THE DAY/NIGHT SWITCH OF THE CIRCADIAN CLOCK OF *SYNECHOCOCCUS ELONGATUS* AND HYDROGEN BONDS OF DNA AND RNA

A Dissertation

by

YONG-ICK KIM

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2008

Major Subject: Biochemistry

THE DAY/NIGHT SWITCH OF THE CIRCADIAN CLOCK OF *SYNECHOCOCCUS ELONGATUS* AND HYDROGEN BONDS OF DNA AND RNA

A Dissertation

by

YONG-ICK KIM

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Pingwei Li
	Andy C. LiWang
Committee Members,	Susan S. Golden
	J. Martin Scholtz
Head of Department,	Gregory D. Reinhart

December 2008

Major Subject: Biochemistry

ABSTRACT

The Day/Night Switch of the Circadian Clock of *Synechococcus Elongatus* and Hydrogen Bonds of DNA and RNA. (December 2008)

Yong-Ick Kim, B.S., Sung Kyun Kwan University;

M.S., Sung Kyun Kwan University

Co-Chairs of Advisory Committee: Dr. Pingwei Li
Dr. Andy C. LiWang

The circadian oscillator of the cyanobacterium Synechococcus elongatus is composed of only three proteins, KaiA, KaiB, and KaiC, which together with ATP can generate a self-sustained ~24 hour oscillation of KaiC phosphorylation for several days. KaiA induces KaiC to autophosphorylate whereas KaiB blocks the stimulation of KaiC by KaiA, which allows KaiC to autodephosphorylate. We propose and support a model in which the C-terminal loops of KaiC, the "A-loops", are the master switch that determines overall KaiC activity. When the A-loops are in their buried state, KaiC is an autophosphatase. When the A-loops are exposed, however, KaiC is an autokinase. The data suggest that KaiA stabilizes the exposed state of the A-loops through direct binding. We also show evidence that if KaiA cannot stabilize the exposed state KaiC remains hypophosphorylated. We propose that KaiB inactivates KaiA by preventing it from stabilizing the exposed state of the A-loops. Thus, KaiA and KaiB likely act by shifting the dynamic equilibrium of the A-loops between exposed and buried states, which shifts the balance of autokinase and autophosphatase activities of KaiC. A-loop exposure

likely moves the ATP closer to the sites of phosphorylation and we show evidence in support of how this movement may be accomplished.

Density functional theory calculations of isolated Watson–Crick A:U and A:T base pairs predict that adenine ^{13}C2 trans-hydrogen bond deuterium isotope shifts due to isotopic substitution at the pyrimidine H3, $^{2h}\Delta^{13}$C2, are sensitive to the hydrogen-bond distance between the N1 of adenine and the N3 of uracil or thymine, which supports the notion that $^{2h}\Delta^{13}$C2 is sensitive to hydrogen-bond strength. Calculated $^{2h}\Delta^{13}$C2 values at a given N1–N3 distance are the same for isolated A:U and A:T base pairs. Replacing uridine residues in RNA with 5-methyl uridine and substituting deoxythymidines in DNA with deoxyuridines do not statistically shift empirical $^{2h}\Delta^{13}$C2 values. Thus, we show experimentally and computationally that the C7 methyl group of thymine has no measurable affect on $^{2h}\Delta^{13}$C2 values. Furthermore, $^{2h}\Delta^{13}$C2 values of modified and unmodified RNA are more negative than those of modified and unmodified DNA, which supports our hypothesis that RNA hydrogen bonds are stronger than those of DNA. It is also shown here that $^{2h}\Delta^{13}$C2 is context dependent and that this dependence is similar for RNA and DNA.

DEDICATION

This work is dedicated to my wife, Young Mi Jeong, and twins, Eugene Young Kim and Daniel Junyoung Kim. Without their help, none of this would have been possible. I give thanks to my family.

ACKNOWLEDGEMENTS

I would like to thank Drs. Susan S. Golden and Patricia J. LiWang and David P. Giedroc for all their advice for science and life. I had helpful discussion with Drs. Sunbae Lee, Li Zhang and Mario A. Pennella, as well as with Young Mi Jeong and Nai-Wei Kuo. Funding for these projects was provided by the National Institutes of Health (GM064576), the Advanced Research Program project (000517-0003-2006), University of California at Merced startup funds and the Robert A. Welch Foundation Grant (A-1471). I give special thanks to Dr. David Giedroc for assistance with his fluorometer. The NMR instrumentation at Texas A&M University was supported by the National Science Foundation Grant DBI-9970232 and by the Texas Agricultural Experiment Station. Part of this work has been published in the *Proceedings of the National Academy of Science of the United States of America* volume 105, pages 12825-12830 and in the *Journal of Biomolecular NMR* volume 34, pages 229-236.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Circadian Clocks

Diverse organisms from bacteria to mammals display metabolic and behavioral rhythms with a circadian (~24 h) period that are matched to the earth's day and night cycle (1, 2). These circadian rhythms are the result of an endogenous biological circadian clock. The evolutionary significance of circadian clock systems remains unclear but they have been found to be important for reproductive fitness in cyanobacteria (3), plants (4, 5), and flies (2). Their presence in virtually all light perceiving organisms strongly suggests that they have played and continue to play a fundamental role in evolution. In spite of the near-universal importance of endogenous biological circadian clocks, the structural and biochemical bases of their function is unexplored territory.

There are three essential and fundamental characteristics of circadian clocks (6-8).

1. They continue to generate robust circadian rhythms in the absence of external time cues. This is the primary evidence showing that circadian rhythmicity is intrinsic.

This dissertation follows the style and format of the *Proceedings of the National Academy of Science of the United States of America.*

2. The phase of the clock is reset by external cues. The circadian clock can be entrained to appropriate environmental cycles such as light/dark and low/high temperature signals.
3. The circadian clock shows temperature compensation. Unlike chemical reactions, the period of the clock is resistant to variations in temperature within physiological limits.

All known circadian clocks are composed of at least three components.

1. Input pathway: synchronizes biological time to external time.
2. Central oscillator: generates the circadian rhythm.
3. Output pathway: transmits the rhythm to downstream effectors.

The motivation here is to significantly advance our understanding of the biochemical and structural bases of a central oscillator of *Synechococcus elongatus* PCC 7942, a cyanobacterium and model organism.

Cyanobacterial Circadian Clocks

Cyanobacteria are among the oldest organisms and the only prokaryotes known that exhibit circadian rhythms (9-11). Clock genes are widely distributed in cyanobacteria (12) and are an outstanding model system for the circadian clock because of their simplicity (11, 13, 14). *Synechococcus* RF-1 is the cyanobacterium first shown to exhibit a circadian

rhythm, where a rhythmic diurnal nitrogen fixation that persisted for at least 4 days could be entrained by a 12-hour dark/light pattern (15). *Thermosynechococcus elongatus* BP-1 (16) and *Synechocystis* sp. PCC 6803 (17) are used as model organisms, but the most studied and best understood circadian clock is that of *Synechococcus elongatus* PCC 7942 (18), hereafter referred to simply as *S. elongatus*. In *S. elongatus*, the circadian clock imposes rhythmicity on cell division, nitrogen fixation, photosynthesis, amino acid uptake, carbohydrate synthesis, and respiration (10). Essentially all gene expression and chromosome compaction is rhythmically regulated by the circadian clock in the *S. elongatus* (19, 20).

Circadian rhythms in *S. elongatus* PCC 7942 can be monitored by introducing the luciferase gene as a bioluminescence reporter for gene expression (18, 21). Many rhythmic phenotypes were isolated by monitoring bioluminescence from chemically mutagenized colonies of *S. elongatus* (22, 23). These mutations were mapped to only three genes shown to be essential for the generation of a circadian rhythm: *kaiA*, *kaiB* and *kaiC*, where *kaiB* and *kaiC* share a promoter (24). Nineteen mutations with distinct rhythm phenotypes and a wide range of period lengths or arrhythmia were mapped to *kaiC* (24). Thirty four mutants with long periods or arrhythmic phenotypes were mapped to *kaiA* (25), while most *kaiB* mutations showed a short period phenotype (24).

The central oscillator of *S. elongatus* consists of three genes, *kaiA*, *kaiB,* and *kaiC*, with KaiA (284 residues), KaiB (102 residues), and KaiC (518 residues) (24). Although all three *kai* genes were shown to be necessary for the generation of circadian rhythms in *S. elongatus* (24, 26, 27), marine cyanobacteria *Prochlorococcus* lack the *k*aiA gene (12, 28,

29), indicating that cyanobacteria have evolved different circadian clock mechanisms. Indeed, *P. marinus* PCC 9511 displays robust 24-h rhythms of DNA replication and gene expression when it is incubated in alternating 12 h light:12 h dark periods, but unlike with *S. elongatus*, the rhythms rapidly damp out under continuous light (28). The central oscillator genes are not required for viability and do not affect growth rates of *S. elongatus* in non-competitive situations, but in a competition the *S. elongatus* strain with a free-running rhythm most similar to that of the environmental light/dark cycle dominates (30, 31).

