Molecular Dynamics Simulation of the KaiC Protein

A Thesis submitted in partial satisfaction of the requirements
for the degree of Master of Science

in

Quantitative and System Biology

by

Theo Crouch

Committee in charge:

Professor Andy LiWang, Chair
Professor Michael Colvin
Acting Associate Professor Eva De Alba
Associate Professor Jing Xu

2019
The Thesis of Theo Crouch, II is approved, and it is acceptable in quality and form publication on microfilm and electronically:

__________________________________________________________________________

__________________________________________________________________________

__________________________________________________________________________

__________________________________________________________________________

Chair

University of California, Merced

2019
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>BACKGROUND</td>
<td>2</td>
</tr>
<tr>
<td>Previous molecular dynamics simulations of the KaiC clock system</td>
<td>5</td>
</tr>
<tr>
<td>SPECIFIC AIM</td>
<td>5</td>
</tr>
<tr>
<td>Hypothesis</td>
<td>5</td>
</tr>
<tr>
<td>AIM</td>
<td>7</td>
</tr>
<tr>
<td>Significance</td>
<td>7</td>
</tr>
<tr>
<td>RESEARCH STRATEGY</td>
<td>7</td>
</tr>
<tr>
<td>Methods</td>
<td>7</td>
</tr>
<tr>
<td>Atomistic simulation</td>
<td>7</td>
</tr>
<tr>
<td>Specific Aim</td>
<td>9</td>
</tr>
<tr>
<td>ATOMISTIC MD RESULTS</td>
<td>10</td>
</tr>
<tr>
<td>Structure Stability</td>
<td>10</td>
</tr>
<tr>
<td>ATP Accessibility</td>
<td>11</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>18</td>
</tr>
<tr>
<td>CONCLUSION</td>
<td>19</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>KaiC</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Circadian Clock Protein</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>Initial and preliminary structural analyses</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>Additional structural analyses</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>Hydrogen Bonds (Hbond)</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>Hydrogen Bonds</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>Hydrogen Bonds</td>
<td>13</td>
</tr>
<tr>
<td>8</td>
<td>Solvent Accessible Surface Area (SASA)</td>
<td>14</td>
</tr>
<tr>
<td>9</td>
<td>Root Means Square Fluctuation (RMSF)</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>RMSF</td>
<td>15</td>
</tr>
<tr>
<td>11</td>
<td>Minimum Distances (Mindist)</td>
<td>16</td>
</tr>
<tr>
<td>12</td>
<td>Mindist</td>
<td>17</td>
</tr>
<tr>
<td>13</td>
<td>Radial Distribution Function (RDF)</td>
<td>17</td>
</tr>
<tr>
<td>14</td>
<td>RDF</td>
<td>18</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.</td>
<td>Completed MD Simulations</td>
<td>9</td>
</tr>
<tr>
<td>Table 2.</td>
<td>Core command line operations us</td>
<td>10</td>
</tr>
</tbody>
</table>
INTRODUCTION

The circadian clock protein, KaiC, is a well-characterized system that is an integral part of the clock protein complex, KaiABC, of cyanobacteria (1, 2). This is a well-defined structure that has been reported to have functional interactions amongst its constituent parts (3, 4). The KaiC hexamer forms a pair of homologous domains that look like stacked rings, one on top of the other (5). The two domains, made up six segments from six identical protein chains, are called the C1 and C2 ring (CI and CII hereafter, respectively) (5, 6). This paired stack is thought to slightly separate and then reconnect cyclically over the day and night 24-hour period within the KaiABC clock protein complex (7, 8). The ring structures contain binding sites for adenosine triphosphates (ATPs) that appear to be locked in place at the interfaces between where each of the six monomeric structures meets (12 ATPs, 6 per ring) (See Figure 1) (5). The interfaces within this protein complex are also areas of importance because these are the sites of KaiC’s autophosphorylation during the progression of the clock cycle (9, 8). The clock protein’s 24-hour period is choreographically driven by the binding of KaiA to the CII ring resulting in the phosphorylation of residues in the CII ring followed by further phosphorylation of adjacent residues causing the unbinding of KaiA from the CII ring; and the binding of KaiB to the CI ring and KaiA binding to KaiB, after it unbinds from the CII ring (8).

Figure 1. KaiC. The stacked ring conformation is a hexameric structure of identical monomers. The ATPs are at the interfaces of the monomers. Image created with VMD 1.9.3. Visual Molecular Dynamics.

