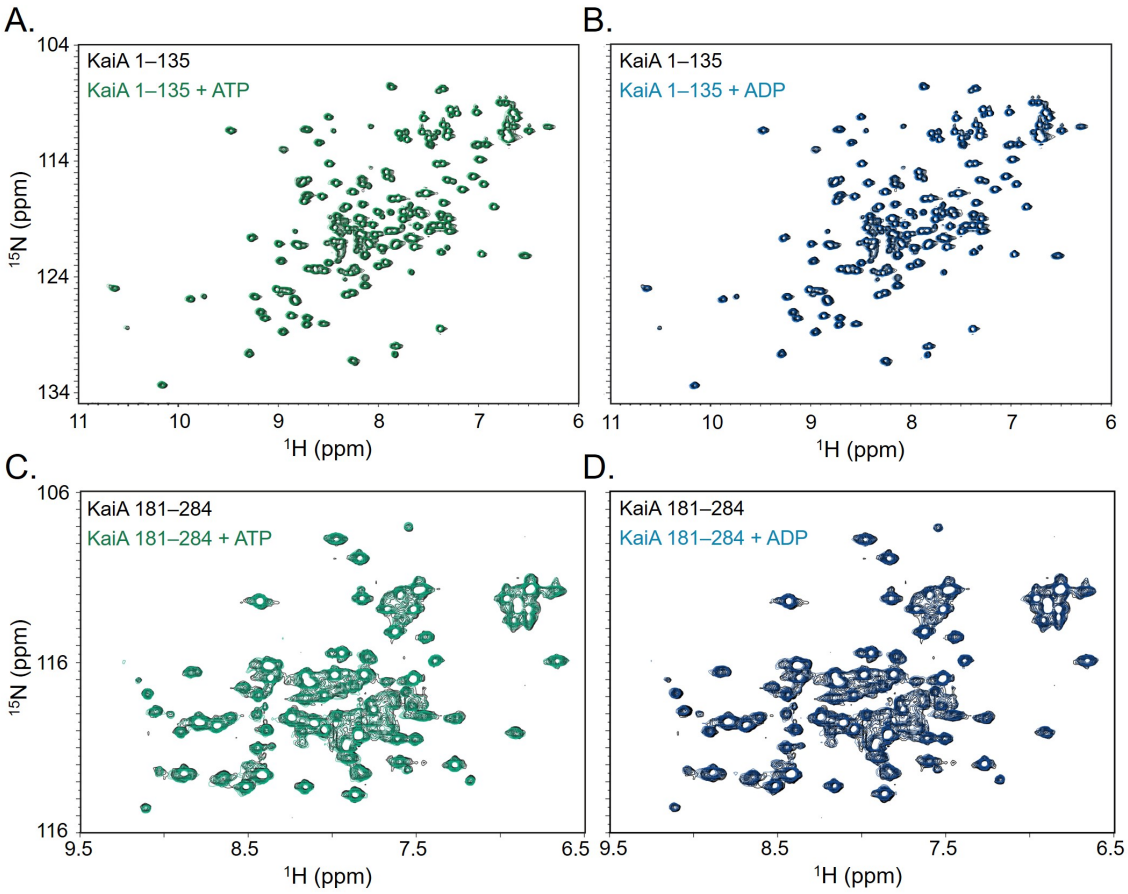


Figure S2.6: No evidence of direct nucleotide-KaiA interaction. ^{1}H-^{15}N HSQC spectra of the N-terminal fragment (residues 1–135) of KaiA in the presence and absence of ATP (A) or ADP (B) show no significant differences in chemical shifts, while spectra of the C-terminal fragment (residues 181–284) show subtle line broadening in the presence of ATP (C) and ADP (D), suggesting weak, if any, interaction between the nucleotide and the C-terminal fragment. Given these results, we do not include any direct KaiA-nucleotide interaction in the model.

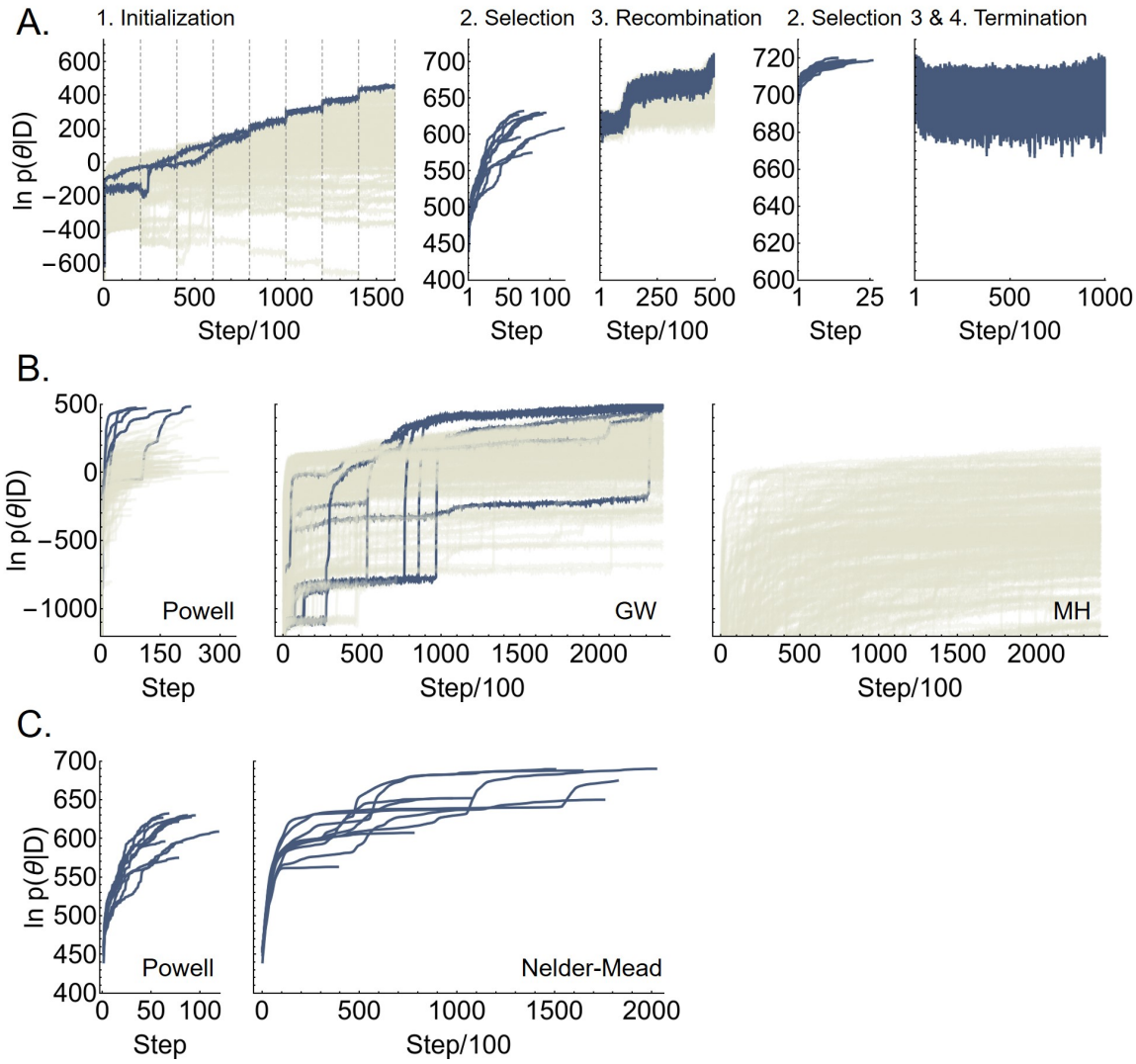


Figure S2.7: Performance of the fitting procedure. A) The time evolution of the log posterior values over the four steps of the fitting procedure (see Materials and Methods). For step 1 and 3, the individual Markov chains that do not produce walkers used in the next step are shown in beige. B) A comparison of the performance of Powell's method, a derivative-free numerical optimization method, Goodman-Weare (GW) ensemble MCMC method, and conventional Metropolis-Hastings (MH) algorithm with a Gaussian trial distribution. For the Metropolis-Hastings algorithm the covariance matrix of the trial distribution is given by the global covariance of the fit (i.e., the last step in panel A), scaled by a factor of 0.005 to give an average acceptance rate of 19.8%. A set of 224 walkers drawn from the prior distribution are used to initialize the simulations for all three methods; the 224 walkers are evolved independently for the Powell's and Metropolis-Hastings methods, and in an ensemble for the Goodman-Weare method. Chains that do not reach log posterior above 450 are shown in beige. C) A comparison of the performance of Powell's method with the Nelder-Mead simplex-based numerical optimization method. The simulations are initialized using the same walkers as in step 2 of A).

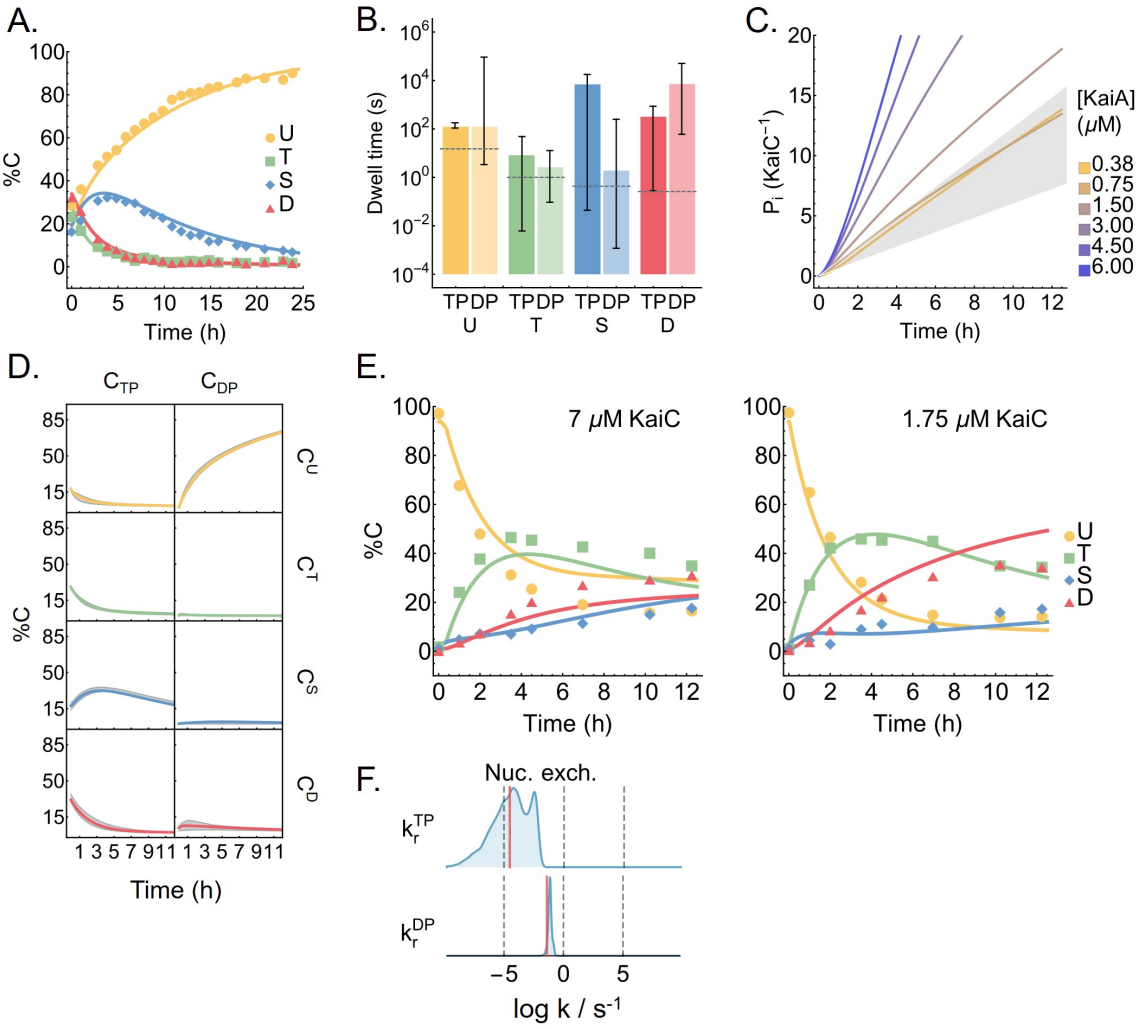


Figure S2.8: Behavior of the model. A. Model fit to the dephosphorylation dataset [17]. B. The best fit KaiA dwell time as a function of KaiC phosphoform and nucleotide-bound state. The error bars represent the 95% posterior interval, and the dashed lines represent the experimental measurements [24, 25], which did not resolve the nucleotide-bound states. C. Inorganic phosphate production per KaiC monomer over the course of a phosphorylation reaction. The gray region represents the experimental bounds on the KaiC hydrolysis rate with 1.2 µM KaiA and no KaiA [26]. D. The kinetics of the dephosphorylation reaction in the absence of KaiA, broken down into the eight individual KaiC states. The gray region represents the 95% posterior interval. Refer to Fig. S2.5A for the KaiC state names. E. The predicted phosphorylation kinetics at 7 and 1.75 µM KaiC, both at 100% ATP and 1.5 µM KaiA, compared to experimental measurements. Note that these two time series are not part of the training set. F. The posterior distributions for k_r^{TP} and k_r^{DP}, the dissociation rates for ATP and ADP, respectively, in an early iteration of the model. The horizontal axis is on a base 10 log scale. The long tail to the left of the posterior distribution for k_r^{TP} suggests that the model can be simplified by setting the rate to zero.

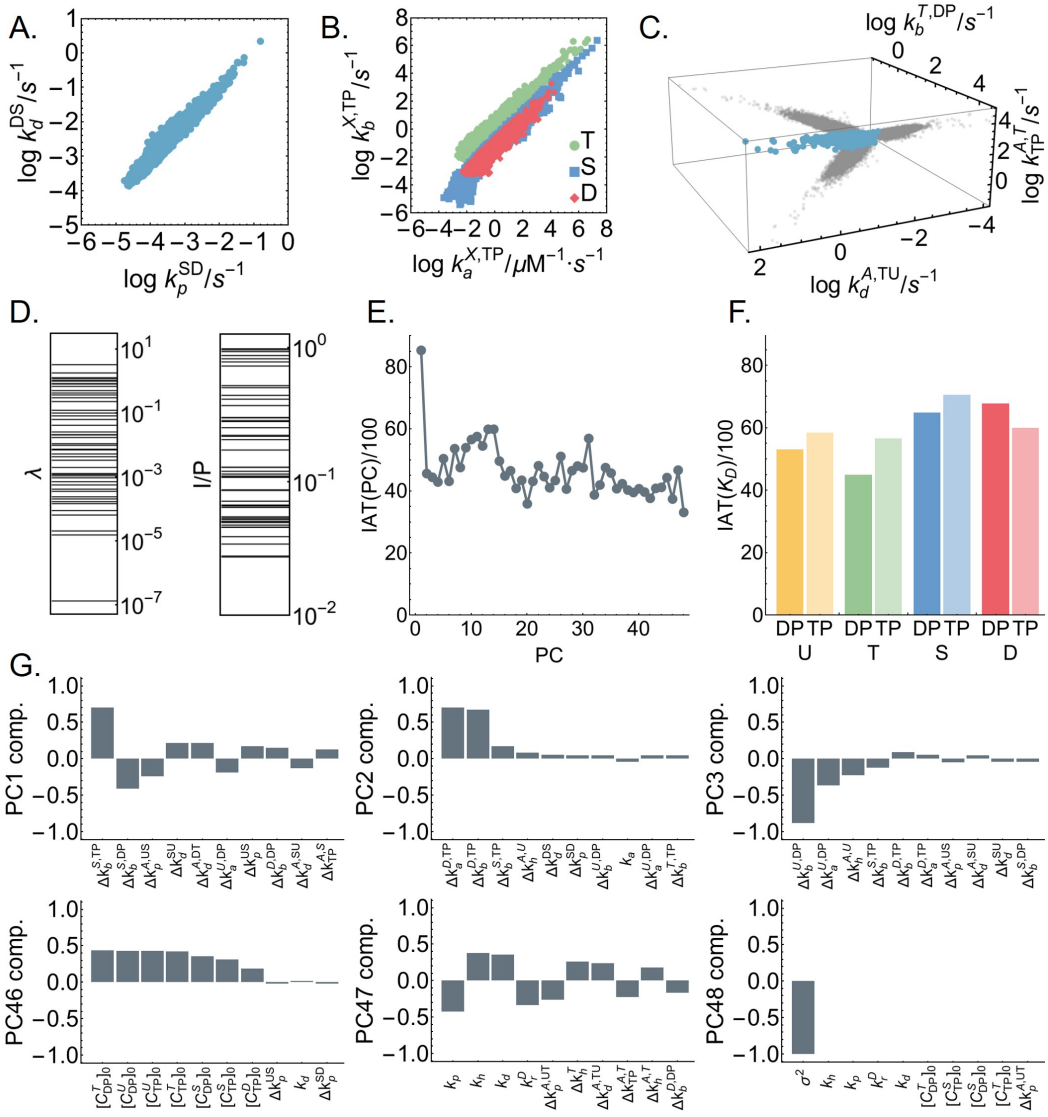


Figure S2.9: Correlation structure in the MCMC ensemble. A), B), and C) show parameters with correlation coefficients larger than 0.9. In B), "X" represents the KaiC phosphoforms. In C), the projections of the 3D scatter plot onto pairwise correlations are shown in gray. D) The principal component/eigenvalue spectrum of the covariance matrix (left), and the alignment of the principal components with the coordinates (right). Here, I denotes the intersection of the principal component ellipoid with the coordinates and P denotes the projection of the principal components onto the corresponding coordinates [55]. E) The integrated autocorrelation time for the 48 principal components (PC); the principal components are indexed from the largest to the smallest. The integrated autocorrelation time is calculated using an automated windowing procedure [137] from the autocorrelation function averaged over the ensemble. F) The integrated autocorrelation time for the KaiA dissociation constants as a function of KaiC phosphoform and nucleotide-bound states. G) The ten largest vector components, ordered by absolute value, for the first and last three principal components.

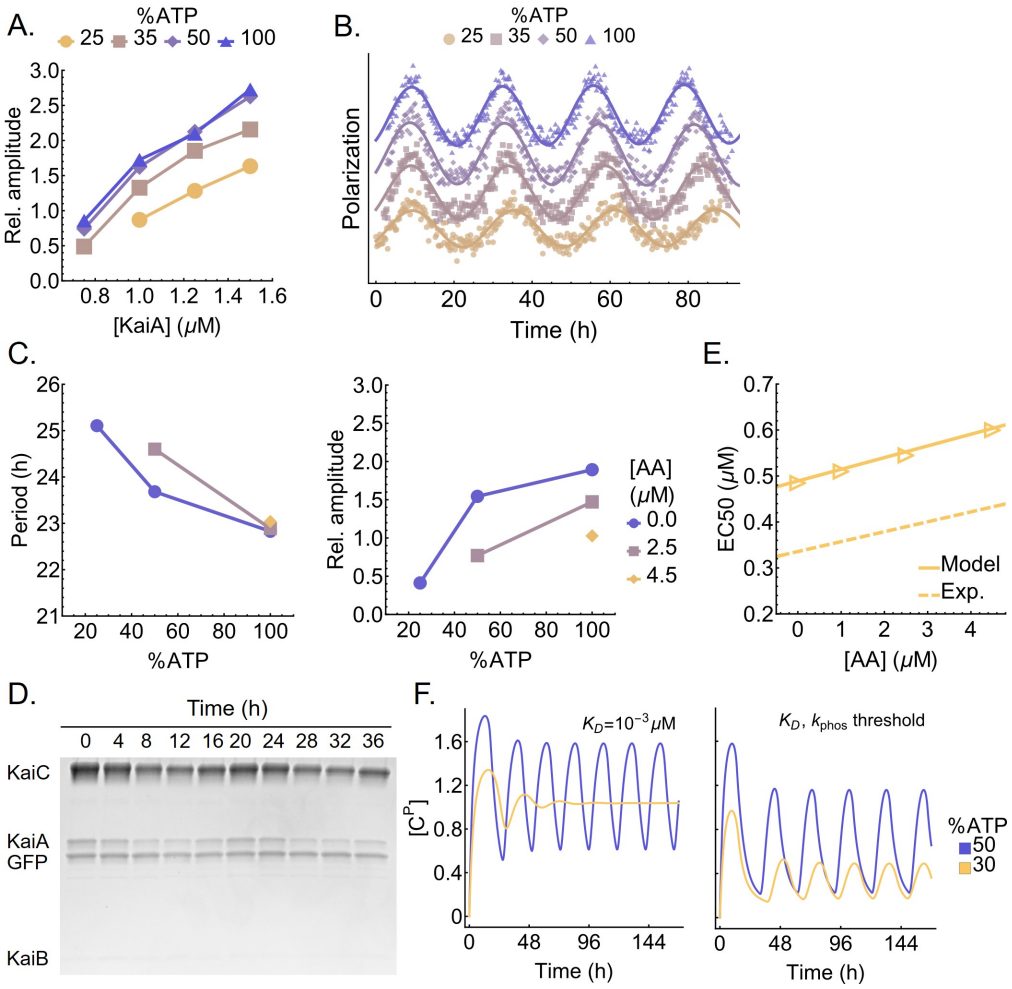


Figure S2.10: The metabolic compensation property of the Kai oscillator. A. Fluorescence polarization measurements of oscillatory reactions at various [KaiA] and %ATP are fit to curves of the form $\mathrm{FP}(t) = A\cos(2\pi T^{-1}t + \phi) + bt + c$ to extract the normalized amplitude ($100A/c$; dimensionless) of the oscillator; see Fig. 2.2D for the periods (T) of the same reactions. Reactions with an amplitude $A < 0.5$ are considered to be non-oscillatory. B. Representative traces demonstrating the effect of %ATP at 1.25 μM KaiA; The polarization data are shifted vertically to avoid overlaps and horizontally to align the first peaks. C. Fluorescence polarization measurements of oscillatory reactions at three %ATP conditions in the presence of 0.0, 2.5, and 4.5 μM KaiC S431A/T431A (AA). The period (left) and amplitude (right) of the reactions are extracted using the same curve fit method as for panel A.; non-oscillatory reactions are not shown. D. SDS-PAGE gel image of the supernatant from the KaiB-FLAG immunoprecipitation experiment. E. Simulated KaiC AA titration experiment using the modified Phong model. The points are model predictions and the solid line is a linear fit. The dotted line is the linear fit to the experimental results (see Fig. 2.3G). The simulations are carried out at 100% ATP, 1.5 μM KaiA, and $K_D = 10^{-3}$ μM condition. F. A Comparison of the metabolic compensation property of the Phong model without (left) or with (right) a phosphorylation threshold at $K_D = 10^{-3}$ μM. The model exhibits phase decoherence at low %ATP without a phosphorylation threshold.

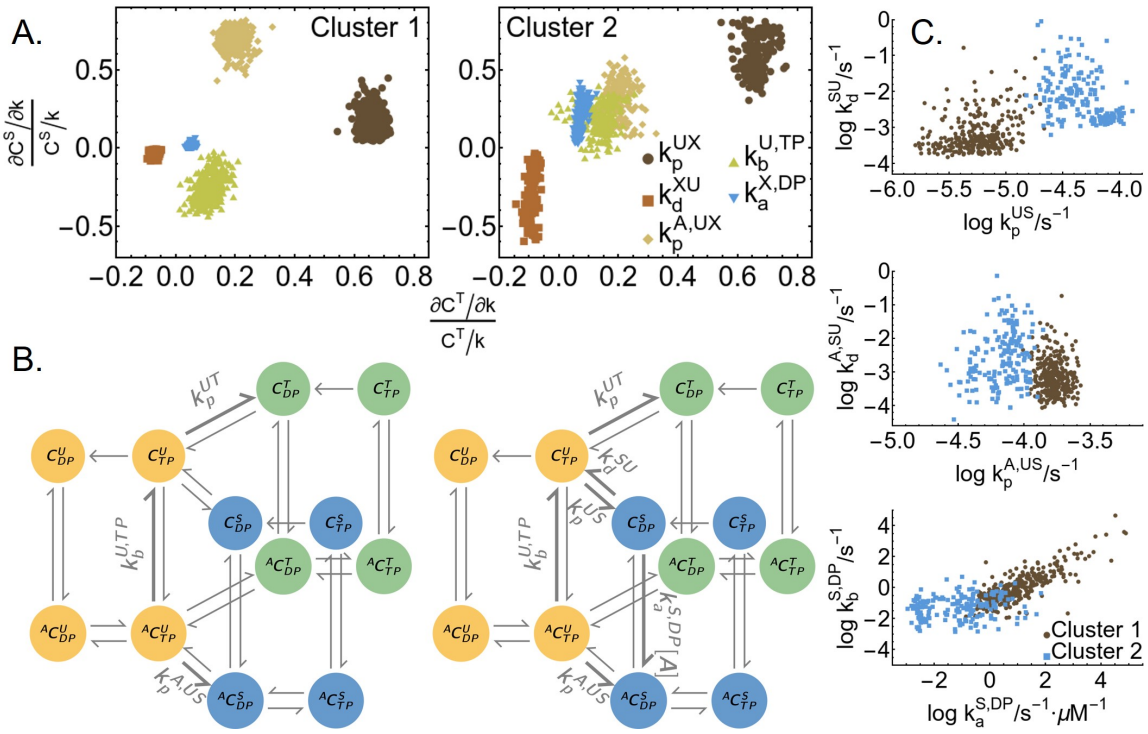


Figure S2.11: The mechanism of kinetic ordering is not well-constrained. A) Spectral clustering on the relative sensitivity of the T and S phosphoform concentrations at $t = 1$ h to rate constants in the T- and S-site models, respectively. Only the parameters with significant (> 0.2) relative sensitivities in either cluster are shown in the plot. "X" stands for either the T (horizontal axis) or S (vertical axis) phosphoform. The sensitivities are calculated using 500 sampled parameter sets chosen randomly from the ensemble. The clustering analysis was done using the FindClusters function in Mathematica 12.0. B) Model diagrams that highlight the reactions that have the highest relative sensitivities in the first (left) and second (right) clusters; the D phosphoform is not shown for clarity. C) Selected model parameter values in the two clusters. A comparison with blue distributions in Fig. S2.5B indicates that the clustering based on sensitivity can be mapped onto the modes of the posterior distribution.

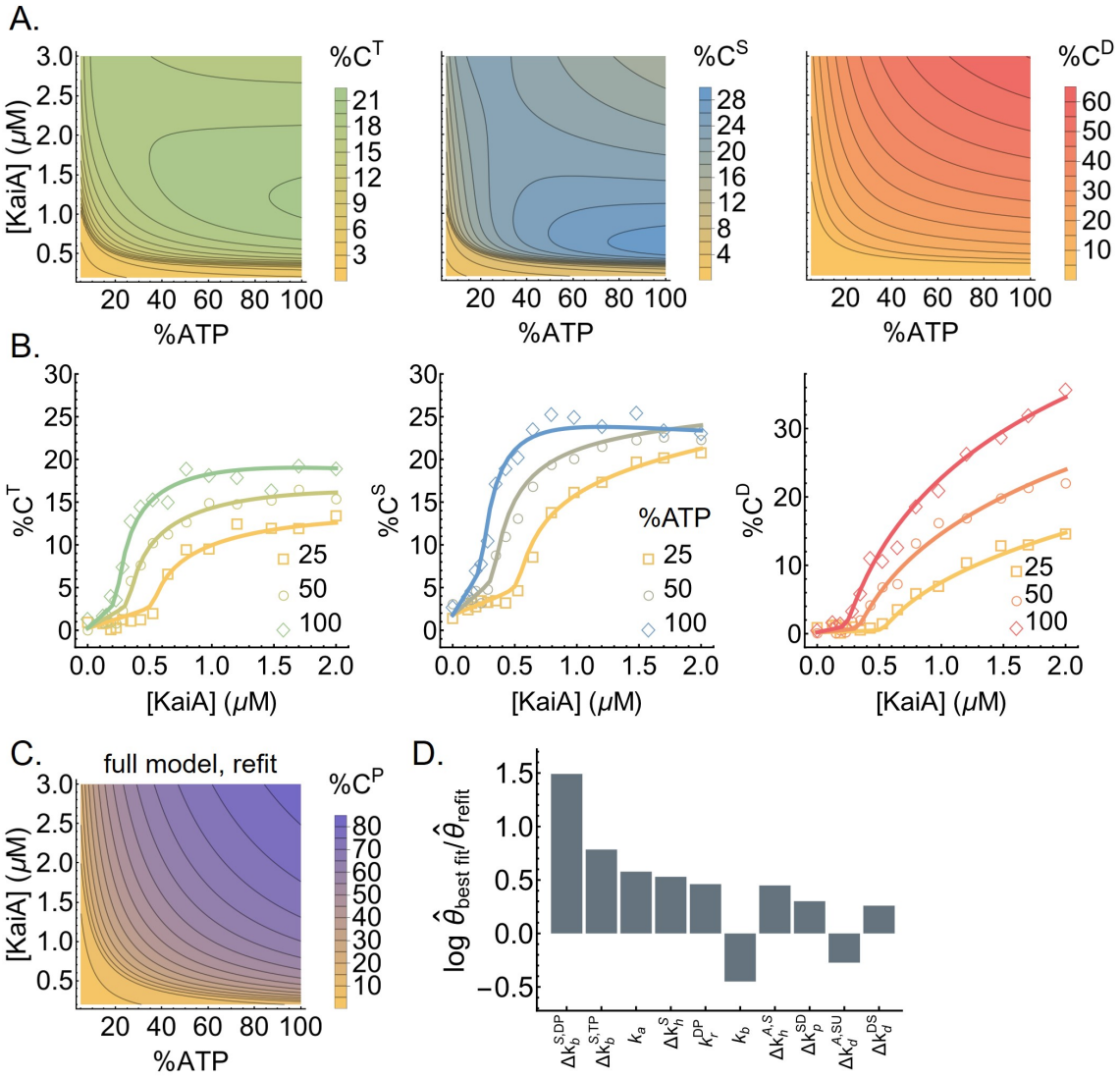


Figure S2.12: KaiC stimulus-response relations. A. The steady-state stimulus-response relations for T, S, and D phosphoforms predicted by the model. B. The experimentally determined stimulus-response functions of the T, S, and D phosphoforms at three %ATP conditions; the curves are based on refitting the best fit to the steady-state measurements. C. The model-predicted stimulus-response relation of the total steady-state KaiC phosphorylation level as a function of %ATP and [KaiA] after refitting to the steady-state measurements. D. The differences in the log parameter values (base 10) of the best fit before and after refit. The differences are ordered by magnitude and only the 10 parameters (in the multiplicative-factor scheme) with the largest changes are shown.

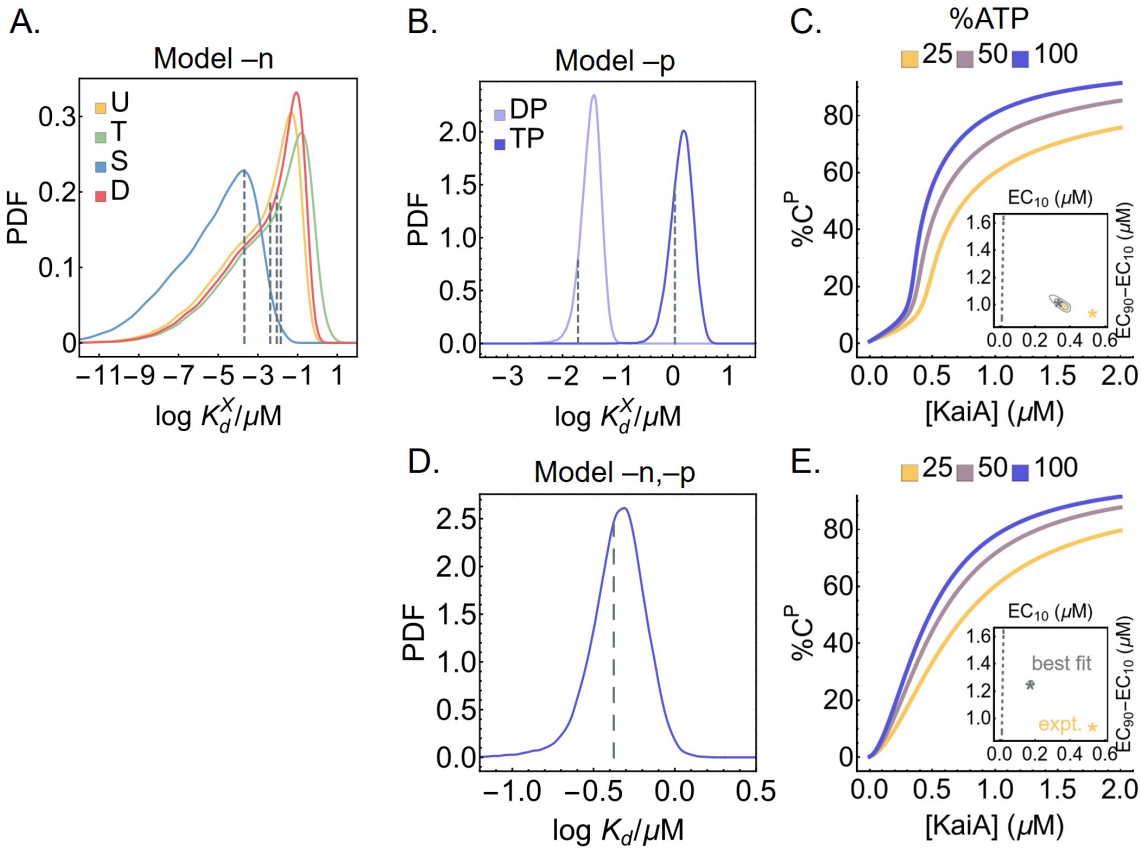


Figure S2.13: KaiA binding affinities of simplified models. A. The posterior distributions for the KaiA dissociation constants as a function of KaiC phosphoform in model –n, where the KaiA on/off rates are decoupled from the nucleotide-bound states of KaiC. The dashed lines represent the best fit. B. Similar to A., but for model –p, where the KaiA on/off rates are decoupled from the KaiC phosphoform; the dashed lines represent the best fit. C. Cross sections of the stimulus-response relation at three %ATP, computed using model –p. The inset represents posterior distribution for the shape measures of the stimulus-response function at 25% ATP. The contours represent the 68% and 95% HDRs, and the gray star represents the model best fit. The shape of the stimulus-response function is quantified using two metrics: EC10, which quantifies threshold-like behavior, and EC90 – EC10, which quantifies switch-like behavior. The shape measures of the experimentally-determined stimulus-response function at 25% ATP is shown as the yellow star. The dashed line represents $(\mathrm{EC10}, \mathrm{EC90} - \mathrm{EC10}) = (K/9, 80K/9)$, which characterizes the shape of a hyperbolic stimulus-response function $[\mathrm{A}]/(K + [\mathrm{A}])$ that has no switching or thresholding. D. Similar to A., but for model –n,–p, where there is a single KaiA on/off rate in the model. E. Similar to C., but for model –n,–p.