KaiA is a positive regulator of *kaiC* expression and KaiC suppresses its own (*kaiBC*) gene expression (24). These observations supported a model where the *kai* genes form an autoregulatory mechanism based on negative feedback. Originally, it was thought that these three genes formed a transcription-translation-derived oscillatory (TTO) feedback loop (24), as is still the currently accepted as the model for eukaryotic central oscillators (32-34). In this TTO model, rhythmic transcription is essential for producing and maintaining a self-sustaining circadian oscillation. Continuous overexpression of KaiC represses not only *kaiBC* expression but practically all promoter regions in the organism (35). Therefore, the rhythmic accumulation and the degradation rates of KaiC are thought to be important for circadian timing. Essentially, all the promoters in cyanobacteria are under circadian control (19, 36). Therefore, KaiC is thought to be a promoter-nonspecific, genome-wide gene expression regulator. A temporal increase in the amount of KaiC expression also resets the phase of the rhythm (24, 37). The mechanism of the transcriptional regulation of *kaiBC* is not fully understood yet, though many possible

mechanisms are proposed (38). One of the possible mechanisms is that global gene expression is regulated by rhythmic chromosome compaction (20).

Yeast two-hybrid experiments showed that KaiA, KaiB, and KaiC interact with each other in *S. elongatus* (39). These experiments and an *in vitro* interaction assay revealed that polypeptide segments in the two domains of KaiC, termed CI and CII respectively, physically interact with KaiA (26). KaiC associates with KaiA during the subjective day and all three Kai proteins form large complexes with unknown stoichiometries during the subjective night (27, 40). This observation and the fact that several rhythm phenotypes are caused by mutations in the *kai* genes implies that the circadian cycle is driven by oscillatory interactions between the Kai proteins.

ATP promotes hexamerization of KaiC with the self association of CI and CII domains forming a double doughnut shape for the KaiC particle (41-43). Each half of KaiC contains a Walker's A motif that bind ATP, an imperfect Walker's motif B and a conserved catalytic glutamate residue (E78 and E318 in the CI and CII halves, respectively), suggesting that KaiC has kinase and phosphatase activities (24, 43). Various mutations of these motifs abolish ATP binding and severely disrupt circadian function (44) indicating that the motifs are critical to clock function. The CI domain is responsible for hexamerization, whereas the CII domain is involved in KaiC autophosphorylation at S431 and T432 (45-48).

The phosphorylation of KaiC rhythmically oscillates between hypophosphorylated and hyperphosphorylated forms *in vivo*, setting the period of the clock (42, 44, 49). KaiC phosphorylation increases during the subjective day and decreases during the subjective

night with a circadian rhythm (27), and this phosphorylation in turn regulates KaiC's activity (35). The KaiC hexamer has multiple possible phosphorylation states ($2^{12} = 4096$). However, if phosphorylation is a cooperative process only a fraction of these states would be biologically relevant with distinct biochemical characteristics (41, 50). KaiA and KaiB modulate the phosphorylation of KaiC *in vitro* and *in vivo*: KaiA enhances KaiC autokinase activity and this effect is antagonized by KaiB (42, 49, 51, 52). KaiA has two domains and it is the C-terminal domain that is necessary and sufficient for interactions with KaiC (52) and inducing KaiC autophosphorylation, whereas the N-terminal domain alone has no effect (52). One or two KaiA dimers can interact with one KaiC hexamer (53). KaiB alone has no affect on KaiC phosphorylation but hinders KaiA stimulation of KaiC, thereby triggering autodephosphorylation (54). Circadian localization of KaiB to the soluble and insoluble fractions of *S. elongatus* suggests that subcellular localization is an important factor for KaiB function *in vivo* (51).

Surprisingly, *S. elongatus* has a robust circadian rhythm and oscillation of KaiC phosphorylation in constant dark (in this condition transcription and translation are suppressed in cyanobacteria) or in the presence of translation inhibitors (55). This result suggested that TTO feedback process, which is considered essential for circadian timing for various eukaryotic systems (2), is not necessary to generate circadian rhythms in *S. elongatus*. This major discovery by the laboratory of Takao Kondo implied that the central oscillator of *S. elongatus* is completely proteinaceous. His laboratory then quickly showed that a robust self-sustained oscillation of KaiC phosphorylation can be reconstituted in a test tube with KaiA, KaiB, KaiC, and ATP (56). The ratio of phosphorylated KaiC (P-

KaiC) and non-phosphorylated KaiC (NP-KaiC) in this system shows temperature-compensated rhythms with a period of approximately 24 hours (56). Both S431 and T432 of KaiC are required for circadian rhythm generation. When one of the two residues is mutated, the circadian rhythm is destroyed *in vivo* and the mutants no longer generate oscillatory phosphorylation *in vitro* (47, 57, 58). These results indicate that the phosphorylation cycle of KaiC is the central pacemaker of the cyanobacterial clock. Phosphorylation at these two sites should not be independent and the detailed mechanism of the correlation between the two sites has been proposed (57, 58). In this model proposed by the Kondo laboratory, the phosphorylation states in KaiC oppositely regulate the phosphorylation and dephosphorylation reactions at the adjacent residue (57, 58). Phosphorylation of T432 stimulates phosphorylation at S431. S431 phosphorylation is a cue to switch from phosphorylation to dephosphorylation. Subsequently, dephosphorylation occurs on the T432, resulting in a form of KaiC that is only serine-phosphorylated (S-KaiC). S-KaiC recruits KaiB to the KaiA-KaiC complex and prevents KaiA from activating KaiC phosphorylation. Then, S431 dephosphorylates, KaiB dissociates, and KaiC returns to an unphosphorylated state to begin a new phosphorylation cycle (57, 58). Thus, the distribution of KaiC phosphoforms over a ~24 hour period proceeds as ST-KaiC (unphosphorylated) → SpT-KaiC (T432 phosphorylated) → pSpT-KaiC (S431 and T432 phosphorylated) → pST-KaiC (S431 phosphorylated) → ST-KaiC. In a significantly different model proposed by the laboratory of Erin O'Shea, phosphorylation at S431 does not change switch KaiC from an autokinase to an autophosphatase, but serves to recruit KaiB to inactivate KaiA (47, 57, 58). As will be shown in Chapter II, our results are more

consistent with this latter model. In addition to autokinase and autophosphatase activities, KaiC shows an extremely low and temperature compensated ATPase activity that has been proposed to define the period of the cyanobacterial circadian system (59).

In vivo and *in vitro* rhythmic KaiC phosphorylation is macroscopic over several cycles, which is possible only if the ensemble Kai proteins are acting in synchrony (60). The mechanism of establishing and maintaining synchrony is likely central to all biological clocks and here is an outstanding opportunity to discover how one system accomplishes the task. Prior to any experimental data, it was demonstrated mathematically that monomer exchange or shuffling between KaiC hexamers would be sufficient to maintain synchrony (61), soon afterwards, monomer shuffling was observed experimentally (54, 62). Because shuffling of monomers was shown to be most frequent during the dephosphorylation period, KaiB may involve an important role to exchange monomers (54).

The structures of the Kai proteins were unknown, but would clearly provide a wealth of functional insights. High-resolution structures for KaiA (40, 52, 63-65), KaiB (40, 66, 67) and KaiC (50) proteins are known. The X-ray crystal structure of KaiA of *S. elongatus* reveals a domain-swapped homodimer where the N-terminal domain of one subunit is packed against the C-terminal domain of the other subunit (65). The C-terminal domain, which is the dimerization domain, adopts a four-helix bundle fold (63, 68). The structure of KaiB from *Anabaena sp.* PCC7120 reveals a thioredoxin-like fold homodimer (40) while from *Synechocystis* PCC6803 and *T. elongatus* show formation of an unusual homotetramer, which are comprised of two asymmetrical dimers (66, 67). However, a KaiB appears to bind to KaiC as a dimer during the KaiC phosphorylation cycle (54). Thus, the

change of oligomeric states between tetramer and dimer may be important for proper clock function.

The amino acid sequences of KaiC showed little similarities to proteins of known structure, thus thwarting structure prediction efforts at the time. Examination of the sequence of KaiC showed two similar domains of 260 residues each with similarities to bacterial helicases (69). A crystal structure of the KaiC homohexamer from *S. elongatus* has double-doughnut shape with a central pore and 12 ATP molecules bound at CI-CI and CII-CII interfaces (50). The N-terminal domain, CI, is responsible for the ATPdependent hexamerization of KaiC (70) whereas the C-terminal domain, CII, by itself is monomeric in solution even at high concentrations (Kim & LiWang, unpublished results). CII, additionally, has autokinase and autophosphatase activities at S431 and T432 (47, 50).