KaiA binding to the dangling C-terminus A-Loop is a critical state that seems to govern the rigidity of the CII-CI stacked configuration (10, 8). Although the specific steps in the KaiABC clock have been mapped, the structural basis of the changes causing, and caused by, the phosphorylation and protein binding are not fully understood. The aim of this project was to use molecular dynamics (MD)
simulations to fill in some structural details of the key states of KaiC during the clock cycle.

BACKGROUND

The day and night, 24-hour period is not simply a phenomenon attributed to the movement of the celestial bodies, the rotation of the earth around its axis, it is a critical rhythm that all living organisms use to adjust to daily changes in the environment (11, 12, 13). Unicellular and higher organisms, including us humans, have a biochemical affinity to the day-night 24-hour (circadian) rhythm (14, 15). Our study of the human biochemical circadian rhythms has produced general criteria for what constitutes such a clock, as well as experimental data giving us the basis of the biochemical interactions causing the rhythms. For example, Burgoyne (14) reports on the very specific criteria for a bona fide circadian rhythm, and Konda et al. (15) re-states these as follows:

1. Persistence in constant conditions,
2. Phase resetting by light/dark signals,
3. Temperature compensation of the period.

After these criteria for circadian rhythms were determined, the basis of this biochemical mechanism did not become experimentally obtainable for over a decade. Luciferase reporter experimentation finally produced bioluminescence rhythm expression that satisfied these criteria for circadian rhythms (15). Surprisingly, this rhythm was suspected and confirmed in prokaryotes, namely cyanobacteria (15, 16). This discovery opened the door to more chemically detailed studies because prokaryotic cells are smaller and experimentally more accessible to discovery and analysis than eukaryotic cells (15).

Cyanobacteria specifically, *Synechococcus* sp. PCC7942, was shown to have a circadian clock rhythm by an experiment using a luciferase reporter gene coupled with continuous and automated monitoring of the bioluminescence (15). In this strain, the bacterial luciferase genes with its bioluminescence capability of the clock-controlled expression of the *psb*AI gene made it possible to track the luminescence of the gene being reported, *psb*AI (15). Moreover, the endogenous nature of this gene and the neutral site at which it was inserted into the bacterial DNA made cyanobacteria the favorable cells to demonstrate by monitoring the clock cycle (15).

Dogma, not experimentation, led to the original skepticism that cyanobacteria, a prokaryote, could have circadian clocks because only eukaryotes were thought to have a circadian rhythm (15). The fact that no prokaryotes were previously discovered with a circadian clock and the complexity of the protein circuit necessary for a circadian clock made this discovery in cyanobacteria a surprise, but once discovered, it made possible the efficient and more rapid approach to discovery in circadian clock research and experimentation (15).

Furthermore, genetic experimentation was used to discover and elucidate the structural basis of the circadian clock system of cyanobacteria (17). Specifically, mutagenesis experimentation was used to confirm the relevant loci in the DNA of cyanobacteria (15, 17). Strains identified by mutagenesis were found to have varying cyclic rhythms further confirming the loci and genes of interest
identifying the proteins responsible for the circadian rhythm in cyanobacteria Synechococcus sp. Strain PCC 7942 (18, 15). Other methods of experimentation, synchronous growing, and diazotrophic synchronous growing, were also used in confirming the cyclic activity of the marine bacteria, cyanobacteria 19, 20, 21, 22). Using the effects of nitrogen-fixing, Mitsui was able to confirm the cycle of various biological functions, for example, cell growth and the temporal separation of oxygen and nitrogen activity in the marine bacteria (20, 21, 16). Further solidifying the role of cyanobacteria as the preeminent model clock system for experimental discovery and advancing the art of the science.

Luciferase reporter experimentation coupled with mutagenesis was used for conclusive confirmation and consequently the naming of the proteins, KaiA, KaiB, and KaiC (28). It was determined experimentally that all three discovered proteins - KaiA, KaiB, and KaiC - were required to bring about the circadian clock cycle in the Synechococcus sp. (22). Moreover, it was determined that spatial segregation by growth experiments was not enough to explain the basis of circadian rhythms. The conclusion that all three proteins, KaiA, KaiB, and KaiC, are necessary and sufficient to elicit cyclic oscillation remains to be true as evidenced by current reports on the cyanobacterial circadian clock (23). Additionally, finding that the different mutants of the KaiABC circadian protein cluster exhibit different rhythms also continues to be true (30, 25, 26) over a decade later. However, with our renewed understanding of the autophosphorylation intrinsic to the KaiABC cluster, very specific mutagenesis to binding sites gives us a better understanding of KaiABC interactions (24, 25). In addition to its role in the circadian rhythm, there has also been considerable research on the cyanobacteria to learn more about the mechanical details of the photosynthesis and nitrogenase activities in the protein cluster (12).