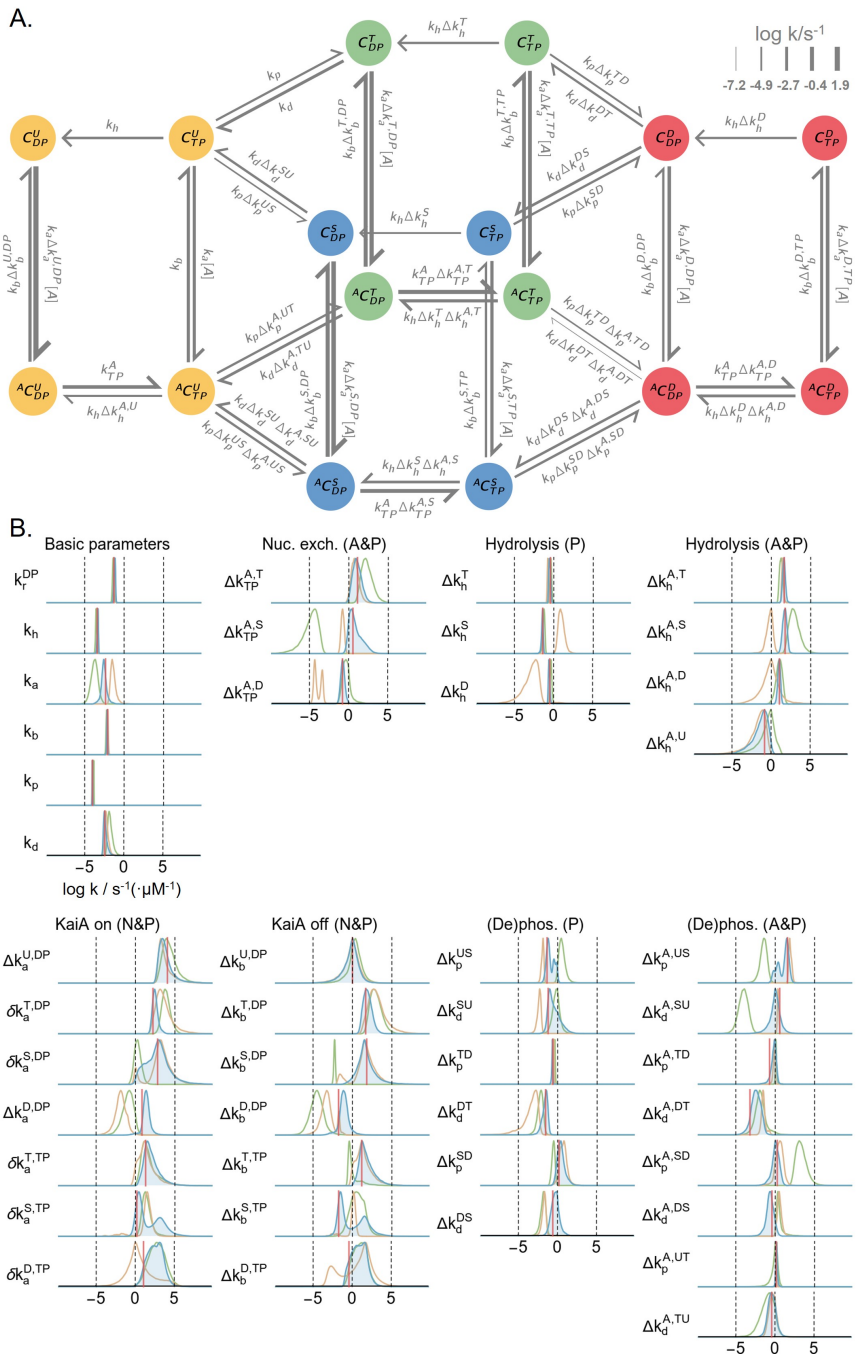


Figure S2.14: Overview of the model with the multiplicative-factor parameterization scheme. Panels A) and B) are analogous to those in Fig. S2.5, but the rate constants are represented as products of the factors that are actually optimized in the MCMC simulations. In B), the δk parameters are fixed parameters determined by detailed balance conditions. The parentheses denote state-dependent effects; A: KaiA-bound state, P: phosphoform, N: nucleotide-bound state. The horizontal axis has a log scale (base 10). The rate constants have the unit s^{-1} (or $s^{-1}\cdot\mu M^{-1}$ for the KaiA on rate constant k_a) and the multiplicative factors are dimensionless. See Materials and Methods for further description of the detailed balance conditions and the model parameterization method.

2.8 Supplementary Tables

Table S2.6: Effects of KaiA on KaiC function

Model	Log likelihood			Bayes factor
	Phosphorylation	Dephosphorylation	Hydrolysis	
Full model	422.9	346.8	–0.8	1
–P*	390.0	338.6	–0.8	1.4
–H†	355.5	308.8	–0.9	2.5
–P,–H	278.2	283.6	–0.2	7.7

* –P: KaiA binding decoupled from (de)phosphorylation rates.
† –H: KaiA binding decoupled from KaiC hydrolysis rates.

2.9 Data and Software Availability

The code used to perform and analyze the MCMC simulations, as well as data from the inference run, are available on github at https://github.com/luhong88/KaiAC_MCMC.

2.10 Acknowledgments

We thank Connie Phong and Haneul Yoo for their protein samples, Jonathan Weare for helpful discussions, and Steven Redford for critical readings of the manuscript. This work was supported by National Science Foundation award MCB-1953402, National Institutes of Health awards GM107369, GM107521, and EY025957, Department of Energy Office of Advanced Scientific Computing Research contract DE-AC02-06CH11347 and award DE-SC0014205, and a Howard Hughes Medical Institute-Simons Foundation Faculty Scholarship (to MJR). AL was also supported by the Center for Cellular and Biomolecular Machines at University of California, Merced (NSF Grant HRD-1547848). Computations were performed on resources provided by the University of Chicago Research Computing Center, and the Extreme Science and Engineering Discovery Environment [138] (NSF Grant ACI-1548562) Bridges (PSC) computing nodes through allocation TG-MCB180007.

2.11 Future Directions

An objective of the line of research demonstrated in this work is to eventually extend such model fitting procedure to account for the full Kai oscillator. A direct extension of the ODE model put forward in this work is computationally intractable due to the combinatorial complexity of the system as one considers additional degrees of freedom associated with the states of CI and KaiB. As such, it is better to build upon existing models of the full oscillator, such as the Paijmans model [40], and then fit the extended model to experimental data. In this section, I outline four aspects of the Paijmans model that can be further elaborated upon in future works. The equation numbers in this section refer to those in [40].

KaiA affinity and KaiC nucleotide-bound state In the Paijmans model, the KaiA binding affinity to each KaiC hexamer is determined by $\Delta G_{\text{bind}}^{\text{CII.KaiA}}$, where each monomer i contributes an additive term $\delta g_{\text{bind}}^{\text{CII.KaiA}}(\text{X}_i)$ depending on its phosphorylation state X (see Eq. 4). Given our KaiAC modeling result, the coupling between nucleotide-bound state and affinity is critical for generating ultrasensitivity in KaiC phosphorylation that is missing in the Paijmans model. To introduce this effect, we replace all the $\delta g_{\text{bind}}^{\text{CII.KaiA}}(\text{X}_i)$ terms with $\delta g_{\text{bind}}^{\text{CII.KaiA}}(\text{X}_i, \text{N}_i)$, where $\text{N}_i \in \{\text{DP}, \text{TP}\}$ denotes the nucleotide-bound state of the monomer i in the CII domain. This effectively introduces four more parameters to the model. In the original model, $\delta g_{\text{bind}}^{\text{CII.KaiA}}(\text{U})$ is by definition set to zero, here we define the reference state to be $\delta g_{\text{bind}}^{\text{CII.KaiA}}(\text{U, TP}) = 0$. Correspondingly, we redefine $k_{\text{on, 0}}^{\text{CII.KaiA}}$ and $k_{\text{off, 0}}^{\text{CII.KaiA}}$ to be the KaiA (un)binding rates when all monomers of an active hexamer are in the U, TP states.

Due to detailed balance, the differential KaiA binding affinity results in changes in the relative stabilities of KaiC phosphoforms in the presence or absence of KaiA. In the

original Paijmans model, this relation is given by Eq. 8:

$$\delta g^{0}_{\mathrm{XY}} - \delta g^{\mathrm{KaiA}}_{\mathrm{XY}} = \delta g^{\mathrm{CII.KaiA}}_{\mathrm{bind}}(\mathrm{Y}) - \delta g^{\mathrm{CII.KaiA}}_{\mathrm{bind}}(\mathrm{X})$$

To account for the coupling between KaiA affinity and KaiC nucleotide-bound state, we replace Eq. 8 with the following set of relations:

$$\delta g^{0}_{\mathrm{UT}} - \delta g^{\mathrm{KaiA}}_{\mathrm{UT}} = \delta g^{\mathrm{CII.KaiA}}_{\mathrm{bind}}(\mathrm{T,\ DP}) - \delta g^{\mathrm{CII.KaiA}}_{\mathrm{bind}}(\mathrm{U,\ TP})$$

$$\delta g^{0}_{\mathrm{US}} - \delta g^{\mathrm{KaiA}}_{\mathrm{US}} = \delta g^{\mathrm{CII.KaiA}}_{\mathrm{bind}}(\mathrm{S,\ DP}) - \delta g^{\mathrm{CII.KaiA}}_{\mathrm{bind}}(\mathrm{U,\ TP})$$

$$\delta g^{0}_{\mathrm{TD}} - \delta g^{\mathrm{KaiA}}_{\mathrm{TD}} = \delta g^{\mathrm{CII.KaiA}}_{\mathrm{bind}}(\mathrm{D,\ DP}) - \delta g^{\mathrm{CII.KaiA}}_{\mathrm{bind}}(\mathrm{T,\ TP})$$

$$\delta g^{0}_{\mathrm{SD}} - \delta g^{\mathrm{KaiA}}_{\mathrm{SD}} = \delta g^{\mathrm{CII.KaiA}}_{\mathrm{bind}}(\mathrm{D,\ DP}) - \delta g^{\mathrm{CII.KaiA}}_{\mathrm{bind}}(\mathrm{S,\ TP})$$

Eq. 9 is updated accordingly.

In the event that the resulting model is too hard to fit/constrain, one possible simplification scheme is to couple KaiA binding affinity to the nucleotide-bound state of KaiC only, similar to model –p. Another possibility is to introduce a single $\delta g^{\mathrm{CII.KaiA}}_{\mathrm{bind}}(\mathrm{DP})$ to account for the effect of ADP relative to the ATP bound states of KaiC subunits.

Representation of KaiB fold switch The Paijmans model does not have an explicit representation of KaiB, and the KaiB binding kinetics is represented by effective first-order rate constants (see Table 3). This is a fine assumption, as the period and amplitude of the oscillator is not strongly dependent on KaiB concentration [139] and its oligomerization reactions serve to stabilize the monomeric KaiB concentration, which is the species that participates in the clock reaction. However, there are still several aspects of KaiB function that remains unclear, and it would be useful to introduce an explicit representation of KaiB into the model to allow specific hypothesis testing. For example, is the free energy of ATP hydrolysis in the CI domain directly coupled to KaiB fold switch? Is the timescale

of KaiB fold switch determined by the ATPase activity of KaiC? When KaiB is in the monomeric state, is the ground state vs. fold-switched state more stable?

To introduce explicit representation of KaiB is fairly straightforward:

$$\mathrm{KaiB}_4^{\mathrm{gs}} \underset{k_{41}}{\overset{k_{14}}{\rightleftharpoons}} 4\mathrm{KaiB}^{\mathrm{gs}} \underset{k_{\mathrm{gs}}}{\overset{k_{\mathrm{fs}}}{\rightleftharpoons}} 4\mathrm{KaiB}^{\mathrm{fs}}$$

Note that we do not model the KaiB dimer state explicitly. Experimental analysis shows that the KaiB dimeric state is sparsely populated over a wide range of total KaiB concentrations [23]; we therefore apply the quasi-steady state approximation and eliminate the explicit representation of the dimer. Here, $\mathrm{KaiB}^{\mathrm{fs}}$ is the only species capable of binding to KaiC. We further assume that KaiB dissociate from CI while remaining in the fold-switched conformation and then revert back to the ground state with a rate of k_{gs}.

Cooperative KaiB binding In the original Paijmans model, there is no cooperative KaiB binding. Instead, to set the timescale of the transition from phosphorylation to dephosphorylation, the model assumes slow and independent KaiB binding to KaiC, and that no KaiA can be sequestered before a full ring of six KaiB monomers are bound.

These assumptions can be relaxed with an improved representation of KaiB cooperative binding. Native MS analysis [23, 78] shows that the only detectable KaiB binding stoichiometry with KaiC is $\mathrm{KaiB}_1\mathrm{C}_6$ and $\mathrm{KaiB}_6\mathrm{C}_6$. Based on this observation, one can introduce a simple scheme for cooperativity as

$$\mathrm{CI}_6 + 6\mathrm{KaiB}^{\mathrm{fs}} \underset{k_{\mathrm{off}}^{\mathrm{CI.KaiB}}}{\overset{k_{\mathrm{on}}^{\mathrm{CI.KaiB}}}{\rightleftharpoons}} \mathrm{CI}_6 \cdot \mathrm{KaiB}^{\mathrm{fs}} + 5\mathrm{KaiB}^{\mathrm{fs}} \underset{k_{\mathrm{off}}^{\mathrm{CI.5KaiB}}}{\overset{k_{\mathrm{on}}^{\mathrm{CI.5KaiB}}}{\rightleftharpoons}} \mathrm{CI}_6 \cdot \mathrm{KaiB}_6^{\mathrm{fs}}$$

The KaiB on/off rates need to depend on the active vs. inactive state of CI. It is likely that in the active state, the $\mathrm{CI}_6 \cdot \mathrm{KaiB}_6^{\mathrm{fs}}$ state is highly unfavorable, such that the $\mathrm{KaiB}_1\mathrm{C}_6$ stoichiometry is observed during the early dephosphorylation stage when KaiC is only starting to switching to the inactive state, while the $\mathrm{KaiB}_6\mathrm{C}_6$ stoichiometry dominates

afterwards. This description is consistent with the fact that native MS analysis using teKaiC-DD does not observe the $KaiB_1C_6$ stoichiometry [140].

A more rigorous description considering all possible intermediate states is possible should the current scheme be inadequate.

KaiA inactivation kinetics The Paijmans model does not consider the inactive conformation of KaiA [77]. Here we introduce a simple description of KaiA inactivation

$$\mathrm{KaiA_2^{act}} \underset{k_{\mathrm{act}}}{\overset{k_{\mathrm{inact}}}{\rightleftharpoons}} \mathrm{KaiA_2^{inact}}$$

where the active state can bind to the CII domain, while the inactive state can bind to the CI domain (in the presence of bound KaiB). Furthermore, we allow for dissociation of a $KaiA_2B_1$ complex during the late phosphorylation stage as a single unit [78]; e.g.,

$$\mathrm{CI_6 \cdot KaiB^{fs} \cdot KaiA_2^{inact}} \underset{k_{\mathrm{on}}^{\mathrm{CI.KaiAB}}}{\overset{k_{\mathrm{off}}^{\mathrm{CI.KaiAB}}}{\rightleftharpoons}} \mathrm{CI_6 + KaiB^{fs} \cdot KaiA_2^{inact}} \underset{k_{\mathrm{on}}^{\mathrm{KaiAB}}}{\overset{k_{\mathrm{off}}^{\mathrm{KaiAB}}}{\rightleftharpoons}} \mathrm{CI_6 + KaiB^{fs} + KaiA_2^{inact}}$$

Again, the KaiAB on/off rates (e.g., $k_{\mathrm{on}}^{\mathrm{CI.KaiAB}}$, $k_{\mathrm{on}}^{\mathrm{CI.5KaiAB}}$) from CI need to depend on the active vs. inactive state of CI.

This description of KaiA inactivation can be made more detailed to allow for alternative (in)activation and (un)binding pathways (e.g., monomeric KaiA has been detected during the late dephosphorylation stage [78]). However, given the level of experimental characterization available for the inactivation process the current scheme should be sufficient.

CHAPTER 3

MOLECULAR DYNAMICS SIMULATIONS OF NUCLEOTIDE RELEASE FROM THE CIRCADIAN CLOCK PROTEIN KAIC REVEAL ATOMIC-RESOLUTION FUNCTIONAL INSIGHTS

This chapter was published as an article in PNAS (doi:10.1073/pnas.1812555115) and appears below with minor formatting modifications.

Lu Hong[a], Bodhi P. Vani[b], Erik H. Thiede[b], Michael J. Rust[c,d,e], and Aaron R. Dinner[b,d,f]

[a]Graduate Program in Biophysical Sciences, University of Chicago

[b]Department of Chemistry, University of Chicago

[c]Department of Molecular Genetics and Cell Biology, University of Chicago

[d]Institute for Biophysical Dynamics, University of Chicago

[e]Institute for Genomics and Systems Biology, University of Chicago

[f]James Franck Institute, University of Chicago

3.1 Abstract

The cyanobacterial clock proteins KaiA, KaiB, and KaiC form a powerful system to study the biophysical basis of circadian rhythms, because an in vitro mixture of the three proteins is sufficient to generate a robust ∼24-h rhythm in the phosphorylation of KaiC. The nucleotide-bound states of KaiC critically affect both KaiB binding to the N-terminal domain (CI) and the phosphotransfer reactions that (de)phosphorylate the KaiC C-terminal domain (CII). However, the nucleotide exchange pathways associated with transitions among these states are poorly understood. In this study, we integrate recent advances in molecular dynamics methods to elucidate the structure and energetics of the pathway for Mg·ADP release from the CII domain. We find that nucleotide release

is coupled to large-scale conformational changes in the KaiC hexamer. Solvating the nucleotide requires widening the subunit interface leading to the active site, which is linked to extension of the A-loop, a structure implicated in KaiA binding. These results provide a molecular hypothesis for how KaiA acts as a nucleotide exchange factor. In turn, structural parallels between the CI and CII domains suggest a mechanism for allosteric coupling between the domains. We relate our results to structures observed for other hexameric ATPases, which perform diverse functions.

3.2 Introduction

Circadian clocks are biochemical oscillators found in organisms in all kingdoms of life. In general, a circadian oscillator is characterized by its ability to generate self-sustaining oscillations with a near-24-hour period in the absence of external environmental cues. Furthermore, the period of the oscillation is largely insensitive to temperature and metabolic perturbations over the physiological range [37]. Circadian rhythms are important for maintaining regular cellular and physiological activities, and defects in the circadian clock can result in reduced fitness [141] and various health issues [142–146].

The core clock of the cyanobacterium *Synechococcus elongatus* PCC 7942 presents a unique opportunity for elucidating the biophysical basis of circadian rhythms [141, 147], because its molecular oscillations can be reconstituted in vitro from purified proteins [12], and they show all of the hallmark properties of circadian rhythms described above. The clock is based on periodic interactions among three soluble proteins: KaiA, KaiB, and KaiC. These interactions give rise to near-24-hour oscillations in the phosphorylation pattern of KaiC, and these oscillations are coupled to the external environment in part through the relative concentrations of ATP and ADP [28, 29, 148].

KaiC is a homohexamer [16] (Fig. 3.1A) within the RecA/DnaB superfamily [58, 149], which includes a large variety of ATPases in archaea, bacteria, and eukaryotes [149]. Each subunit of *S. elongatus* KaiC has 519 residues that are organized into two

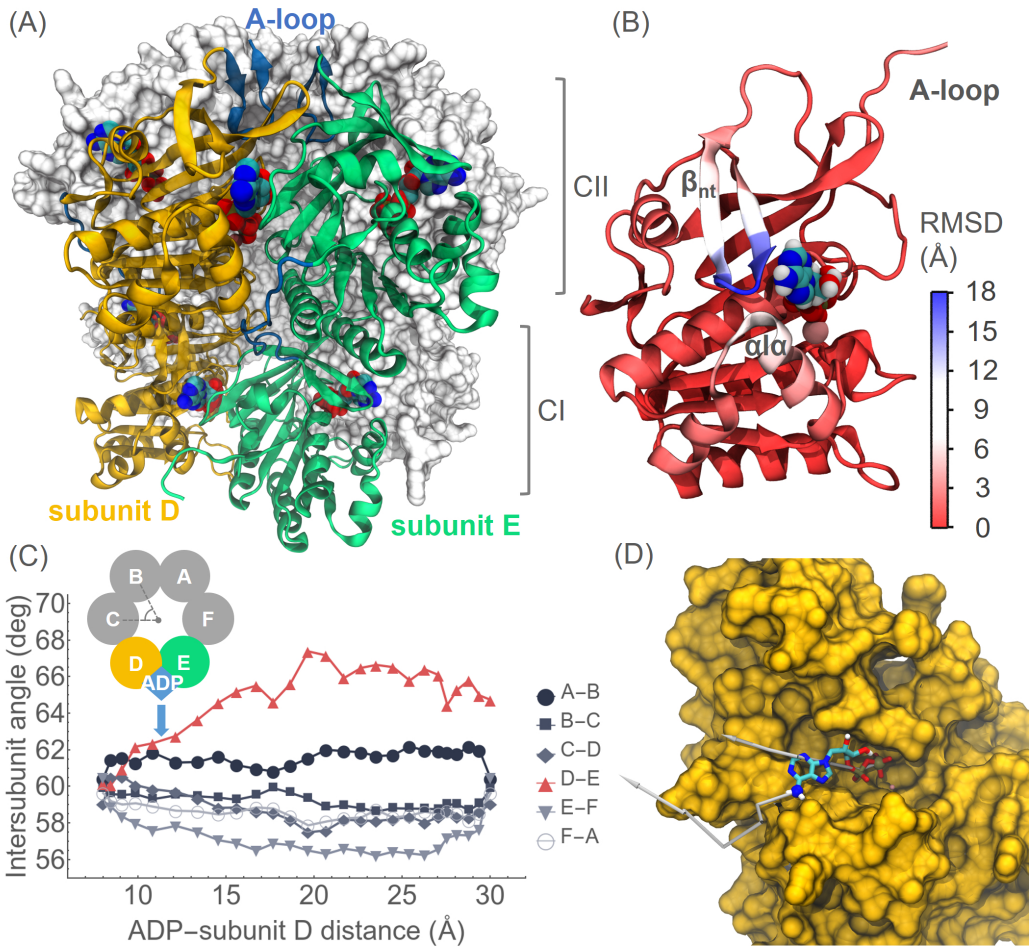


Figure 3.1: The CII nucleotide release pathway involves both local and global conformational changes. (A) KaiC is a homohexamer with active sites located at the subunit interfaces. The hexamer forms a double-ring architecture; the N-terminal ring is termed the CI domain, and the C-terminal ring is termed the CII domain. The secondary structures of subunits D and E are shown as yellow and green ribbon diagrams, respectively, while the other four subunits are represented by a white van der Waals surface. The CI–CII linker (residues 248 to 261) and the A-loop (residues 486 to 498) of subunits D and E are colored blue. (B) A β-hairpin (β_{nt}; residues 468 to 483) and an α-loop-α motif ($\alpha l\alpha$; residues 321 to 342) show the biggest displacements in the locally enhanced sampling simulations. The color represents the per-residue root-mean-square deviation (RMSD) between initial and final protein backbone conformations, mapped onto the initial structure. (C) The string calculation reveals concerted motion of the six CII intersubunit angles during ADP dissociation at the subunit D–E active site. The inset shows a schematic of the CII domain, illustrating the intersubunit angle variables and the ADP–subunit distance variables employed in the simulation. (D) The nucleotide moves radially outward in the string pathway; the arrows indicate the movement of N6 and Pβ atoms in ADP. In (A), (B), and (D), the Mg·ADP atoms are colored by element (C, cyan; N, blue; O, red; P, gold; H, white; Mg, pink).

domains formed by a tandem repeat of a common ATPase fold [150]; the N-terminal and C-terminal domains are known as CI and CII, respectively. Thus the hexameric assembly has a two-tiered ring structure. Both the CI and CII domains hydrolyze ATP, and the CII domain can also reversibly transfer phosphoryl groups between amino acids S431 and T432 and bound nucleotides (i.e., from ATP and to ADP) [38, 39, 59]. KaiC is a relative of the AAA+ superfamily [149], and each KaiC globular domain shares close structural similarity to F_1-ATPase, part of the F_OF_1-ATP synthase [39]. Many of the proteins that are evolutionarily related to KaiC use mechanochemical coupling to do work, and it is an open question how the mechanisms of action of these ATPases relate to those of the KaiC family, which uses ATP hydrolysis for signal transduction and timing.

Stimulation by KaiA promotes KaiC phosphorylation during the day [12, 15, 17, 24, 60, 61, 80], and sequestration of KaiA by KaiB leads to dephosphorylation during the night [22, 28, 77, 83]. However, the molecular mechanism of KaiA's action is unclear. The activity of KaiA was initially defined by its ability to stimulate KaiC autokinase activity [15]. Kinetic studies indicate that this activity also inhibits the rate of dephosphorylation [17]. Recent work shows that KaiA can act as a nucleotide exchange factor for CII, which otherwise very slowly releases ADP with an estimated rate less than 0.1 h^{-1} [19]. These observations, together with the fact that CII is a reversible phosphotransferase, suggest that KaiA promotes phosphorylation by enabling exchange of ADP for ATP. Since ATP cannot be used as a phosphotransfer acceptor in the dephosphorylation reaction, the ability of KaiA to regulate the nucleotide bound state in CII provides a unifying hypothesis that could explain both stimulation of phoshorylation and inhibition of dephosphorylation.

Thus, the mechanism of nucleotide exchange in CII is key to understanding the regulation of KaiC phosphorylation. Indeed, control of nucleotide exchange is likely used throughout the clock cycle. The binding of KaiB to the homologous CI domain requires the ability of CI to go through a catalytic cycle [28, 76], and structural studies of the

KaiB·KaiC complex show that ADP is retained in the CI active site when KaiB is bound [77, 78]. However, accessing this binding-competent state in CI requires S431 phosphorylation in the distal CII domain [17, 21, 79]. Taken together, these results suggest that the CII ring influences nucleotide exchange in the CI ring, just as KaiA influences nucleotide exchange in CII.

To the best of our knowledge, a nucleotide exchange pathway has not been previously characterized for any homohexameric ATPase in the RecA/DnaB or AAA+ superfamilies. However, structures of these proteins in certain nucleotide-bound states show that they can take on conformations that lack C_6 symmetry. Examples include the staggered conformation of ClpX resulting from a rigid-body rotation of the small α and large $\alpha\beta$ domains [151], the split washer conformation at the apo interface of NSF [152], or the staircase-like DNA-binding hairpin structures of the E1 helicase [153, 154]. These structures have been interpreted to mean that hydrolysis proceeds sequentially or stochastically, rather than concertedly, around the rings of these proteins [155]. Although there are subtly asymmetric structures of the CI domain of KaiC with both ATP- and ADP-bound active sites [156], it is not clear whether the full molecule functions in a symmetric (i.e., concerted) or asymmetric fashion and, if the latter, how its structures relate to those of other homohexameric ATPases.

What is the molecular mechanism of nucleotide exchange, and how is it influenced by KaiA? How does the phosphorylated S431 (pS431) form of CII allow ADP accumulation in CI? In principle, molecular dynamics (MD) simulations can provide atomic-resolution answers to these questions. However, the size of the system and the timescale of nucleotide exchange prohibit direct treatment by conventional methods. To overcome this challenge, we implemented a pipeline of methods for enhancing the sampling of molecularly rare events. The results of our simulations provide a detailed characterization of the structure and energetics of the ADP release pathway in CII and give new insight into how KaiC conformational changes during nucleotide release are coupled to KaiA binding.

Implications for other homohexameric ATPases are discussed.

3.3 Results

3.3.1 Nucleotide release in CII involves both local and intersubunit conformational changes

Our goal was to simulate nucleotide release and to understand how it is coupled to changes in KaiC structure. In KaiC, the central channel is sufficiently large to accommodate a free nucleotide, so it is not immediately obvious from the crystal structure whether the nucleotide exits radially inward or outward. We thus performed the simulation in stages that progressed from exploratory and qualitative to physically well-controlled and quantitative, as follows. We first determined the general direction of release using a method known as locally enhanced sampling [157–159], which floods the protein with non-interacting copies of Mg·ADP that weakly interact with the protein and activate local conformational dynamics to accelerate the release. Given the general direction of release, we then sought to identify a manageable number of collective variables (CVs) that could be used to characterize and control sampling along the pathway. We used these CVs to generate an initial guess for the release pathway with steered molecular dynamics (SMD) [160], which introduces an artificial force to drive the release. We then used the string method [161, 162] to refine this guess to a minimum free energy path without artificial driving forces. Finally, the free energy of the release pathway was characterized using umbrella sampling [163]. While our conclusions are based only on these final simulations, to explain the CVs that we use to analyze the dynamics, we briefly review the findings of the full series of calculations (further detailed in Methods) in the remainder of this section and the following one.

In both CI and CII, the active sites are located at the subunit interfaces, for a total of 12 sites per hexamer. Our calculations are based on a WT KaiC crystal structure [Protein

Data Bank (PDB) ID: 3DVL] [164], which we modified to be in the phosphorylated T432-only (pT432) form in all six subunits and with all bound ATPs replaced by ADPs in the CII domain. Because the structure has almost perfect C_6 symmetry, we arbitrarily chose the active site at the interface of subunits D and E for the release simulations. The locally enhanced sampling simulations show that Mg·ADP is released radially outward (Fig. 3.1 and Fig. S3.6); this is consistent with further analyses of the alternative possibility of inward release, which appears highly unfavorable (see Supplementary Results and Fig. S3.7). For steric reasons the outward release requires an increase in the distance between the β-hairpin at residues 468 to 483 (which we denote β_{nt}) and an α-loop-α motif ($\alpha l \alpha$) at residues 321 to 342, both of which are located near the exit point of the release pathway (Fig. 3.1B and Fig. S3.8). It is worth noting that β_{nt} is directly connected to the A-loop (residues 486 to 498), which is implicated in KaiA binding. For these reasons, we included the distance between the β_{nt} and $\alpha l \alpha$ centers of mass as a CV.

To enable global structural changes, we also selected as CVs the six angles formed by each pair of adjacent subunits in the CII ring (Fig. 3.1C Inset). This choice of CVs is motivated by two observations. First, SAXS experiments indicate that the radius of gyration of KaiC changes depending on the phosphorylation state of the protein, and the variations in this parameter are attributed to contraction and expansion of the CII ring [165]. Second, other multimeric nucleotide-binding proteins have been shown to exhibit changes in solvent accessibility upon nucleotide binding at subunit interfaces [151, 152, 166–173]. Two final CVs that we used to parameterize the release pathway were the distances between the center of mass of the ADP and the centers of mass of the active-site residues of the D and E subunits, respectively. We note that the two adjacent subunits do not contribute equally to the active site; specifically, the nucleotide makes more extensive contact with subunit D (Fig. 3.1D and Fig. S3.9). A summary of the CVs is given in Table S3.1.

Among the nine CVs that we identified, we found that the subunit D–E angle

and ADP–subunit D distance provided the best control in SMD, and we used them to construct an initial guess for the release pathway. Specifically, we opened the intersubunit angle, moved the Mg·ADP to increase the distance, and then closed the angle in a stepwise manner. This pathway was then refined using the string method in the full nine-dimensional CV space (Fig. S3.10). In the refined pathway, the intersubunit angle opens and the Mg·ADP dissociates in a concerted fashion (Fig. 3.1C). With the exception of the subunit A–B angle, which is diametrically opposed to the vacated active site, the angles between the rest of the subunits are compressed, especially the subunit E–F angle. This implies that ADP release in CII does not occur concurrently around the ring. No clear trends are observed for the CI intersubunit angles (Fig. S3.11).

In summary, our simulations show that Mg·ADP dissociation from CII is coupled to global conformational changes in that domain, which involves an asymmetric opening of the subunit angles. Below, we further analyze the pathway from the string method to understand the energetics of nucleotide release.

3.3.2 The energetics of Mg·ADP release are dominated by salt bridge interactions

We used umbrella sampling to characterize the free energy in the (extended) neighborhood of the string pathway. In this method, copies of the system are restrained to different values of the CVs to enforce sampling, and the contributions from different copies are then reweighted to obtain the correct equilibrium statistics (see Methods). Because the computational cost of umbrella sampling grows exponentially with the number of CVs, we again focused on the two-dimensional space defined by the ADP–subunit D distance and the subunit D–E angle. Care was taken to ensure convergence, as discussed in Methods.

The free energy as a function of these CVs is shown in Fig. 3.2A. As the ADP–subunit D distance increases from 10 to 13 Å, there is a steep rise in the free energy to a plateau that is about 10 kcal/mol above the bound state, consistent with the fact that

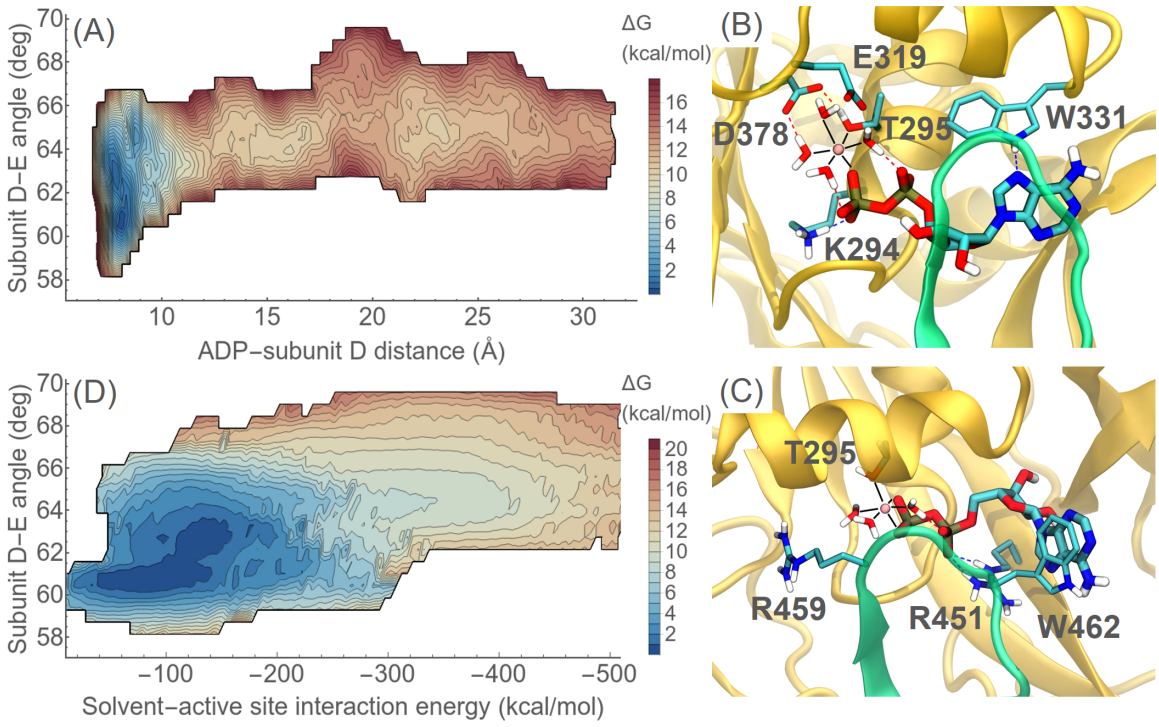


Figure 3.2: Electrostatic interactions dominate the energetics of nucleotide release. (A) The free energy profile of the release process, calculated using umbrella sampling. (B) and (C) show the interaction between Mg·ADP and key active site residues. The solid lines show the coordination of the Mg^{2+} ion, and the dotted lines represent hydrogen-bond interactions. Colors are the same as in Fig. 1D. (B) A representative nucleotide-bound structure, from (8.5 Å, 60.2°) in (A). (C) A representative structure from the region of rapid increase in free energy, specifically (10.4 Å, 64.0°) in (A). (D) Alternative projection of the umbrella sampling data in (A). The CVs are the subunit D–E angle and the electrostatic interaction energy between solvent and active-site residues K294, T295, E319, and D378.

nucleotide binding is required for the hexamerization of KaiC [16]. This rise coincides with the disruption of electrostatic interactions in the active site: the sidechains of T295, E319, and D378 directly and indirectly determine the coordination of the Mg^{2+} ion, and the sidechain of K294 makes a salt bridge to the terminal phosphate group of the ADP (Fig. 3.2B). Among these residues, K294 and T295 are Walker A residues, D378 is a Walker B residue, and E319 is a catalytic glutamate [39]; the importance of such conserved residues for nucleotide binding and catalysis is well-established for many AAA+ and RecA-like ATPases [155]. The free energy plateau for ADP–subunit D distances greater than 10

Å is centered on subunit D–E angle values of about 64°, compared to 60° observed in the crystal structure. Although this opening permits additional water molecules into the active site, maximum solvation is not achieved until after loss of the protein-nucleotide interactions (Fig. S3.12).