The C-terminal 22 residues of KaiC, which directly bind to KaiA, were partly resolved in a recent crystal structure (71). The only high-resolution complex structure of KaiA and KaiC is the NMR solution structure of a complex between the dimeric C-terminal KaiA domain and two C-terminal 32-residue KaiC peptides from the cyanobacterium *T. elongatus* BP-1 (64). The peptides are bound inside the grooves on KaiA with mainly hydrophobic interaction (64) and contact both subunits of the dimer, thereby explaining the biological significance of KaiA dimerization. Recently, a model for the *T. elongatus* BP-1 KaiA–KaiC (ThKaiAC) complex was derived by single-particle EM reconstruction (71). In this model, a KaiA dimer is tethered to a KaiC hexamer through binding to a single C-terminal peptide of KaiC (71). The suggested mechanism of KaiC phosphorlyation is that the tethered KaiA dimer may have a transitory interaction with a KaiCII ATP binding cleft

and thusly enhance the autokinase activity (71). We will show in Chapter II that autokinase activity is not explained by this hypothetical interaction. Recently, a 3D model of a KaiB and KaiC complex was developed from negative-stain and cryo-electron microscopy in which two KaiB dimers are bound to the CII dome of a KaiC hexamer (72). It is proposed that a segment near the C-terminus of KaiC found looped in the X-ray crystal structure is extended in the KaiBC complex. We will show evidence in Chapter II that this is not the case. Still, structural details of the KaiA-KaiB-KaiC complex are unknown. Despite this progress in structural characterization, the biochemical and biophysical basis of the rhythmic assembly/disassembly dynamics and phosphorylation of KaiC are still poorly understood. Furthermore, rhythmic phosphorylation of KaiC may not be the only oscillatory mechanism in cyanobacteria. The phosphorylation cycle of KaiC is apparently integrated with a transcription-translation oscillator (73) to maintain stable 24-h cellular rhythms in *S. elongatus* genes (19, 35).

This dissertation focuses on the question of how KaiA and KaiB switch KaiC from an autokinase to an autophosphatase, which is central to the generation of the circadian rhythm of this central oscillator of *S. elongatus*. It will be shown here that KaiA stabilizes a segment of KaiC in a position that likely moves bound ATP molecules close to the sites of phosphorylation whereas KaiB prevents KaiA from doing so. The N-terminal domain of KaiA regulates KaiA-KaiC affinity as well as being required by KaiB to block KaiA. The motivation for the work presented here arose from examination of the NMR and X-ray structures of circadian clock proteins of this organism.

Hydrogen bonds of DNA and RNA

It is established that the hydrogen bond is of central importance in nucleic acid structure and function (74). The hydrogen bond allows nucleic acids to form a wide variety of structures in addition to the antiparallel double helix of Watson-Crick base pairs. DNA and RNA molecules can form hairpins, stable non-Watson-Crick base pairs, parallel stranded duplexes, triplexes, quadruplexes, junctions, hybrid duplexes, and a variety of folded tertiary structures (74-79). In addition, nucleic acids can form complexes with proteins, drugs, and metal ions through specific geometrically arranged patterns of hydrogen bonds (80-87). Their various structures modulate processes such as gene regulation, translation, mitosis, and chromosome stability (88-90). The large variety of structures and functions of DNA and RNA are achievable in large part because nucleosides contain several hydrogen bonding functional groups.

Understanding DNA and RNA function, and malfunction, relies on a clear picture of the intricate hydrogen bonding interactions. Yet, the only method that will directly and quantitatively measure site-specific hydrogen bond length (not angle) is to measure scalar $^{15}N\cdots^{15}N$ trans-hydrogen bond couplings, $^{2h}J_{NN}$ (91, 92). This technique requires the uniform isotopic enrichment of nucleic acid polymer with ^{15}N, which is expensive and technically difficult. Furthermore, $^{2h}J_{NN}$ is determined from peak intensities, which are determined not only by $^{2h}J_{NN}$ but also complicated by relaxation rates and noise, yielding uncertainties of several tenths of hertz. Most commonly, in hydrogen bonds are inferred from the spatial proximity of donor and acceptor atoms after the structure has been solved

by single crystal X-ray diffraction crystallography or NMR spectroscopy (93). This approach is indirect.

NMR evidence for the presence of individual hydrogen bonds comes from protection of labile protons from hydrogen exchange with bulk solvent (94, 95). However, hydrogen exchange requires unfolding or breathing of the molecule and therefore is sensitive to numerous other factors as well (96-98). The proton chemical shift can be sensitive to hydrogen bonding (99), but is also influenced by ring current and other effects (100). As a result, "standard" hydrogen bond distances are routinely assumed and used in nucleic acid structure determination by NMR, and hydrogen bond parameters such as propeller twist are often poorly determined (101). Furthermore, current molecular biophysics techniques to study hydrogen bonding in biomolecules yield data which are not usually straightforward to interpretation (102). Thermal denaturation profiles yield bulk properties and a comparison of molecules differing at a single site can be complicated by non-local perturbing effects of the mutation. Infrared, raman, and solid state nuclear quadrupole resonance spectroscopy are excellent techniques for small molecules, but spectral resolution is a major problem when investigating nucleic acid polymers.

As interactions with modified bases, proteins, drugs, nucleic acids, and metals are mediated by hydrogen bonds of varying angles and lengths, it is necessary to obtain quantitative parameters for a complete understanding of the relationship between the structure of DNA, RNA, their complexes, and function. Optimally, the technique should not require isotopic enrichment and yet allow exquisitely precise measurements.

It was recently shown that the ^{13}C2 chemical shift of adenosine residues of A:T base pairs of double stranded DNA, measured at natural abundance ^{13}C, experience a trans-hydrogen bond isotope effect ($^{2h}\Delta^{13}$C2) upon substitution of ^{1}H for ^{2}H at the imino site of the complementary thymine (103). This isotope effect is a direct gauge of the hydrogen bond interaction since it is transmitted through space from the hydrogen bond donor and detected at the hydrogen bond acceptor. The $^{2h}\Delta^{13}$C2 can be thought of as a two-bond deuterium isotope shift, with one covalent bond and one hydrogen bond. Theory based on small molecule experiments show that the two-bond deuterium isotope shift reflects the anharmonicity (asymmetry) of the, say, N–H vibrational potential, which increases in anharmonicity as the length of a weak hydrogen bond decreases (104). These $^{2h}\Delta^{13}$C2 values were correlated with the isotropic chemical shift of the base paired imino proton, suggesting that $^{2h}\Delta^{13}$C2 are indeed a measure of the N–H···H hydrogen bond length. $^{2h}\Delta^{13}$C2 is measured from DNA samples dissolved in a buffer of 50% H_2O/50% D_2O and is defined as the chemical shift difference δ^{13}C2(H) – δ^{13}C2(D), and were found ranging from -55 to -42 ppb. The wide line widths resulting from ^{13}C-decoupled HSQC or HMQC type experiments gave poor resolution of the ^{13}C2(H) and ^{13}C2(D) resonances, necessitating the use of the line-narrowing TROSY type NMR experiment (103).

Interestingly, in 2004 it was still unknown whether there are differences between the hydrogen bonds of DNA and RNA, even though much is known about their structural, thermodynamic, and functional differences. Early NMR experiments showed that RNA protons in general relax much slower than those of DNA (105). However, a comparison of $^{2h}J_{NN}$ values of non-isosequential DNA and RNA did not reveal any differences within the

experimental uncertainty of tenths of hertz (91, 92). A comparison of the highest resolved X-ray crystal structures of DNA and RNA showed their hydrogen bonds to be equal within the experimental uncertainty of several hundredths of an Angstrom (106). Thus, either DNA and RNA hydrogen bonds were the same, or differences would only be revealed by an exquisitely sensitive method.

The advantage of $^{2h}\Delta^{13}C2$ was that it depended only on the determination of peak positions, which the most precisely determined observable of NMR, unlike peak intensities used for determination of $^{2h}J_{NN}$. Measurements of a series of isosequential DNA and RNA oligomers showed that RNA consistently had larger $^{2h}\Delta^{13}C2$ magnitudes (106). Thus, it was concluded that the A:U hydrogen bonds of RNA are shorter than those of A:T hydrogen bonds of DNA.

Differences in hydration between DNA and RNA may be the reason for this difference in $^{2h}\Delta^{13}C2$. Hydrogen bonds are stronger under conditions of low water activity (107), and RNA is less hydrated than DNA (108). Indeed, dehydration of DNA often changes it from the B-form conformation to the A-form conformation, the conformation of double helical RNA (74).

However, a subsequent report showed computationally that differences between $^{2h}\Delta^{13}C2$ of DNA and RNA could be explained simply by the chemical difference between thymine and uracil (109). Specifically, the calculations showed that the electron-donating character of the C7 methyl group of thymine reduces the magnitude of $^{2h}\Delta^{13}C2$ of DNA. This paper apparently invalidated the experimentally determined differences in $^{2h}\Delta^{13}C2$ between DNA and RNA. The objective of the second part of this dissertation is to

determine experimentally the impact of the chemical difference between thymine and uracil on $^{2h}\Delta^{13}C2$.

CHAPTER II

THE DAY/NIGHT SWITCH IN KAIC, A CENTRAL OSCILLATOR COMPONENT OF THE CIRCADIAN CLOCK OF CYANOBACTERIA*

Introduction

Virtually all light-perceiving organisms display circadian (≈24-h) rhythms in their gene activity, metabolism, physiology, and behavior in anticipation of and preparation for daily swings in sunlight and ambient temperature (2, 110). These robust biological rhythms are the result of an endogenous clock called the circadian clock and have been identified in cyanobacteria (9, 18). In the cyanobacterium *Synechococcus elongatus*, the clock is important for reproductive fitness (3, 31), and expression of the entire genome is under clock control (19, 35).