Other discoveries solidified cyanobacteria as the model system of choice for clock systems. Cyanobacteria clock proteins were shown to be resistant to change in oscillation due to intracellular activity (noise), such as cell divisions and cellular proximity (crowding by the other cells). The “noise resistance” property is a discovery that solidifies the identification of this system as a “true” circadian clock. Experimentally, it was demonstrated that clock phase shifts can occur among colonies of cyanobacteria, but the periods were essentially the same which were clearly mapped occurrences within different populations of cyanobacteria oscillators (27).

Examining the phasing phenomena further, more observations were made to determine if intercellular coupling was an influencer of the circadian rhythm, in other words, does population coupling change the clock cycle. Intercellular coupling was experimentally demonstrated to have no permanent change to the oscillators in cyanobacteria (28). Each population was observed to resist intercellular influences to alter the circadian rhythm. Amdaoud et al. reported that strong coupling indeed can force a distribution of oscillator phases between the cyanobacterial oscillator populations, however, the coupling is known experimentally to usually be very weak. The authors report that the cyanobacterial
clock proteins act as “dynamic noise-resistant networks.” Further demonstrating the robustness of the cyanobacteria circadian clock.

Forming a basis for the structural molecular changes during the cycle requires experimentation on the structural interactions of the three proteins. This work was initially focused on the KaiC and KaiB interaction, since KaiB was thought to bind to KaiC, forming the KaiBC complex. Subsequent studies have considered all physical interactions among the KaiA, KaiB, and KaiC proteins, the KaiABC complex, (29). In this experimentation, KaiA appeared to be a positive regulator of KaiBC expression (29). Additionally, over-expressed KaiC completely suppresses KaiBC formation rhythmicity (30; 29). These experiments made direct interactions amongst the KaiABC proteins, but it was not yet clear experimentally which of the CII and CI rings of KaiC interacted with the other proteins (29).

These reported interactions informed us that there is KaiB and KaiC association, but these original KaiABC interaction data was misleading. Kitayama et al., experimentally following this same focus on the interactions of KaiA and KaiB in relation to the KaiBC complex (31). KaiB was reported to attenuate KaiA and this led to the suggestion that KaiB activates dephosphorylation of KaiC (31, 32). Nishiwaka et al. also reported data suggesting the KaiA-KaiB association cycled with Serine-431 phosphorylation (31). In studies mapping the KaiC phosphorylation and dephosphorylation cycle using immunoprecipitation experimentation followed by SDS-Page (sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis) (32), it was erroneously concluded that KaiB negating KaiA proceeds by KaiC recruiting KaiB to bind KaiBC to KaiA stopping phosphorylation.

These reports progressed to what was thought to be the binding site of KaiB to KaiC. Pattanayak et al. reported that KaiB binds to the CII ring, which was subsequently found to by the wrong binding site of KaiB to KaiC (33). The evidence of their findings for the KaiB binding site is the EM (electron microscopy) analysis results obtained using the gold labeled C-terminal KaiC showing KaiB lodged closely to KaiA on the CII ring of KaiC, and the SAXS (small angle X-ray scattering) analysis results revealing a protrusion of KaiB binding to the CII ring C-terminal peptides (33).

It was determined that KaiC facilitates (or enhances) KaiA-KaiB interaction. More specifically, with the CI ring being sufficient to enhance KaiA-KaiB interaction as reported by Iwasaki et al., this finding gave direction to how each domain of KaiC, CII and CI, differ functionally (29).

Although it is now understood that KaiC is in specific phosphorylated states before interactions take place, KaiA and KaiB interactions continue to be reported to interact with KaiC (34, 35, 36, 37, 8, 23). KaiA-KaiB interactions with KaiC, namely the CI domain, was reported to induce dephosphorylation of KaiC; moreover, KaiC enhanced KaiA-KaiB interaction that is otherwise weak; and KaiA does not interact alone with the CI domain (37). Additionally, it was reported that KaiAB disassociated with KaiC, once the dephosphorylation of the CII domain was complete (23).
The aim of our work is to help elucidate the basis of the form versus function paradigm where circadian clocks are concerned. To this end we are performing molecular dynamics simulations of different mutants of the KaiC hexamer to give insight into the structural conformational basis of the autonomous cycle of the KaiABC complex of the cyanobacterial circadian clock. For example, describing the effects of protein mutations and truncation on its structure and flexibility, especially in the ATP binding sites, is expected to deliver new information to help explain previous experimental findings and possibly to suggest new experimental studies of this complex biochemical machine.