In addition to the interactions listed above, two arginine residues on the periphery of the active site also exhibit strong electrostatic interactions with Mg·ADP for ADP–subunit D distances up to 16 Å (Fig. 3.2C). R451 forms salt bridges to each phosphate group in turn as the ADP makes its way out of the protein. The interaction of Mg·ADP with R459 from subunit E contrasts with those discussed so far in that it is net repulsive. Both R451A and R459A abolish oscillation in vivo [174], confirming the importance of these two residues. In particular, R459 is an arginine finger, a common motif required for sensing the presence of γ-phosphate groups during ATP hydrolysis [155, 174].

In all the nucleotide release simulations that we performed, the Mg^{2+} ion and ADP dissociated from the active site together, even though no restraint was applied to enforce either the Mg–ADP or Mg–active site distance. However, it has been suggested for F_1-ATPase that the Mg^{2+} ion leaves prior to ADP dissociation [175], so we also considered such a two-step release mechanism. To test this possibility, we ran a free energy calculation to determine the standard binding free energy of the Mg^{2+} ion to a CII active site with bound ADP using the double annihilation method [176] (see Methods and Figs. S3.13 and S3.14). The result of the calculation suggests that the standard binding free energy of Mg^{2+} is at least –34 kcal/mol, which is very strong. We thus disfavor the two-step release mechanism.

Mg·ADP release is regulated by tryptophan residues. Two tryptophan residues close to the protein surface also make extensive contact with the adenine base throughout the release process (Figs. 3.2B and 3.2C). The sidechain of W331 forms hydrogen bond interactions first with the adenine base, and then with the phosphate groups as ADP moves out of the active site. The geometry in proximity of W462 is suggestive of π-

stacking, although the force field does not explicitly include such effects. Previous experiments point to the importance of these residues; the mutations W331F and W462F lead to abnormal oscillations in vitro [79], and W331A traps KaiC in a constitutively phosphorylated state [60].

Evolutionarily, W331 and W462 are highly conserved among KaiC proteins found in KaiABC clusters in the cyanobacterial phylum, a clade known as KaiC1 (Table S3.1) [14]. W331 is also conserved among KaiC3 homologs, which are functionally poorly characterized [177]. Because of the proposed role of KaiA as a nucleotide exchange factor [19] and the observation that KaiA-mediated phosphorylation has only been found among KaiC1 homologs [14], the simulation results are consistent with the notion that the tryptophan residues, in particular W462, may play a role in KaiA-regulated nucleotide exchange in the CII domain.

3.3.3 Subunit angle motion is coupled to quaternary conformational changes

In the crystal structure of the KaiC protein employed in the present study, all six CII intersubunit angles assume equilibrium values around 60° (the difference between the largest and smallest angle is no more than 0.3°). The umbrella sampling calculation, however, reveals an additional asymmetric metastable state where the subunit interface from which the nucleotide is released opens to approximately 64°, while the nucleotide remains bound (Fig. 3.2A). This metastable state is separated from the (essentially symmetric) global minimum by a free energy barrier that is less than 2 kcal/mol, indicating that interconversion between the states is thermally accessible. To understand the significance of the large-angle asymmetric state, we projected the umbrella sampling data onto a new CV space consisting of the subunit D–E angle and the electrostatic interaction energy of water with the active site residues discussed above (K294, T295, E319, and D378) (Fig. 3.2D). This free energy surface indicates that the subunits

with a larger angle allow stronger interactions between active site residues and solvent, facilitating nucleotide release. The existence of asymmetric quaternary states that are thermodynamically stable in CII is consistent with the observation of a mixture of loose and tight subunit interfaces in a high-resolution CI crystal structure (PDB ID: 4TLA) [156], although the largest angle observed therein is only 61.7°.

Given the existence of multiple metastable states distinguished by the intersubunit angles, we sought to determine if there are additional large-scale quaternary conformational changes that are associated with the release process but are not described by the angles. To detect such motions, we applied principal component analysis (PCA) to the Cα atoms in the structures obtained from umbrella sampling, without reweighting, after aligning all structures with respect to the CII domain of subunit D. The first two principal components, PC1 and PC2, account for most of the variance (43% and 28%, respectively) (Fig. S3.15A). Structures representative of the extremes of these two PCs are shown in Fig. 3.3A and B. PC1 represents the formation of a split washer structure with a helical pitch up to about 11 Å (Fig. 3.3C) in the CII domain, and a concomitant rigid-body tilt in the CI domain. The split in the CII ring is located at the subunit D–E interface, the site of Mg·ADP dissociation in the simulation. PC2 represents a concerted, circular compression motion in both domains in response to the angle-widening motion at the subunit D–E interface.

Given that the PCs are correlated with the subunit D–E angle, which possesses at least two metastable states in the nucleotide bound state (Fig. 3.2A), we hypothesized that motions in the direction of the PCs are accessible without nucleotide release. To test this hypothesis, we measured motion along the aforementioned PCs in a 100 ns SMD simulation of the bound state, with the artificial force in the direction of subunit D–E angle widening (Fig. S3.15B). Although the PCs explore a smaller range of values, the intersubunit angle variations in the nucleotide bound state clearly result in changes to the first two PCs, which remain the most dominant modes.

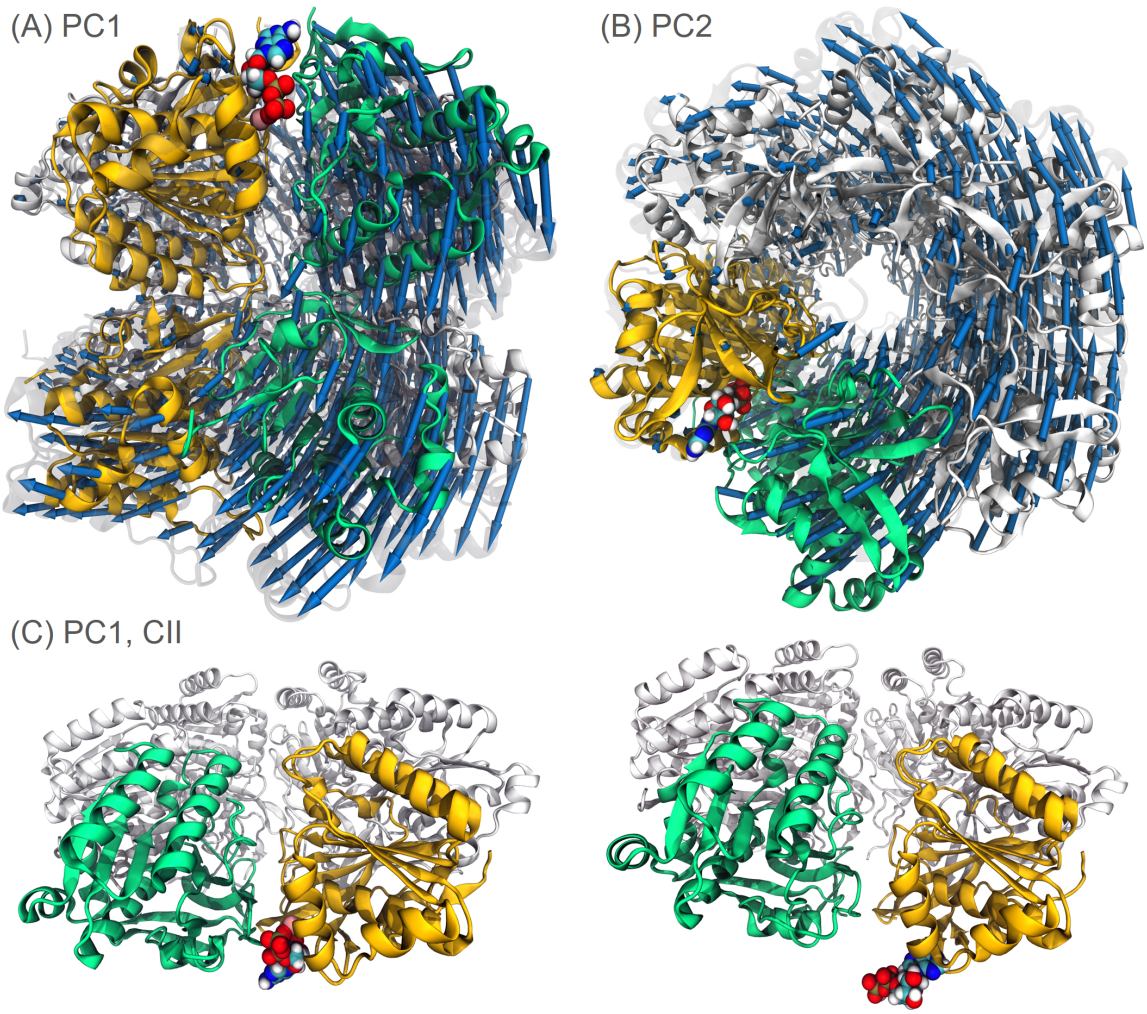


Figure 3.3: Principal component analysis (PCA) reveals large-scale conformational changes during nucleotide release. (A) Principal component (PC) 1 represents the formation of a split washer structure in CII, and (B) PC2 represents a circular compression in both domains. In (A) and (B), arrows show the Cα difference vectors between a pair of structures representing the extremes of motion along the PC, aligned to minimize the difference in the CII domain of subunit D. Structures in (A) are chosen to have similar PC2 values, and structures in (B) are chosen to have similar PC1 values, such that the differences shown approximate motions in only the PC of interest. (C) Conformations of the CII domain in the low PC1 (left) and high PC1 (right) structure, respectively. The side-by-side comparison indicates that the PC1 eigenvector points in the direction of increased helical pitch in the CII domain.

A-loop conformation is coupled to the intersubunit angle. We now use our simulation results to elucidate possible mechanisms of KaiC regulation. As discussed in the Introduction, an outstanding question is how KaiA promotes phosphorylation. Previous

work suggests that KaiA binding stabilizes structures known as the A-loops (residues 486 to 498, preceding the C-terminal tails of KaiC) in an extended conformation [60, 178]. Thus, we explored whether the umbrella sampling simulations resulted in structures with extended A-loops, with a view toward understanding the impact of KaiA binding on the KaiC structure.

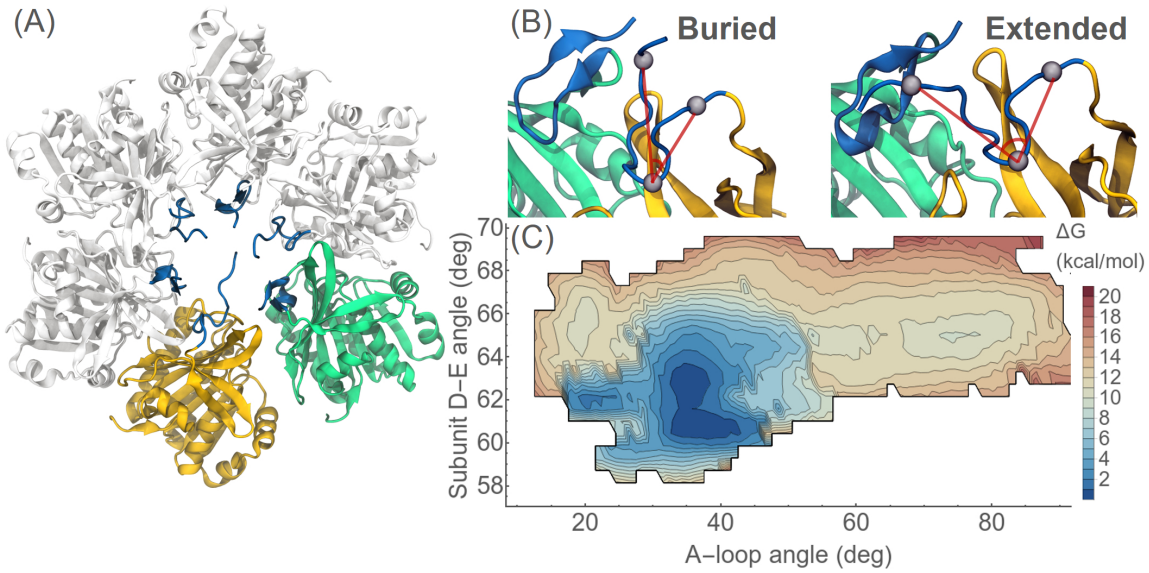


Figure 3.4: The A-loop (residues 486 to 498) is coupled to nucleotide release. (A) The A-loop (blue) of subunit D (yellow) samples an extended conformation. (B) The structure of the subunit D A-loop in buried and extended states can be quantified using the angle formed by the Cα atoms at residues 486, 490 and 496, which are represented as gray spheres. The values of this A-loop angle are 33.6° and 75.8° in the buried and extended states shown in (B). The subunit D–E angle values in the buried and extended structures are 62.9° and 65.7°, respectively. (C) Projection of the umbrella sampling data onto the CV space consisting of the A-loop angle and the subunit D–E angle shows that a more extended subunit D A-loop conformation is correlated with a larger subunit D–E angle.

We observed that the A-loop in subunit D sampled an extended conformation during nucleotide release (Fig. 3.4A), albeit to a lesser extent than that in an NMR structure of a KaiA fragment in complex with the KaiC C-terminus [60]. We only observe this state in subunit D (Fig. 3.4A and Fig. S3.16), which suggests that it is a specific response to Mg·ADP dissociation. To quantify this behavior, we calculated the angle formed by three atoms at the beginning, middle, and end of the A-loop (Fig. 3.4B) and projected the free

energy onto this A-loop angle and the subunit D–E angle. The resulting surface shows that the A-loop is more likely to sample the extended state as the subunit D–E angle widens (Fig. 3.4C). Because we expect allostery to be bidirectional (reflecting the fact that free energies are state functions) [179], this finding is equivalent to A-loop extension favoring a larger intersubunit angle, or, taking this logic one step further, KaiA binding favoring a larger intersubunit angle. Given the importance of the intersubunit angle for the energetics of nucleotide release, detailed in previous sections, the coupling of KaiA binding to intersubunit angle opening may mechanistically explain its ability to act as a nucleotide exchange factor.

The simulation hints at two other potential mechanisms that could couple the A-loop to the active site. First, the A-loop may be both sterically and electrostatically coupled to the active site via a β-hairpin (residues 436 to 459, which we denote β_{AL}) (Fig. S3.17A). β_{AL} is located directly between the active site and the A-loop and harbors R451, which can interact electrostatically with the ADP, as well as E444, which forms a salt bridge with R496 in the A-loop. More specifically, β_{AL} is involved in contraction of subunit D (Fig. S3.17B and C) as the ADP–subunit D distance reaches around 10–11 Å, a point at which the free energy increases sharply (Fig. 3.2A). This contracted state is characterized by 1) a subtle movement of $\alpha l\alpha$, β_{AL}, and the A-loop, 2) a decrease in the distance between the R451 sidechain on β_{AL} and the α-phosphate group on ADP from approximately 8.6 Å to 4.5 Å, and 3) a remodeling of the release channel due to the nucleotide movement. Among these conformational changes, the movement of ADP towards R451 has the most obvious consequences for the energetics of the release pathway. β_{AL} thus connects the binding affinity of ADP (through the R451–ADP interaction) with the A-loop conformation.

Second, the simulation suggests that A-loop conformation is coupled to the active site phosphorylation state through the 422-loop [180] (Fig. S3.18), so called because the A422V mutation alters the amplitude and phase-resetting behavior of the oscillator in

vivo [181]. Our results suggest that the 422-loop couples to the active site via a water-mediated hydrogen-bond network between pT432 and D417, a residue on the 422-loop. This interaction, however, can be disrupted as the 422-loop moves away from the active site and forms hydrophobic contacts with the A-loop. In the release simulation, these two states appear to exist in a dynamic equilibrium, which presumably can be shifted by the phosphorylation state of KaiC and the binding of KaiA, thus making the 422-loop a potential feedback mechanism that communicates the KaiC phosphorylation state back to the A-loop to modulate KaiA binding. This idea is consistent with the observations that, even in the absence of KaiB, KaiC phosphorylation is less sensitive to KaiA in the pS431 state [21] and the binding affinity of KaiA to KaiC is reduced in the pS431 and doubly phosphorylated (pT432/pS431) states [66].

3.3.4 CI/CII interaction is mediated by a CI β-hairpin

As discussed in the Introduction, KaiB binding to CI is an important feedback process that allows oscillations to occur. The formation of the KaiB·KaiC complex appears to be determined by the retention of ADP in CI after hydrolysis [77, 78] and is critically dependent on the S431 phosphorylation of the distal CII domain [17, 21, 79]. How the phosphorylation state in CII is communicated to CI to allow KaiB binding is a major unresolved question. Based on gel filtration experiments that monitor the binding of separated CI and CII domains with phosphomimetic mutations, it has been suggested that, when CII is in the pS431 state, it "stacks" more tightly with CI, even in the absence of the linker that tethers the two domains together [67]. Because ADP does not accumulate in the isolated CI domain, but ATP hydrolysis still occurs [19], a plausible hypothesis is that phosphorylation-dependent interactions between CII and CI inhibit nucleotide release from CI. We thus examined the CI/CII interactions in our umbrella sampling simulation data to determine if there were structures that could provide hints about how information is communicated between CI and CII.

We observed changes in linker-independent CI/CII contact, as measured by their van der Waals interaction energy (Fig. S3.19A). We specifically did not include the linker in the energy calculation to mimic the CI/CII domain stacking experiment [67]. Because the PCA results, detailed in a previous section, revealed motions between the two domains, especially along PC1, we hypothesized that the CI/CII contact energy should change with the values of the PCs. However, we found that its variance could not be explained by projection onto the ten largest PCs. Instead, changes in CI/CII contact energy are largely mediated by interactions between the CII domain and six β-hairpin structures (residues 203 to 226, which we denote $\beta_{\mathrm{I/II}}$), one in each KaiC subunit (Fig. S3.19B). $\beta_{\mathrm{I/II}}$ is part of the β-sheet structure at the hydrophobic core of each CI globular domain, and its secondary structure can become frayed near the turn region. When the two domains separate, the turn region of $\beta_{\mathrm{I/II}}$ moves away from the CII domain while leaving the rest of the CI/CII interface largely unchanged (Fig. S3.19C and D). Salt bridges connect neighboring $\beta_{\mathrm{I/II}}$ hairpins: R216–E221 is stable throughout our simulations, and E214–R217 is favored by reduced CI/CII contact. The simulations thus suggest new targets for further study of CI/CII interactions.

3.4 Discussion

Our simulations can be synthesized into a new interpretation of the molecular events in the oscillator cycle. Near dawn, KaiC is largely unphosphorylated [12]; while phosphorylation can occur, it is not sustained because the ADP retained in the CII active site allows rapid dephosphorylation [38, 39]. As KaiA is released from inhibitory complexes, it binds the A-loop and stabilizes an extended conformation [60], favoring the opening of subunit interfaces and exchange of ADP for ATP in solution. This shifts the equilibrium towards phosphorylation which occurs first on T432 and then on S431 [17]. During this part of the cycle, CI slowly hydrolyzes ATP [156], but the product ADP molecules can be rapidly exchanged with ATP in solution [19], possibly because the CI/CII linker is in

a favorable conformation similar to the extended A-loop (see discussion below). Phosphorylation on T432 favors A-loop extension, encouraging further phosphorylation [67]. However, subsequent phosphorylation of S431 eventually results in conformational changes in the CI/CII linker and the $\beta_{\mathrm{I/II}}$ hairpin, among others, that restrict the ability of CI to release ADP. Inhibition of ADP release in CI when S431 is phosphorylated may explain why phosphomimetic mutants show decreased ATP hydrolysis at steady state [26]. As CI ATPase activity causes ADP to accumulate, the B-loop adopts an extended conformation that allows KaiB to bind [77]. KaiB cooperatively assembles in a ring-like structure on the CI domain [78], which may further restrict nucleotide release from CI. KaiB bound to KaiC adopts an alternative fold which can then bind to and inhibit KaiA [22, 77, 78]. Sequestration of KaiA limits nucleotide exchange in CII, and CII tends to dephosphorylate [12], first T432 and then S431 [17], returning KaiC to the beginning of the cycle.

This model raises the question of how Kai systems that lack KaiA, such as those from the marine *Prochlorococcus* cyanobacterial species [182], are able to build up their KaiC phosphorylation levels [183]. A plausible answer comes from the fact that such KaiC sequences are truncated near the A-loop, compared with *S. elongatus* KaiC [183]. Previous experiments on *S. elongatus* KaiC demonstrated that truncating the A-loop can result in KaiA-independent, constitutive hyperphosphorylation [60], which suggests that an analogous mechanism may allow *Prochlorococcus* KaiC to autophosphorylate without KaiA. Our simulations thus make the prediction that Kai oscillators that lack KaiA should have KaiCs with active sites that are more accessible to molecules in solution, obviating the need for KaiA-like nucleotide exchange factors. Relative solvent accessibilities of different KaiCs should be measurable with hydrogen-deuterium exchange mass spectroscopy [23].

A striking feature of the simulation structures is their broken C_6 symmetry, especially the split washer and ring compression motion observed in the PCA. This is consistent with a recent crystal structure of the CI domain with a mixture of bound ATP and ADP,

which resulted in heterogeneous subunit structures that correlate with the hydrolysis states of the active sites [156]. In addition to nucleotide binding, KaiA binding is another potential source of KaiC asymmetry. Biochemical studies suggest that one dimer of KaiA is sufficient to stimulate the phosphorylation of a KaiC hexamer [134]. Given that one KaiA dimer can bind to at most two KaiC subunits [60, 61], it likely activates only part of the ring at a time. This model is consistent with the asymmetric conformations observed for other ring-shaped oligomeric ATPases, especially those in the AAA+ and RecA/DnaB superfamilies [155, 184, 185].

The split washer structure of KaiC is also found in AAA+ ATPases such as the bacterial replicative helicase DnaB [186], the Simian virus 40 replicative helicase Ltag [187], as well as NSF, an eukaryotic protein involved in SNARE complex disassembly during membrane trafficking [152]. In DnaB, NSF and KaiC, the split in the hexameric ring is located at an apo subunit interface, suggesting that the split interface is implicated in nucleotide exchange. The parallel between KaiC and NSF is especially striking, because similar to KaiC, the D1 and D2 domains of NSF also assemble into double-tiered ring structures with distinct functions. These split washer structures are reminiscent of the helical structures of the RecA filament, where the helical pitch is dependent on the states of the nucleotides bound to the subunits [166, 168]. However, the helical pitches observed in the non-filamentous proteins are much smaller. Taken together, it is likely that the hexameric ATPases can access similar quaternary conformational spaces despite their significant sequence and functional divergences. This idea raises the possibility that KaiC may be able to access other symmetry-broken states that are not explored in our simulations but are observed in other hexameric ATPases, such as a conformation with C_2 symmetry (i.e., a dimer of trimers), observed in the unfoldase ClpX [151], or a conformation with C_3 symmetry (i.e., a trimer of dimers), observed in the replicative helicase DnaC of *Geobacillus kaustophilus HTA426* [188]. Analysis of cryo-EM data without symmetry averaging may reveal such states. Alternatively, a bis-His motif [189] or

disulfide bond can be engineered at the subunit interface to limit the angle motion, which should lead to a decrease in the nucleotide release rate.

An outstanding question for KaiC is whether its asymmetric features propagate around the ring in an ordered fashion, as is believed for F_1-ATPase, or in a stochastic fashion, as is believed for ClpX [190]. An indirect way of distinguishing these possibilities would be a mutant doping [132] or covalent linking [190] experiment where WT and CII-catalytically-dead subunits are mixed in a single KaiC hexamer. Because a stochastic mechanism implies independent catalytic cycles at individual active sites, the hybrid protein should still be able to function in the full oscillator. On the other hand, a sequential/rotary mechanism would predict abolition of oscillation, because the dead subunits would prevent propagation of activity around the ring. Indeed, it is already known that catalytically dead CII subunits can be efficiently phosphorylated when mixed with WT CII [110].

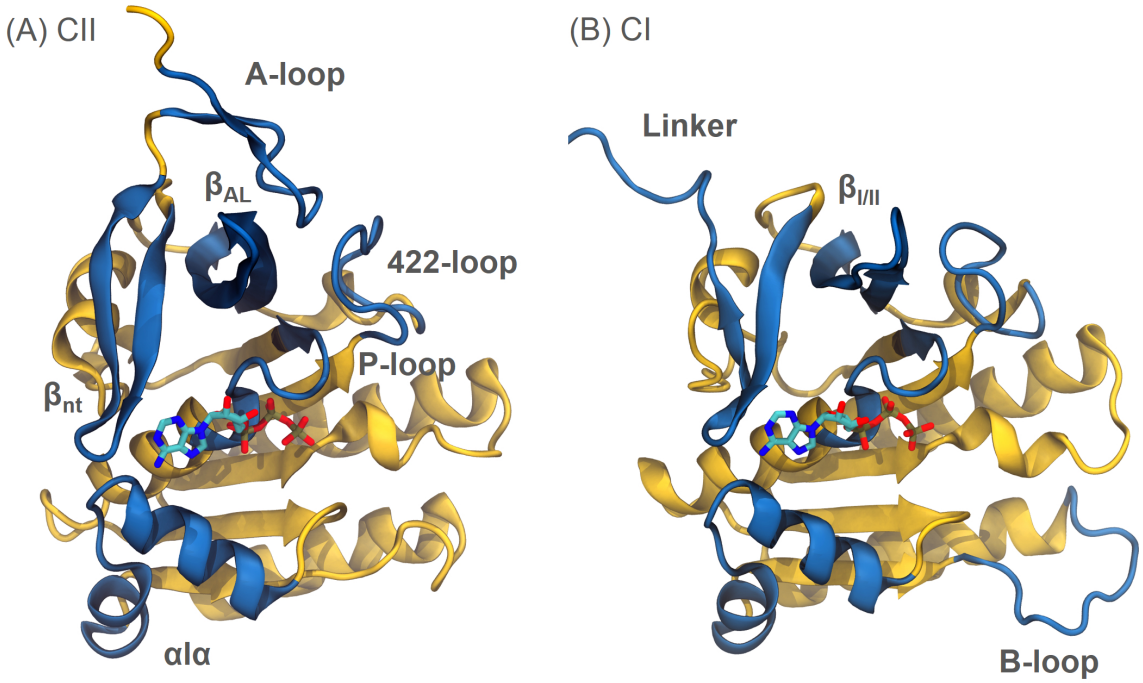


Figure 3.5: CI and CII share similar conformations. In (A) and (B), Secondary structural elements identified by the simulations as important for nucleotide release from CII (A) and their homologs in CI (B) are colored in blue. In each case, the structure shown is the domain from the D subunit of the crystal structure (PDB ID: 3DVL).

We also speculate here on how our analysis of the nucleotide release pathway in CII may shed light on the regulation of nucleotide release in CI. Given that the CI/CII double ring structure likely arose as a result of gene duplication [150], it is unsurprising that the CI and CII structures share many similarities (Fig. 3.5). In particular, the CI structural element homologous to the A-loop is the linker between the CI and CII domains, which adopts a conformation in the crystal structure that is reminiscent of an NMR A-loop structure in complex with KaiA [60, 61]. Given the importance of the A-loop in regulating nucleotide release in CII, this apparent analogy raises the possibility that CII may act on CI through the linker in a way that is similar to how KaiA acts on CII through the A-loop. If KaiA is a nucleotide exchange factor for CII, then it follows that CII may act as a nucleotide exchange factor for CI and regulate its intersubunit angle opening. This proposal makes the prediction that when CII is in a pS431 state that favors KaiB binding, the CI/CII linker may adopt an alternative conformation that is more similar to the buried A-loop in CII. Furthermore, the βAL and $\beta_{\mathrm{I/II}}$ hairpins, two structural features that our simulations suggested deserve further study, are also homologous in the two domains (Fig. 3.5). For example, it would be interesting to disrupt contacts at the CI/CII interface by mutating residues in the R216–E221 and E214–R217 salt bridges, in the hope of decoupling the two rings from each other, which should result in dysregulation of KaiB binding and/or KaiA sequestration. Because β_{AL} likely couples the A-loop to the active site, this parallel suggests that there may be unexplored similarities between the mechanism of CI/CII interaction and KaiA·KaiC interaction.

3.5 Conclusions

In this paper, we implemented a pipeline of enhanced sampling methods for molecular dynamics to elucidate and quantitatively characterize the pathway for ADP release from KaiC, a homohexameric ATPase that serves as the central component of the cyanobacterial circadian clock. Nucleotides bind at the interface of KaiC subunits, and

the energetics of its release are dominated by the interplay of active site solvation with electrostatic interactions among charged active site residues, the Mg^{2+} ion, and the ADP phosphate groups. The calculations suggest that the C_6 symmetry of the hexamer is broken during operation, with the interface opening in a split washer fashion. We identify specific interactions that couple this opening to extension of the A-loop, providing a molecular mechanism for the role of KaiA as a nucleotide exchange factor for KaiC. Parallels between structural elements in the two homologous domains of the protein suggest the intriguing possibility that CII may serve as a nucleotide exchange factor for CI. Overall, the information that the simulations provide about structures accessible to KaiC reveal fruitful connections to other ATPases.

3.6 Methods

3.6.1 System setup

The initial coordinates of *S. elongatus* KaiC with bound ATP and Mg2+ were obtained from the Protein Data Bank (PBD ID: 3DVL) [164]. The structural water was retained, and the C-terminal tail (residues 499 to 519) in each subunit was deleted to reduce the size of the system. In the 3DVL structure, which has a resolution of 2.8 Å, no Mg^{2+} was present in the CI ATP binding sites. We introduced Mg^{2+} into all six CI nucleotide binding sites using the Mg^{2+} coordinates from a 1.94 Å resolution KaiC CI structure (PDB ID: 4TL7) [156] after aligning the ATP molecules modeled in each chain in the two structures. In 4TL7, the CI Mg^{2+} is coordinated by T53Oγ1, Oγ3 and Oβ2 of ATP, and three water molecules, which are absent in 3DVL. To complete the octahedral Mg^{2+} coordination geometry, we added three water molecules to each CI active site based on the 4TL7 crystal water positions and changed the coordinates of the three phosphate groups in ATP to those in 4TL7. We used a similar procedure to model Mg·ADP in the CII nucleotide binding sites, using the Mg·ADP coordinates in a 1.8 Åresolution, ADP-

containing CI structure as a reference (PDB ID: 4TLA) [156].

The subunits in the 3DVL crystal structure are in the doubly phosphorylated (pT432/pS431) state. In all subunits, we deleted the phosphate group on S431 and retained the phosphate group on T432 to mimic the pT432 phosphorylation state. Together, the nucleotides and phosphorylation states mimic a KaiC with CI in the pre-hydrolysis state and CII in a kinase product state during the initial phosphorylation phase of the oscillator.

The model system was solvated in a cubic box of TIP3P water molecules, and the final dimensions of the box were 132 Å × 132 Å × 132 Å. The system was neutralized, and then brought to a final concentration of 150 mM KCl, which represents 194 K^+ ions and 122 Cl^- ions. The total number of atoms in the system was 216,892. The system was setup using CHARMM-GUI [191–193]. All molecular visualizations were made in VMD 1.9.2 [194]. Further details on the system setup are provided in Supplementary Methods.

3.6.2 Minimization and equilibration

Unless otherwise specified, all MD simulations were performed using GROMACS 5.1.4 [195], with the CHARMM36 force field parameters for the protein (with CMAP correction) [196–198], TIP3P water [199], ions [200], and the nucleotides [201]. The energy of the system was minimized with the steepest descent method, until the largest force in the system was smaller than 1000.0 kJ/mol·nm. The system was equilibrated in the NVT ensemble for 25 ps with a 1 fs integration timestep, followed by the NPT ensemble for 10 ns with a 2 fs integration timestep. The positions of all non-hydrogen atoms in the protein and its ligands were constrained during energy minimization and the NVT equilibration runs to stabilize the protein structure and enforce the octahedral Mg^{2+} coordination geometry. The temperature of the simulations was set at 303.15 K using the modified Berendsen thermostat with a time constant of 1.0 ps. For the NPT simulations, the pressure was set at 1 bar using Parrinello-Rahman coupling with a time constant of

5.0 ps. Periodic boundary conditions were employed, and the particle-mesh Ewald method [202] was used to calculate electrostatic forces with a real space cut-off distance of 1.2 nm. The Lennard-Jones forces were smoothly switched off from 0 to 1.2 nm. All bonds to hydrogen atoms were constrained using the LINCS algorithm [203] in the NVT and NPT simulations.

3.6.3 Locally enhanced sampling

Locally enhanced sampling [157–159] was used initially to explore the structural changes at the active sites during nucleotide release (Fig. 3.1B and Fig. S3.6). All locally enhanced sampling simulations were performed using the implementation available in NAMD 2.9 [204] with CHARMM36 force field parameters. The structure and topology files required for the simulations were generated using the psfgen plugin for VMD 1.9.2 [194]. Specifically, each of the six Mg·ADP bound to CII active sites were duplicated 200 times. The Mg·ADP copies at each active site do not interact with each other, while their intra- and inter-molecular interactions were scaled down by a factor of 200. To preserve the structural integrity of the ADP molecules, the force constants for all bond, angle, dihedral angle and improper dihedral angle terms for ADP were manually multiplied by 200; in addition, an artificial bond with a force constant of 60,000.0 kcal/mol·Å^2 and equilibrium bond length of 2.91 Å was added between the P and P3 atoms to prevent the α- and β-phosphate groups from collapsing onto each other. To preserve the interaction between each Mg^{2+} and its associated ADP, another artificial bond with a force constant of 40,000.0 kcal/mol·Å^2 and equilibrium bond length of 2.50 Å was added between the P3 and MG atoms. Further details on the simulation setup are provided in Supplementary Methods.