Only three proteins make up the central oscillator of the *S. elongatus* clock (24): KaiA, KaiB, and KaiC. The high-resolution structures of all three proteins are known (40, 50, 52, 64-66). Remarkably, the central oscillator can be reconstituted in a test tube (56). A simple mixture of the three proteins and ATP regenerates the sustained cellular ≈24-h rhythm of KaiC phosphorylation. Temperature compensation and mutant rhythm phenotypes observed *in vivo* are also reproduced by this clock-in-a-test tube.

---

* Reprinted with permission from "The day/night switch in KaiC, a central oscillator component of the circadian clock of cyanobacteria." by Kim, Y.-I., Dong, G., Carruthers, C.W. Jr., Golden, S.S. LiWang, A.C. (2008) *Proc Natl Acad Sci USA*. **105**, 12825-12830. Copyright 2008 National Academy of Sciences, U.S.A. All *in vivo* experiments were performed by Dr. Guogang Dong in the laboratory of Dr. Susan S. Golden (Texas A&M University, Department of Biology).

An ensemble of KaiC molecules exhibits both autokinase and autophosphatase activities (49, 51, 52). In *S. elongatus*, the balance between these two activities swings back and forth on a daily basis. *In vivo*, KaiC is hyperphosphorylated during the night and hypophosphorylated during the day (42, 55). There are two phosphorylation sites on KaiC, S431 and T432 (47, 48). The distribution of KaiC phosphoforms over a ≈24 h period proceeds as ST-KaiC (unphosphorylated) → SpT-KaiC (T432 phosphorylated) → pSpT-KaiC (S431 and T432 phosphorylated) → pST-KaiC (S431 phosphorylated) → ST-KaiC (57, 58), although individual KaiC molecules do not necessarily go all of the way around the loop (58). Alone, the autophosphatase activity of KaiC is dominant; KaiA shifts the balance of activities from autophosphatase to autokinase (42, 49, 52). However, the pST-KaiC phosphoform recruits KaiB, and, together, they inactivate KaiA in KaiABC complexes (57, 58, 60). With insufficient levels of active KaiA in solution, the ensemble of KaiC molecules begins to dephosphorylate. KaiA activity resumes once enough pST-KaiC has decayed to ST-KaiC [implying that not all KaiC molecules make it back to the ST-KaiC state before being restimulated by KaiA to autophosphorylate (58)]. Monomer shuffling between KaiC hexamers may play a role in synchronization (62, 111).

The phosphorylation cycle of KaiC is apparently integrated with a transcription-translation oscillator (73) to maintain stable 24-h cellular rhythms in *S. elongatus* genes (19, 35). Thus, it is of central importance to understand how KaiA and KaiB shift the relative autokinase and autophosphatase rates in KaiC. Here, we present evidence that a segment of residues near the C terminus of each KaiC subunit determines which activity is dominant. We propose that when these ‘‘A-loops’’ are buried, KaiC is an auto-phosphatase. However,

when the A-loops are exposed, KaiC is an autokinase. We suggest that there is a dynamic equilibrium between the buried and exposed states of the A-loops, so that an ensemble of KaiC molecules exhibits both activities simultaneously. In the absence of other proteins, KaiC is both an autokinase and autophosphatase, with the latter activity dominant over the former (42, 49, 52). In this case, according to our model the dynamic equilibrium favors the buried state of the A-loops. We propose that KaiA stabilizes the exposed state, thereby increasing the autokinase rate relative to that of the autophosphatase. We think that KaiB acts by preventing this KaiA-mediated stabilization.

Results

The Equilibrium Position of the A-Loop Determines the Steady-State Phosphorylation Level of KaiC: Shown in Figure 2.1*A* is the NMRstructure of the complex between the C-terminal domain of KaiA, KaiAC, and a peptide derived from C-terminal residues 488–518 of KaiC, both from *Thermosynechococcus elongatus* (64). The x-ray crystal structure of *S. elongatus* KaiC (Figure 2.1*B*) shows that residues 488–497 (magenta) are buried in a looped conformation (50). Here, we refer to this stretch of residues as the A-loop. ‘‘Tail’’ residues 498–519 are not resolved in this x-ray crystal structure but are seen to protrude from the top of KaiC in a later study (71). The homologous solvent-exposed tail residues from *T. elongatus* are colored yellow in Figure 2.1*A*. The A-loop + tail segment of KaiC ($KaiC^{AL+\,tail}$) from *S. elongates* and *T. elongatus* are 61% identical and 84% similar.

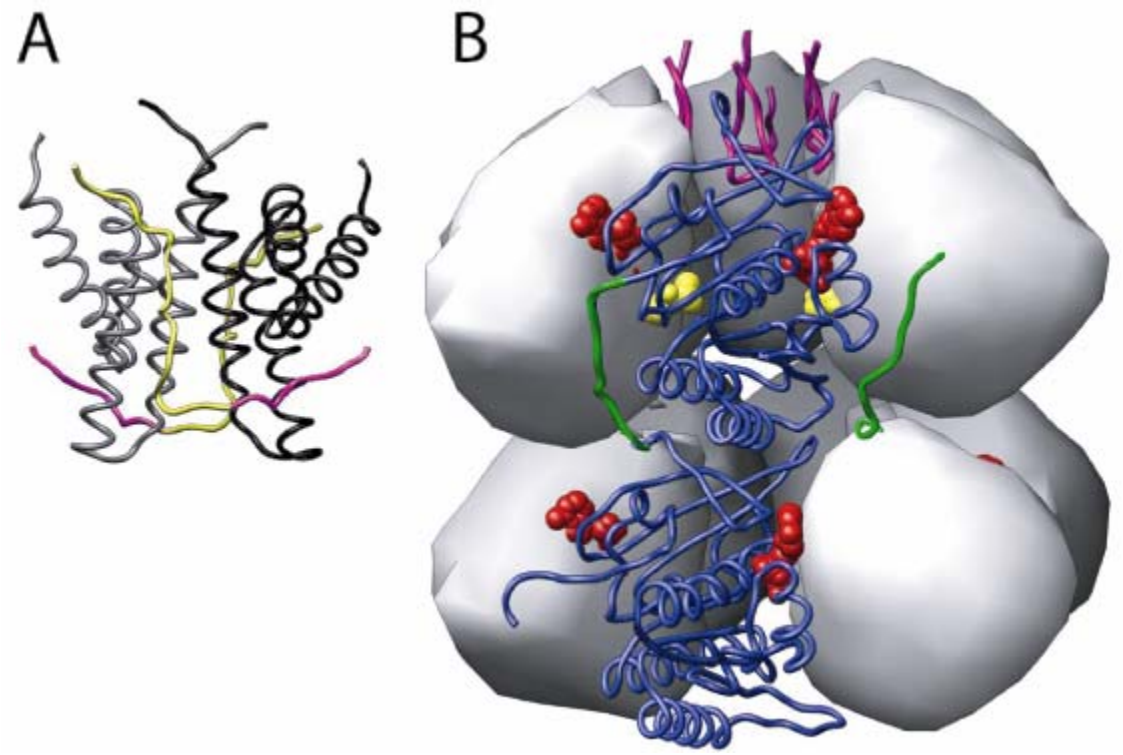

Figure 2.1. The NMR structure of the *T. elongatus* KaiA^{C}– KaiC$^{AL+tail}$ complex (*A*) and x-ray crystal structure of KaiC from *S. elongatus* (*B*). (*A*) The two subunits of the KaiA^{C} dimer of the KaiA^{C}–KaiC$^{AL+tail}$ complex (PDB ID code 1suy) are shown as black and gray ribbons. The two bound KaiC$^{AL+tail}$ peptides are shown as ribbons, with the A-loop segment colored magenta and the tail segment colored yellow. (*B*) The N- and C-terminal domains of five subunits of KaiC (PDB ID code 1tf7) are shown as gray surfaces, whereas those of the remaining subunit are shown as a blue ribbon. The polypeptide segment connecting the two domains is shown as a green ribbon. The A-loops are shown as magenta ribbons. The solvent-exposed C-terminal residues 498–519 are unresolved in this structure. ATP molecules are red and S431 and T432 are yellow. All figures of Kai protein structures were created using the program Chimera (112).

In addition, residues in *T. elongatus* KaiA^{C} that interact with KaiC$^{AL+\ tail}$ according to the NMR structure share a high level of identity with *S. elongatus* KaiA^{C} (64). In the present study, all experiments were performed using *S. elongatus* proteins, except for one set of fluorescence anisotropy experiments shown in the figure on page 30.

A comparison of Figure 2.1 *A* and *B* suggested that A-loops have two states, buried as seen in Figure 2.1*B* and exposed as implied by Figure 2.1*A*. In our model, the A-loops in an ensemble of KaiC molecules exist in a dynamic equilibrium between buried and exposed states, which determines the steady-state level of phosphorylation of KaiC. If the equilibrium favored the buried state, the rate of autodephosphorylation would have been faster than that of autophosphorylation. In contrast, if the exposed state were favored, then the autokinase rate would be faster, leading to a high steady-state phosphorylation level for KaiC. According to our model, KaiA increased KaiC phosphorylation levels by stabilizing the exposed state of the A-loops by directly binding to KaiC$^{AL+\ tail}$.