**Previous molecular dynamics simulations of the KaiC clock system:**

Molecular dynamic simulations have previously been successfully performed and results reported on the KaiABC clock system. Villarreal et al. also performed MD simulations related to the KaiB binding to the KaiC CII ring, which was based on the earlier KaiABC model that involved direct interactions of these proteins which is different that the working model we are developing and studying. In Villarreal’s study molecular dynamics is used to fit KaiB in 16 distinct starting orientations to KaiC. From their results they proposed that KaiB functionally “push” KaiC into the dephosphorylation state because KaiB structurally binds and blocks the ATP cleft on the CII ring (38).

Dong et al. performed MD simulations on the interaction of KaiA-KaiC, with a specific focus on KaiA’s conformational change before and after binding, not on KaiC. The relatively short (10 ns) simulations were completed on KaiA predicted binding sites binding to the C-terminus tail (alone) of KaiC to determine what structural effects occur when KaiA binds, but the simulations do not include anything for KaiC (39). There were some conformational changes in KaiA observed during the simulation as a result of the binding (39).

Another group, Hong et al., used non-equilibrium MD simulations to develop insights on ADP release and exchange with the ATP. Simulating the release of nucleotides from the KaiC structure at a single location of interfacing monomers in the CII ring hexameric domain resulted in their proposed idea that the release occurs because of changes in local and internal atomic conformations (40).

The goal of our study is very different than these earlier studies that focused on the details of the inter-protein contacts or the dynamics of ADP release. Our goal is to characterize the structure and dynamics of the 5 distinct forms of KaiC used as in vitro models of the steps in the clock cycle: the wild-type KaiC hexamer, hexameric KaiC with the 10 C-terminal amino acids truncated, as well as the T432E, S431E/T432E, and S432E mutants with and without the C-terminal truncations (8).

**SPECIFIC AIM**

**Hypothesis:** Our hypothesis is that the effects of KaiA binding and subsequent amino acid phosphorylation are to induce changes in the structure and flexibility at distant points in KaiC communicated by through-protein interactions and that the changes within the CII ring occur rapidly enough to by observed in half-microsecond scale molecular dynamics simulations.
Circadian clocks play important roles in health and environmental microbiology so there is value in elucidating their function at the molecular level. However, fundamental limitations in the experimental approaches now available limit the amount of chemical detail that can be measured in a complex protein machine like the KaiABC clock. The chemical details are essential to having a full understanding of circadian clocks. Our strategy is to address the problem using all atom molecular dynamics simulations at the half-microsecond time scale that we believe will be long enough to distinguish structural and dynamical differences between the key KaiC states during the circadian cycle.

KaiC is a part of a widely studied circadian (day, 24 hours) clock system that operates through a coordinated sequence of ATP binding, residue phosphorylation, and accessory protein binding. Experimental studies have mapped out a hypothesized set of steps in the KaiC mechanism, but the basis of structural details is not accessible (Fig. 2).

The goal of this research is to use atomistic molecular dynamics simulations to fill in these structural details. This will allow us to quantify some of the proposed mechanisms within the protein complex. The results can be used to propose future experiments.

The circadian clock protein of cyanobacteria has been reported to be the most ideal circadian clock system to study. Consisting of only three proteins, KaiC, KaiA and KaiB, this circadian clock will oscillate in vitro as the three Kai proteins interact in a large heteromultimeric complex. KaiC interaction with KaiA followed by KaiC-KaiB interaction has been reported to have a close relationship with the autophosphorylation of Thr 432 then Ser 431 followed by autodephosphorylation of Thr 432 then Ser 431. The binding of KaiA to the alternatively dangling A-loop of the KaiC CII ring is speculated to cause a structural change that allows the sequential interaction of adenosine triphosphates (ATPs) with the CII ring (Fig. 2). KaiB, binding to the KaiC CI ring, sequesters KaiA which
subsequently allows for deautophosphorylation. The dephosphorylation of the CII ring is proposed to cause the CII ring to loosen and CII and CI ring to unstack followed by the release of KaiB and KaiA. Lastly, the Kai proteins return to the beginning of the circadian, 24-hour, cycle. This project has involve the simulations of the eight distinct KaiC states shown in Figure 2.