3.6.4 Steered MD (SMD)

Constant-velocity SMD simulations [160] were done using the moving restraint functionality in PLUMED 2.3 [205], which applies a time-dependent bias of the form $U(x) = \frac{1}{2}\kappa(x - x_0(t))^2$ to a given CV x; here κ is the force constant, and $x_0(t)$ is the center of the biasing potential. The collective variables used in the SMD and subsequent string and umbrella sampling calculations are defined in Table S3.1.

As described in Results, we used SMD to construct an initial pathway for the string calculations; this pathway was generated in three steps. First, to open the intersubunit angle, a harmonic bias with a force constant of 100,000 kJ/mol·rad^2 was applied to the subunit D–E angle, and the center of the potential was moved linearly, over the course of 1 ns, from 1.04 rad to 1.23 rad. Second, to pull out the ADP, a harmonic bias with a force constant of 1,000 kJ/mol·nm^2 was applied to the ADP–subunit D distance, and the center of the potential was moved linearly from 0.8 nm to 3.0 nm over the course of 1 ns. During the ADP pulling simulation, the subunit D–E angle was restrained at 1.23 rad. Finally, the subunit D–E angle was closed from 1.23 rad to 1.13 rad using the same protocol as that in the first step, while the ADP–subunit D distance was restrained at 3.0 nm.

We note that splitting the active site residues by their subunit membership and steering on ADP–subunit D distance and/or ADP–subunit E distance independently provided better control of nucleotide position in SMD simulations (and subsequent string and umbrella sampling simulations) than a single CV measuring distance to the center of mass of all active site residues. This is because the value of the intersubunit angle variable affects the overall center of mass of the active site residues; tracking the distances to the residues in the two subunits separately largely decouples the intersubunit angle from the progress of the Mg·ADP out of the protein.

3.6.5 String method

To initialize the images of the string, we selected 24 structures that are roughly equally spaced along the SMD pathway in the CV space of subunit D–E angle and ADP–subunit D distance and then equilibrated them for 5 ns with restraints in the same CV space. The pathway was then refined by the string method with fixed end points, as described in [161], implemented as an in-house Python wrapper for GROMACS and PLUMED. Briefly, at each iteration of the calculation, each structure was equilibrated with a harmonic restraint centered on the images for 20 ps. The string path was then updated via

$$x_i^n(\tau + \Delta\tau) = x_i^n(\tau) - \Delta\tau \sum_{j=0}^{8} P_{ij}(x^n(\tau)) \frac{\partial F(x^n(\tau))}{\partial x_j^n}$$

Here, the ith CV evaluated at the nth image is denoted as $x_i^n(\tau)$, and τ tracks the progress of the gradient descent in the CV space. The local gradient ∇F was computed as the average displacement from each image during a 50 ps production run, scaled by the force constants. The matrix P removes the component of ∇F parallel to the string at each image. The string was propagated with a constant multiplier $\Delta\tau = 0.0001$. The updated string is linearly interpolated and smoothed before the next iteration. We did not apply any correction for non-linear dependences between the CVs.

The string calculation employed nine collective variables, described in Table S3.1. The force constants used in the production runs were 10,000 kJ/mol·rad^2 for the angle variables and 2,000 kJ/mol·nm^2 for the distance variables. The force constants used in equilibration were twice as large. We found that 30 iterations of the string simulations were sufficient for generating a qualitatively reasonable pathway, which we used to initialize the umbrella sampling (see Supplementary Methods, Figs. S3.20 and S3.21).

3.6.6 Umbrella sampling and analysis

We performed umbrella sampling calculations [163] using the harmonic restraint functionality in PLUMED 2.3. Images from the string simulations were used as starting structures, and additional windows were added iteratively using structures generated from existing windows until overlapping coverage was achieved for the free energy basin surrounding the minimum free energy pathway in the two-dimensional subspace of the ADP–subunit D distance and subunit D–E angle. The force constant on the ADP–subunit D distance ranged from 2,000 to 60,000 kJ/mol·nm^2, and the force constant on the subunit D–E angle ranged from 10,000 to 600,000 kJ/ mol·rad^2, depending on the local free energy gradient. Each window was simulated for either 5 ns or 3 ns, depending on the area of the CV space covered, and the first 1 ns was discarded for subsequent analysis. Overall, a total of 3.18 μs of simulation time distributed over 849 windows was used to construct the free energy surface.

EMUS [206] was used to construct the free energy surface (Fig. S3.22A) and estimate the error associated with the free energy (Fig. S3.22C). WHAM [207, 208] yielded comparable results (Fig. S3.22B). The method for calculating averages conditioned on the CV coordinates (Figs. S3.12, S3.19A, and S3.21) is detailed in the Supplementary Methods. To find the free energy projection onto additional CVs not used in the umbrella sampling (Figs. 3.2D and 3.4C), we reran the EMUS calculation with a fictitious restraint on the additional CVs with zero force constant, and then averaged over the original CVs that were not of interest.

3.6.7 Mg^{2+} standard binding free energy calculation

The standard binding free energy of Mg^{2+} to a CII active site populated with ADP is estimated indirectly via a thermodynamic cycle using the double annihilation method; the method is outlined in [176], and the discussion below uses the notation therein. The starting structure was obtained from an equilibrated system with CII-bound

Mg·ADP. The Mg^{2+} to be removed is restrained at its equilibrium position using a set of harmonic biasing potentials ($\Delta G_{\mathrm{t}}^{\mathrm{site}}$) to eliminate the wandering ligand problem [209] before decoupling from the ligand-binding pocket ($-\Delta G_{\mathrm{int}}^{\mathrm{site}}$) using alchemical free energy perturbation (AFEP) [210]. The harmonic positional restraints are then removed ($-RT \ln F_t C^\circ$; discussed further in the Supplementary Methods). A separate cubic TIP3P water box (60 Å each side) with one Mg^{2+} and 150 mM KCl was used to calculate the coupling free energy in water ($\Delta G_{\mathrm{int}}^{\mathrm{bulk}}$). Taken together, the standard binding free energy is

$$\Delta G_{\mathrm{bind}}^{\circ} = -\Delta G_{\mathrm{t}}^{\mathrm{site}} + \Delta G_{\mathrm{int}}^{\mathrm{site}} + RT \ln F_t C^\circ - \Delta G_{\mathrm{int}}^{\mathrm{bulk}}$$

The values for each individual term contributing to the standard binding free energy are listed in Fig. S3.13, along with a thermodynamic cycle illustrating their relations. With the exception of $\Delta G_{\mathrm{int}}^{\mathrm{site}}$, the free energy calculations appeared well converged (Fig. S3.14). The value of $\Delta G_{\mathrm{bind}}^{\circ}$ reported in the Results section is based on the value of $\Delta G_{\mathrm{int}}^{\mathrm{site}}$ in the forward AFEP transformation, which most favors the alternative mechanism. Further details of the simulation protocol are provided in Supplementary Methods.

3.6.8 Interaction energy calculations

Energy calculations for the analyses in Fig. 3.2 and Figs. S3.12, S3.19A and B were done with the NAMDEnergy 1.4 plugin for VMD using the same force field as that in the simulation. The energy calculation used a switch function with a distance of 10 Å and a cutoff distance of 12 Å for nonbonded interactions and did not employ periodic boundary conditions.

3.6.9 Principal component analysis (PCA)

The PCA of KaiC structures (Fig. 3.3 and Figs. S3.15 and S3.21) was done using the NMWiz 1.0 plugin for VMD [211]. KaiC structures obtained from umbrella sampling

were downsampled to every 1 ns and aligned by their CII subunit D structures. Only the Cα positions for residues 20 to 497 in all six subunits were used for the analysis. No reweighting was performed for the PCA. The projection of each structure onto a principal component was calculated by finding the dot product of the normalized principal component eigenvector with the displacement vector of each structure from the average umbrella sampling structure (without reweighting).

3.6.10 Structural and sequence alignment

All structural alignment (Fig. 3.5) was done using STAMP [194, 212, 213], and all sequence alignment (Table S3.2) was done using CLUSTAL Omega [214, 215]. The sequences used in the analysis in Table S3.2 and their annotation were directly obtained from datasets published with [14], which are available at https://dx.doi.org/10.6084/m9.figshare.3823899.v3.

3.7 Supplementary Results

3.7.1 Radially inward release of ADP is unfavorable

We sought to assess whether there are alternative release pathways that were unexplored by the locally enhanced sampling simulations because they are slightly disfavored energetically relative to the pathway discussed in the main text. Based on the geometry of the active site (Fig. S3.9) and an inward release pathway proposed for P_i release after hydrolysis in the CI domain [156], we considered a radially inward release pathway, where a nucleotide would be released into the KaiC central channel and subsequently escape through the exit at the CI ring opening (release through the CII end of the channel is sterically hindered due to the presence of the A-loops).

To generate an initial guess for the inward release pathway, we used adiabatic biased MD (ABMD) [216] with the bias on the distance between the center-of-mass of the ADP

molecule and the CII ring while holding the subunit D–E angle open (see Supplementary Methods). We found that ABMD was less likely than SMD [160] to distort the protein structure in the absence of a good choice of bias coordinate, presumably because it takes advantage of spontaneous thermal fluctuations in the system. In the ABMD simulation, the ribose and adenine moieties rotate with the phosphate groups in place such that the base enters the central channel first (Fig. S3.7). This motion is likely enabled by the fact that the hydrophobic protein-ligand interactions are much weaker than the electrostatic interactions, which results in differential mobility of the charged and hydrophobic moieties of the ligand during release. To accommodate this motion, the Walker A motif (i.e., the P-loop) at the binding site and the two adjacent 422-loops rearrange.

Similarly to our treatment of the outward release pathway, we refined the inward release pathway with the string method and computed free energies with umbrella sampling (see Supplementary Methods). Subunit D–E angle widening and nucleotide movement are concurrent in the resulting pathway (Fig. S3.7A). However, the free energy barrier for reaching the first intermediate state on the inward release pathway is approximately 50 kcal/mol (Fig. S3.7B), compared with 10 kcal/mol for the outward pathway. This suggests that the inward pathway does not contribute significantly to release under physiological conditions.

3.8 Supplementary Methods

3.8.1 System setup

The protonation states of the titratable sidechains were determined using PROPKA 3.1 [217, 218] at a pH = 8, representing standard buffer conditions in in vitro biochemical assays of KaiC. The phosphate groups on T432 were modeled as dianionic phosphate groups. All histidine residues were modeled as neutral, with a proton at the Nδ position; the tautomers were determined by inspecting the histidine sidechain environments in

the crystal structure. Four lysine residues in each KaiC subunit, K224, K232, K457 and K465, as well as one glutamate residue, E357, were modeled as neutral. The geometry of the E318 sidechain indicates that it forms a hydrogen bond with the phosphate group on T432 in the protein crystal structure, suggesting that either pT432 or E318 is protonated. However, the pKa estimates for the catalytic E318 varied widely between 10.2 in subunit A and 2.1 in subunit B of the crystal structure, likely due to subtle variations in the distance between the E318 sidechain and the Mg^{2+} ion. This makes it difficult to unambiguously determine the relative acidity of pT432 and E318. We thus chose, arbitrarily, to make the E318 sidechains protonated in the simulation. Overall, the protonation states set according to this procedure were found to be insensitive to ATP/ADP binding in the CII domain.

3.8.2 Locally enhanced sampling

The locally enhanced sampling simulations were carried out using periodic boundary conditions with cubic computational boxes. Particle-mesh Ewald was used to compute electrostatic interactions with 1.2 nm cutoff in real space and a spline order of 6 and 144 grid points along all three spatial dimensions. The Lennard-Jones forces were smoothly switched off from 1.0 to 1.2 nm. The non-bonded pair list was built with a cutoff distance of 1.6 nm. The simulations were conducted in the NPT ensemble with a reference temperature of 303 K and a reference pressure of 1 atm. A Langevin thermostat [219] with a damping coefficient of 1 ps^{-1} was used. Isotropic volume fluctuations were controlled via the Nosé-Hoover Langevin piston method [220, 221] with an oscillation period of 50 fs and a damping decay time of 25 fs. All bonds to hydrogen atoms were constrained using the SHAKE algorithm [222]. The energy of the system was minimized with the default algorithm in NAMD for 1000 steps, and then the system was equilibrated for 50,000 steps with a 0.5 fs time step. The production run was 0.4 ns with a 1 fs time step.

3.8.3 String method

The convergence of the string in the CV space was initially assessed by visual inspection of projections onto 2D subspaces (Fig. S3.10), which suggested that all major conformational rearrangements are complete well within 30 iterations. To verify this conclusion, we ran an additional 70 iterations. Only the β_{nt}–$\alpha l\alpha$ distance variable showed further relaxation (Fig. S3.20).

To assess the relaxation of conformational degrees of freedom outside the CV space, we computed projections onto the first two principal components (see Results in the main text) and compared the results with the corresponding umbrella sampling average (Fig. S3.21). We chose to investigate these two particular degrees of freedom because they are biologically relevant and involve large-scale global conformational changes not directly controlled in the string calculation. Consistent with the CVs, most of the relaxation of these degrees of freedom occurs in the first 10 iterations; subsequent iterations result mainly in fluctuations around the umbrella sampling average.

It is important to emphasize that we draw all our quantitative conclusions from the umbrella sampling, and we have taken care to ensure its convergence (see Methods in the main text). The main purpose of the string is to generate a qualitatively reasonable pathway for the choice of CVs and initialization of the umbrella sampling. For this purpose, it appears that 30 iterations are sufficient.

3.8.4 Umbrella sampling and analysis

Averages against the stationary distribution were estimated as

$$\langle f \rangle = \int f(x)\pi(x)\,dx = \sum_{i=0}^{848} z_i \int \frac{f(x)}{\sum_{k=0}^{848} \psi_k(x)} \pi_i(x)\,dx$$

where π_i is the partition function of each window , $\psi_i(x) = \exp[-U(x)/RT]$ and $U(x)$ is the harmonic restraining potential. In practice, we replace z_i with the estimates

from the EMUS estimator $\bar{z}_i$ and approximate the integrals using umbrella sampling trajectory averages. In particular, this identity can be used to estimate the expectation of the quantity f, conditioned on the system being in a set of configurations S, which, in this case, represents a bin in the 2D CV space. Specifically, we can rewrite the conditional expectation of f as

$$\langle f|x \in S\rangle = \frac{\langle f \mathbb{1}_S\rangle}{\langle \mathbb{1}_S\rangle}$$

where is the indicator function defined by

$$\mathbb{1}_S(x) = \begin{cases} 1, & x \in S \\ 0, & \text{otherwise} \end{cases}$$

3.8.5 Mg^{2+} standard binding free energy calculation

The simulation setup was similar to that used in the locally enhanced sampling simulations, except that the time step used in all simulations was 2 fs, and for the bulk-phase simulations, a spline order of 6 and 72 grid points was used for the particle-mesh Ewald method.

Six restraints were defined on the distance between Mg^{2+} and Oβ2 of ADP, Oγ1 of T295, and the oxygen atoms on four water molecules to preserve the octahedral coordination geometry. The force constants for all restraints are 10 kcal/mol·Å^2, and the centers were chosen based on the average distances over a preliminary simulation. The free energy change due to the addition of the positional restraints ($\Delta G_{\mathrm{t}}^{\mathrm{site}}$) was calculated using thermodynamic integration (TI) [223]; specifically, the force constants for all restraints were simultaneously incremented from 0 to 10 kcal/mol·Å^2 according to the formula $k_\lambda = 10\lambda^4$, with values set to 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99, 0.999, 0.9999, 1. The process was then reversed to assess convergence (Fig. S3.14A). Finally, $\partial G/\partial\lambda$ for each restraint was linearly interpolated and then integrated and summed

together to obtain . Each $\Delta G_{\mathrm{t}}^{\mathrm{site}}$ state was simulated for 0.4 ns after 80 ps equilibration.

AFEP [210] was used to estimate $\Delta G_{\mathrm{int}}^{\mathrm{site}}$ and $\Delta G_{\mathrm{int}}^{\mathrm{bulk}}$. Van der Waals interactions were turned off with a soft-core radius-switching coefficient of 6.0 $\mathrm{Å}^2$ from $\lambda = 0$ to $\lambda = 0.2$ and a step size of 0.05, and then electrostatic interactions were turned off linearly from $\lambda = 0.2$ to $\lambda = 1$ with a step size of 0.01. The calculations were then reversed to assess convergence (Fig. S3.14B). Each λ step was simulated for 0.1 ns after 10 ps equilibration. The simulations were post-processed using ParseFEP 2.0 [224], and the BAR estimator [225] was used to determine $\Delta G_{\mathrm{int}}^{\mathrm{site}}$ and $\Delta G_{\mathrm{int}}^{\mathrm{bulk}}$. Electrostatic finite-size effects were corrected using the method proposed in [226].

The term $-RT \ln F_t C^\circ$ [176] accounts for the contribution to the free energy change from the translational degrees of freedom of the Mg^{2+} ion after all positional restraints are removed and it is decoupled from the protein in the standard state. Specifically, $C^\circ = N_A/1$ mL $\approx 1/1661$ $\mathrm{Å}^{-3}$ is a conversion factor to standard state concentration (1 M) and F_t is defined in this case as

$$F_t = \int_{\mathbb{R}^3} \exp\left[-\sum_{i=1}^{6} \frac{1}{2} k_i \left(\|\mathbf{r} - \langle \mathbf{r}_i \rangle\| - \|\langle \mathbf{r}_i \rangle\|\right)^2 / RT\right] dV$$

where k_i is the force constant of a Mg^{2+} positional restraint, and $\langle \mathbf{r}_i \rangle$ is the corresponding position vector of one of the six atoms used to define the Mg^{2+} positional restraints, averaged over an equilibrium simulation, assuming that the equilibrium Mg^{2+} position is the origin. The integral was evaluated numerically over boxes centered at the origin with successively larger dimensions until fluctuations in the results were within 10^{-7} $\mathrm{Å}^3$.

3.8.6 Determination of a hypothetical radially inward release pathway

The adiabatic biased MD (ABMD) [216], implemented in the ABMD bias functionality in PLUMED 2.3 [205], was used in conjunction with SMD [160] to generate an initial pathway for inward release into the KaiC central channel. ABMD adds a time-dependent

bias to a given CV x to steer the CV to a target value x_{fin} similar to the action of a Brownian ratchet. In practice, the CV is transformed to $X(t) = (x(t) - x_{\text{fin}})^2$, and a half umbrella bias potential is applied to $X(t)$:

$$U(X) = \begin{cases} \frac{1}{2}\kappa \left(X(t) - X_m(t)\right)^2, & X(t) > X_m(t) \\ 0, & X(t) \leq X_m(t) \end{cases}$$

where κ is the force constant, and $X_m(t) = \min_{\tau<t} X(\tau)$. A key feature of ABMD is that the bias potential does not do work on the system, because the potential is updated only when its value is zero.

To generate the initial pathway, we started from a structure selected from the SMD simulation (see Method in the main text) in which the subunit D–E angle was opened to 1.17 rad. ABMD was then applied on the ADP–CII center distance (see Table S3.1 for the definition of the CV) from the initial distance of 3.07 nm to the target distance of 0.0 nm with a force constant of 1,000.0 kJ/mol·nm^4, while the subunit D–E angle was restrained at 1.17 rad using a constant harmonic bias with a force constant of 100,000 kJ/mol·rad^2. The ADP–CII center distance reached the target value after 5 ns.

The initial pathway was refined using the string method and umbrella sampling, similarly to the treatment of the outward release pathway (see Methods in the main text). For the string calculation, 15 structures from the SMD and ABMD simulations were selected to construct a two-step initial pathway where the subunit D–E angle was first opened to 1.17 rad and then the ADP–CII center distance was reduced to 0.0 nm. The structures were equilibrated for 5 ns with restraints in the 2D CV space, and then evolved using the string method in a 7-dimensional CV space consisting of the ADP–CII center distance variable and the six intersubunit angles for 30 iterations. The force constants used in the production runs were 20,000 kJ/mol·nm^2 for the distance variable and 10,000 kJ/mol·rad^2 for the angle variables. The string was propagated with a constant multiplier $\Delta\tau = 0.00005$. Images from the string calculation were used as starting structures for

umbrella sampling on the ADP–CII center distance variable. Each window was simulated for 5 ns, and the force constant ranged from 4,000 to 60,000 kJ/mol·nm^2 depending on the local free energy gradient. In total, 90 windows were used to construct the free energy profile.

3.9 Supplementary Figures

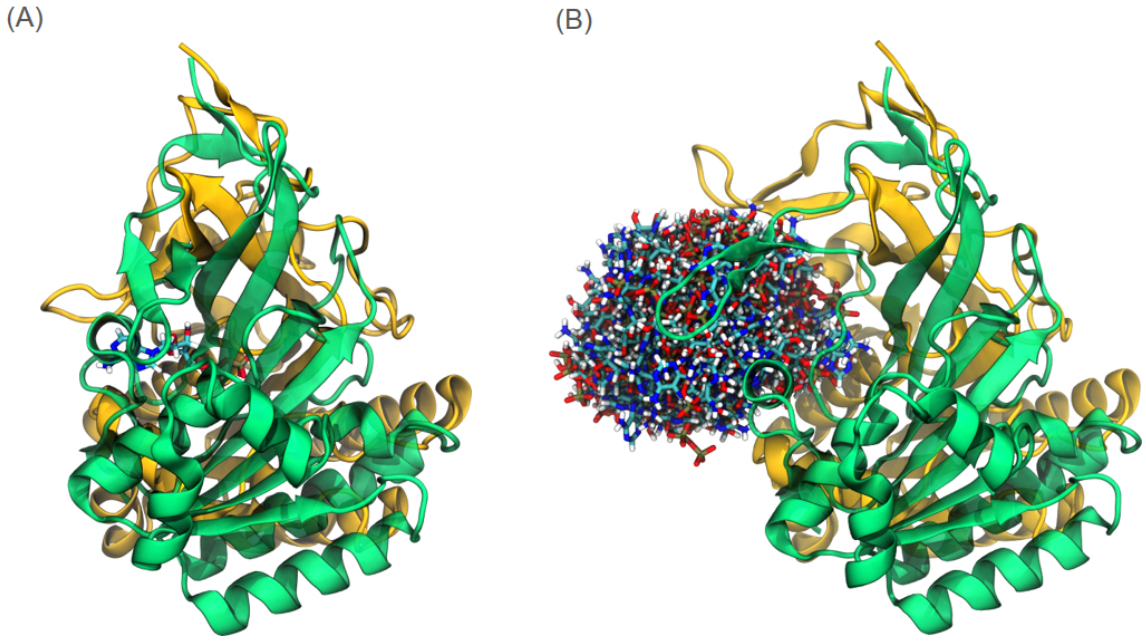


Figure S3.6: Locally enhanced sampling indicates that the release pathway for Mg·ADP points radially outwards, rather than inwards towards the central channel of KaiC. (A) The initial structure. (B) The final structure at 400 ps, with 200 replicas. Subunit D is shown in ribbon diagram in yellow, and subunit E is shown in green. Mg·ADP is colored by element, as in Fig. 3.1D.

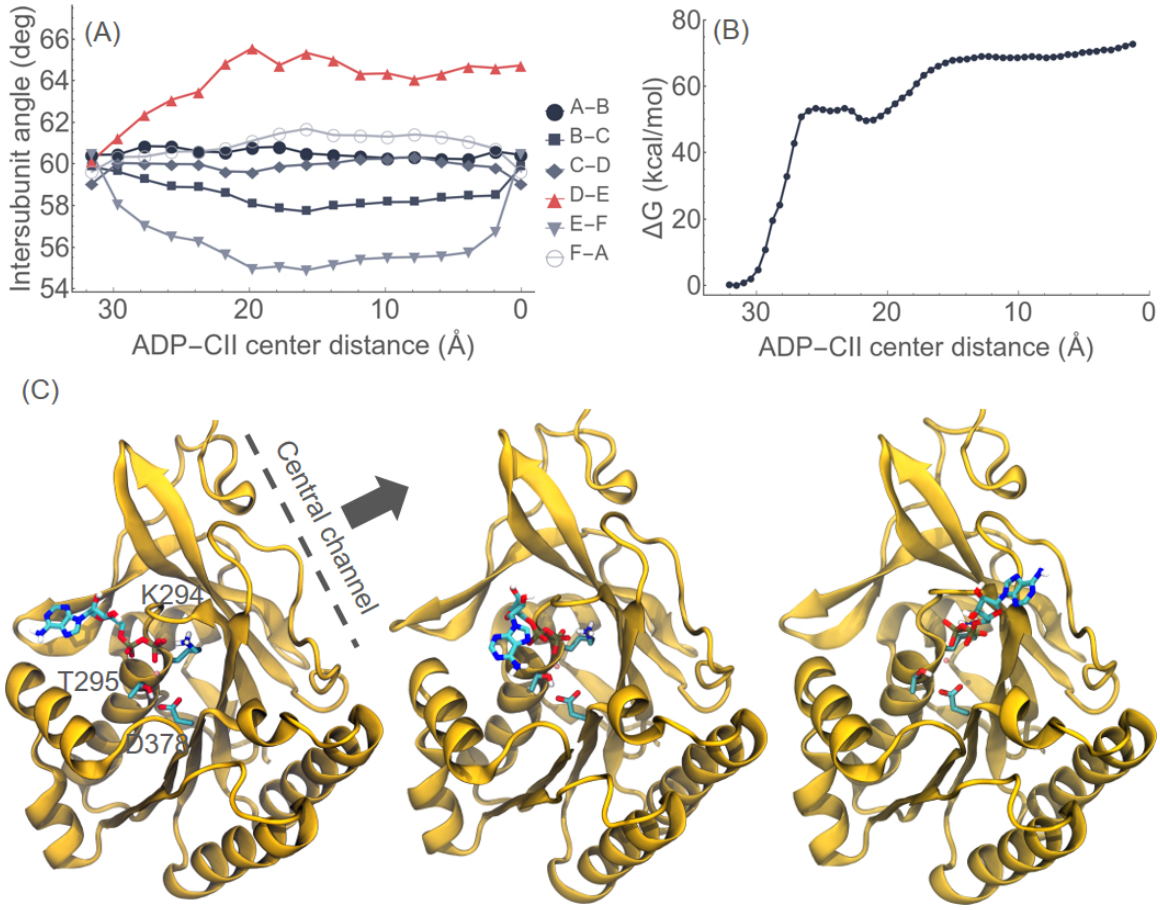


Figure S3.7: The radially inward Mg·ADP release pathway is highly unfavorable. (A) CII intersubunit angle changes during inward release. (B) The free energy profile for inward release, calculated using umbrella sampling. (C) The inward release pathway involves a rotational and bending motion of the hydrophobic moieties of ADP; the three snapshots represent structures at ADP–CII center distance of, from left to right, 31.9 Å, 25.7 Å and 17.3 Å.

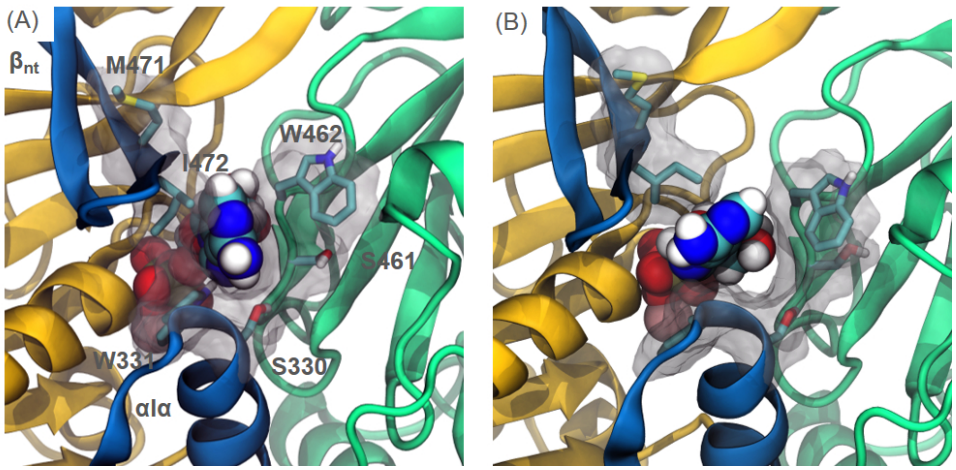


Figure S3.8: The movements of a local gating element near the exit point of the release pathway, consisting of a β-hairpin (β_{nt}; residue 468 to 483) and an α-loop-α motif ($\alpha l \alpha$; residues 321 to 342). The structures are obtained from the string simulations. The distance between the centers of mass of βnt and αlα takes the values of 11.0 Å in (A) and 11.4 Å in (B). The residues in close contact with the nucleotides are shown both in licorice and surface representation. The ADP–subunit D distance in (A) and (B) are 8.6 Å and 9.5 Å, respectively, representing structures of the protein very early in the release process.

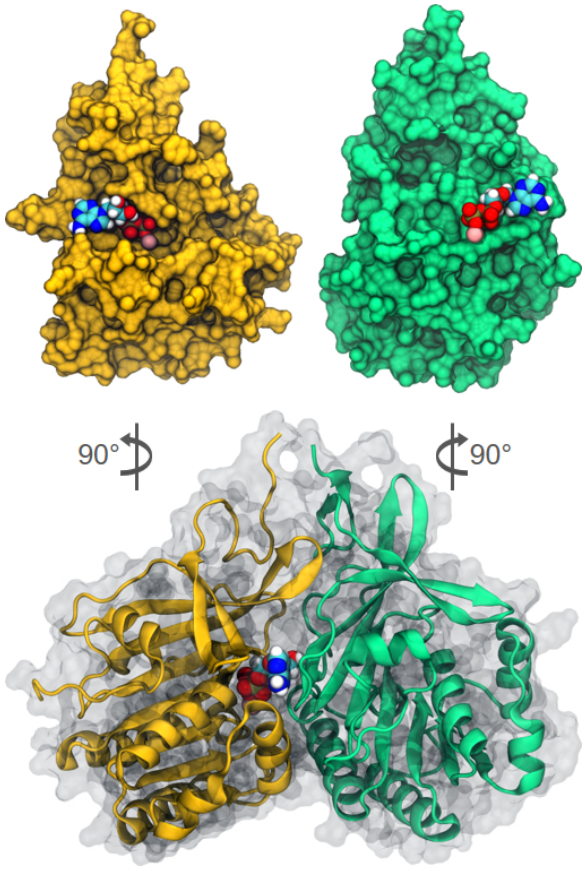


Figure S3.9: The geometry of the CII active site located at the subunit D–E interface. Subunit D is shown in yellow, and subunit E is shown in green. Note that the nucleotide makes more extensive contact with subunit D than subunit E.

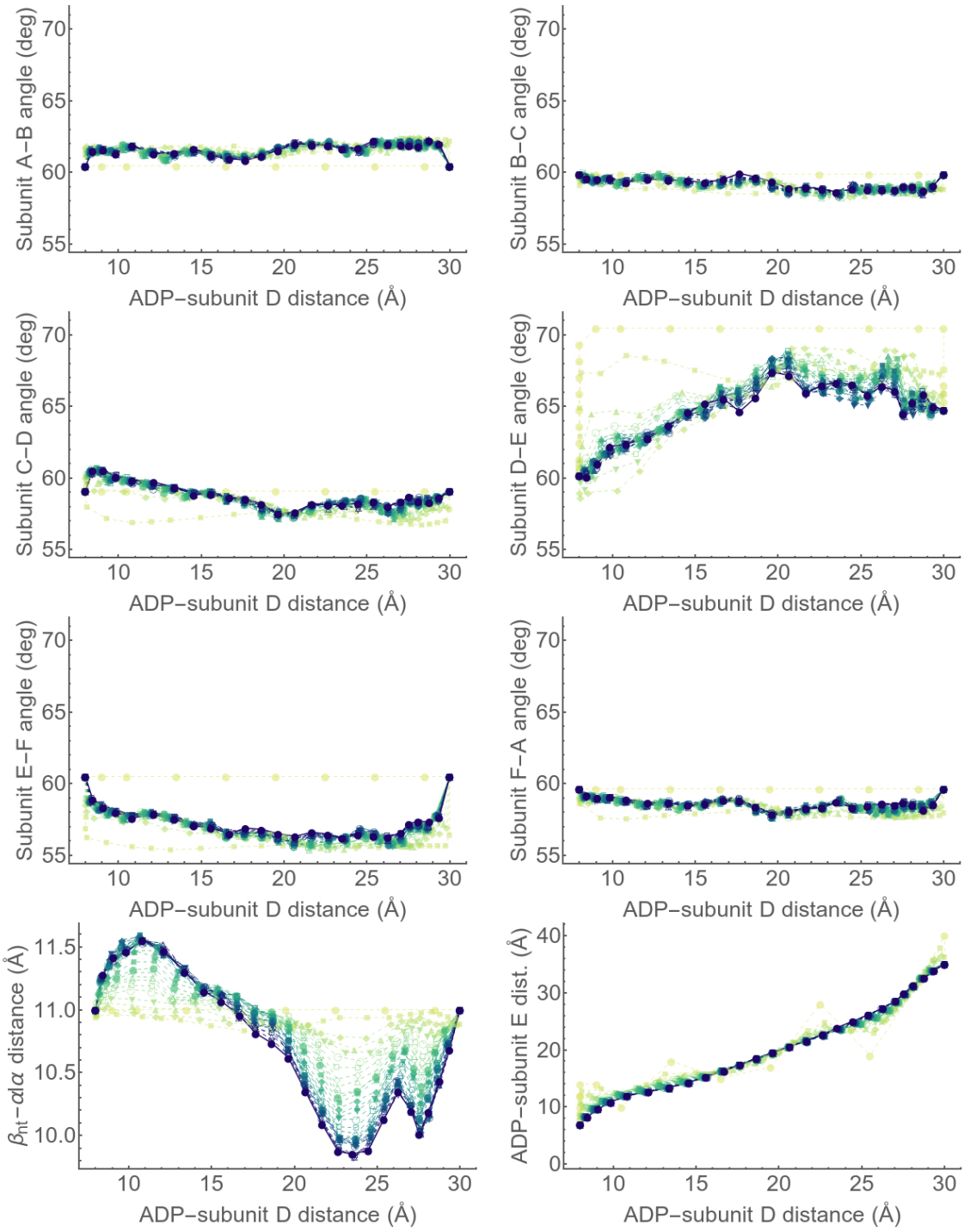


Figure S3.10: Convergence of the string simulations. The nine-dimensional string is represented in eight two-dimensional subspaces, with the ADP–subunit D distance degree of freedom used as a common progress variable. The color gradient represents the progress of the string, with yellow being the first iteration and blue the thirtieth.