As a test of our model, we made a variant of KaiC to mimic the exposed state of A-loops. This variant, KaiC487, was truncated after residue 487 and, therefore, missing the KaiC$^{AL+tail}$ segment. As seen in Figure 2.2*A* (green ○), KaiC487 by itself was constitutively 100% phosphorylated. A recent study has shown that the steady-state phosphorylation level for *S. elongatus* KaiC under increasing concentrations of KaiA does not exceed ≈85% (113). Another study has shown, for *T. elongatus* proteins, that a ratio of one KaiA dimer to one KaiC hexamer is enough to reach saturation level of phosphorylation, which is also <100% (53). KaiA cannot induce 100% phosphorylation of a population of KaiC molecules

because, even under saturating conditions, A-loops probably still sample the buried state to a minor extent.

Our model predicted that reducing the ability of KaiA to stabilize the exposed state of the A-loops should lead to a lower steady-state level of KaiC phosphorylation. We, therefore, produced a KaiC variant, KaiC497, which was truncated after residue 497, so that it had the A-loop but was missing the solvent-exposed tail. The steady-state phosphorylation levels of KaiC497 alone and in the presence of KaiA were similarly low (Figure 2.2 *A* and *B*, black ▲ and ▼), which suggests that the dynamic equilibrium of the A-loop in both cases is shifted toward the buried state. A KaiC variant similar to KaiC497 is unable to form a complex with KaiA, as determined by electrophoretic mobilityshift assays (71). Truncations partway into the A-loop resulted in KaiC variants that were hyperphosphorylated (Figure 2.2*A*, green ×, blue ■, and red ◇), although to a lesser extent than KaiC487 (green ○). To a minor extent, these truncated A-loops may still sample the buried state, thus leading to <100% steady-state phosphorylation levels. In addition to deletion experiments, KaiC variants containing point mutations in or bordering the A-loop were also made to test our model. In the structure of KaiC, the side chains of residues E487 and T495 appeared to be hydrogen-bonded to each other (Figure 2.3), suggesting that the buried state of the A-loop was stabilized by this interaction. Indeed, the E487A substitution created a KaiC variant that was 100% constitutively phosphorylated, whereas the T495A variant was 80% constitutively phosphorylated (Figure 2.4).

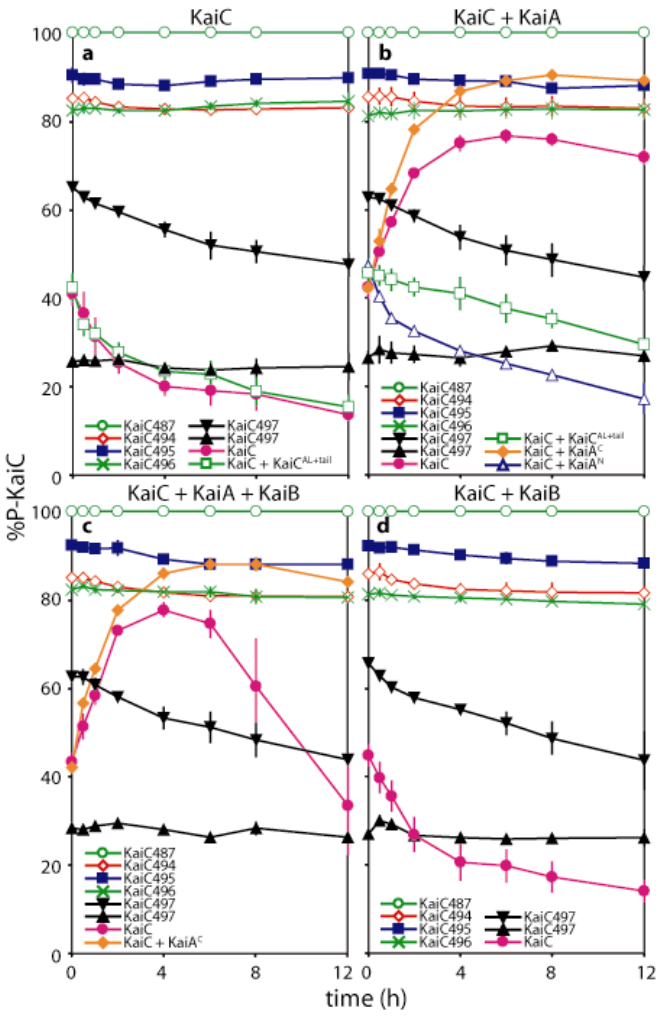

Figure 2.2. Phosphorylation of KaiC and KaiC variants for KaiC alone (*A*), KaiC + KaiA (*B*), KaiC + KaiA + KaiB (*C*), and KaiC + KaiB (*D*). For all images: pink ●, KaiC; black ▼ and ▲, KaiC497; green ×, KaiC496; blue ■, KaiC495; red ◇, KaiC494; green ○, KaiC487. In *A*, green □ indicates KaiC + $KaiC^{AL+tail}$ (500 μM). In *B*, green □ indicates KaiC + $KaiC^{AL+tail}$ (500 μM) + KaiA; blue △, KaiC + $KaiA^{N}$ only; and orange ◆, KaiC + $KaiA^{C}$ only. In *C*, orange ◆ indicates KaiC + $KaiA^{C}$ + KaiB. Each data point is the mean of two independent experiments. Solid lines connect data points for visual clarity. Images of the stained polyacrylamide gels are shown in Appendix A. Assignment of phosphorylated and unphosphorylated KaiC bands resolved by PAGE was determined from lambda phosphatase assays (Appendix A). All KaiC samples were fresh or from freshly frozen samples except for data indicated by black ▲ symbols, which are from KaiC497 reaction mixtures after incubation at 30°C for 18 h ± KaiA ± KaiB. It should be noted that the phosphorylation levels of KaiC487, KaiC494, KaiC495, and KaiC496 after incubation at 30°C for 18 h ± KaiA ± KaiB remained similar to those after the initial 12 h (data not shown).

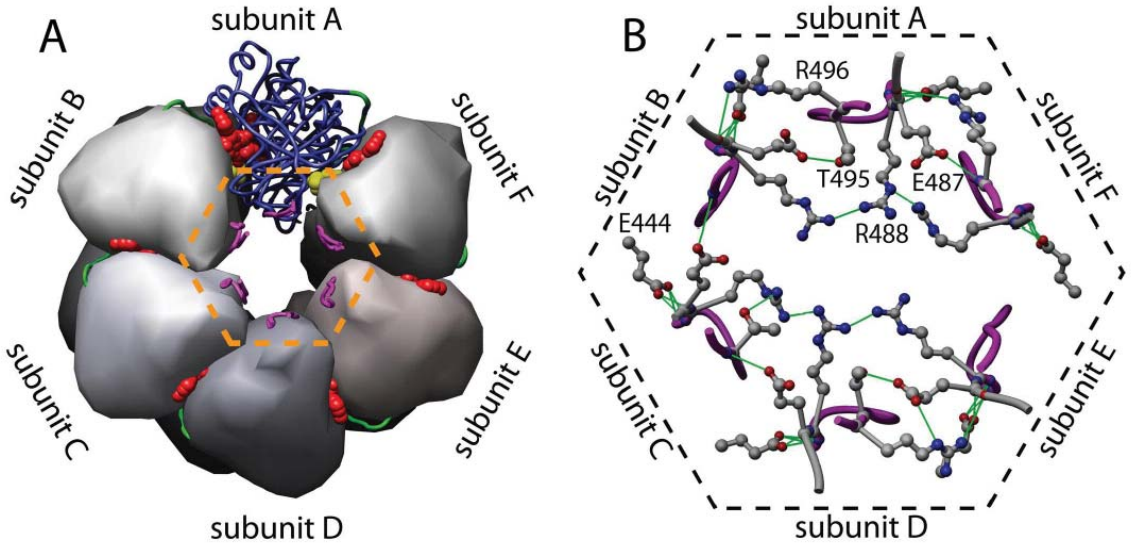

Figure 2.3. Putative hydrogen bonds linking adjacent A-loops. (*A*) Top-down view of KaiC. The coloring scheme is identical to that in Fig. 1*B*.(*B*) Expansion of the A-loop region. Putative intersubunit hydrogen bonds involving A-loop residues are shown as green lines.