**AIM:** Using all-atom molecular dynamics simulation, determine the structural effects of truncation and Glutamate mutations in KaiC to validate that protein modifications mimicking KaiA binding and amino acid phosphorylation cause changes in the local flexibility and structure of the CII ring and near ATP binding on time scales more rapid than 1 microsecond.

Experimentally, studies of the KaiC complex have shown that residue phosphorylation at Serine (Ser 431) and Threonine (Thr 432) sites can be functionally mimicked by glutamate (E) substitution and that KaiA binding to the KaiC CII ring can be mimicked by truncating the first 10 residues from the C-terminal end of KaiC. We will use half-microsecond (µs) scale all-atom molecular dynamics simulations to determine how these changes affect the local structure and fluctuations of KaiC. An expected outcome of this aim is that we will be able to quantify conformational changes that occur on this time scale as well as changes in the local flexibility near the CII ring ATP binding sites.

The KaiC simulations are moderately large, over 270,000 atoms in the KaiC protein complex, including the surrounding water and counter-ions. Our locally available computers, limit the length of our simulations to less than a microsecond (41). To achieve much longer time scales would require the use of more approximate simulation methods. Such “coarse-grained simulations” allow much longer simulation time scales (up to 1 millisecond) but at the expense of reduced fidelity. The goal of this study was to use the most accurate all-atom classical simulations available, at the expense of a more limited simulation time.

**Significance:**

The proposed research demonstrates how available computational tools are revolutionizing our understanding of biochemistry. The project takes necessary steps to provide functional experimentation to available MD modeling and simulation techniques. There is potential for discovering the structural basis of the mechanism of the circadian clock of cyanobacteria.

**RESEARCH STRATEGY**

**Methods**

**Atomistic simulation:**

Simulation-enabled biology and biochemistry provide meaningful insights into biomolecular systems (42). In this project, we will be using all-atom molecular dynamics that have been well established for studying protein structural properties, including protein folding (43). We are simulating structural flexibility of already folded proteins after introducing mutations. Because the full chemical details of biomolecular systems are not accessible by experimental means (42), molecular simulations have become a useful method for accessing details that are unclear.

Molecular dynamics (MD) calculates the potential energy of the physical system and uses the gradient of the energy (i.e. the interatomic forces) to
dynamically evolve the system in time (43). The set of equations and parameters, the interactions, used to calculate the forces amongst our atoms is collectively called a force field, which includes terms for all standard non-bond interatomic forces, van der waals interactions, electrostatic interactions, as well as parameters for bonded interactions including the magnitude and stiffness of bond lengths, angles, and torsional angles (44). Note that most force fields also make the assumption that all intra- and intermolecular interactions can by broken down into 2-body functions which allows practical simulations of large chemical systems. Such force fields implemented in classical MD simulations have been well documented to yield good results for a number of molecular properties. Since we are not simulating full protein folding, which is not yet practical for the large protein or protein complexes we considered here, the starting point for our simulations is limited to the experimental data and the crystal protein structure obtained from the Protein Data Bank (8, 45).

A limitation of classical MD is that it does not include any form of bond breaking of formation, including proton exchanges. Nevertheless, for many questions in biochemistry and biophysics, such simulations have been well established to yield accurate and useful results (42, 44).

Since we are limited to simulating systems that are vastly smaller than the in vitro or in vivo environments they exist in, we need to include boundary effects in our simulations. This challenge is met in our work by utilizing a periodic container that repeats “virtual” copies of itself equally and infinitely in all directions. Surface effects are eliminated.

In general, the goal of so-called in silico research is not to discover fundamental new biochemical paths of protein functions (although this may be possible in the future), but instead to provide a detailed chemical picture of the clock mechanism inferred from previous experimental data. Therefore, the goal of our simulations was to either corroborate or refute some of the experimental details of the proposed mechanism in the circadian clock system. A disagreement in our simulation results tells us to try alternative simulation techniques. For example, if we cannot find there are differences in how the two rings stack, work with experimental collaborators that can disambiguate which of the results are true. With NMR (Nuclear Magnetic Resonance), distances can be calculated and, likewise, with MD (or a different MD approach), we can analyze the simulations to find the explicit amino acid distances that we may predict will not change.