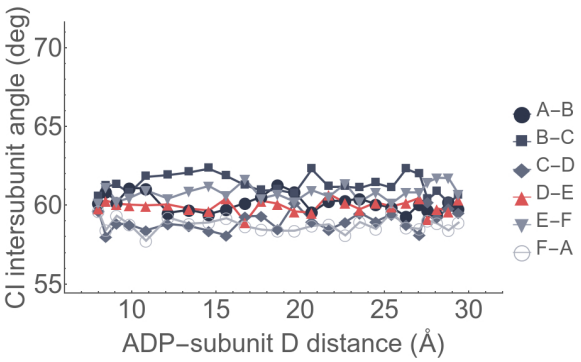


Figure S3.11: The CI intersubunit angles are decoupled from the CII release process. The figure shows the motion of the six CI intersubunit angles along the final string path, plotted as a function of the ADP–subunit D distance.

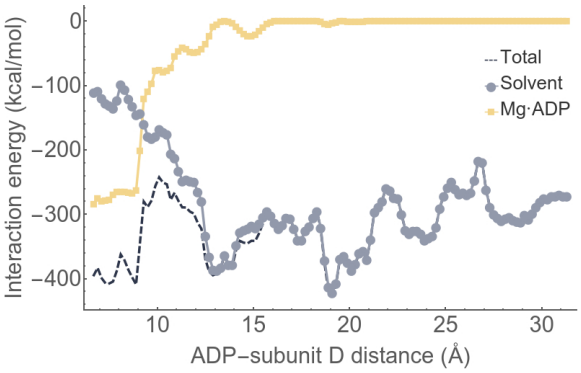
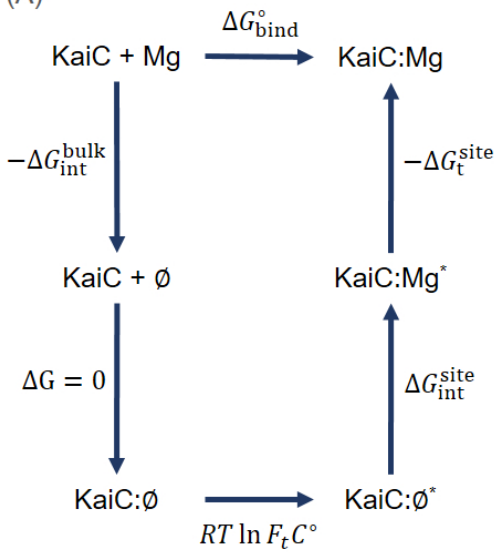


Figure S3.12: The average electrostatic interaction energy between a group of active site residues (K294, T295, E319 and D378) and Mg·ADP (yellow), the solvent (gray), or both (dashed), conditioned on the ADP–subunit D distance. There is a rapid loss of interaction between the active site residues and Mg·ADP around 10 Å, which is not immediately compensated by the solvent.

(A)



(B)

Term	Estimate (kcal/mol)	Method
$-\Delta G^{\text{site}}_{\text{t}}$	-0.13	TI
$\Delta G^{\text{site}}_{\text{int}}$	-514.09, -465.82	AFEP
$\Delta G^{\text{site}}_{\text{int}}$ [correction]	-21.14	Rocklin et al. (14)
$RT \ln F_t C^{\circ}$	-5.97	analytical (15)
$-\Delta G^{\text{bulk}}_{\text{int}}$	423.93	AFEP
$-\Delta G^{\text{bulk}}_{\text{int}}$ [correction]	35.25	Rocklin et al.
$\Delta G^{\circ}_{\text{bind}}$	-33.88, -82.15	

Figure S3.13: Free energy simulations indicate that Mg^{2+} leaves with ADP. (A) The thermodynamic cycle used to determine the standard binding free energy of Mg^{2+} in a CII active site ($\Delta G^{\circ}_{\text{bind}}$). Notationally, "+" indicates that Mg^{2+} is solvated in the bulk phase, ":" indicates that Mg^{2+} is in complex with ADP in the active site, "*" indicates that the position of Mg^{2+} is restrained, and "∅" indicates that Mg^{2+} is decoupled from the environment. (B) The free energy change associated with each step in the thermodynamic cycle. TI stands for thermodynamic integration, and AFEP stands for alchemical free energy perturbation. The two values listed for are the results of the forward and backward transformations, respectively (Fig. S3.14). See Methods for further details.

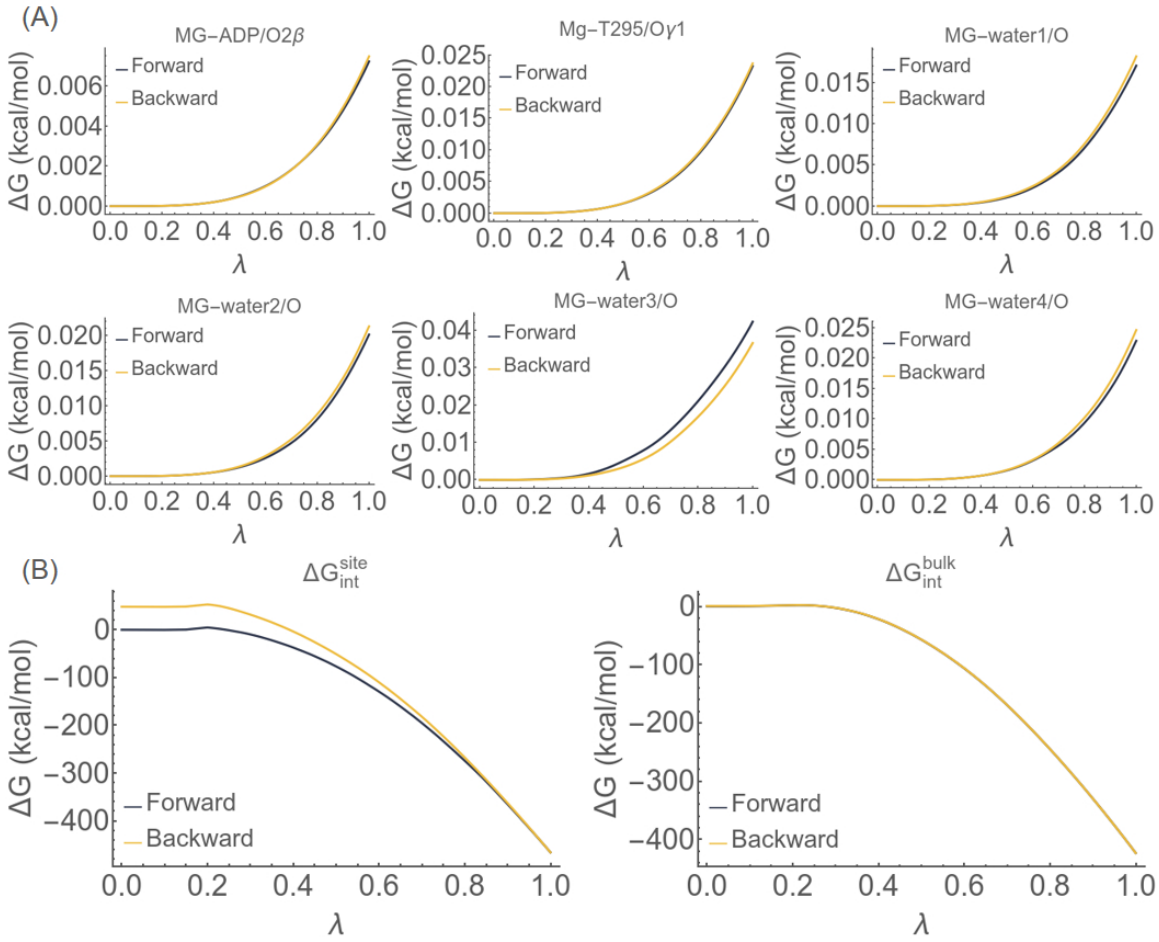


Figure S3.14: The reversibility of the thermodynamic integration (A) and alchemical free energy perturbation (B) calculations. (A) The cumulative free energy change of imposing a harmonic distance restraint between Mg^{2+} and each of the six atoms participating in the Mg^{2+} coordination geometry (forward) and then removing the restraint (backward). (B) The cumulative free energy change of coupling the Mg^{2+} to the environment through nonbond interactions (forward) and then decoupling the Mg^{2+} (backward), in either the CII active site (left) or the bulk phase (right). In both (A) and (B) λ represents the switching variable. See Methods for further details.

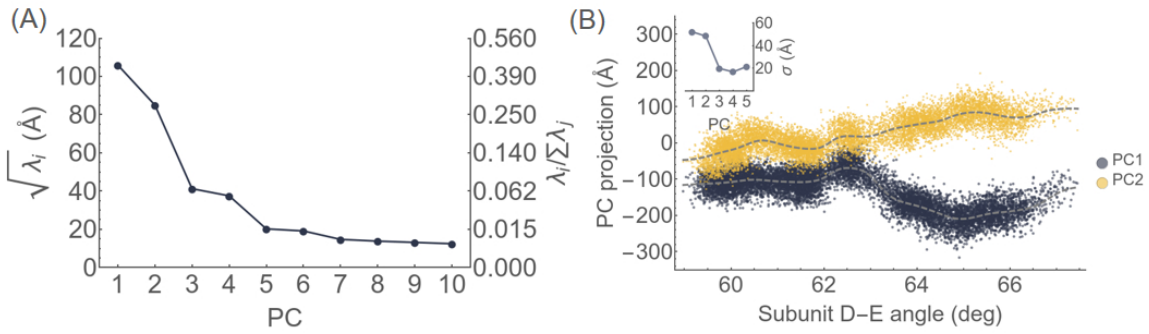


Figure S3.15: Principal component analysis. (A) The square root of the largest 10 eigenvalues. The eigenvalues represent the variance along the corresponding principal component (PC). The largest spectral gap occurs between PC2 and PC3, which indicates that the first two PCs represent the most significant large-scale conformational changes. The vertical scale on the right shows the variance explained by each PC as a fraction of the total variance in Cα positions. (B) The motion along PC1 and PC2 when steering the subunit D–E angle at constant velocity over 100 ns; structures are saved every 10 ps. The force constant used in the SMD is 100,000 kJ/mol·rad^2. The inset shows the standard deviation of projections onto the first five PCs over the steering simulation. Both the eigenvalues in (A) and the projections in (B) are calculated using the Cα atom displacement vectors. The dashed curves represent the results of Nadaraya-Watson kernel regression with the radial basis function kernel [227, 228]. The gamma parameter (i.e., the bandwidth for the radial basis function) was automatically chosen within the range $[10^{-1}, 10^{1}]$ to minimize the mean-squared-error of leave-one-out cross-validation.

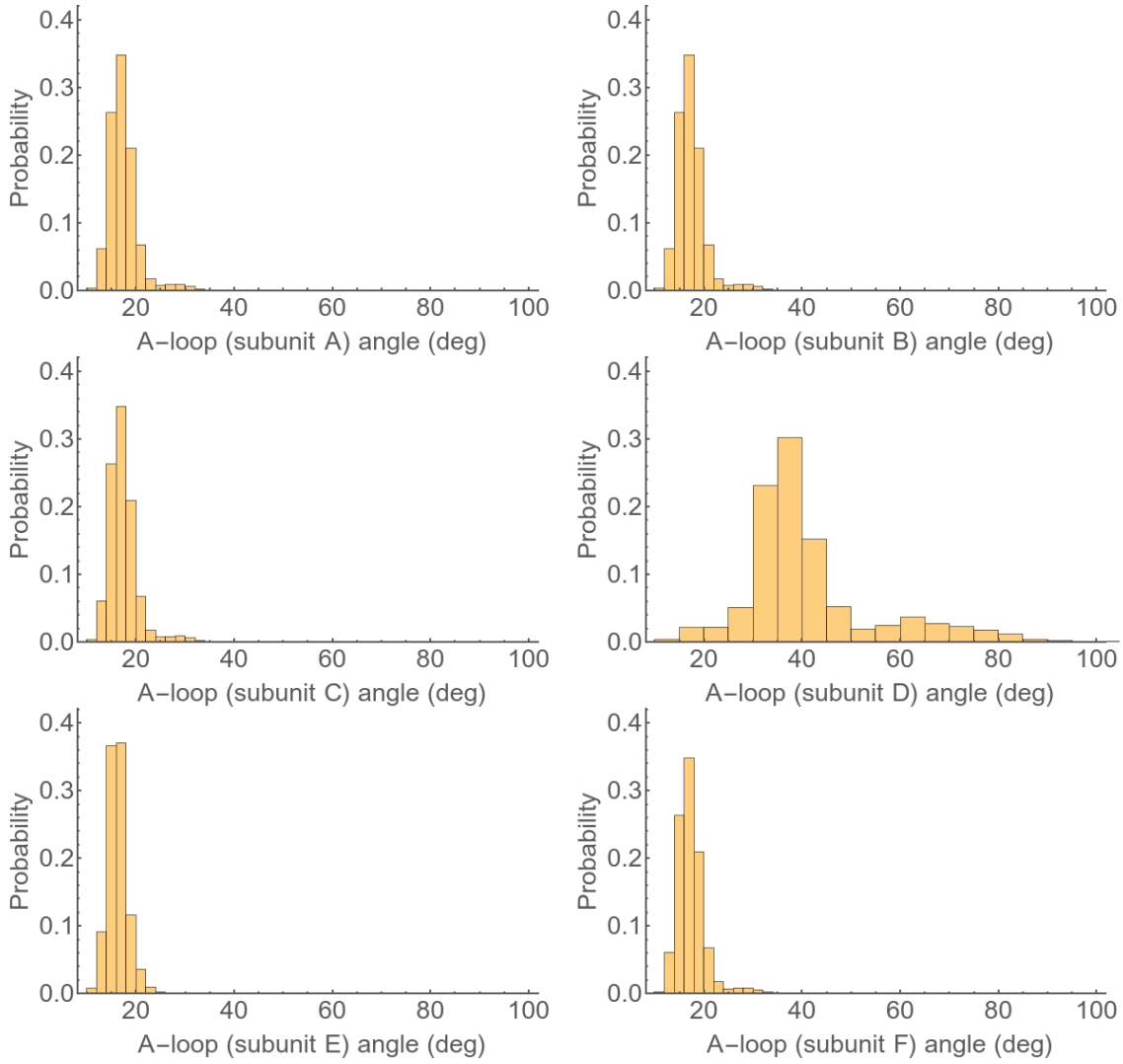


Figure S3.16: The distribution of the A-loop angle, defined as the angle formed by the Cα atoms of residues 486, 490, and 496. Panels correspond to different subunits as indicated. Data are from the umbrella sampling simulations, without reweighting. The A-loop angle in subunit A, B, C, E and F rarely, if ever, exceeds 40°, while the subunit D A-loop angle can reach up to 100°. The differential flexibility of the A-loop suggests that the A-loop extension observed in subunit D is a specific response to nucleotide release.

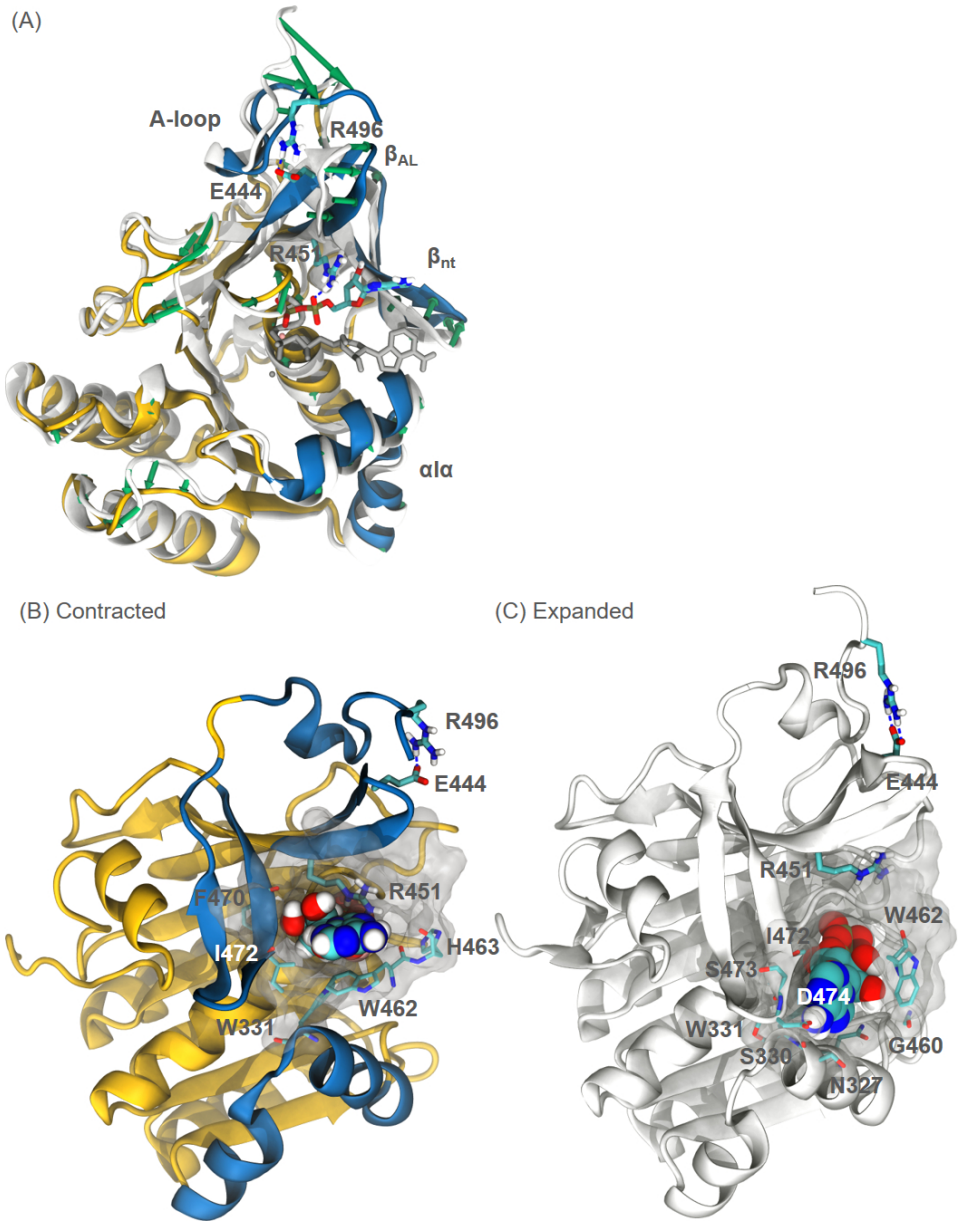


Figure S3.17: The A-loop is coupled to a contraction motion in subunit D through an adjacent β-hairpin (β_{AL}; residues 436 to 459). (A) Superposition and (B, C) side-by-side comparison of the contracted (yellow/blue) and expanded (white) states observed in the umbrella sampling simulations. Cα displacements larger than 1.5 Å are indicated by green arrows in (A). Key secondary structural elements are highlighted in blue: A-loop (residues 486 to 498), β_{AL}, β_{nt} (residues 468 to 483) and $\alpha I \alpha$ (residues 321 to 342). Key residues that couple the A-loop to the nucleotide through β_{AL} (R451, E444 and R496) are shown in licorice representation. (B) In the contracted state, the nucleotide is in contact with W331, R451, W462, H463, F470 and I472. (C) In the expanded state, the nucleotide is in contact with S330, W331, N327, I472, S473, D474, G460 and W462.

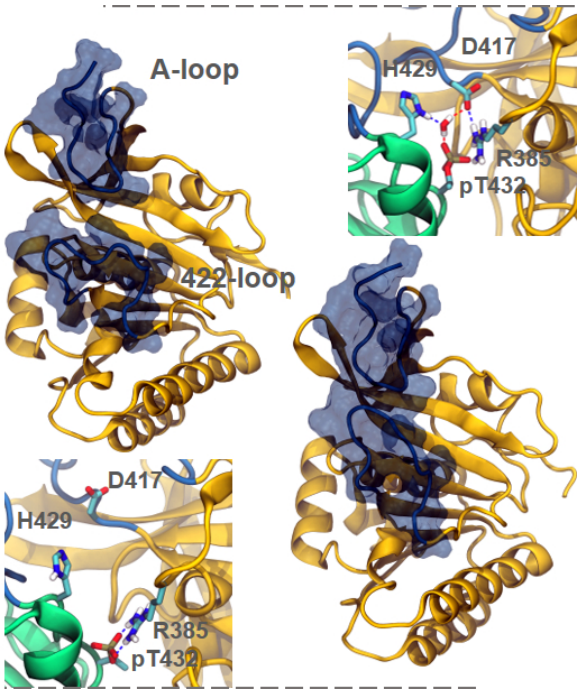


Figure S3.18: The A-loop (residues 486 to 498) is coupled to the active site phosphorylation state via the 422-loop. When the A-loop and 422-loop are not in close contact, D417 on the 422-loop is more likely to form a water-mediated hydrogen bond network with pT432 as well as the adjacent H429 and R385 (upper right). However, when the 422-loop is in contact with the A-loop, the increased distance between the 422-loop and the active site disrupts the hydrogen bond network (lower left).

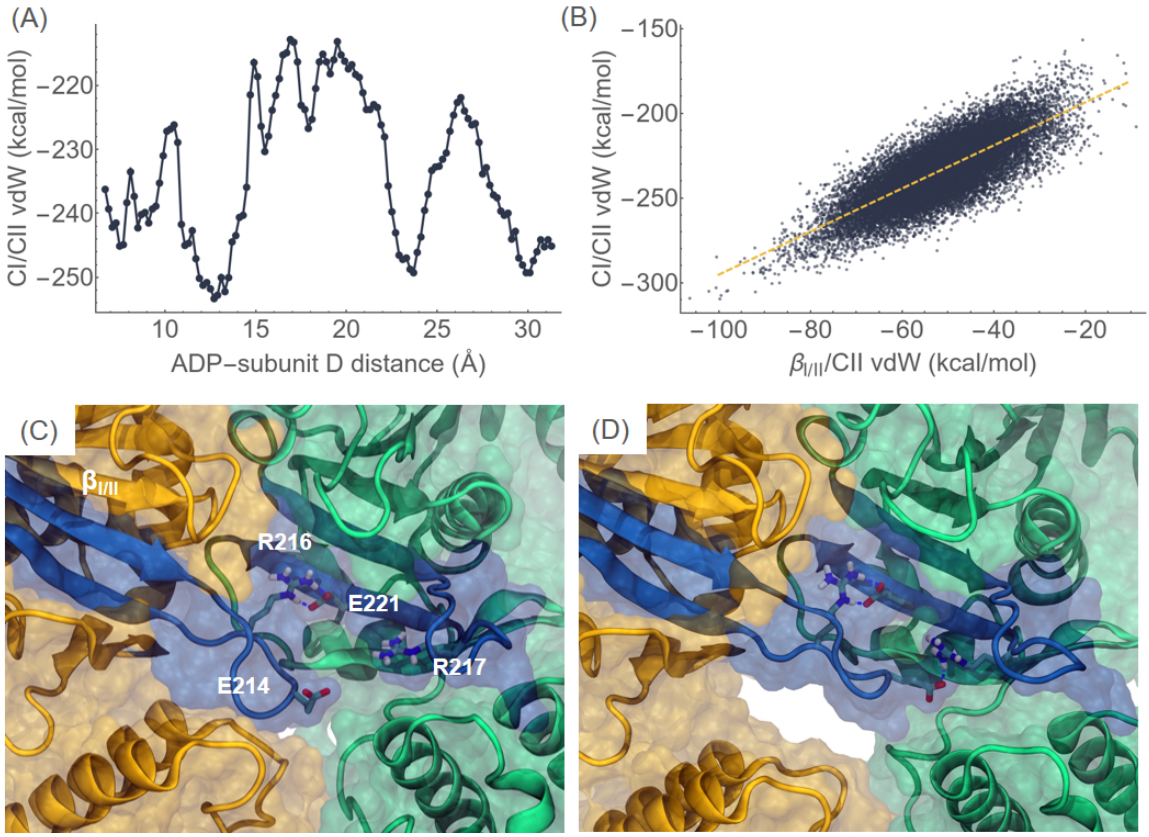


Figure S3.19: CI/CII contact is mediated by the motion of a CI β-hairpin ($\beta_{\mathrm{I/II}}$; residues 203 to 226). (A) The average CI/CII van der Waals (vdW) interaction energy conditioned on the ADP–subunit D distance, calculated from umbrella sampling, shows fluctuations on the order of 40 kcal/mol. (B) The variance of the CI/CII vdW interaction energy is correlated with the $\beta_{\mathrm{I/II}}$/CII vdW interaction energy (including $\beta_{\mathrm{I/II}}$ in all six subunits). The scatter plot represents the umbrella sampling results, sampled every 100 ps, and the best linear fit line without reweighting (vdW[CI/CII] = 168.3 kcal/mol + 1.27 vdW[$\beta_{\mathrm{I/II}}$/CII]) is shown in yellow. (C) and (D) Structures representative of strong and weak CI/CII interactions, respectively. Subunit D (yellow) and E (green) are shown in surface representation, with a ribbon diagram of the secondary structure. The $\beta_{\mathrm{I/II}}$ structures in both subunits are rendered in blue, and the two salt bridge interactions that couple neighboring $\beta_{\mathrm{I/II}}$ structures, R216–E221 and R217–E214, are shown in licorice representation.

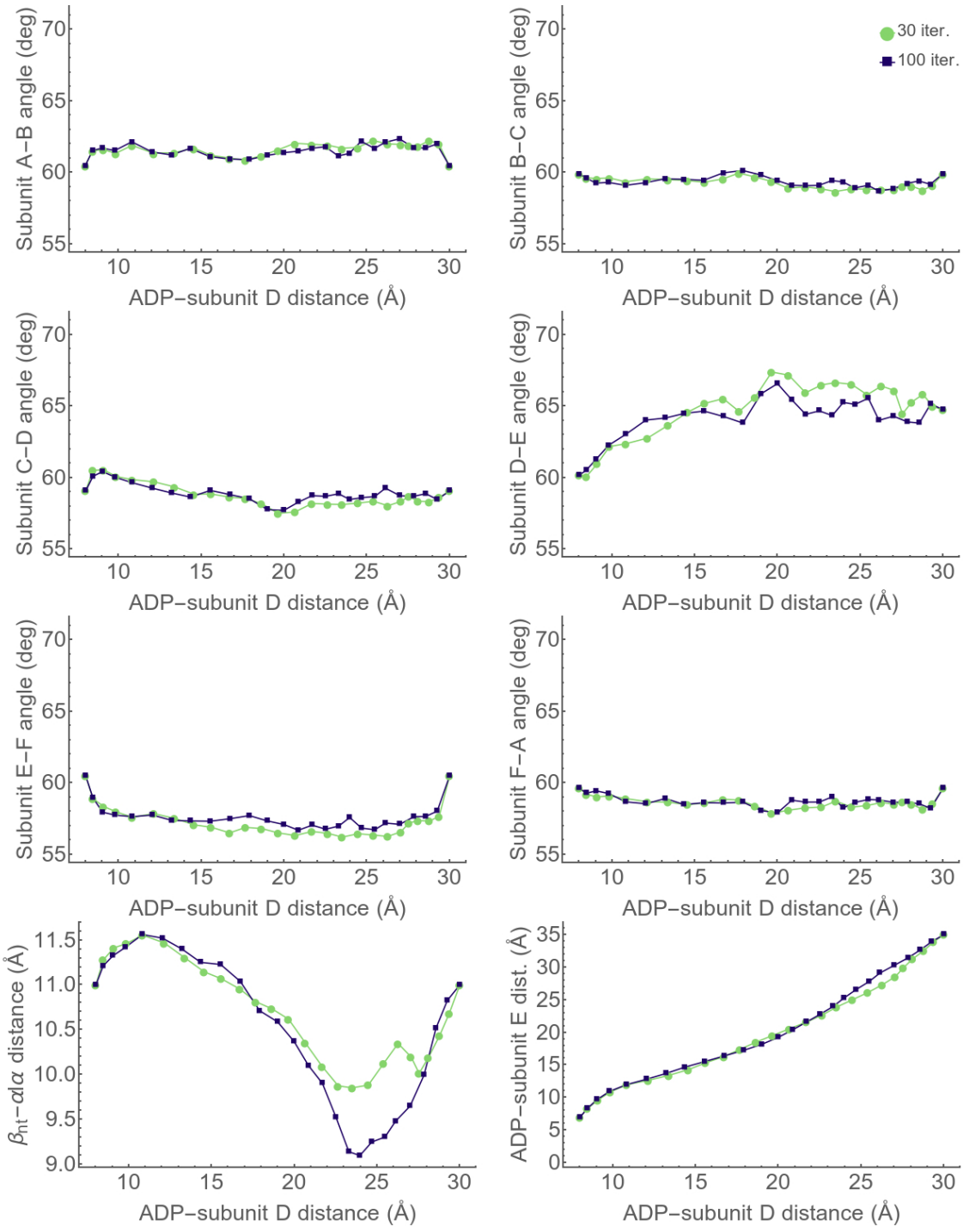


Figure S3.20: Comparison of the 30th (green circles) and 100th (blue squares) iterations of the string calculation. The nine-dimensional string is represented in eight two-dimensional subspaces, with the ADP–subunit D distance degree of freedom used as a common progress variable.

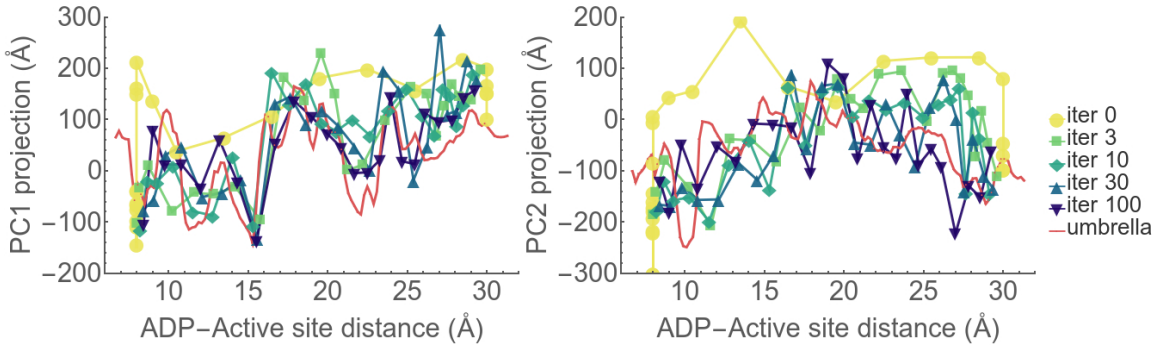


Figure S3.21: The string calculations projected onto the first two principal components (PC). The red curves represent the umbrella sampling averages conditioned on the ADP–active site distance, included as a reference curve for assessing string convergence.

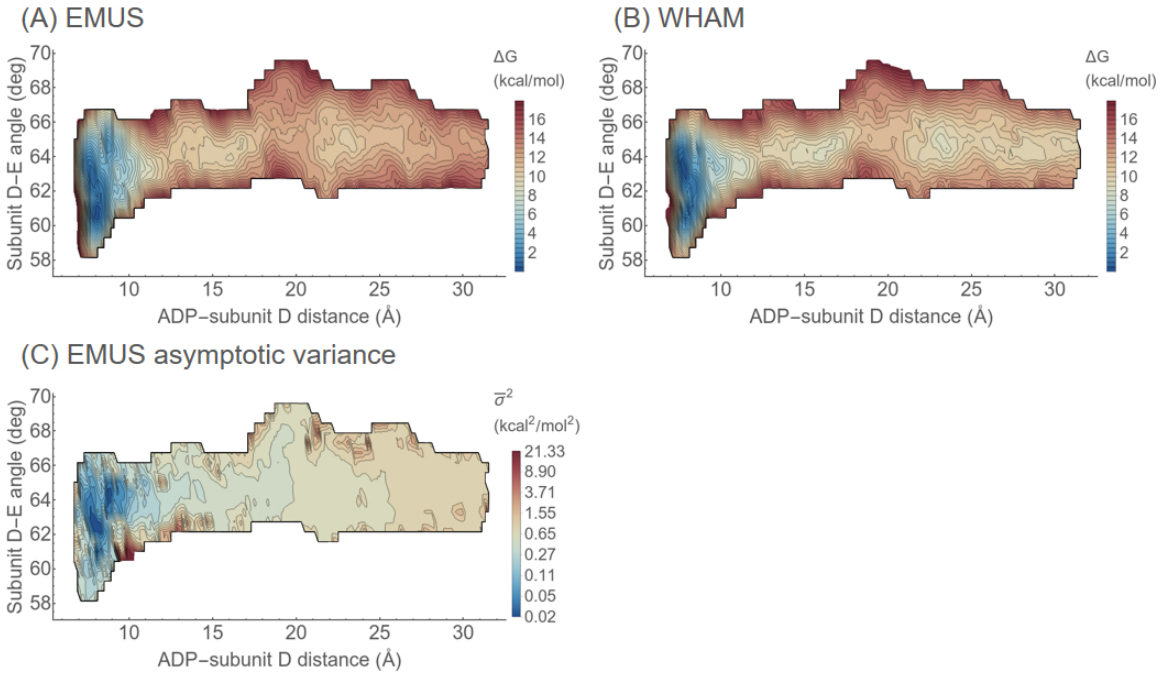


Figure S3.22: The free energy profile for the nucleotide release pathway is estimated using EMUS (A) and WHAM (B); the two methods produce comparable results. The asymptotic variance of the EMUS estimator is plotted with a logarithmic scale in (C). The error analysis suggests that the free energy profile has sufficiently converged, with the exception of a small region near (11 Å, 61°).

3.10 Supplementary Tables

Table S3.1: The collective variable (CV) definitions. The first section of the table defines atom groups, which are then used in the second section to define the nine collective variables employed in the string calculation.

Group	Atoms
ADP	All non-hydrogen atoms of the ADP molecule at the CII subunit D–E active site
active site (subunit D)	Subunit D Cα atoms on residues 286, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 317, 318, 319, 378, 379, 413, 414, 415, 440, 442, and 451
active site (subunit E)	Subunit E Cα atoms on residues 432, 435, 436, 437, 455, 456, 457, 458, 459, and 460
$\alpha l \alpha$	Subunit D Cα atoms on residues 329 to 334
β_{nt}	Subunit D Cα atoms on residues 473 to 476
CII	Cα atoms on residues 260 to 498 in all subunits
subunit X[a]	Cα atoms on residues 260 to 498 in subunit X

Collective variable (CV)	Type	Groups
ADP–subunit D distance	Center-of-mass (COM) distance	ADP, active site (subunit D)
ADP–subunit E distance	COM distance	ADP, active site (subunit E)
β_{nt}–$\alpha l \alpha$ distance	COM distance	$\alpha l \alpha$, β_{nt}
Subunit A–B angle	COM angle	subunit A, CII, subunit B
Subunit B–C angle	COM angle	subunit B, CII, subunit C
Subunit C–D angle	COM angle	subunit C, CII, subunit D
Subunit D–E angle	COM angle	subunit D, CII, subunit E
Subunit E–F angle	COM angle	subunit E, CII, subunit F
Subunit F–A angle	COM angle	subunit F, CII, subunit A
ADP–CII center distance[b]	COM distance	ADP, CII

[a] X = A, B, C, D, E, F

[b] The ADP–CII center distance variable is only employed for the hypothetical radially inward release pathway (see the Supplementary Results and Methods).