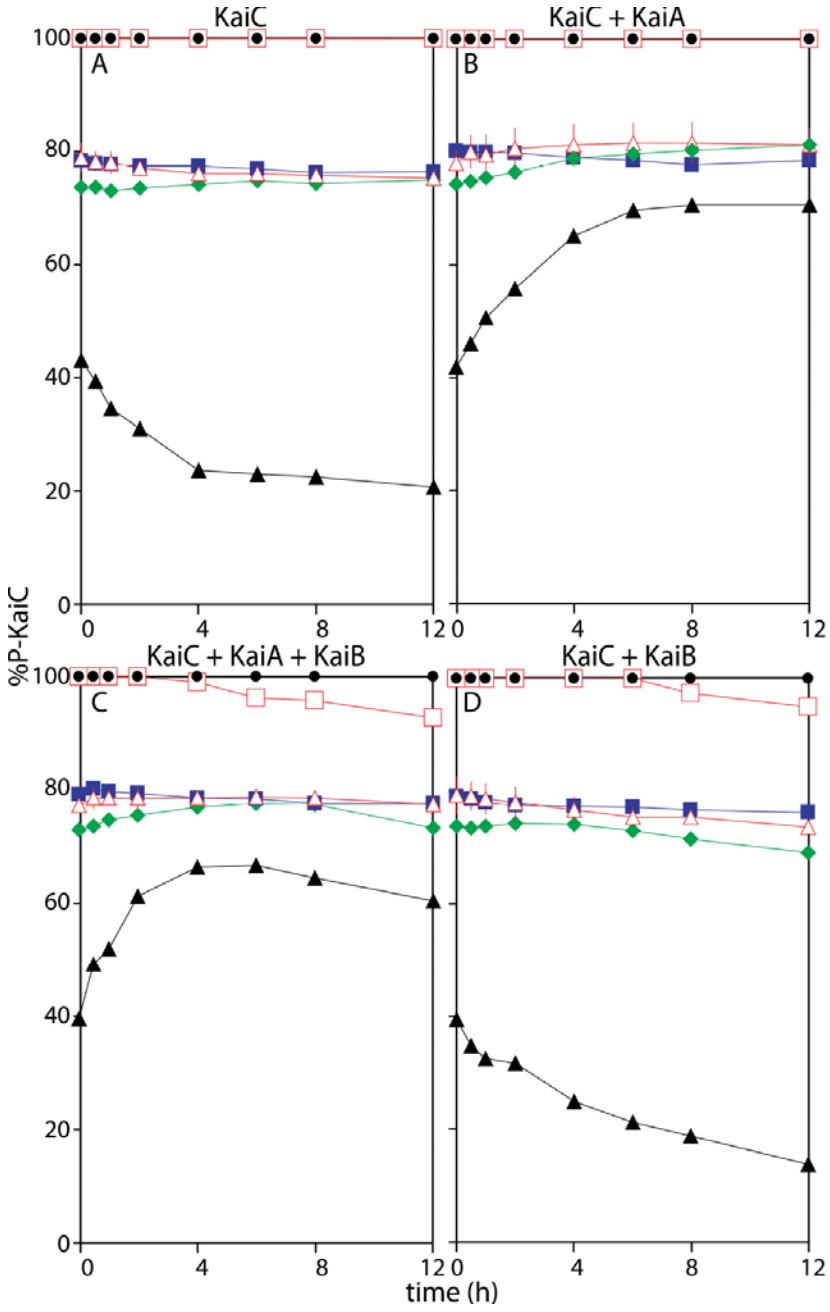

Figure 2.4. Phosphorylation kinetics of KaiC variants for KaiC alone (*A*), KaiC + KaiA (*B*), KaiC + KaiA + KaiB (*C*), and KaiC + KaiB (*D*). For all images: red □ indicates KaiC E487A; black ▲, KaiC D474A; red △, KaiC T495A; green ◇, KaiC W331A; black ●, KaiC E444D; blue ■, KaiC I472A. Solid lines connect data points for visual clarity. Images of the stained polyacrylamide gels are shown in Appendix A. All KaiC samples were fresh or from freshly frozen samples. Each data point is from a single experiment, except for those of T495A which are an average from two experiments.

KaiC487 and KaiC497 expressed in *S. elongatus* were, respectively, hyper- and hypophosphorylated as well (Figure 2.5); moreover, KaiC487 remained hyper-phosphorylated even when KaiA was absent and KaiB was present (Figure 2.5*A*, lanes 4–6), and KaiC497 remained hypophosphorylated when KaiA was present and KaiB was absent (lanes 7–9). In addition, ectopic expression of either KaiC487 or KaiC497 in a WT *S. elongatus* background had a dominant negative effect, abolishing circadian rhythmicity (Figure 2.5*B*), even though they had opposite states of phosphorylation. In contrast, rhythmicity was preserved when WT KaiC was expressed in a WT background, with a slightly longer free-running period than normal (Figure 2.5*B* and Figure 2.6).

KaiC497 and KaiC496, which differ only in the presence or absence of I497, had opposite steady-state levels of phosphorylation (Figure 2.2*A*, black ▲ and green ×). This result suggests that I497, the terminal residue of the A-loop, plays a critical role in stabilizing the buried position of the A-loop. It can be seen from the structure of KaiC (Figure 2.7) that I497 is apparently part of an intrasubunit hydrophobic cluster that includes residues V443, I445, I467, F483, F486, I489, I490, P494, and T495. Perhaps the removal of I497 in KaiC496 was sufficient to destabilize the cluster and thereby shift the dynamic equilibrium of the A-loop toward the exposed state.

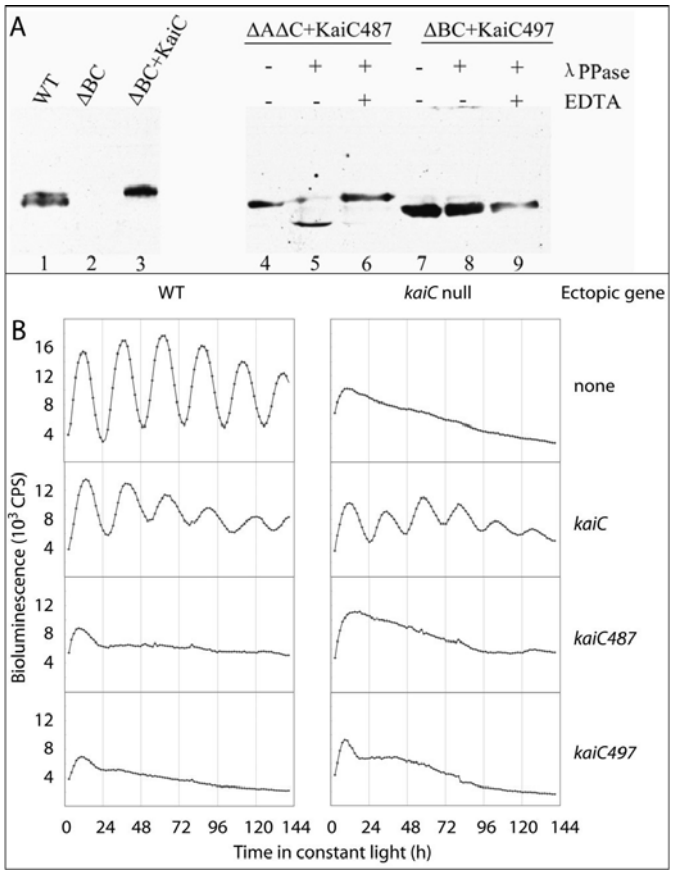

Figure 2.5. Phosphorylation states and functional analysis of KaiC487 and KaiC497 *in vivo*. (*A*) KaiC487 and KaiC497 are, respectively, constitutively hyper- and hypophosphorylated independently of KaiA in *S. elongatus*. Immunoblots of KaiC and its truncated variants were detected in soluble protein extracts from strains that express: lane 1, WT KaiC; lane 2, no KaiB or KaiC; lane 3, WT KaiC from an ectopic site in a *kaiBC* null background (KaiA is present); lanes 4–6, KaiC487 from an ectopic site in a *kaiA kaiC*-null background (KaiB is present); lanes 7–9, KaiC497 in a *kaiBC*-null background (KaiA is present). The phosphorylation status of KaiC variants, whose mobility differs from WT was determined by λ phosphatase treatment with (lanes 6 and 9) or without (lanes 5 and 8) the inhibitor EDTA. Total cell extracts from 10 ml of OD_{750} = 0.5 cyanobacterial cultures were prepared and further analyzed by immunoblotting (114) with polyclonal KaiC antiserum (115) at 1:2,000 dilution. Treatment with λ phosphatase was performed according to the recommendation of the manufacture (New England BioLabs). Briefly, 100 μg of total protein was incubated with 400 units of λ phosphatase in 50 μl of total volume at 30°C for 30 min. EDTA was used at a final concentration of 50 mM to inhibit the phosphatase reaction. (*B*) Ectopic expression of KaiC487 or KaiC497 abolishes circadian rhythmicity in WT *S. elongatus*. WT KaiC, KaiC487, and KaiC497 are expressed from the native *kaiBC* promoter in a WT (*Left*) or *kaiC* null (*Right*) background. All strains harbor a bioluminescence reporter gene driven by the *kaiBC* promoter.