Finally, we should note that there are a number of possible limitations, even using state of the art MD simulations for such a large and complex protein machine. The most significant drawback is the limited half-microsecond time scale, which means that our simulations will simply not “see” biochemical processes happening at longer time scales. Next, there is a risk that the forcefield describing the interatomic forces may simply not be accurate enough to predict the changes in structure and dynamics between the different KaiC states, although a long history of MD research suggests that we will get useful results. Dramatically increasing the time scales of the simulations to reveal longer time structural changes would require using lower-fidelity models.
**Specific Aim:** Using all-atom simulation, determine the structural effects of truncation and glutamate mutations in KaiC.

We have performed atomistic simulations on different states of the KaiC protein. Utilizing a standard toolset for MD – Linux OS, operating system; GROMACS 2016, GROningen Machine for Chemical Simulations; VMD 1.9.3, Visual Molecular Dynamics); R, statistical software – we performed 500 nanosecond (ns) (i.e. ½ microsecond) simulations on the KaiC hexamer, Protein Data Bank PDB Id 1TF7 and force field AMBER99 SB-ILDN. In total, we performed 8 MD simulations, one on the wild-type full KaiC protein structure and seven separate simulations of KaiC mutants – phosphomimetics S431E, T432E, S431E+T432E, and a 10 C-terminal AA truncated structure of the wild-type and the phosphomimetics models (E, glutamic acid; S, serine; T, threonine; See Figure 2). Notably, the full KaiC model represents the pre-KaiA state with bonded ATPs. The Threonine mutated to Glutamate (T432E) represents a post-KaiA state of the KaiC protein complex. Serine and Threonine mutated to glutamate (S431E+T432E) represent distinctly a pre-KaiB condition. The truncated mutant simulates KaiA grabbing the A-loop of the CI Ring of KaiC. Lastly, Serine only mutated to glutamate (S431E) represents the post-KaiB condition.

<table>
<thead>
<tr>
<th>Simulations</th>
<th>KaiC Models</th>
<th>Mimicked Phosphorylation States</th>
<th>Ligand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>Wildtype</td>
<td>TE (pT432)</td>
<td>SETE (pS431 pT432)</td>
</tr>
<tr>
<td>Trunc</td>
<td>Wildtype</td>
<td>TE (pT432)</td>
<td>SETE (pS431 pT432)</td>
</tr>
</tbody>
</table>

**Table 1.** Completed MD simulations (E, glutamic acid; S, serine; T, threonine; p, phosphorylated). MD environmental settings: Force field, AMBER99 SB-ILDN; Water model, TIP3P; PDB ID 1TF7; 277,127 atoms.

The data is accumulated during the period of the simulation in the form of trajectories which run for several weeks on a dedicated server with GPU acceleration. With this required run-time for a system as large as ours, periodic checks over the accumulating data were performed to safeguard against numerical errors that can occur in such simulations. The MD simulation produce a record of the exact position of every atom at very high time resolution, in this case every...
10 ps. The resulting record is called the trajectory or the “movie” and gives a complete record of the system structure and dynamics over the period simulated.

<table>
<thead>
<tr>
<th>Command Prompt</th>
<th>Brief Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gmx command suffixes (7 total)</td>
<td>GROMACS, Version 2016.3</td>
</tr>
<tr>
<td>1. DSSP</td>
<td>Define Secondary Structure of Protein, analyzes structural changes of a protein by computing the secondary structure elements for each time frame.</td>
</tr>
<tr>
<td>2. Hbond</td>
<td>Counts the number of hydrogen bonds between all possible donors and acceptors (amongst designated group of atoms).</td>
</tr>
<tr>
<td>3. MinDist</td>
<td>Computes the distance between one group and a number of other groups. Both the minimum distance (between any pair of atoms from the respective groups) and the number of contacts within a given distance.</td>
</tr>
<tr>
<td>4. RDF</td>
<td>Radial Distribution Function counts the number of particles within concentric rings around another particle.</td>
</tr>
<tr>
<td>5. RMS</td>
<td>Compares two structures (initial or designated structure location with the same structure’s new location) by computing root mean squared deviation.</td>
</tr>
<tr>
<td>6. RMSF</td>
<td>Computes the root mean square fluctuation of atomic positions (and local structure flexibility) in the trajectory after fitting to a reference frame.</td>
</tr>
</tbody>
</table>

Table 2. Core command line operations used in this project.