Table S3.2: Conservation of nucleotide gating tryptophan residues (W331 and W462) in KaiC CII domains. The sequences used in the analysis and their annotation were directly obtained from datasets published with [14].

Phylum[a]	Homolog	Hits	W331	W462	Identity[b]
cyanobacteria	KaiC1	100	97	95	0.77±0.10
	KaiC2	15	0	0	0.348±0.007
	KaiC3	24	21	0	0.50±0.09
α-proteobacteria	KaiC1	5	0	0	0.23±0.01
	KaiC2	26	0	0	0.33±0.02
	KaiC3	94	7	0	0.26±0.07
GNS bacteria	KaiC1	0	0	0	N/A
	KaiC2	1	0	0	0.32
	KaiC3	9	5	0	0.41±0.15
euryarchaeotes	KaiC1	132	0	2	0.24±0.06
	KaiC2	111	0	0	0.25±0.06
	KaiC3	206	0	4	0.25±0.05

[a] Only phyla that contain proteins with conserved CII tryptophan residues are included from the original dataset.

[b] Identity is calculated based on comparison to KaiC in *Synechococcus Elongatus* PCC 7942. The values reported are the mean with standard deviation.

3.11 Acknowledgements

We thank Andy LiWang and Glen Hocky for critical readings of the manuscript. Personnel were supported by National Institutes of Health (NIH) Awards GM109455-02, GM107369 and EY025957, and by a MolSSI Software Fellowship (to EHT). Computations were performed on resources provided by the University of Chicago Research Computing Center, by NIH Award 1S10OD018495-01 for the Beagle computer at the University of Chicago and Argonne National Laboratory, as well as by the National Science Foundation through the Extreme Science and Engineering Discovery Environment (XSEDE) [138], under grant number ACI-1548562, through allocation TG-MCB180007 giving access to Comet (SDSC) and Bridges (PSC) computing nodes.

CHAPTER 4

CHARACTERIZATION OF THE KAIB FOLDSWITCH PATHWAY

4.1 Introduction

KaiB, a component of the Kai oscillator, functions primarily through its binding interactions with KaiA and KaiC. The binding of KaiB to KaiC mediates the negative feedback loop that suppresses KaiC phosphorylation at night. KaiB binds to the B-loop, an unstructured loop region on the CI domain of KaiC [20, 77], in a cooperative fashion [23, 140], and the binding interaction appears to require the ATP hydrolysis product, ADP, in the CI nucleotide-binding pocket [28]. Once bound to KaiC, KaiB can sequester KaiA from its site of action at CII, thus leading to KaiC auto-dephosphorylation [17]. When KaiB is not bound to KaiC, it exists in an equilibrium between tetrameric, dimeric, and monomeric states, although the monomeric state dominates in the micromolar regime [23]. KaiB is also a metamorphic protein with two known folds called the ground state (gs) and the foldswitched (fs) state, and it appears to interact with KaiA and KaiC only in the latter conformation [22].

Currently, several questions remain about KaiB's function. First, model estimates of the KaiB foldswitch and binding reactions suggest that it is slow, on the timescale of hours [22], comparable to the kinetics of ATP hydrolysis in CI [26]. This raises the question of exactly how the timescale of the switch to the dephosphorylation phase relates to the kinetics of KaiB foldswitch and CI ATP hydrolysis. Is one of them rate-limiting, or do both contribute to setting the timescale? Does KaiC play any role in the kinetics of KaiB foldswitch? For example, does KaiB foldswitch become rapid once KaiC becomes available for binding? Second, as one of the better documented examples of metamorphic proteins [229–231], KaiB provides a unique opportunity to probe the biophysical basis of multiple, stable protein folds, and the pathway(s) that allow their interconversion.

Overall, it is clear that a detailed characterization of the KaiB foldswitch pathway is critical for understanding the kinetics of the Kai oscillator, and may prove useful in future engineering efforts to design oscillators with altered periods.

This chapter describes some preliminary results from our attempt to characterize KaiB foldswitch using molecular dynamics (MD) simulations. We started with steered MD (SMD) to generate an initial guess for the foldswitch pathway from the gsKaiB monomer structure. The initial pathway was refined using the string method, and the FAST method, an adaptive sampling method, was used to explore the KaiB conformational space. This is an ongoing study, and further details will be made available in future publications.

4.2 Methods

4.2.1 Structural model for fsKaiB

Currently there are several NMR and crystal structures of the fsKaiB based on the *Thermosynechococcus elongatus* KaiB [77], but no such structures exist for the *S. elongatus* KaiB. The cryo-EM structure of the *S. elongatus* KaiABC complex (PDB ID: 5N8Y) [78] resolves the backbone configuration of the fsKaiB in complex with KaiA and KaiC. Based on this structure (subunit G), the Rosetta Packer algorithm (with standard energy weights) was used to generate five models for the side chain rotamers [232]. The five models were virtually identical and agreed well with existing *T. elongatus* fsKaiB structures. As such, one structure was chosen arbitrarily as a reference model for fsKaiB for the purpose of subsequent steered MD and string calculations.

4.2.2 System setup

We used CHARMM-GUI [191–193] to setup the MD simulations. The initial protein structure was chosen to be subunit A from a crystal structure of *S. elongatus* gsKaiB

tetramer (PBD ID: 3DVL) [233]. The system was solvated with TIP3P water molecules, and the final dimensions of the box after equilibration were 73.16 Å × 73.16 Å × 73.16 Å. 36 K^+ ions and 36 Cl^- ions were added to the water box for a final concentration of 150 mM KCl. The total number of atoms in the system was 39,551. All molecular visualizations were made in VMD 1.9.3 [194].

4.2.3 Minimization and equilibration

All MD simulations were performed using GROMACS 2019.4 [195], with the CHARMM36m force field parameters for the protein (with CMAP correction) [196–198, 234], TIP3P water [199], and ions [200]. Steepest descent was used for energy minimization, until the largest force in the system was smaller than 1000.0 kJ/mol·nm. The system was equilibrated first in the NVT ensemble for 25 ps with a 1 fs integration timestep, followed by the NPT ensemble for 5 ns with a 2 fs integration timestep, and lastly in the NVT ensemble for 5 ns using the average box size from the NPT equilibration. Restraints were applied to all non-hydrogen atoms in the protein during energy minimization and the initial NVT equilibration run, with a force constant of 400 kJ/mol·nm^2 for the backbone atoms and 40 kJ/mol·nm^2 for the sidechain atoms. The velocity Verlet algorithm was used to integrate the equations of motion. The temperature of the simulations was set at 303.15 K using Nóse-Hoover chain [235] with a time constant of 1.0 ps, and the protein and the non-protein atoms were coupled to the thermostat separately. For the NPT equilibration, Parrinello-Rahman coupling was used to set the pressure at 1 bar, with a time constant of 5.0 ps. Periodic boundary conditions were employed, and the particle-mesh Ewald method [202] was used to calculate electrostatic forces with a real space cut-off distance of 1.2 nm. The Lennard-Jones forces were smoothly switched off from 1.0 to 1.2 nm. The LINCS algorithm was used to constrain bonds to hydrogen atoms [203].

4.2.4 Collective variables

In total, 40 collective variables (CVs) are used to control KaiB foldswitch. All simulations involving CVs are done with Plumed 2.5.3 [205]. These CVs, along with their definitions, are given in Table 4.1. First, we explain the CV naming convention. All CV names have the same "type[groups;reference]" format. The groups, or structural elements, employed by the CV are comma-separated. The semicolon notation, when necessary, indicates the KaiB fold referred to by the preceding structural elements. In addition, we use "/" and "+" to represent the union of two secondary structural elements of the same type or different types, respectively. For example, the CV coor[β3+α2/3,β4;gs] employes two groups of atoms: the first includes β3, α2, and α3 and the second group includes β4; all of them refer to the gsKaiB structure. Lastly, if a structural element is disucssed by itself and is present in both KaiB folds (i.e., the N-terminal segment of the protein), the semicolon notation is omitted.

The CVs coor[β1,β4;gs] and coor[β3+α2/3,β4;gs] are defined as the sum of all pairwise contact functions between two groups of atoms, and the contact function between each pair of atoms i and j is defined

$$s_{ij} = \frac{1 - \left(\frac{r_{ij} - d_0}{r_0}\right)^n}{1 - \left(\frac{r_{ij} - d_0}{r_0}\right)^m} \tag{4.1}$$

where r_{ij} is the distance between i and j, $n = 6$, $m = 2n$, and the values of r_0 and d_0 are given in Table 4.1. This function is implemented in Plumed as the COORDINATION CV.

The CVs $\text{RMSD}_{\text{C}\alpha}$[NT,CT;fs] and RMSD_{bb}[CT;fs] are dRMSD functions computed as the root-mean-square deviation of pairwise atom distances from a reference structure:

$$\text{dRMSD} = \sqrt{\frac{1}{N(N-1)} \sum_{i \neq j} \left[r_{ij} - r_{ij}^{\text{ref}} \right]^2}. \tag{4.2}$$

In particular, $\mathrm{RMSD}_{C\alpha}$[NT,CT;fs] is an inter-dRMSD CV, which means that the indices i and j iterate over two different groups of atoms. The dRMSD functions use an upper and lower cutoff distance to accelerate computation; the cutoff values are given in Table 4.1.

The CVs dihe[α2;fs], dihe[α3;fs], and dihe[β3/4;fs] are the cosine-transformed sums of dihedral angle differences from a reference structure:

$$s_{\alpha\beta} = \frac{1}{2} \sum_i \left[1 + \cos\left(\varphi_i - \varphi_i^{\mathrm{ref}} \right) \right] \tag{4.3}$$

In Plumed the $s_{\alpha\beta}$ function is implemented as the ALPHABETA CV.

For the CVs defined by the dRMSD and $s_{\alpha\beta}$ functions, the reference structure is defined to be the cryo-EM structure of fsKaiB (PDB ID: 5N8Y; subunit G) [78].

4.2.5 Steered MD

Starting with the final NVT equilibration structure, an initial KaiB foldswitch pathway was generated with a 5-ns constant-velocity steered MD (SMD) [160] using the MOVING-RESTRAINT function of Plumed, which for each CV x applies a time-dependent harmonic bias $U(x) = \frac{1}{2}\kappa(x - x_0(t))^2$; here κ is the force constant, and $x_0(t)$ is the center of the bias. How the system was steered using the 40 CVs is described in Table 4.2.

Table 4.1: Collective variables for the KaiB foldswitch SMD and string calculations

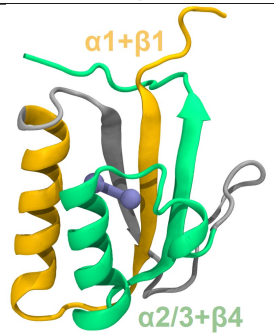
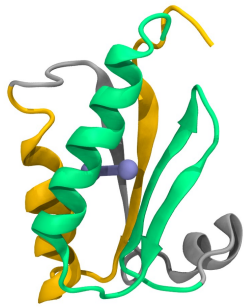
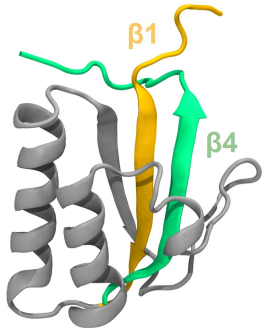
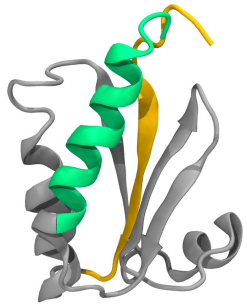
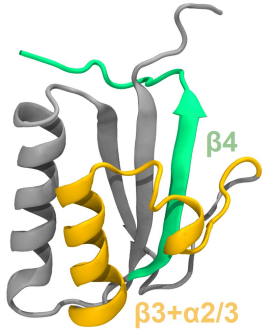
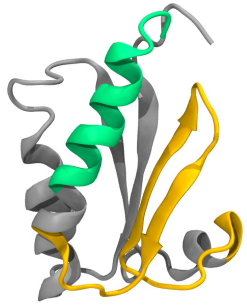
Name	Type	Definition	Structural reference	
			gs	fs
d[α1+β1,α2/3+β4;gs]	COM distance	Cα, residues 1–33 Cα, residues 60–102		
coor[β1,β4;gs]	s^*	Cα, residues 1–15 Cα, residues 85–102		
coor[β3+α2/3,β4;gs]	s^{**}	Cα, residues 55–83 Cα, residues 87–102		

Table 4.1: Collective variables for the KaiB foldswitch SMD and string calculations (continued)

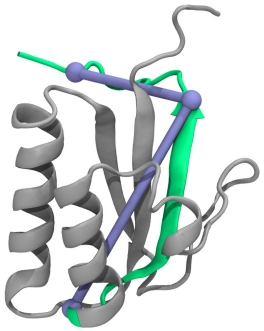
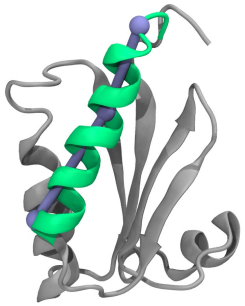
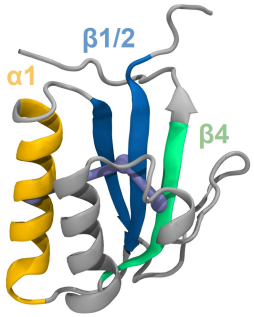
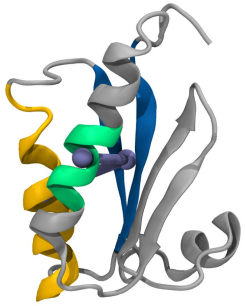
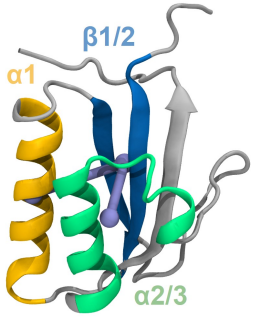
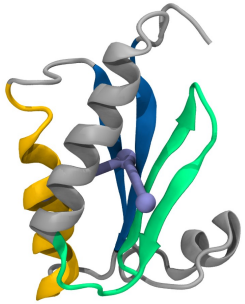
Name	Type	Definition	Structural reference	
			gs	fs
∠[CT loop;gs]	angle	Cα, residues 83, 93, 100		
∠[α1,β1/2,β4;gs]	COM angle	Cα, residues 18–33 Cα, residues 6–13, 39–45 Cα, residues 86–92		
∠[α1,β1/2,α2/3;gs]	COM angle	Cα, residues 18–33 Cα, residues 6–13, 39–45 Cα, residues 62–82		

Table 4.1: Collective variables for the KaiB foldswitch SMD and string calculations (continued)

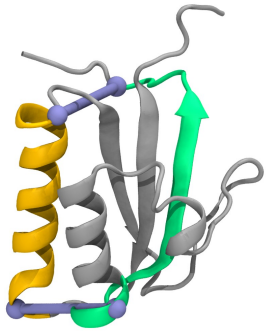
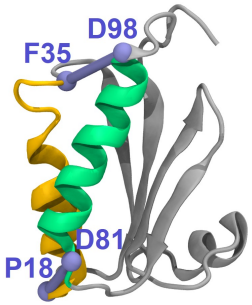
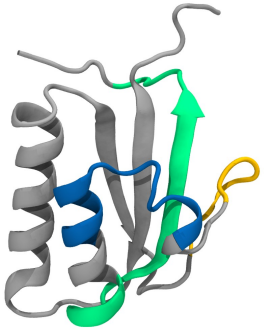
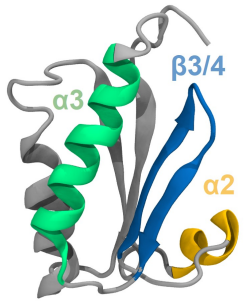
Name	Type	Definition	Structural reference	
			gs	fs
d[D81Cα,P18Cα] d[D98Cα,F35Cα]	distance	—		
dihe[α2;fs]	$s_{\alpha\beta}$	ψ and ϕ, residues 51–57		
dihe[α3;fs]	$s_{\alpha\beta}$	ψ and ϕ, residues 81–98		
dihe[β3/4;fs]	$s_{\alpha\beta}$	ψ and ϕ, residues 63–76		

Table 4.1: Collective variables for the KaiB foldswitch SMD and string calculations (continued)

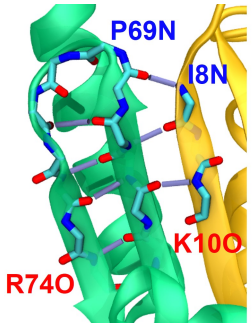
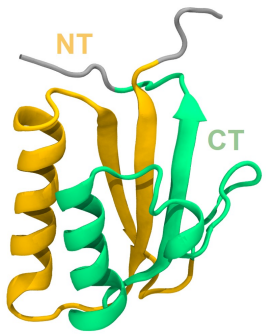
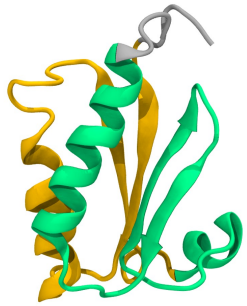
Name	Type	Definition	Structural reference	
			gs	fs
d[L64O,I75N] d[A65O,K10N] d[A67N,I8O] d[L68O,I8N] d[P71O,A67O] d[R73N,K66O] d[R73O,K66N]	distance	—		
$RMSD_{C\alpha}$[NT,CT;fs]	dRMSD (inter)†	Cα, residues 6–50 Cα, residues 51–98		
$RMSD_{bb}$[CT;fs]	dRMSD‡	backbone, residues 51–98		

Table 4.1: Collective variables for the KaiB foldswitch SMD and string calculations (continued)

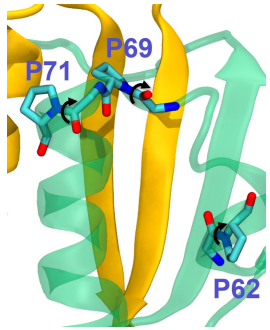
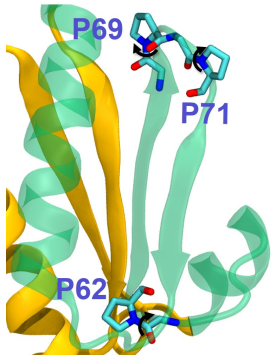
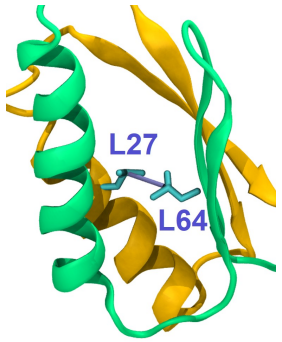
Name	Type	Definition	Structural reference	
			gs	fs
ω[P62] ω[P69] ω[P71]	dihedral	for residue n, the angle is defined by O($n-1$), C($n-1$), N(n), and Cα(n)		
d[L64Cγ,L27Cγ]	distance	—		

Table 4.1: Collective variables for the KaiB foldswitch SMD and string calculations (continued)

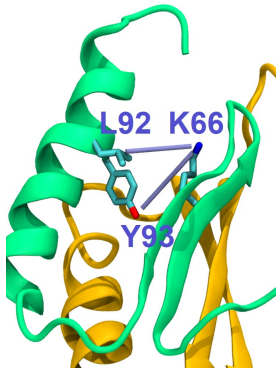
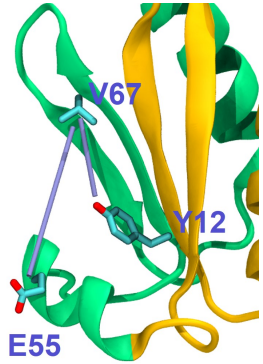
Name	Type	Definition	Structural reference	
			gs	fs
d[K66Nζ,L92Cγ] d[K66Nζ,Y93Oη]	distance	—		
d[V67Cβ,Y12Oη] d[V67Cβ,E55Cδ]	distance	—		

Table 4.1: Collective variables for the KaiB foldswitch SMD and string calculations (continued)

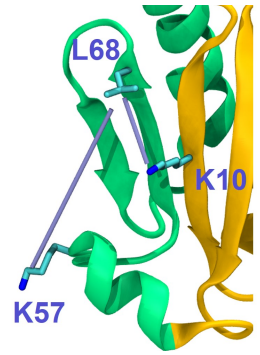
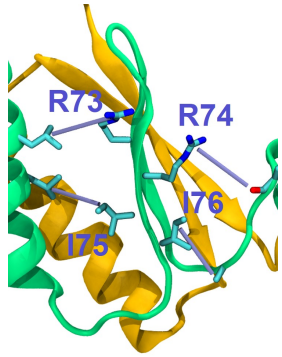
Name	Type	Definition	Structural reference	
			gs	fs
d[L68Cγ,K10Nζ] d[L68Cγ,K57Nζ]	distance	—		
d[R73Cζ,L92Cγ] d[R74Cζ,K57O] d[I75Cγ1,E89Cγ] d[I76Cγ1,A60Cβ]	distance	—		

Table 4.1: Collective variables for the KaiB foldswitch SMD and string calculations (continued)

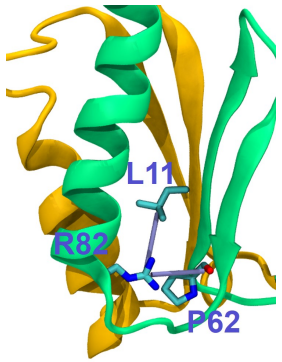
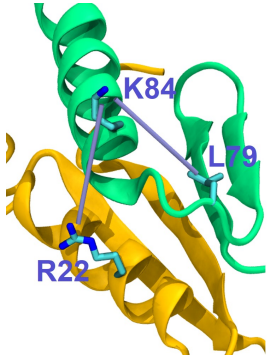
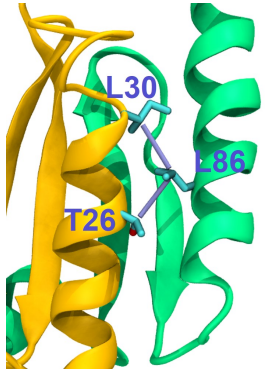
Name	Type	Definition	Structural reference	
			gs	fs
d[R82Cζ,L11Cγ] d[R82Cζ,P62O]	distance	—		
d[K84Nζ,R22Cζ] d[K84Nζ,L79Cγ]	distance	—		

Table 4.1: Collective variables for the KaiB foldswitch SMD and string calculations (continued)

Name	Type	Definition	Structural reference	
			gs	fs
d[L86Cγ,T26Cβ] d[L86Cγ,L30Cβ]	distance	—		

* $d_0 = 0.8$ nm, $r_0 = 0.3$ nm

** $d_0 = 0.7$ nm, $r_0 = 0.3$ nm

† 0 nm lower cutoff; 5 nm upper cutoff

‡ 0 nm lower cutoff; 2 nm upper cutoff

4.2.6 String method

60 structures that were roughly equally spaced along the steered foldswitch pathway were chosen to initialize the string method [161] after a 1 ns equilibration with restraints in the same 40-dimensional CV space. In each iteration, the initial structure for each image was steered (constant velocity) towards the center of the image over 15 ps with 1.2x the force constants listed in Table 4.2, which was then followed by another 15 ps of equilibration with a constant restraint at the image center. If the final structure was not within the Voronoi cell of the image or over $3\sqrt{RT/\kappa}$ away from the image center, the process was repeated with all the force constants further multiplied by a factor of 1.2. Here R is the gas constant and T is the simulation temperature. Once the equilibration steps were successful, the force constants were reduced to the original values (Table 4.2) over 15 ps, which was then followed by a 65-ps production run with constant restraints; the last 50 ps was used to approximate the local free energy gradient ∇F.

At the end of each iteration, the center of each image x_i in the CV space was updated according to

$$x_i(\tau + \Delta\tau) = x_i(\tau) - \Delta\tau P(x(\tau))\nabla F(x_i(\tau)). \tag{4.4}$$

The projection matrix P removes the component of ∇F parallel to the string at each image. If the image is at either end of the string, the matrix P is not computed. Here, τ tracks the progress of the gradient descent in the CV space. The string was propagated with a constant multiplier $\Delta\tau = 0.0001$. The updated string was linearly interpolated and smoothed (with a smoothing factor $s = 0.4$; see [161]) before the next iteration. Correction for non-linear dependences between the CVs was not applied.

Convergence of the string was monitored using the Hausdorff distance from the initial string [236]. The string method was implemented as an in-house Python wrapper for Gromacs and Plumed.

4.2.7 FAST method

A modified version of the FAST method [237, 238], an adaptive sampling method, was used to explore the KaiB conformational space beyond the pathway identified by the string method. For each iteration, a set of 60 initial structures were used to generate 10-ns unbiased trajectories. The trajectories were aggregated and clustered using the regular space clustering algorithm [239] implemented in PyEMMA [240]. The structures were clustered in the space of all pairwise Cα distances. The minimal cluster center distance $d_{\min}$ was chosen using a bisection algorithm such that the number of clusters fell between 2 to 5 times the number of initial structures. A random structure from each cluster was chosen to be the "center" of the cluster for the following analysis.

The FAST method scores each cluster center i using the function

$$r(i) = \overline{\phi}(i) + \alpha\overline{\psi}(i) + \beta\chi(i) \tag{4.5}$$

This function has three components and two weights, α and β. The first term, $\overline{\phi}(i)$, is the directed component that favors the optimization of certain structural metrics; here we define $\overline{\phi}(i) = [x_{\max} - x(i)]/[x_{\max} - x_{\min}]$ for the CV $x = \mathrm{RMSD_{bb}}$[CT;fs] to favor the minimization of dRMSD difference from the fsKaiB structure. The second term, $\overline{\psi}(i)$, is the undirected component that favors exploration of new configurations and is defined to be $\overline{\psi}(i) = [C_{\max} - C_i]/[C_{\max} - C_{\min}]$ where C_i is the number of structures in cluster i. Lastly, the third term, $\chi(i)$, penalizes clusters that are structurally similar to existing clusters:

$$\chi(i) = \frac{1}{N}\sum_j \left(1 - e^{-r_{ij}^2/2d_{\min}^2}\right) \tag{4.6}$$

where r_{ij} is the distance between the centers of clusters i and j in the space of all pairwise Cα distances. This third term is computed iteratively to pick out the best clusters to seed the next iteration using the following procedure:

1. Rank all clusters using the first two terms of the scoring function $r(i)$, which we denote as $\rho(i) = \overline{\phi}(i) + \alpha\overline{\psi}(i)$.

2. Compute the third term $\chi(i)$ for each cluster; the index j in (4.6) iterates over clusters from all past iterations. Denote this value as $\chi_{\text{old}}(i)$. The highest ranked cluster is the cluster i that maximizes $\rho(i) + \beta\chi_{\text{old}}(i)$. If this is the first iteration, the highest ranked cluster is simply the cluster i that maximizes $\rho(i)$. Add i to an index set S.

3. Compute the third term $\chi(i)$ for each cluster, but let the index j in (4.6) iterate over the set S. Denote this value as $\chi_{\text{new}}(i)$. The highest ranked cluster is the cluster $i \notin S$ that maximizes $\rho(i) + \beta[\gamma\chi_{\text{old}}(i) + (1-\gamma)\chi_{\text{new}}(i)]$. If this is the first iteration, the highest ranked cluster is the cluster $i \notin S$ that maximies $\rho(i) + \beta\chi_{\text{new}}(i)$. Add i to the index set S. Here, $\gamma \in [0, 1]$ is a weight factor.

4. Repeat the previous step until S reaches the number of clusters required to seed the next iteration (in this case 60).

For the current simulations, we chose $\alpha = 2$, $\beta = 2$, and $\gamma = 0.3$. The structures from the last iteration of the string method were used to seed the first iteration of the FAST method. The FAST method was implemented as an in-house Python wrapper for Gromacs and Plumed.

Table 4.2: SMD and string setup for the KaiB foldswitch simulation

Function	Collective variable	SMD			String
		t (ns)	κ (kJ/mol)*	x_0	κ (kJ/mol)
Dissociate CT from NT	d[α1+β1,α2/3+β4;gs]	0.0–2.0 2.0–2.4 2.4–	4000/nm^2 4000–0/nm^2 0/nm^2	0.7–2.0 nm 2.0 nm 2.0 nm	2400/nm^2
Dissociate β4;gs from β1	coor[β1,β4;gs]	0.0–2.0 2.0–2.4 2.4–4.0 4.0–	500–4000 4000–400 400–0 0	80–0 0 0 0	100
Dissociate β4;gs from β3+α2/3;gs	coor[β3+α2/3,β4;gs]	1.0–1.4 1.4–1.8 1.8–	500–4000 4000–0 0	43–0 0 0	
Reposition the CT loop (part of α3;fs)	∠[CT loop;gs]	0.0–0.8 0.8–2.4 2.4–	0–2500/rad^2 2500–50/rad^2 50/rad^2	1.3–3.14 rad 3.14 rad 3.14 rad	1500/rad^2
Reposition β4;gs (part of α3;fs)	∠[α1,β1/2,β4;gs]	2.0–2.8 2.8–2.88 2.88–	0–5000/rad^2 5000–0/rad^2 0/rad^2	1.8–0.0 rad 0.0 rad 0.0 rad	2400/rad^2
Prevent entanglement between β4;gs and α2/3;gs	∠[α1,β1/2,α2/3;gs]	2.0–2.8 2.8–2.88 2.88–	0–5000/rad^2 5000–0/rad^2 0/rad^2	1.2 rad	
Position α3;fs relative to α1	d[D81Cα,P18Cα] d[D98Cα,F35Cα]	2.0–2.8 2.8–4.4 4.4–	0–5000/nm^2 5000–0/nm^2 0/nm^2	1.455 nm 1.322 nm	3000/nm^2
Enforce fs 2° structures	dihe[α2;fs]	2.0–4.0 4.0–	0–10000 10000	19–36 36	250
	dihe[α3;fs]	2.0–4.0 4.0–	0–10000 10000	8–14 14	
	dihe[β3/4;fs]	2.0–4.0 4.0–	0–12000 12000	14–28 28	300

Table 4.2: SMD and string setup for the KaiB foldswitch simulation (continued)

Function	Collective variable	SMD			String
		t (ns)	κ (kJ/mol)*	x_0	κ (kJ/mol)
Enforce β3/4;fs 2° structures Enforce contact between β2 and β3;fs	d[L64O,I75N]	3.2–4.8	0–500/nm^2	0.262 nm	300/nm^2
	d[A65O,K10N]			0.303 nm	
	d[A67N,I8O]			0.321 nm	
	d[L68O,I8N]	4.8–	500/nm^2	0.261 nm	
	d[P71O,A67O]			0.401 nm	
	d[R73N,K66O]			0.316 nm	
	d[R73O,K66N]			0.312 nm	
Enforce fs 2°/3° structures	$\text{RMSD}_{\text{C}\alpha}$[NT,CT;fs]	3.6–4.8	0–50000/nm^2	0.6–0.0 nm	1200/nm^2
		4.8–	5000/nm^2	0.0 nm	
	RMSD_{bb}[CT;fs]	3.6–4.8	0–50000/nm^2	0.5 nm	
		4.8–	5000/nm^2	0.0 nm	
Proline isomerization	ω[P62]	2.0–2.8	0–500/rad^2	0–3.1 rad	300/rad^2
	ω[P69]	2.8–	500/rad^2	3.1 rad	
	ω[P71]	2.0–2.8	0–500/rad^2		
		2.8–	500/rad^2	0.0 rad	
Sidechain repacking	d[L64Cγ,L27Cγ]	2.4–4.0	0–250/nm^2	2.780 nm	150/nm^2
		4.0–	250/nm^2	0.492 nm	
	d[K66Nζ,L92Cγ]	2.4–4.0	0–250/nm^2	2.430 nm	
		4.0–	250/nm^2	0.509 nm	
	d[K66Nζ,Y93Oη]	2.4–4.0	0–250/nm^2	2.020 nm	
		4.0–	250/nm^2	0.527 nm	
	d[V67Cβ,Y12Oη]	2.4–4.0	0–250/nm^2	2.490 nm	
		4.0–	250/nm^2	0.780 nm	
	d[V67Cβ,E55Cδ]	2.4–4.0	0–250/nm^2	2.470 nm	
		4.0–	250/nm^2	1.297 nm	

Table 4.2: SMD and string setup for the KaiB foldswitch simulation (continued)

Function	Collective variable	SMD			String
		t (ns)	κ (kJ/mol)*	x_0	κ (kJ/mol)
	d[L68Cγ,K10Nζ]	2.4–4.0	0–250/nm^2	2.440 nm	
		4.0–	250/nm^2	0.889 nm	
	d[L68Cγ,K57Nζ]	2.4–4.0	0–250/nm^2	1.510 nm	
		4.0–	250/nm^2	2.035 nm	
	d[R73Cζ,L92Cγ]	2.4–4.0	0–250/nm^2	2.040 nm	
		4.0–	250/nm^2	0.598 nm	
	d[R74Cζ,K57O]	2.4–4.0	0–250/nm^2	3.030 nm	
		4.0–	250/nm^2	0.444 nm	
	d[I75Cγ1,E89Cγ]	2.4–4.0	0–250/nm^2	1.500 nm	
		4.0–	250/nm^2	0.522 nm	
	d[I76Cγ1,A60Cβ]	2.4–4.0	0–250/nm^2	2.470 nm	
		4.0–	250/nm^2	0.602 nm	
	d[R82Cζ,L11Cγ]	2.4–4.0	0–250/nm^2	2.730 nm	
		4.0–	250/nm^2	0.650 nm	
	d[R82Cζ,P62O]	2.4–4.0	0–250/nm^2	3.010 nm	
		4.0–	250/nm^2	0.506 nm	
	d[K84Nζ,R22Cζ]	2.4–4.0	0–250/nm^2	1.800 nm	
		4.0–	250/nm^2	1.570 nm	
	d[K84Nζ,L79Cγ]	2.4–4.0	0–250/nm^2	1.230 nm	
		4.0–	250/nm^2	1.104 nm	
	d[L86Cγ,T26Cβ]	2.4–4.0	0–250/nm^2	1.890 nm	
		4.0–	250/nm^2	0.333 nm	

Table 4.2: SMD and string setup for the KaiB foldswitch simulation (continued)

Function	Collective variable	SMD			String
		t (ns)	κ (kJ/mol)*	x_0	κ (kJ/mol)
	d[L86Cγ,L30Cβ]	2.4–4.0	0–250/nm^2	2.090 nm	
		4.0–	250/nm^2	0.474 nm	

$*$ An "–" indicate that the value of κ (or x_0) is to be linearly interopated over the corresponding time span.