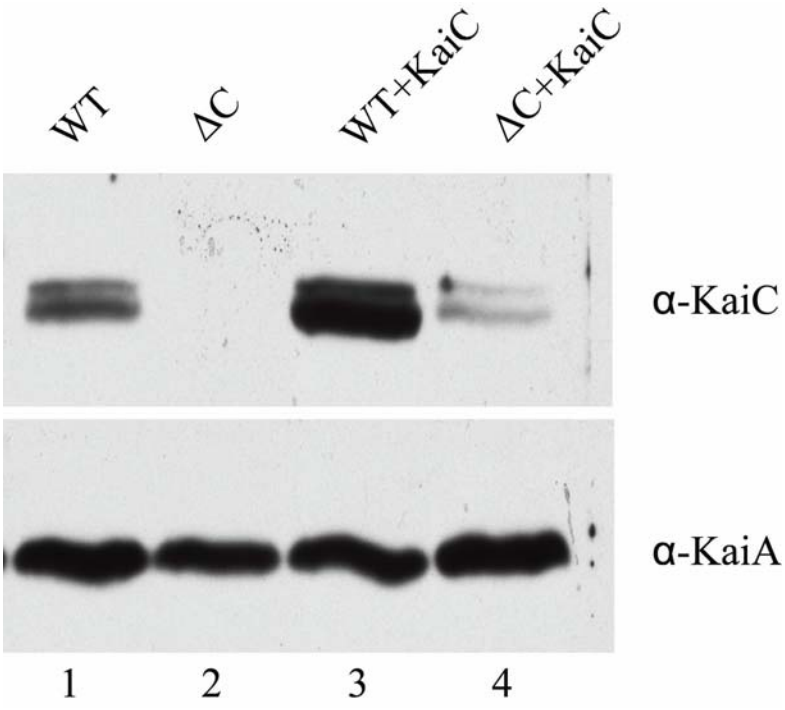

Figure 2.6. The circadian period of the gene expression rhythm correlates with intracellular KaiC abundance in *S. elongatus*. Immunoblots of KaiC (*Upper*) detected in soluble protein extracts from strains that express KaiC in the following backgrounds (with period values, *n* = 12): 1, WT KaiC (25.02 ± 0.05 h); 2, no KaiC (arrhythmic); 3, WT KaiC from an ectopic site in aWT background (25.43 ± 0.07 h); 4, WT KaiC from an ectopic site in a *kaiC* null (23.33 ± 0.09 h). KaiA (*Lower*) was detected in the same samples as an internal loading control.

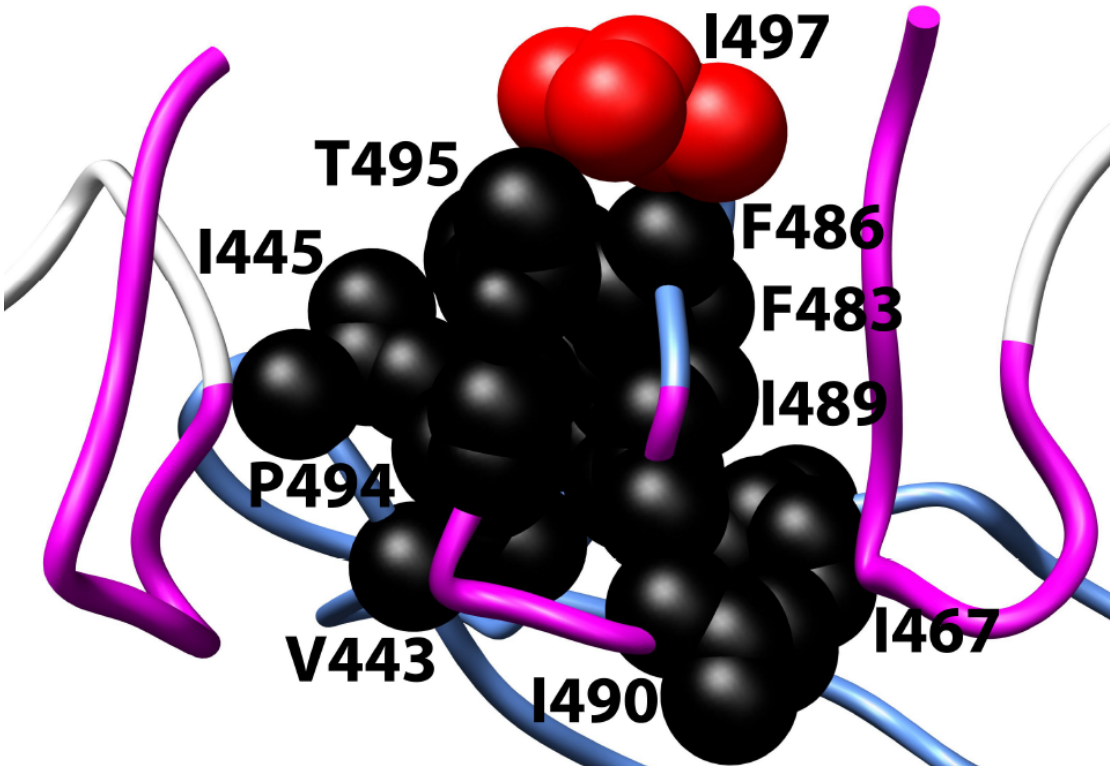

Figure 2.7. Putative hydrophobic cluster at the A-loop. This partial view of the KaiC hexamer (PDB ID code 1tf7) has the A-loops colored magenta. The atoms of the side chains of residues of the putative hydrophobic cluster are colored black, except for I497 which is red.

KaiA Activates KaiC Autophosphorylation by Directly Binding to the A-Loop + Exposed Tail Segment: The structure of the $KaiA^{C}$–$KaiC^{AL+tail}$ complex has been solved for the polypeptides from the thermophilic species *T. elongatus*, which are more stable under NMR conditions (Figure 2.1*A*) (64). To test whether this interaction is similar for the *S. elongatus* homologs, we carried out fluorescence anisotropy assays. The fluorescence anisotropy of a fluorophore is sensitive to the size of the macromolecule to which it is attached. An increase in the fluorescence anisotropy of a labeled peptide after the addition of unlabeled protein can reflect the formation of complexes. As seen in Figure 2.8*A*, *S. elongatus* KaiA and $KaiC^{AL+tail}$ interacted, although to a weaker extent than for the *T. elongatus* proteins (Figure 2.8*B*). As a further test of the KaiA–$KaiC^{AL+tail}$ interaction, we added *S. elongates* $KaiC^{AL+tail}$ peptides to a mixture of *S. elongatus* KaiA + KaiC. As shown by the green □ symbols in Figure 2.2*B*, $KaiC^{AL+tail}$ blocked KaiA-induced KaiC autophosphorylation, which suggests that this peptide was competing with KaiC for KaiA binding. The K_D values observed here were much higher than those reported for full-length KaiA and KaiC binding for *S. elongatus* (2.5 μM) (54) and *T. elongatus* (1.3 μM) (53). We think that these discrepancies arise from (*i*) the higher local concentration of $KaiC^{AL+tail}$ segments on a KaiC hexamer, and (*ii*) the possibility that the KaiA-KaiC interaction is more extensive than that observed in the NMR structure (Figure 2.1*A*).