After the completion of our MD simulations, we commenced post-simulation analysis using command-line operations in a Bash script (and script programming). We performed the following specific analyses on all 7 KaiC models and the Wild-type model, (See Table 1):

- DSSP, secondary structure analysis
- Hbond, hydrogen bonds vs. time, ATP – AA; ATP – water
- MinDist, minimum distance, Ion – ATP interaction
- RDF, Radial Distribution Function, Ion – ATP
- MS, (RMSD) Root Means Square Deviation vs. time
- RMSF, Root Means Square Fluctuation vs. AA, amino acid residues
- SASA, Solvent Accessible Surface Areas, ATP

(for brief descriptions, See Table 2).

Atomistic MD Results:
Structure Stability

The overall structural analyses show no major problems with the structural integrity of the system, the hexamers stay together. The Define Secondary Structure of Protein (DSSP) analysis and the Root Mean Square Deviation (RMSD) analysis indicates no dramatic change in protein structure (Fig. 3). Additional structural analysis, RMSD of experimentally relevant KaiC models continue to show excellent structural integrity (Fig. 4). There is no dramatic change in the protein structure of the additional models.
ATP Accessibility

The next question to address is whether there are differences in the binding and external accessibility of the ATPs for the different mutants. The ATP is more mobile in the binding site of the truncated mutant compared to the other KaiC states. The truncated KaiC mutant shows a different degree of hydrogen bonding between the ATP and the protein and surrounding waters (Fig. 5).

Interestingly, the truncated glutamate mutants show a degree of ATP-protein hydrogen bonding that's intermediate between the non-truncated glutamate mutants and the truncated structure without glutamate mutations (Fig. 6).
Our hydrogen bonds analysis (Hbond) results illustrate how long the hydrogen bonds persist over the duration of the simulation. The truncated mutant has the shortest-lived ATP-protein hydrogen bonds, exchanging most if not all bonds before the 250 ns point in time (Fig. 7). This is consistent with the idea that KaiA binding (which the truncation mutant is mimicking) acts to loosen the ATP in its binding pockets, and also consistent with the observed increase hydrogen bonds between ATP and water, indicating water is now more accessible to the ATP pockets (Fig. 5).

**Figure 4. Additional structural analysis.** Simulation of time steps of experimentally relevant models again demonstrate structural soundness of our simulation.

**Figure 5. Hydrogen Bonds (Hbond).** The count of hydrogen bonds vs. simulation time (500 ns). Truncated KaiC shows a different pattern of ATP-protein hydrogen bonding.
**Figure 6. Hydrogen Bonds.** The truncated glutamate mutants give a different idea about the ATP binding. The truncated mutants show a degree of ATP-protein hydrogen bonding that’s intermediate between the non-truncated glutamate mutants and the truncated structure without glutamate mutations.

**Figure 7. Hydrogen Bonds.** ATP and protein hydrogen bonds. The truncated mutant exchanges its ATP-protein hydrogen bonds more quickly than the other KaiC states.
The ATP Solvent Accessible Surface Area (SASA, Fig. 8) analysis is another metric of how accessible the ATP groups are to the outside water. The Full (unmutated wild-type model) and the threonine glutamate (TE) mutant solvent accessible surface areas are virtually identical. The TE on its own doesn’t seem to open the accessibility to the water for the ATPs. The serine glutamate (SE) mutant is a little more accessible in relation to the solvent than the threonine mutant. The truncated models, including those with SE and SETE mutations are

![SASA - ATP - KaiC-ALL](image)

Figure 8. Solvent Accessible Surface Area (SASA). The truncated models, with the exception of the truncated TE mutant, show the most accessibility to water in the truncated model’s simulation. Surface area measurements show that the simulations of the KaiC models expands and opens notably more than the glutamate non-truncated.

the most highly accessible with the greatest surface area. The truncated model with the TE mutation has an ATP solvent accessible surface area similar to untruncated SETE mutant. With the exception of the TE mutant, the ATPs are much more accessible to the water in the truncated models simulations, which is additional evidence that one effect of KaiA binding is to open up the ATP binding sites in the CII ring of KaiC.

There are big differences in the flexibility in certain regions of the truncated KaiC protein model. Our RMS Residue Fluctuations analysis shows distinct differences in flexibility of the truncated wild-type mutant (peaks in residues 300 to 500, See Fig. 9). The truncated glutamate models show particular differences in flexibility (Fig. 10, red boxed area).
Figure 9. Root Means Square Fluctuation (RMSF). There are notable changes in flexibility among the models (note peaks in graphs 1, 2, 3, and 4).