4.3 Results

As detailed in the Methods section, SMD was used to generate an initial guess for the pathway for KaiB foldswitching. As outlined in Table 4.2, the steered pathway primarily revolves around the movement towards α1 during foldswitch of residues in and near β4;gs, which forms α3;fs. To achieve this, the steered pathway first disrupts the C-terminal segment of the protein fold, both its contacts with the N-terminal part of the protein, especially the ground-state β sheet structure containing β1 and β4;gs, and the contacts within the C-terminal segment. This is then followed by repositioning of the β4;gs residues towards the target position, while preventing entanglement between β4;gs with α2/3;gs; in particular, the interaction between α3;gs and α1, as well as the interactions of the gs C-terminal loop, which wraps around the N-terminal structural elements, needs to be disrupted to allow for β4;gs movement. The repositioning of β4;gs is accompanied by RMSD, backbone dihedral, and distance restraints to enforce the fsKaiB C-terminal secondary and tertiary structures. It is worth noting that attempts to apply straightforward RMSD restraints did not lead to successful foldswitching, presumably due to the complexity of the changes in protein topology.

The initial SMD pathway was relaxed using the string method. The progression of the string method along the 40 CVs used to control the foldswitch process is shown in Table 4.3. Unsurprisingly, the string pathway shows qualitatively similar foldswitch patterns to the steered pathway, in that the dissociation and repositioning of β4;gs takes place before the formation of the correct fsKaiB secondary and tertiary structures (Fig. 4.1). However, what the string pathway demonstrates is that a large portion of the foldswitch pathway goes through molten globules as the topology of the C-terminal segment is rearranged while little to no stable secondary structures exist. Furthermore, the molten globules sampled in the string path are much more collapsed than those in the steered pathway (see, e.g., the string progression for $\mathrm{RMSD}_{\mathrm{C}\alpha}$[NT,CT;fs] and $\mathrm{RMSD}_{\mathrm{bb}}$[CT;fs] in Table 4.3). The convergence of the string method was monitored using the Hausdorff distance

from the initial string (Table 4.3, last panel), which shows that the simuation has yet to converge after the first 300 iterations.

In general, protein folding is characterized by the existence of multiple pathways. As such, the foldswitch pathway identified in the SMD and refined using the string method is likely only one of many competing pathways. Therefore, we have started to further characterize the KaiB foldswitch process and discover potential alternative pathways using the FAST method, with the string results as an initial starting point. The end goal is to leverage recent advances in chemical kinetics, such as the dynamical Galerkin approximation [241], to characterize the kinetics and energetics of the KaiB foldswitch pathway(s) and its intermediates. This computational pipeline can then be applied to KaiB in complex with KaiC to investigate the possible role of KaiC in setting the timescale of KaiB foldswitch.

Table 4.3: KaiB foldswitch string progression. The strings are shown as projected onto each CV space and colored by iteration from blue (iteration 0) to red (iteration 310). The strings are plotted every 10 iterations to prevent cluttering. The last panel shows the Hausdorff distance of each string iteration from the initial string.

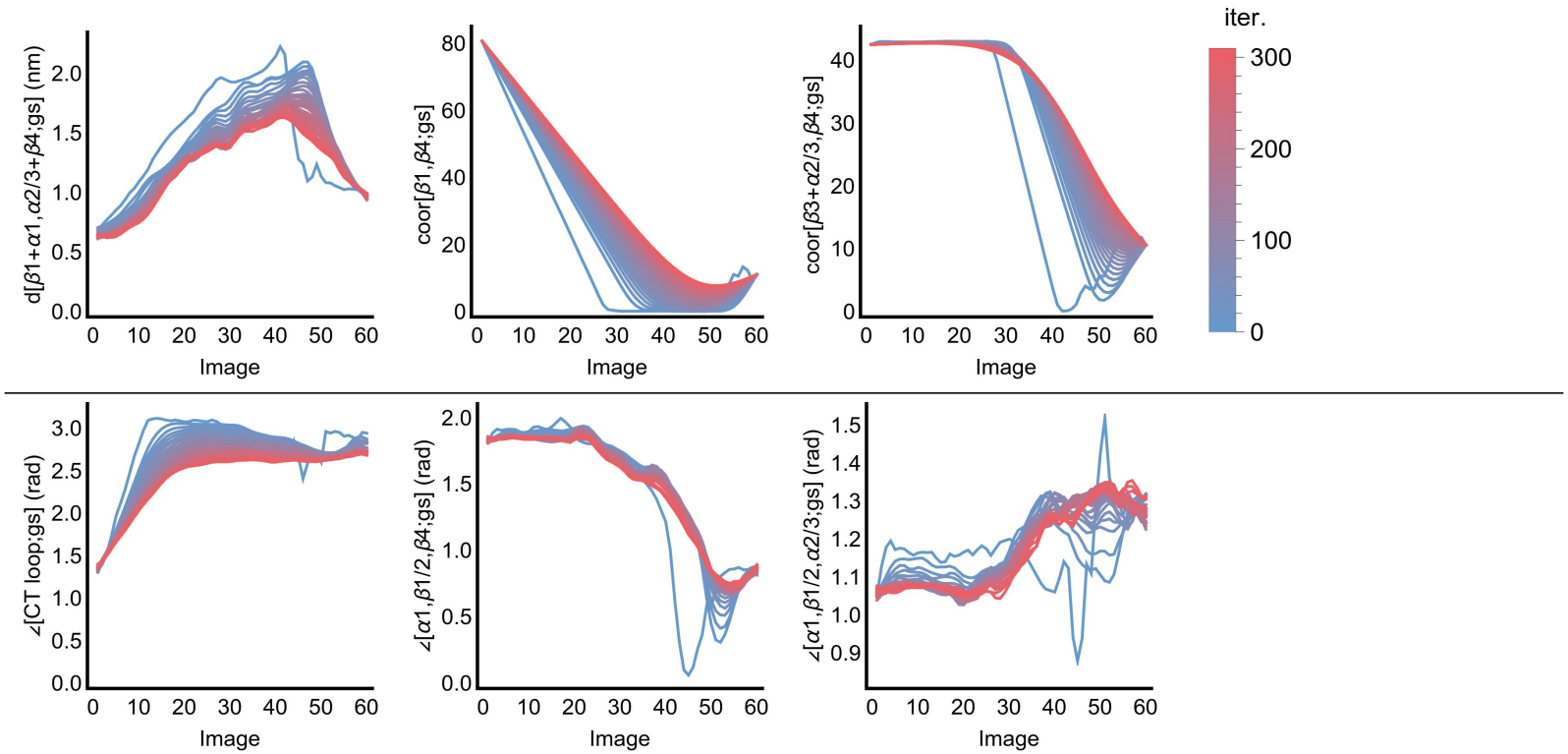


Table 4.3: KaiB foldswitch string progression (continued)

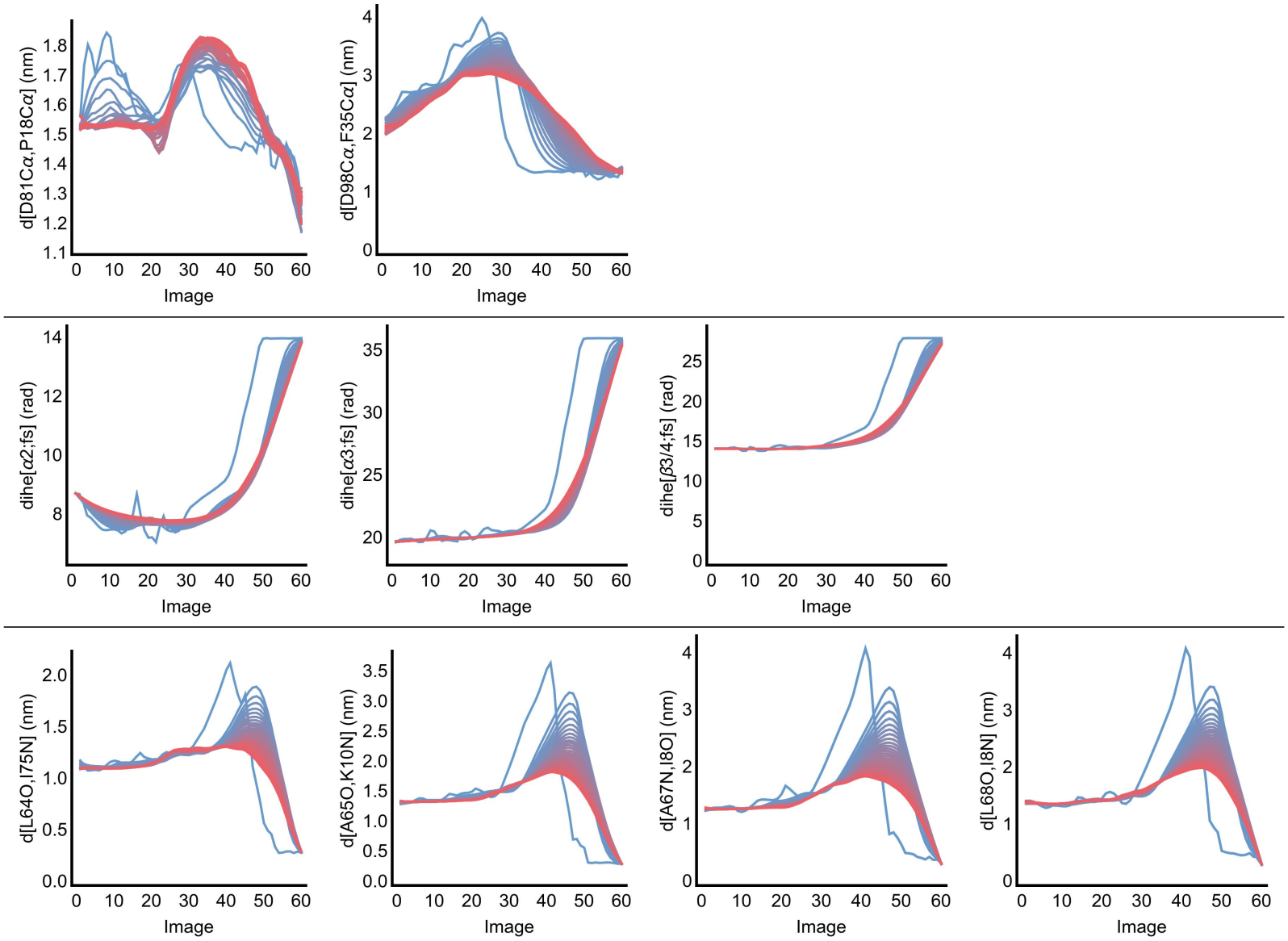


Table 4.3: KaiB foldswitch string progression (continued)

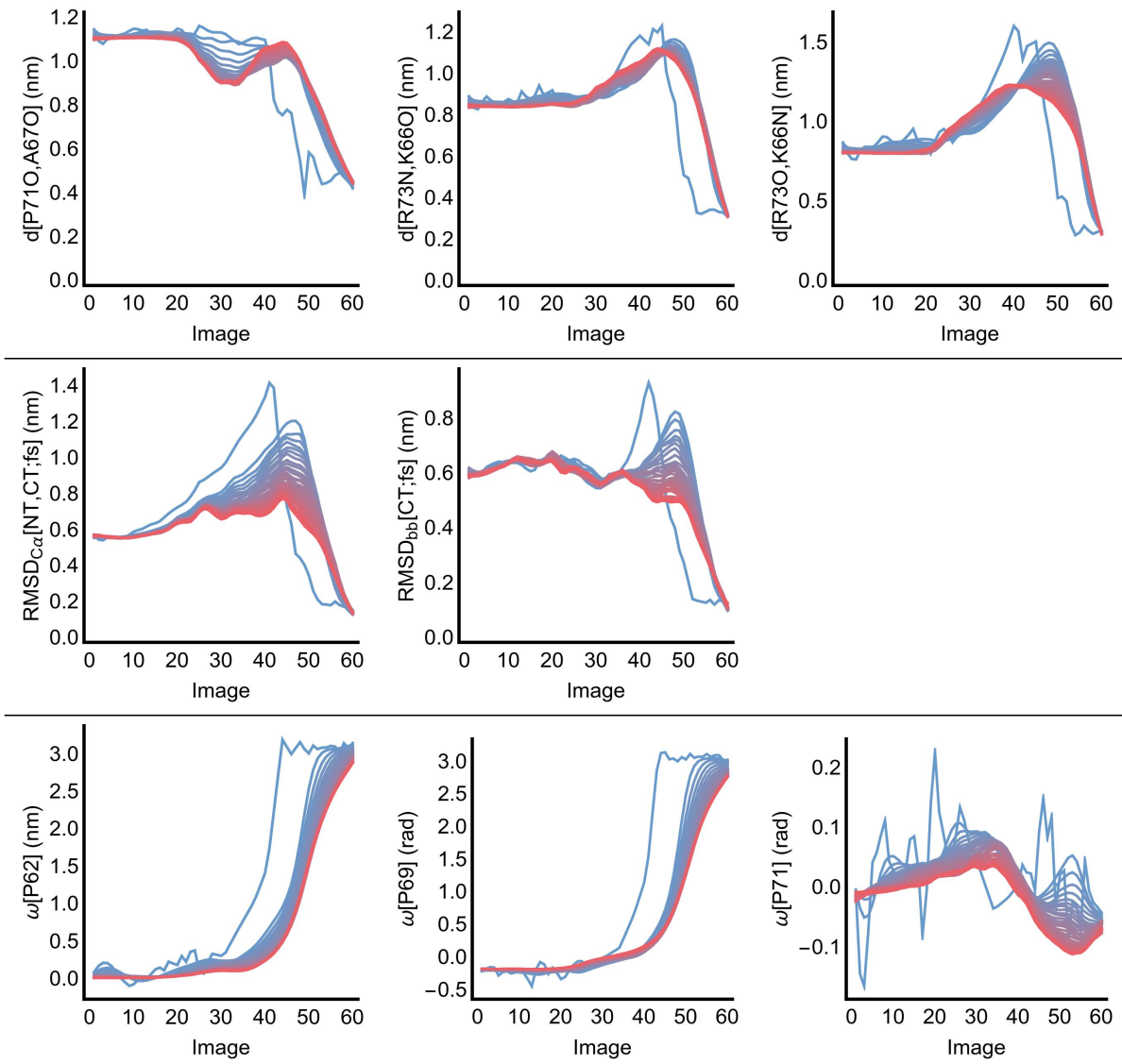


Table 4.3: KaiB foldswitch string progression (continued)

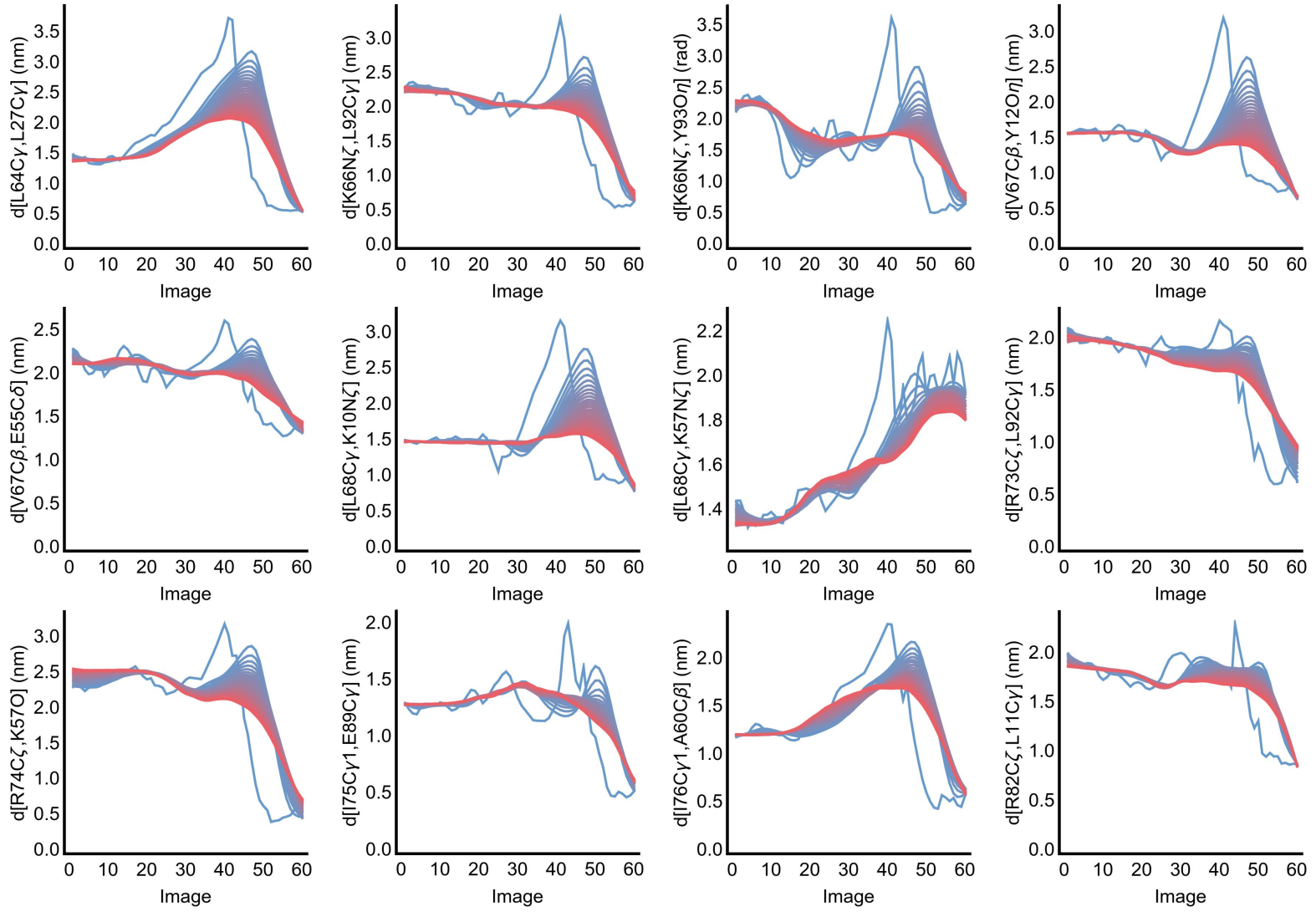
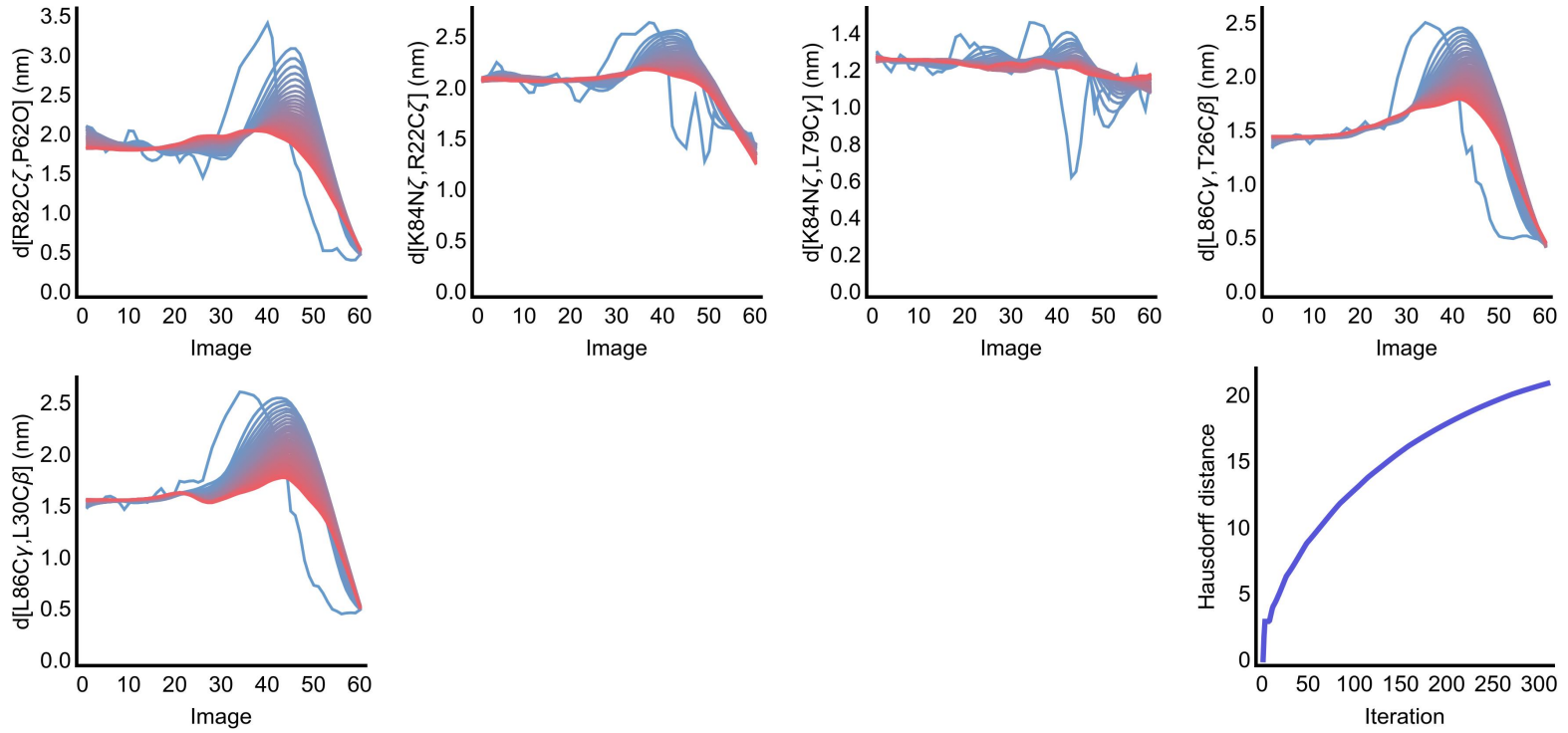


Table 4.3: KaiB foldswitch string progression (continued)



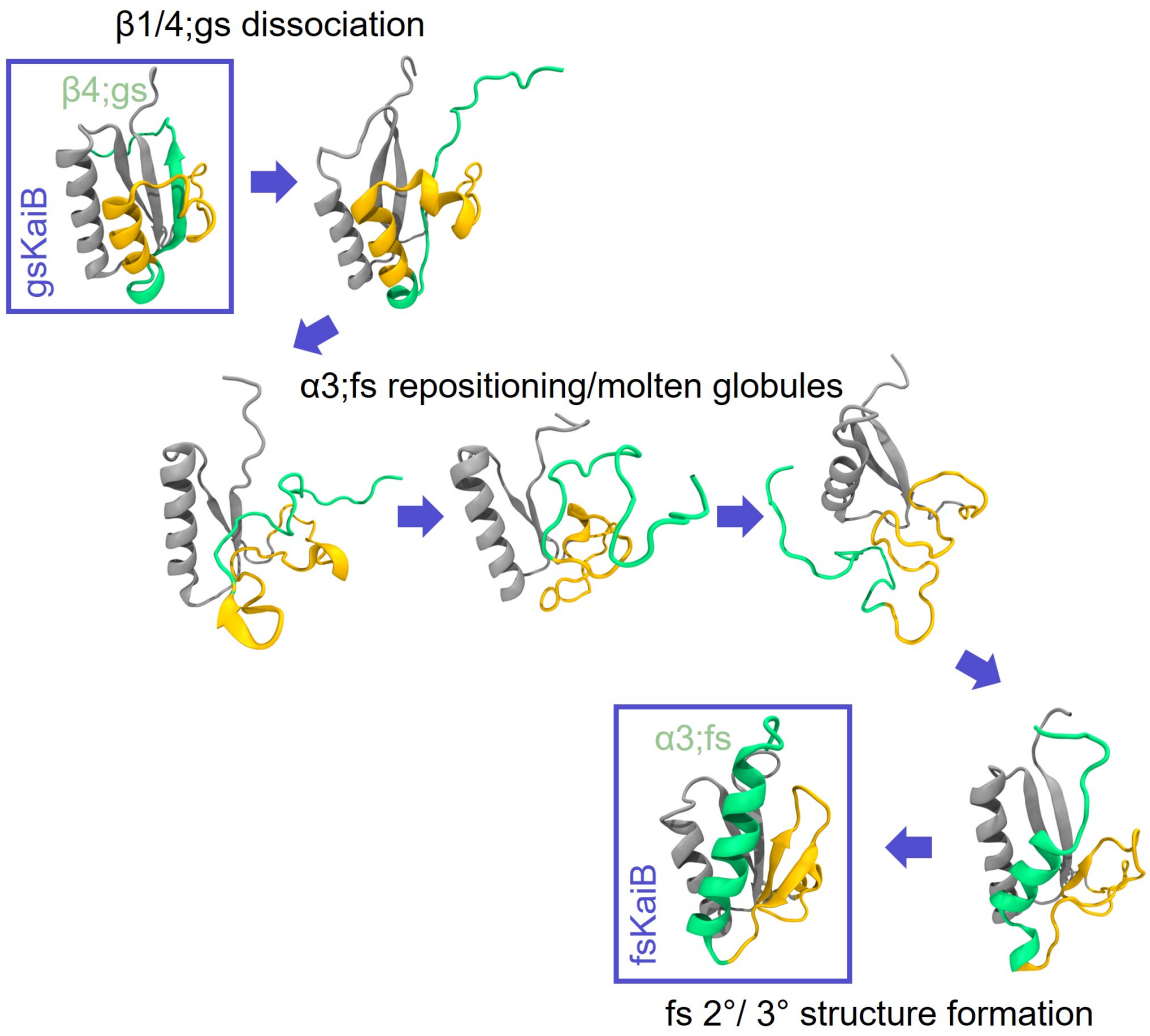


Figure 4.1: Structures of the KaiB foldswitch string pathway. Starting with the gsKaiB structure, the string pathway first unzips the contact between β1 and β4;gs and then repositions β4;gs towards α1 as the pathway goes through molten globules. Lastly, residues around β4;gs refolds into α3;fs as the C-terminal segment adopts the correct fsKaiB secondary and tertiary structures. The gray segment (N-terminal): residues 1–50; the yellow segment: residues 51–80; the green segment: residues 81–102. The structures are chosen from the images 0, 18, 25, 30, 43, 54, and 58.

CHAPTER 5

CONCLUSION AND FUTURE OUTLOOK

The work presented in this thesis was motivated by the question of how the Kai oscillator couples to metabolism. To address this question, we first employed kinetic modeling, along with Bayesian MCMC, to investigate the interaction between KaiA and KaiC (Chapter 2). The resulting model revealed unexpected ultrasensitivity in KaiC phosphorylation due to differential KaiA binding affinities to the ADP- vs. ATP-bound states of KaiC. We argued that the ultrasensitivity plays a role in period stability by acting as a synchronization mechanism at night. Delving into the molecular details, we then used MD simulations to investigate the mechanism of ADP release in KaiC and how the process may be accelerated by KaiA binding (Chapter 3). The simulations suggest that KaiA binding is coupled to a partial dissociation of the subunit interface that forms the nucleotide-binding pockets, thus allowing solvent influx to compensate for lost electrostatic interactions between the protein and ADP. Together, these results provide new insights that augment our current understanding of metabolic compensation.

Much of the work presented in this thesis is guided by the belief that an understanding of the oscillator, and our ability to manipulate it, can only be achieved through a detailed, quantitative analysis of the structural and biophysical basis of the oscillator components and their interactions. In this spirit, I outline here three broad areas of inquiry that could benefit from the approach taken in this thesis and further our understanding of the Kai system.

First, many questions remain of the biochemistry of KaiC. As a homohexameric protein with multiple catalytic activities, the MD simulations presented in this thesis only begin to scratch the complexity of the protein. Is nucleotide exchange in KaiC concerted, sequential, or stochastic? KaiA and KaiC appear to bind mostly with 1:1 stoichiometry [134], but the complex is highly dynamic [164], and forms and dissociates on a very fast timescale [25]; what does this say about the mechanism of nucleotide exchange? What is

the pathway that allows communication between the CI and CII domain? Is intersubunit communication important for accounting for the enzymatic activity of KaiC? Over the phosphorylation cycles, are certain phosphoforms preferentially enriched in subpopulations of KaiC, or are all phosphorylation states uniformly distributed in each KaiC hexamer?

Second, the exact molecular mechanism underlying the negative feedback loop is unclear. Specifically, what sets the timescale of KaiB binding, ATP hydrolysis in CI or KaiB foldswitch? How exactly does free energy dissipation in KaiC stabilize the foldswitch KaiB conformation? What is the basis of the KaiB cooperative binding? Why does KaiB binding exhibit an initial burst phase followed by slower binding kinetics [22]? For cyanobacterial species that do not have KaiA, what is the function of KaiB binding? It is known that KaiA adopts an inactivated conformation when sequestered in an inhibitory complex with KaiB [77] and can disassemble into KaiA monomers as KaiB dissociates from KaiC at the late dephosphorylation stage [78]; very little is known about the structural basis of such transitions, and whether they contribute to the slowness of the oscillator.

Lastly, the question of what sets the 24-h period of the oscillator is still unanswered. It is likely that the period of the oscillator is dominated by a few slow steps in the system. Can one write down an equation for the period in terms of the rates of such slow steps? Currently, point mutants in KaiC have been identified that can elongate the period of the oscillator by up to a factor of six, but mutants that accelerate the oscillator to a comparable extent are not known. This suggests a separation of timescale that breaks down as the oscillator becomes faster, where more and more steps contribute to the bottleneck that limits the oscillator speed. What is the timescale hierarchy in the system? What is the range of rate constants the system can tolerate and still generate oscillatory dynamics? Does the range become smaller as the oscillator becomes faster? Moreover, metabolism is not the only environmental factor that the oscillator has to contend with, and how the oscillator period remains insensitive to temperature perturbations (i.e.,

temperature compensation) is another major unresolved issue. Theoretical speculations of the mechanism of temperature compensation abound, but which model is correct? Are the various forms of compensation related to the more general phenomenon of (near-)perfect adaptation [242]?

Over the past decade or so, the Kai oscillator has attracted significant experimental and theoretical interests partly due to its simplicity. However, these efforts have only highlighted how much we still do not understand about the oscillator, let alone being able to rationally manipulate its behavior. Hopefully, the preceding discussion could serve as a useful point of reference to stimulate further studies of the Kai oscillator. As more becomes known about the oscillator, it is also important to ask what features of the oscillator are evolutionary contingencies and what embody general principles required for oscillation, a question that I attempted to tackle in Chapter 1. Such a distinction likely cannot be made prior to a thorough, quantitative examination of the oscillator, an approach advocated in this thesis, but an analysis of this question will also hinge on a more comparative and phenomenological examination of clock biology. Such a synthesis would go a long way in making circadian biology a more robust, predictive science.

REFERENCES

[1] Novák B, Tyson JJ (2008) Design principles of biochemical oscillators. *Nature Reviews Molecular Cell Biology* 9(12):981–991.

[2] Swan JA, Golden SS, LiWang A, Partch CL (2018) Structure, function, and mechanism of the core circadian clock in cyanobacteria. *Journal of Biological Chemistry* 293(14):5026–5034.

[3] 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.

[4] Zwicker D, Lubensky DK, ten Wolde PR (2010) Robust circadian clocks from coupled protein-modification and transcription–translation cycles. *Proceedings of the National Academy of Sciences* 107(52):22540–22545.

[5] Qin X, Byrne M, Xu Y, Mori T, Johnson CH (2010) Coupling of a Core Post-Translational Pacemaker to a Slave Transcription/Translation Feedback Loop in a Circadian System. *PLOS Biol* 8(6):e1000394.

[6] Liu Y, et al. (1995) Circadian orchestration of gene expression in cyanobacteria. *Genes & Development* 9(12):1469–1478.

[7] Takai N, et al. (2006) A KaiC-associating SasA-RpaA two-component regulatory system as a major circadian timing mediator in cyanobacteria. *Proceedings of the National Academy of Sciences* 103(32):12109–12114.

[8] Gutu A, O'Shea EK (2013) Two antagonistic clock-regulated histidine kinases time the activation of circadian gene expression. *Molecular cell* 50(2):288.

[9] Grobbelaar N, Huang TC, Lin HY, Chow TJ (1986) Dinitrogen-fixing endogenous rhythm in Synechococcus RF-1. *FEMS Microbiology Letters* 37(2):173–177.

[10] Mitsui A, et al. (1986) Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically. *Nature* 323(6090):720–722.

[11] Sweeney BM, Borgese MB (1989) A Circadian Rhythm in Cell Division in a Prokaryote, the Cyanobacterium Synechococcus Wh78031. *Journal of Phycology* 25(1):183–186. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0022-3646.1989.00183.x.

[12] Tomita J, Nakajima M, Kondo T, Iwasaki H (2005) No Transcription-Translation Feedback in Circadian Rhythm of KaiC Phosphorylation. *Science* 307(5707):251–254.

[13] Nakajima M, et al. (2005) Reconstitution of Circadian Oscillation of Cyanobacterial KaiC Phosphorylation in Vitro. *Science* 308(5720):414–415.

[14] Schmelling NM, et al. (2017) Minimal tool set for a prokaryotic circadian clock. *BMC Evolutionary Biology* 17:169.

[15] Iwasaki H, Nishiwaki T, Kitayama Y, Nakajima M, Kondo T (2002) KaiA-stimulated KaiC phosphorylation in circadian timing loops in cyanobacteria. *Proceedings of the National Academy of Sciences* 99(24):15788–15793.

[16] Hayashi F, et al. (2003) ATP-induced hexameric ring structure of the cyanobacterial circadian clock protein KaiC. *Genes to Cells* 8(3):287–296.

[17] 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.

[18] Nishiwaki T, et al. (2007) A sequential program of dual phosphorylation of KaiC as a basis for circadian rhythm in cyanobacteria. *The EMBO Journal* 26(17):4029–4037.

[19] 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. *Proceedings of the National Academy of Sciences* 111(12):4455–4460.

[20] Chang YG, Tseng R, Kuo NW, LiWang A (2012) Rhythmic ring–ring stacking drives the circadian oscillator clockwise. *Proceedings of the National Academy of Sciences* 109(42):16847–16851.

[21] Lin J, Chew J, Chockanathan U, Rust MJ (2014) Mixtures of opposing phosphorylations within hexamers precisely time feedback in the cyanobacterial circadian clock. *Proceedings of the National Academy of Sciences* 111(37):E3937–E3945.

[22] Chang YG, et al. (2015) A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science* 349(6245):324–328.

[23] 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* 111(4):1379–1384.

[24] Kageyama H, et al. (2006) Cyanobacterial Circadian Pacemaker: Kai Protein Complex Dynamics in the KaiC Phosphorylation Cycle In Vitro. *Molecular Cell* 23(2):161–171.

[25] Mori T, et al. (2018) Revealing circadian mechanisms of integration and resilience by visualizing clock proteins working in real time. *Nature Communications* 9(1):3245.

[26] Terauchi K, et al. (2007) ATPase activity of KaiC determines the basic timing for circadian clock of cyanobacteria. *Proceedings of the National Academy of Sciences* 104(41):16377–16381.