Figure 10. RMSF. The truncated glutamate KaiC models showed more stability than the truncated wild-type simulations (Fig. 9), but slightly more flexibility than the wild-type.
Highly charged molecules like ADP and ATP are usually coordinated by counter-ions which also play a role in their mobility into and out of protein pockets, as noted in the recent study by Hong et al. (40). Two analysis techniques were used to determine the distribution of sodium ions (Na+) around the ATP molecules bound in the KaiC CII ring. The first simply kept track frame by frame of the closest approach of a Na+ to ATP oxygen atoms. This analysis we refer to as “Mindist” the other analysis tool was to calculate radial distribution function of the Na+ around the ATP oxygens. This is a more costly analysis than Mindist but provides a complete picture of the location of the sodiums around the ATP molecules.

Our Mindist calculations (Fig. 11 & 12) show there is a clear effect of the SE and TE mutations. Sodium ions cluster around the ATPs in the glutamate mutants. The Mindist measurements show more sodium atoms nearer to the ATP in the glutamate mutants. The truncated and wild-type KaiC, lacking a glutamate mutation, show no close contacts between sodium ions and the ATP oxygens. This is an interesting result since the ATP charge, but the addition of one or two additional negative charges through the glutamate mutations (and presumably phosphorylation in the actual clock system) significantly increase the Na+ concentration near the ATP, which should facilitate the binding of ATP and the unbinding of ADP.

**Figure 11. Minimum Distances (Mindist).** Sodium and ATP.
Figure 12. **Mindist.** Sodium and ATP distances in the wild-type and truncated wild-type show no sodium ions bound near the ATP binding pockets. Whereas all the glutamate mutants showed a notable attraction by the sodium ions.

Figure 13. **Radial Distribution Function (RDF).** Sodium count is a normalized number of particles in a concentric ring of radius around the oxygen atoms of the ATPs in the CII ring models. Peaks near zero show clustering of ions near the ATP oxygens.
Simulating KaiA binding to the A-Loop (of KaiC) by truncating the last 10 C-Terminus AA give a vivid outcome. Experimentally, KaiA pulls the A-Loop (C-Terminus residues) away from KaiC, not allowing the A-Loop to tuck down into KaiC. Without the A-Loop available in the truncated mutants, our simulations appear to mimic this effect. Our simulations show accessibility to the ATPs by other atoms (here water) is increased, thus the number of water-ATP hydrogen bonds in the truncated mutants is greater than the other mutant models’ simulations (Fig. 5, ATP-water hydrogen bonds).

Observing solvent mobility and accessibility, one clear effect is that the glutamate mutants pull in sodium ions. How many sodiums are around the central oxygen (O) atoms of the ATPs? Sodium ions are found near the oxygen of the ATPs of the protein conformations with the glutamate mutations. The opposite appears to be true for the full (wild-type) and truncated mutant (Fig. 13 & 14).

**SUMMARY**

Our simulations of the 8 distinct forms of the KaiC hexamer show distinct differences for a number of properties. The truncation mutants, designed to mimic the effect of KaiA binding to the KaiC CII ring, have the effect of loosening the ATP binding sites as shown by reduced protein-ATP and increased ATP-water hydrogen bonding. This effect is also evident in the increased solvent accessible
surfaces of the bound ATPs in the truncation mutants. More generally the truncation mutation is seen to cause increased flexibility in distinct regions of the protein which appears to be consistent with decreasing the binding of ATPs in the CII ring. A different result is that the inclusion of glutamate mutations near the ATP binding sites increases the clustering of sodium ions near the ATP, which should facilitate ATP binding or ADP unbinding.

CONCLUSION

Our half-microsecond simulations of distinct states of KaiC in the cyanobacteria circadian clock yielded notable differences in the structure and dynamics of the different states. Notably, the binding of KaiA to the CII ring, which we mimicked by truncating the 10 C'-terminal amino acids was found to increase the flexibility of the CII ring and to open up the ATP/ADP binding pockets. Additionally, the phosphorylation of one or two residues near the ATP binding site, which we mimicked with glutamate mutations, was found to increase the sodium ion concentration in the pocket. Such results highlight the value of atomistic MD simulations of large protein complexes, like KaiC, even for time scales limited to sub-microsecond scales.