[27] Zhang D, Cao Y, Ouyang Q, Tu Y (2019) The energy cost and optimal design for synchronization of coupled molecular oscillators. *Nature Physics* pp. 1–6.

[28] Phong C, Markson JS, Wilhoite CM, Rust MJ (2013) Robust and tunable circadian rhythms from differentially sensitive catalytic domains. *Proceedings of the National Academy of Sciences* 110(3):1124–1129.

[29] 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.

[30] Heisler J, Chavan A, Chang YG, LiWang A (2019) Real-Time In Vitro Fluorescence Anisotropy of the Cyanobacterial Circadian Clock. *Methods and Protocols* 2(2):42.

[31] Murayama Y, et al. (2017) Low temperature nullifies the circadian clock in cyanobacteria through Hopf bifurcation. *Proceedings of the National Academy of Sciences* 114(22):5641–5646.

[32] Yoshida T, Murayama Y, Ito H, Kageyama H, Kondo T (2009) Nonparametric entrainment of the in vitro circadian phosphorylation rhythm of cyanobacterial KaiC by temperature cycle. *Proceedings of the National Academy of Sciences* 106(5):1648–1653.

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

[34] Hatakeyama TS, Kaneko K (2012) Generic temperature compensation of biological clocks by autonomous regulation of catalyst concentration. *Proceedings of the National Academy of Sciences* 109(21):8109–8114.

[35] Kidd PB, Young MW, Siggia ED (2015) Temperature compensation and temperature sensation in the circadian clock. *Proceedings of the National Academy of Sciences* 112(46):E6284–E6292.

[36] Paijmans J, Lubensky DK, ten Wolde PR (2017) Period Robustness and Entrainability of the Kai System to Changing Nucleotide Concentrations. *Biophysical Journal* 113(1):157–173.

[37] Johnson CH, Stewart PL, Egli M (2011) The Cyanobacterial Circadian System: From Biophysics to Bioevolution. *Annual Review of Biophysics* 40(1):143–167.

[38] Nishiwaki T, Kondo T (2012) Circadian Autodephosphorylation of Cyanobacterial Clock Protein KaiC Occurs via Formation of ATP as Intermediate. *Journal of Biological Chemistry* 287(22):18030–18035.

[39] Egli M, et al. (2012) Dephosphorylation of the Core Clock Protein KaiC in the Cyanobacterial KaiABC Circadian Oscillator Proceeds via an ATP Synthase Mechanism. *Biochemistry* 51(8):1547–1558.

[40] Paijmans J, Lubensky DK, ten Wolde PR (2017) A thermodynamically consistent model of the post-translational Kai circadian clock. *PLOS Computational Biology* 13(3):e1005415.

[41] Wasserman L (2000) Bayesian Model Selection and Model Averaging. *Journal of Mathematical Psychology* 44(1):92–107.

[42] MacKay DJC, Kay DJCM (2003) *Information Theory, Inference and Learning Algorithms*. (Cambridge University Press).

[43] Flaherty P, et al. (2008) A Dual Receptor Crosstalk Model of G-Protein-Coupled Signal Transduction. *PLOS Computational Biology* 4(9):e1000185.

[44] Klinke DJ (2009) An empirical Bayesian approach for model-based inference of cellular signaling networks. *BMC Bioinformatics* 10(1):371.

[45] Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MP (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface* 6(31):187–202.

[46] Xu TR, et al. (2010) Inferring Signaling Pathway Topologies from Multiple Perturbation Measurements of Specific Biochemical Species. *Science Signaling* 3(113):ra20–ra20.

[47] Schmidl D, Hug S, Li WB, Greiter MB, Theis FJ (2012) Bayesian model selection validates a biokinetic model for zirconium processing in humans. *BMC Systems Biology* 6(1):95.

[48] Eydgahi H, et al. (2013) Properties of cell death models calibrated and compared using Bayesian approaches. *Molecular Systems Biology* 9(1).

[49] Pullen N, Morris RJ (2014) Bayesian Model Comparison and Parameter Inference in Systems Biology Using Nested Sampling. *PLOS ONE* 9(2):e88419.

[50] Mello BA, Pan W, Hazelbauer GL, Tu Y (2018) A dual regulation mechanism of histidine kinase CheA identified by combining network-dynamics modeling and system-level input-output data. *PLOS Computational Biology* 14(7):e1006305.

[51] Higham CF, Husmeier D (2013) A Bayesian approach for parameter estimation in the extended clock gene circuit of Arabidopsis thaliana. *BMC Bioinformatics* 14(10):S3.

[52] Trejo Banos D, Millar AJ, Sanguinetti G (2015) A Bayesian approach for structure learning in oscillating regulatory networks. *Bioinformatics* 31(22):3617–3624.

[53] Martins BMC, Tooke AK, Thomas P, Locke JCW (2018) Cell size control driven by the circadian clock and environment in cyanobacteria. *Proceedings of the National Academy of Sciences* 115(48):E11415–E11424.

[54] Brown KS, Sethna JP (2003) Statistical mechanical approaches to models with many poorly known parameters. *Physical Review E* 68(2):021904.

[55] Gutenkunst RN, et al. (2007) Universally Sloppy Parameter Sensitivities in Systems Biology Models. *PLOS Computational Biology* 3(10):e189.

[56] Ferrell JE, Ha SH (2014) Ultrasensitivity part II: Multisite phosphorylation, stoichiometric inhibitors, and positive feedback. *Trends in Biochemical Sciences* 39(11):556–569.

[57] Johnson CH, Egli M (2014) Metabolic Compensation and Circadian Resilience in Prokaryotic Cyanobacteria. *Annual Review of Biochemistry* 83(1):221–247.

[58] Pattanayek R, et al. (2004) Visualizing a Circadian Clock Protein: Crystal Structure of KaiC and Functional Insights. *Molecular Cell* 15(3):375–388.

[59] Xu Y, et al. (2004) Identification of key phosphorylation sites in the circadian clock protein KaiC by crystallographic and mutagenetic analyses. *Proceedings of the National Academy of Sciences of the United States of America* 101(38):13933–13938.

[60] Kim YI, Dong G, Carruthers CW, Golden SS, LiWang A (2008) The day/night switch in KaiC, a central oscillator component of the circadian clock of cyanobacteria. *Proceedings of the National Academy of Sciences* 105(35):12825–12830.

[61] Pattanayek R, Egli M (2015) Protein–Protein Interactions in the Cyanobacterial Circadian Clock: Structure of KaiA Dimer in Complex with C-Terminal KaiC Peptides at 2.8 Å Resolution. *Biochemistry* 54(30):4575–4578.

[62] Williams SB, Vakonakis I, Golden SS, LiWang AC (2002) Structure and function from the circadian clock protein KaiA of Synechococcus elongatus: A potential clock input mechanism. *Proceedings of the National Academy of Sciences* 99(24):15357–15362.

[63] Hong L, Vani BP, Thiede EH, Rust MJ, Dinner AR (2018) Molecular dynamics simulations of nucleotide release from the circadian clock protein KaiC reveal atomic-resolution functional insights. *Proceedings of the National Academy of Sciences* p. 201812555.

[64] Goodman J, Weare J (2010) Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science* 5(1):65–80.

[65] Foreman-Mackey D, Hogg DW, Lang D, Goodman J (2013) Emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific* 125(925):306.

[66] Tseng R, et al. (2014) Cooperative KaiA–KaiB–KaiC Interactions Affect KaiB/SasA Competition in the Circadian Clock of Cyanobacteria. *Journal of Molecular Biology* 426(2):389–402.

[67] Chang YG, Kuo NW, Tseng R, LiWang A (2011) Flexibility of the C-terminal, or CII, ring of KaiC governs the rhythm of the circadian clock of cyanobacteria. *Proceedings of the National Academy of Sciences* 108(35):14431–14436.

[68] Gomez-Uribe C, Verghese GC, Mirny LA (2007) Operating Regimes of Signaling Cycles: Statics, Dynamics, and Noise Filtering. *PLOS Computational Biology* 3(12):e246.

[69] Ferrell JE, Ha SH (2014) Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity. *Trends in Biochemical Sciences* 39(10):496–503.

[70] Ma L, Ranganathan R (2012) Quantifying the Rhythm of KaiB-C Interaction for In Vitro Cyanobacterial Circadian Clock. *PLOS ONE* 7(8):e42581.

[71] Gunawardena J (2005) Multisite protein phosphorylation makes a good threshold but can be a poor switch. *Proceedings of the National Academy of Sciences* 102(41):14617–14622.

[72] Yunoki Y, et al. (2019) ATP hydrolysis by KaiC promotes its KaiA binding in the cyanobacterial circadian clock system. *Life Science Alliance* 2(3):e201900368.

[73] Buchler NE, Louis M (2008) Molecular Titration and Ultrasensitivity in Regulatory Networks. *Journal of Molecular Biology* 384(5):1106–1119.

[74] Vyshemirsky V, Girolami MA (2008) Bayesian ranking of biochemical system models. *Bioinformatics* 24(6):833–839.

[75] Schwarz G (1978) Estimating the Dimension of a Model. *The Annals of Statistics* 6(2):461–464.

[76] Mutoh R, Nishimura A, Yasui S, Onai K, Ishiura M (2013) The ATP-Mediated Regulation of KaiB-KaiC Interaction in the Cyanobacterial Circadian Clock. *PLOS ONE* 8(11):e80200.

[77] Tseng R, et al. (2017) Structural basis of the day-night transition in a bacterial circadian clock. *Science* 355(6330):1174–1180.

[78] Snijder J, et al. (2017) Structures of the cyanobacterial circadian oscillator frozen in a fully assembled state. *Science* 355(6330):1181–1184.

[79] Mukaiyama A, et al. (2018) Conformational rearrangements of the C1 ring in KaiC measure the timing of assembly with KaiB. *Scientific Reports* 8(1):8803.

[80] Qin X, et al. (2010) Intermolecular associations determine the dynamics of the circadian KaiABC oscillator. *Proceedings of the National Academy of Sciences* 107(33):14805–14810.

[81] Yoda M, Eguchi K, Terada TP, Sasai M (2007) Monomer-Shuffling and Allosteric Transition in KaiC Circadian Oscillation. *PLOS ONE* 2(5):e408.

[82] van Zon JS, Lubensky DK, Altena PRH, ten Wolde PR (2007) An allosteric model of circadian KaiC phosphorylation. *Proceedings of the National Academy of Sciences* 104(18):7420–7425.

[83] Brettschneider C, et al. (2010) A sequestration feedback determines dynamics and temperature entrainment of the KaiABC circadian clock. *Molecular Systems Biology* 6(1):389.

[84] Forger DB, Peskin CS (2003) A detailed predictive model of the mammalian circadian clock. *Proceedings of the National Academy of Sciences* 100(25):14806–14811.

[85] Locke JCW, Millar AJ, Turner MS (2005) Modelling genetic networks with noisy and varied experimental data: The circadian clock in Arabidopsis thaliana. *Journal of Theoretical Biology* 234(3):383–393.

[86] Mirsky HP, Liu AC, Welsh DK, Kay SA, Doyle FJ (2009) A model of the cell-autonomous mammalian circadian clock. *Proceedings of the National Academy of Sciences* 106(27):11107–11112.

[87] Relógio A, et al. (2011) Tuning the Mammalian Circadian Clock: Robust Synergy of Two Loops. *PLOS Computational Biology* 7(12):e1002309.

[88] Kim JK, Forger DB (2012) A mechanism for robust circadian timekeeping via stoichiometric balance. *Molecular Systems Biology* 8(1):630.

[89] Sasai M (2019) Effects of Stochastic Single-Molecule Reactions on Coherent Ensemble Oscillations in the KaiABC Circadian Clock. *The Journal of Physical Chemistry B* 123(3):702–713.

[90] Cao Y, Wang H, Ouyang Q, Tu Y (2015) The free-energy cost of accurate biochemical oscillations. *Nature Physics* 11(9):772–778.

[91] François P, Despierre N, Siggia ED (2012) Adaptive Temperature Compensation in Circadian Oscillations. *PLOS Computational Biology* 8(7):e1002585.

[92] Chew J, Leypunskiy E, Lin J, Murugan A, Rust MJ (2018) High protein copy number is required to suppress stochasticity in the cyanobacterial circadian clock. *Nature Communications* 9(1):3004.

[93] Pittayakanchit W, Lu Z, Chew J, Rust MJ, Murugan A (2018) Biophysical clocks face a trade-off between internal and external noise resistance. *eLife* 7:e37624.

[94] Monti M, Lubensky DK, ten Wolde PR (2018) Robustness of Clocks to Input Noise. *Physical Review Letters* 121(7):078101.

[95] del Junco C, Vaikuntanathan S (2020) Robust oscillations in multi-cyclic Markov state models of biochemical clocks. *The Journal of Chemical Physics* 152(5):055101.

[96] Jolley CC, Ode KL, Ueda HR (2012) A Design Principle for a Posttranslational Biochemical Oscillator. *Cell Reports* 2(4):938–950.

[97] Dovzhenok AA, Baek M, Lim S, Hong CI (2015) Mathematical Modeling and Validation of Glucose Compensation of the Neurospora Circadian Clock. *Biophysical Journal* 108(7):1830–1839.

[98] Pett JP, Korenčič A, Wesener F, Kramer A, Herzel H (2016) Feedback Loops of the Mammalian Circadian Clock Constitute Repressilator. *PLOS Computational Biology* 12(12):e1005266.

[99] O'Neill JS, et al. (2011) Circadian rhythms persist without transcription in a eukaryote. *Nature* 469(7331):554–558.

[100] Edgar R, et al. (2012) Peroxiredoxins are conserved markers of circadian rhythms. *Nature* 485(7399):459–464.

[101] Leloup JC, Gonze D, Goldbeter A (1999) Limit Cycle Models for Circadian Rhythms Based on Transcriptional Regulation in Drosophila and Neurospora. *Journal of Biological Rhythms* 14(6):433–448.

[102] Gonze D, Halloy J, Goldbeter A (2002) Robustness of circadian rhythms with respect to molecular noise. *Proceedings of the National Academy of Sciences* 99(2):673–678.

[103] Leloup JC, Goldbeter A (2003) Toward a detailed computational model for the mammalian circadian clock. *Proceedings of the National Academy of Sciences* 100(12):7051–7056.

[104] Brown SA, Kowalska E, Dallmann R (2012) (Re)inventing the Circadian Feedback Loop. *Developmental Cell* 22(3):477–487.

[105] Gallego M, Virshup DM (2007) Post-translational modifications regulate the ticking of the circadian clock. *Nature Reviews Molecular Cell Biology* 8(2):139–148.

[106] Reischl S, Kramer A (2011) Kinases and phosphatases in the mammalian circadian clock. *FEBS Letters* 585(10):1393–1399.

[107] Zhou M, Kim JK, Eng GWL, Forger DB, Virshup DM (2015) A Period2 Phosphoswitch Regulates and Temperature Compensates Circadian Period. *Molecular Cell* 60(1):77–88.

[108] Fustin JM, et al. (2018) Two CK1δ transcripts regulated by m6A methylation code for two antagonistic kinases in the control of the circadian clock. *Proceedings of the National Academy of Sciences* 115(23):5980–5985.

[109] Reischl S, et al. (2007) β-TrCP1-Mediated Degradation of PERIOD2 Is Essential for Circadian Dynamics. *Journal of Biological Rhythms* 22(5):375–386.

[110] Hayashi F, Iwase R, Uzumaki T, Ishiura M (2006) Hexamerization by the N-terminal domain and intersubunit phosphorylation by the C-terminal domain of cyanobacterial circadian clock protein KaiC. *Biochemical and Biophysical Research Communications* 348(3):864–872.

[111] Transtrum MK, Qiu P (2014) Model Reduction by Manifold Boundaries. *Physical Review Letters* 113(9):098701.

[112] Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58(1):267–288.

[113] Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by Simulated Annealing. *Science* 220(4598):671–680.

[114] Kirkpatrick S (1984) Optimization by simulated annealing: Quantitative studies. *Journal of Statistical Physics* 34(5):975–986.

[115] Beyer HG, Schwefel HP (2002) Evolution strategies – A comprehensive introduction. *Natural Computing* 1(1):3–52.

[116] Powell MJD (1964) An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal* 7(2):155–162.

[117] Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. (Cambridge University Press, Cambridge, UK ; New York), 3 edition edition.

[118] Nelder JA, Mead R (1965) A Simplex Method for Function Minimization. *The Computer Journal* 7(4):308–313.

[119] Langtangen HP, Wang L (2015) Odespy software package.

[120] Eric Jones, Oliphant TE, Pearu Peterson (2001) SciPy: Open Source Scientific Tools for Python.

[121] Lam SK, Pitrou A, Seibert S (2015) Numba: A LLVM-based Python JIT Compiler in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, LLVM '15. (ACM, Austin, Texas), pp. 7:1–7:6.

[122] Dalcín L, Paz R, Storti M (2005) MPI for Python. *Journal of Parallel and Distributed Computing* 65(9):1108–1115.

[123] Dalcín L, Paz R, Storti M, D'Elía J (2008) MPI for Python: Performance improvements and MPI-2 extensions. *Journal of Parallel and Distributed Computing* 68(5):655–662.

[124] Dalcin LD, Paz RR, Kler PA, Cosimo A (2011) Parallel distributed computing using Python. *Advances in Water Resources* 34(9):1124–1139.

[125] Zwanzig RW (1954) High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* 22(8):1420–1426.

[126] Vakonakis I, et al. (2001) Letter to the Editor: Sequence-specific 1H, 13C and 15N resonance assignments of the N-terminal, 135-residue domain of KaiA, a clock protein from Synechococcus elongatus. *Journal of Biomolecular NMR* 21(2):179–180.

[127] 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. *Proceedings of the National Academy of Sciences of the United States of America* 101(30):10925–10930.

[128] Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* 6(3):277–293.

[129] Hayashi F, et al. (2004) Roles of Two ATPase-Motif-containing Domains in Cyanobacterial Circadian Clock Protein KaiC. *Journal of Biological Chemistry* 279(50):52331–52337.

[130] Jeong YM, et al. (2019) Magnesium Regulates the Circadian Oscillator in Cyanobacteria. *Journal of Biological Rhythms* p. 0748730419851655.

[131] Pattanayek R, et al. (2009) Structures of KaiC Circadian Clock Mutant Proteins: A New Phosphorylation Site at T426 and Mechanisms of Kinase, ATPase and Phosphatase. *PLOS ONE* 4(11):e7529.

[132] Kitayama Y, Nishiwaki-Ohkawa T, Sugisawa Y, Kondo T (2013) KaiC intersubunit communication facilitates robustness of circadian rhythms in cyanobacteria. *Nature Communications* 4:2897.

[133] Li C, Chen X, Wang P, Wang W (2009) Circadian KaiC Phosphorylation: A Multi-Layer Network. *PLOS Comput Biol* 5(11):e1000568.

[134] Hayashi F, et al. (2004) Stoichiometric interactions between cyanobacterial clock proteins KaiA and KaiC. *Biochemical and Biophysical Research Communications* 316(1):195–202.

[135] Xu Y, Mori T, Johnson CH (2003) Cyanobacterial circadian clockwork: Roles of KaiA, KaiB and the kaiBC promoter in regulating KaiC. *The EMBO Journal* 22(9):2117–2126.

[136] Murakami R, et al. (2008) ATPase activity and its temperature compensation of the cyanobacterial clock protein KaiC. *Genes to Cells* 13(4):387–395.

[137] Madras N, Sokal AD (1988) The pivot algorithm: A highly efficient Monte Carlo method for the self-avoiding walk. *Journal of Statistical Physics* 50(1):109–186.

[138] Towns J, et al. (2014) XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering* 16(5):62–74.

[139] 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.

[140] Murakami R, et al. (2019) Cooperative Binding of KaiB to the KaiC Hexamer Ensures Accurate Circadian Clock Oscillation in Cyanobacteria. *International Journal of Molecular Sciences* 20(18):4550.

[141] Ouyang Y, Andersson CR, Kondo T, Golden SS, Johnson CH (1998) Resonating circadian clocks enhance fitness in cyanobacteria. *Proceedings of the National Academy of Sciences* 95(15):8660–8664.

[142] Scheer FAJL, Hilton MF, Mantzoros CS, Shea SA (2009) Adverse metabolic and cardiovascular consequences of circadian misalignment. *Proceedings of the National Academy of Sciences* 106(11):4453–4458.

[143] Froy O (2010) Metabolism and Circadian Rhythms—Implications for Obesity. *Endocrine Reviews* 31(1):1–24.

[144] Sigurdardottir LG, et al. (2012) Circadian Disruption, Sleep Loss, and Prostate Cancer Risk: A Systematic Review of Epidemiologic Studies. *Cancer Epidemiology and Prevention Biomarkers* 21(7):1002–1011.

[145] Leonardi GC, et al. (2012) Correlation of the risk of breast cancer and disruption of the circadian rhythm (Review). *Oncology Reports* 28(2):418–428.

[146] Musiek ES, Xiong DD, Holtzman DM (2015) Sleep, circadian rhythms, and the pathogenesis of Alzheimer Disease. *Experimental & Molecular Medicine* 47(3):e148.

[147] Woelfle MA, Ouyang Y, Phanvijhitsiri K, Johnson CH (2004) The Adaptive Value of Circadian Clocks: An Experimental Assessment in Cyanobacteria. *Current Biology* 14(16):1481–1486.

[148] Kim YI, Vinyard DJ, Ananyev GM, Dismukes GC, Golden SS (2012) Oxidized quinones signal onset of darkness directly to the cyanobacterial circadian oscillator. *Proceedings of the National Academy of Sciences* 109(44):17765–17769.

[149] Leipe DD, Aravind L, Grishin NV, Koonin EV (2000) The Bacterial Replicative Helicase DnaB Evolved from a RecA Duplication. *Genome Research* 10(1):5–16.

[150] Dvornyk V, Vinogradova O, Nevo E (2003) Origin and evolution of circadian clock genes in prokaryotes. *Proceedings of the National Academy of Sciences* 100(5):2495–2500.

[151] Glynn SE, Martin A, Nager AR, Baker TA, Sauer RT (2009) Structures of Asymmetric ClpX Hexamers Reveal Nucleotide-Dependent Motions in a AAA+ Protein-Unfolding Machine. *Cell* 139(4):744–756.

[152] Zhao M, et al. (2015) Mechanistic insights into the recycling machine of the SNARE complex. *Nature* 518(7537):61–67.

[153] Enemark EJ, Joshua-Tor L (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* 442(7100):270–275.

[154] Sanders CM, et al. (2007) Papillomavirus E1 helicase assembly maintains an asymmetric state in the absence of DNA and nucleotide cofactors. *Nucleic Acids Research* 35(19):6451–6457.

[155] Sysoeva TA (2017) Assessing heterogeneity in oligomeric AAA+ machines. *Cellular and Molecular Life Sciences* 74(6):1001–1018.

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

[157] Roitberg A, Elber R (1991) Modeling side chains in peptides and proteins: Application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *Journal of Chemical Physics* 95(12):9277.

[158] Simmerling C, Fox T, Kollman PA (1998) Use of Locally Enhanced Sampling in Free Energy Calculations: Testing and Application to the $\alpha \rightarrow \beta$ Anomerization of Glucose. *Journal of the American Chemical Society* 120(23):5771–5782.

[159] Simmerling C, et al. (2000) Combining MONSSTER and LES/PME to Predict Protein Structure from Amino Acid Sequence: Application to the Small Protein CMTI-1. *Journal of the American Chemical Society* 122(35):8392–8402.

[160] Isralewitz B, Gao M, Schulten K (2001) Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology* 11(2):224–230.

[161] Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G (2006) String method in collective variables: Minimum free energy paths and isocommittor surfaces. *Journal of Chemical Physics* 125(2):024106.

[162] Ovchinnikov V, Karplus M, Vanden-Eijnden E (2011) Free energy of conformational transition paths in biomolecules: The string method and its application to myosin VI. *Journal of Chemical Physics* 134(8):085103.

[163] Torrie GM, Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* 23(2):187–199.

[164] Pattanayek R, et al. (2006) Analysis of KaiA–KaiC protein interactions in the cyano-bacterial circadian clock using hybrid structural methods. *The EMBO Journal* 25(9):2017–2028.

[165] Murayama Y, et al. (2011) Tracking and visualizing the circadian ticking of the cyanobacterial clock protein KaiC in solution. *The EMBO Journal* 30(1):68–78.

[166] VanLoock MS, et al. (2003) ATP-Mediated Conformational Changes in the RecA Filament. *Structure* 11(2):187–196.

[167] Xing X, Bell CE (2004) Crystal Structures of Escherichia coli RecA in Complex with MgADP and MnAMP-PNP. *Biochemistry* 43(51):16142–16152.

[168] Chen Z, Yang H, Pavletich NP (2008) Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. *Nature* 453(7194):489–494.

[169] Stock D, Leslie AGW, Walker JE (1999) Molecular Architecture of the Rotary Motor in ATP Synthase. *Science* 286(5445):1700–1705.

[170] Dautant A, Velours J, Giraud MF (2010) Crystal Structure of the Mg·ADP-inhibited State of the Yeast F1c10-ATP Synthase. *Journal of Biological Chemistry* 285(38):29502–29510.

[171] Louet M, Perahia D, Martinez J, Floquet N (2011) A Concerted Mechanism for Opening the GDP Binding Pocket and Release of the Nucleotide in Hetero-Trimeric G-Proteins. *Journal of Molecular Biology* 411(1):298–312.

[172] Dror RO, et al. (2015) Structural basis for nucleotide exchange in heterotrimeric G proteins. *Science* 348(6241):1361–1365.

[173] Jones PM, George AM (2009) Opening of the ADP-bound active site in the ABC transporter ATPase dimer: Evidence for a constant contact, alternating sites model for the catalytic cycle. *Proteins: Structure, Function, and Bioinformatics* 75(2):387–396.

[174] Pattanayek R, Xu Y, Lamichhane A, Johnson CH, Egli M (2014) An arginine tetrad as mediator of input-dependent and input-independent ATPases in the clock protein KaiC. *Acta Crystallographica Section D* 70(5):1375–1390.

[175] Bason JV, Montgomery MG, Leslie AGW, Walker JE (2015) How release of phosphate from mammalian F1-ATPase generates a rotary substep. *Proceedings of the National Academy of Sciences* 112(19):6009–6014.

[176] Wang J, Deng Y, Roux B (2006) Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials. *Biophysical Journal* 91(8):2798–2814.

[177] Wiegard A, et al. (2013) Biochemical analysis of three putative KaiC clock proteins from Synechocystis sp. PCC 6803 suggests their functional divergence. *Microbiology* 159(5):948–958.

[178] Vakonakis I, et al. (2004) NMR structure of the KaiC-interacting C-terminal domain of KaiA, a circadian clock protein: Implications for KaiA–KaiC interaction. *Proceedings of the National Academy of Sciences of the United States of America* 101(6):1479–1484.

[179] Hertig S, Latorraca NR, Dror RO (2016) Revealing Atomic-Level Mechanisms of Protein Allostery with Molecular Dynamics Simulations. *PLOS Computational Biology* 12(6):e1004746.

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

[181] Kiyohara YB, Katayama M, Kondo T (2005) A Novel Mutation in kaiC Affects Resetting of the Cyanobacterial Circadian Clock. *Journal of Bacteriology* 187(8):2559–2564.

[182] Holtzendorff J, et al. (2008) Genome Streamlining Results in Loss of Robustness of the Circadian Clock in the Marine Cyanobacterium Prochlorococcus marinus PCC 9511. *Journal of Biological Rhythms* 23(3):187–199.

[183] Axmann IM, et al. (2009) Biochemical Evidence for a Timing Mechanism in Prochlorococcus. *Journal of Bacteriology* 191(17):5342–5347.

[184] Hwang W, Lang MJ (2013) Nucleotide-Dependent Control of Internal Strains in Ring-Shaped AAA+ Motors. *Cellular and Molecular Bioengineering* 6(1):65–73.

[185] Ma W, Schulten K (2015) Mechanism of Substrate Translocation by a Ring-Shaped ATPase Motor at Millisecond Resolution. *Journal of the American Chemical Society* 137(8):3031–3040.

[186] Itsathitphaisarn O, Wing RA, Eliason WK, Wang J, Steitz TA (2012) The Hexameric Helicase DnaB Adopts a Nonplanar Conformation during Translocation. *Cell* 151(2):267–277.

[187] Cuesta I, et al. (2010) Conformational Rearrangements of SV40 Large T Antigen during Early Replication Events. *Journal of Molecular Biology* 397(5):1276–1286.

[188] Lo YH, et al. (2009) The crystal structure of a replicative hexameric helicase DnaC and its complex with single-stranded DNA. *Nucleic Acids Research* 37(3):804–814.

[189] Arnold FH, Haymore BL (1991) Engineered metal-binding proteins: Purification to protein folding. *Science* 252(5014):1796–1797.

[190] Martin A, Baker TA, Sauer RT (2005) Rebuilt AAA+ motors reveal operating principles for ATP-fuelled machines. *Nature* 437(7062):1115–1120.

[191] Jo S, Lim JB, Klauda JB, Im W (2009) CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophysical Journal* 97(1):50–58.

[192] Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal Of Computational Chemistry* 29(11):1859–1865.

[193] Lee J, et al. (2016) CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation* 12(1):405–413.

[194] Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 14(1):33–38.

[195] Abraham MJ, et al. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2:19–25.

[196] MacKerell AD, et al. (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B* 102(18):3586–3616.

[197] MacKerell AD, Feig M, Brooks CL (2004) Improved Treatment of the Protein Backbone in Empirical Force Fields. *Journal of the American Chemical Society* 126(3):698–699.

[198] Best RB, et al. (2012) Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $X1$ and $X2$ Dihedral Angles. *Journal of Chemical Theory and Computation* 8(9):3257–3273.

[199] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 79(2):926–935.

[200] Beglov D, Roux B (1994) Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *The Journal of Chemical Physics* 100(12):9050–9063.

[201] Pavelites JJ, Gao J, Bash PA, Mackerell AD (1997) A molecular mechanics force field for NAD+ NADH, and the pyrophosphate groups of nucleotides. *Journal of Computational Chemistry* 18(2):221–239.

[202] Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* 98(12):10089–10092.

[203] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* 18(12):1463–1472.

[204] Kalé L, et al. (1999) NAMD2: Greater Scalability for Parallel Molecular Dynamics. *Journal of Computational Physics* 151(1):283–312.

[205] Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G (2014) PLUMED 2: New feathers for an old bird. *Computer Physics Communications* 185(2):604–613.

[206] Thiede EH, Van Koten B, Weare J, Dinner AR (2016) Eigenvector method for umbrella sampling enables error analysis. *The Journal of Chemical Physics* 145(8):084115.

[207] Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1995) Multidimensional free-energy calculations using the weighted histogram analysis method. *Journal of Computational Chemistry* 16(11):1339–1350.

[208] Grossfield A (2013) WHAM: The weighted histogram analysis method.

[209] Gumbart JC, Roux B, Chipot C (2013) Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *Journal of Chemical Theory and Computation* 9(1):794–802.

[210] Mark AE (2002) Free Energy Perturbation Calculations in *Encyclopedia of Computational Chemistry*. (John Wiley & Sons, Ltd).

[211] Bakan A, Meireles LM, Bahar I (2011) ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* 27(11):1575–1577.

[212] Roberts E, Eargle J, Wright D, Luthey-Schulten Z (2006) MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* 7:382.

[213] Russell RB, Barton GJ (1992) Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins: Structure, Function, and Bioinformatics* 14(2):309–323.

[214] Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* 7:539–539.

[215] Li W, et al. (2015) The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic acids research* 43(W1):W580–4.

[216] Marchi M, Ballone P (1999) Adiabatic bias molecular dynamics: A method to navigate the conformational space of complex molecular systems. *The Journal of Chemical Physics* 110(8):3697–3702.

[217] Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of Chemical Theory and Computation* 7(2):525–537.

[218] Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH (2011) Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *Journal of Chemical Theory and Computation* 7(7):2284–2295.

[219] Kubo R, Toda M, Hashitsume N (2003) *Statistical Physics II: Nonequilibrium Statistical Mechanics*. (Springer, New York), 2nd edition edition.

[220] Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* 101(5):4177–4189.

[221] Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics* 103(11):4613–4621.

[222] Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *Journal of Computational Physics* 23(3):327–341.

[223] Kirkwood JG (1935) Statistical Mechanics of Fluid Mixtures. *The Journal of Chemical Physics* 3(5):300–313.

[224] Liu P, Dehez F, Cai W, Chipot C (2012) A Toolkit for the Analysis of Free-Energy Perturbation Calculations. *Journal of Chemical Theory and Computation* 8(8):2606–2616.

[225] Pohorille A, Jarzynski C, Chipot C (2010) Good Practices in Free-Energy Calculations. *The Journal of Physical Chemistry B* 114(32):10235–10253.

[226] Rocklin GJ, Mobley DL, Dill KA, Hünenberger PH (2013) Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects. *The Journal of Chemical Physics* 139(18):184103.

[227] Pedregosa F, et al. (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

[228] Metzen JH (2018) Kernel_regression: Implementation of Nadaraya-Watson kernel regression with automatic bandwidth selection compatible with sklearn.

[229] Murzin AG (2008) Metamorphic Proteins. *Science* 320(5884):1725–1726.

[230] Lella M, Mahalakshmi R (2017) Metamorphic Proteins: Emergence of Dual Protein Folds from One Primary Sequence. *Biochemistry* 56(24):2971–2984.

[231] Porter LL, Looger LL (2018) Extant fold-switching proteins are widespread. *Proceedings of the National Academy of Sciences* 115(23):5968–5973.

[232] Leaver-Fay A, et al. (2011) Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules in *Methods in Enzymology*, Computer Methods, Part C, eds. Johnson ML, Brand L. (Academic Press) Vol. 487, pp. 545–574.

[233] 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. *Journal of Molecular Biology* 425(18):3311–3324.

[234] Huang J, et al. (2017) CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods* 14(1):71.

[235] Martyna GJ, Klein ML, Tuckerman M (1992) Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics* 97(4):2635–2643.

[236] Van Koten B, Luskin M (2019) Stability and Convergence of the String Method for Computing Minimum Energy Paths. *Multiscale Modeling & Simulation* 17(2):873–898.

[237] Zimmerman MI, Bowman GR (2015) FAST Conformational Searches by Balancing Exploration/Exploitation Trade-Offs. *Journal of Chemical Theory and Computation* 11(12):5747–5757.

[238] Zimmerman MI, et al. (2017) Prediction of New Stabilizing Mutations Based on Mechanistic Insights from Markov State Models. *ACS Central Science* 3(12):1311–1321.

[239] Prinz JH, et al. (2011) Markov models of molecular kinetics: Generation and validation. *The Journal of Chemical Physics* 134(17):174105.

[240] Scherer MK, et al. (2015) PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation* 11(11):5525–5542.

[241] Thiede EH, Giannakis D, Dinner AR, Weare J (2019) Galerkin approximation of dynamical quantities using trajectory data. *The Journal of Chemical Physics* 150(24):244111.

[242] Ferrell JE (2016) Perfect and Near-Perfect Adaptation in Cell Signaling. *Cell Systems* 2(2):62–67.