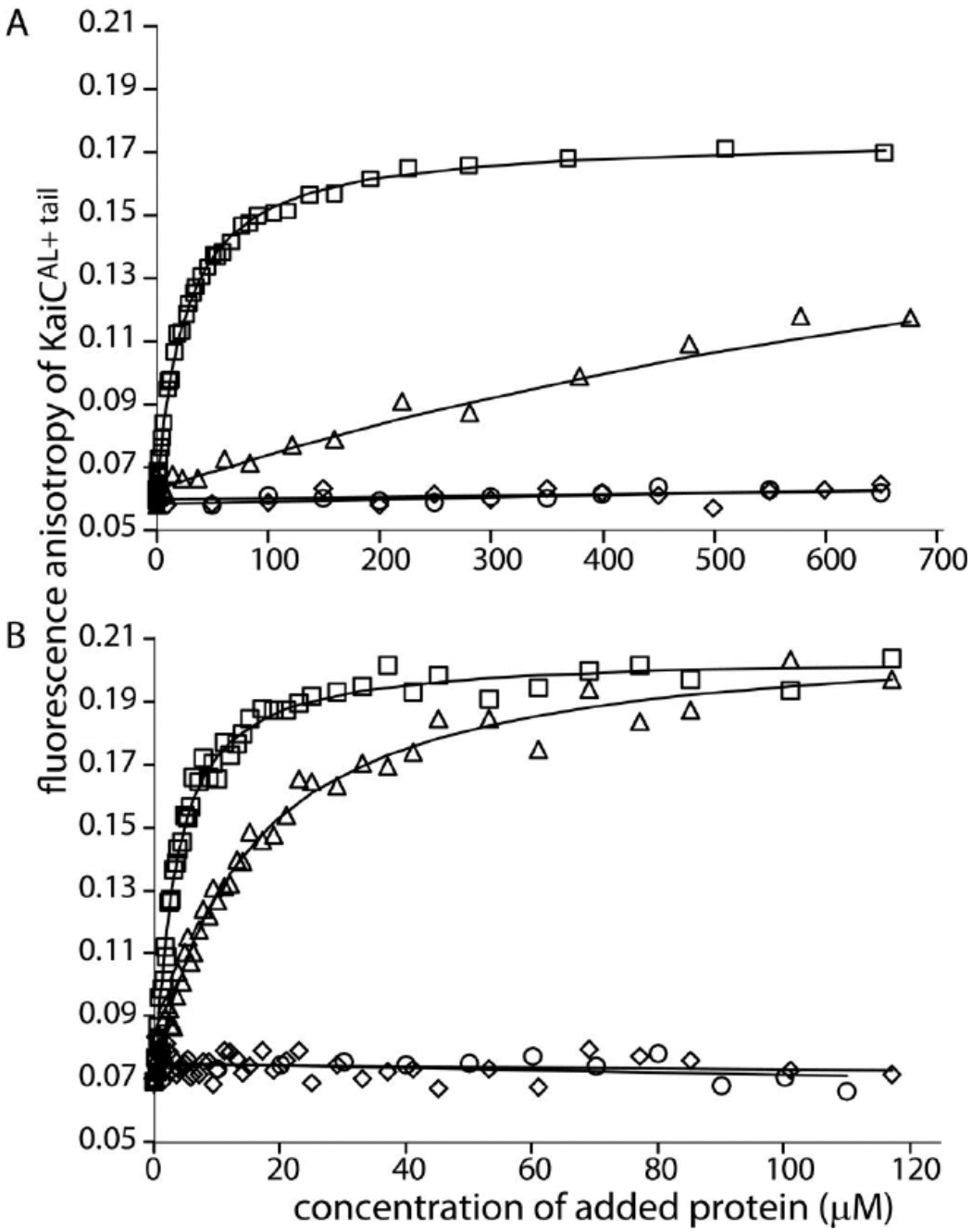

Figure 2.8. Fluorescence anisotropy of 6-iodoacetamidofluorescein-KaiC$^{AL+tail}$ as a function of concentration of KaiA for (A) *S. elongatus* and (B) *T. elongatus* proteins. △, KaiA; □, KaiA^{C}; ◇, KaiA^{N}; and ○, KaiB. Fits of the data yielded the following K_D values: 24.3 ± 0.5 μM, *S. elongatus* KaiA^{C} + KaiC$^{AL+tail}$; 170 ± 13 μM, S. *elongatus* KaiA + KaiC$^{AL+tail}$; 2.9 ± 0.2 μM, *T. elongatus* KaiA^{C} + KaiC$^{AL+tail}$; 14.4 ± 1.1 μM, *T. elongatus* KaiA + KaiC$^{AL+tail}$.

KaiA^N Attenuates the KaiA–KaiC Interaction and Is Important for KaiB Function: The N-terminal domain of KaiA, KaiA^N (residues 1–135), did not have any detectable affinity for KaiC$^{AL+tail}$, as gauged by fluorescence anisotropy experiments (Figure 2.8*A*, ◇). In addition, KaiA^N has no effect on the phosphorylation activity of KaiC (blue Δ, Figure 2.2) (52). Thus, the KaiA–KaiC interaction probably involves only weak, if any, interactions between KaiA^N and KaiC. However, full-length KaiA had a weaker affinity for KaiC$^{AL+tail}$ and enhanced KaiC autophosphorylation less than KaiA^C (Figure 2.8*A*, Δ and □; Figure 2.2*B*, pink ● and orange ◆). KaiA^N may attenuate the KaiA–KaiC interaction by affecting the structure of KaiA^C (64).

Apparently, KaiB inactivates KaiA in a complex with pSTKaiC, which allows the ensemble of KaiC proteins to autodephosphorylate (57, 58). We found that KaiA^N was important for this KaiB function. As shown in Figure 2.2*C*(orange ◆), KaiB had little effect on KaiAC-induced KaiC autophosphorylation. The observation that *S. elongatus* clones that express KaiAC instead of KaiA generate very weak 40-h rhythms (63) may be attributable to the inability of KaiB to inactivate KaiA^C in a complex with pST-KaiC.

KaiB Does Not Interact with the A-Loop: It has been shown that the kinetic ordering of phosphorylation at S431 and T432 for an ensemble of KaiC proteins in the presence of KaiA and KaiB is as follows (57, 58): ST → SpT → pSpT → pST → ST. Apparently, the oscillation depends on the fast buildup of pSpT-KaiC under stimulation by KaiA, followed by the slow inactivation of KaiA on pST-KaiC by KaiB (58). How KaiB achieves this

inactivation is unclear. Does KaiB interact directly with the A-loop? Fluorescence anisotropy experiments did not detect any binding between KaiB and $KaiC^{AL+tail}$ in either *S. elongatus* (Figure 2.8*A*, ○) or *T. elongatus* (Figure 2.8*B*, ○). Furthermore, KaiB by itself does not affect the phosphorylation of KaiC (Figure 2.2*D*) (17, 18, 24). Thus, KaiB likely affects the dynamic equilibrium of the A-loop indirectly by hindering its interaction with KaiA.

A-Loop Displacement Probably Moves ATP Closer to the Sites of Phosphorylation: To the best of our knowledge, the longest distance reported between the γ-phosphate group of ATP and an acceptor oxygen (serine hydroxyl group) is 5.3 Å in bovine cAMP-dependent protein kinase (116), which has since been suggested to be an overestimate (117). In the crystal structure of KaiC, the A-loops are buried, and the γ-phosphate group of ATP is 8.5 and 6.6 Å from the hydroxyl oxygens of S431 and T432, respectively (Figure 2.9), which explains why KaiC by itself had a low level of autokinase activity (Figure 2.2*A*, pink ●). It is, therefore, probable that A-loop exposure significantly repositions ATP closer to the sites of phosphorylation.

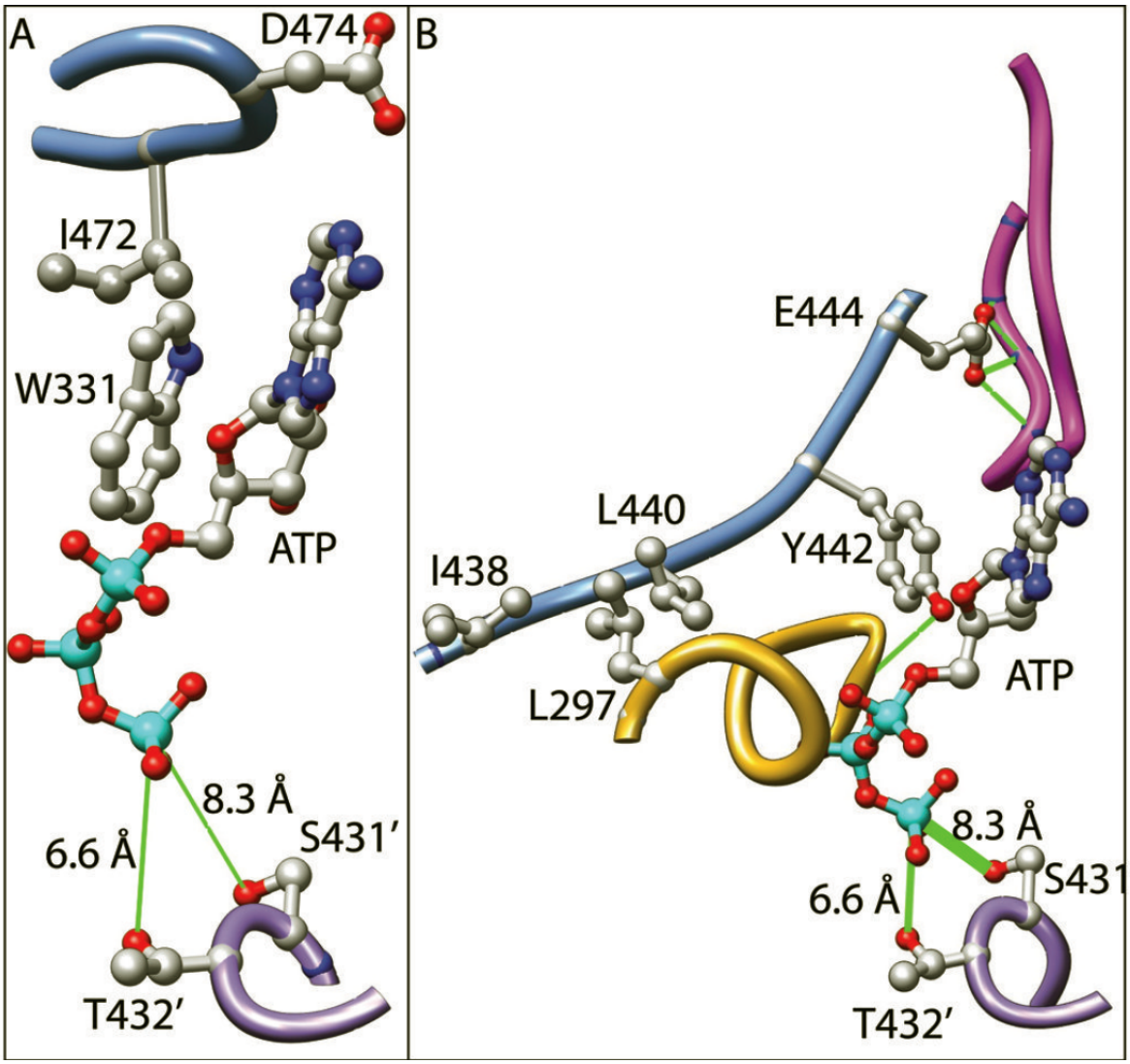

Figure 2.9. Putative interactions between ATP and KaiC as inferred from the X-ray crystal structure (50). Atoms are shown in ball and stick form (gray, carbon; red, oxygen; blue, nitrogen; cyan, phosphorus). The prime (') is used to denote residues from an adjacent subunit. Green lines indicate potential hydrogen bonds and the distances between the γ-phosphorus atom of ATP and the hydroxyl oxygens of S431' and T432'. (A) The structure of KaiC suggests that the position of ATP depends in part on interactions with I472, D474, and W331. (B) The position of the γ-phosphorus atom of ATP may be coupled to the interaction between E444 and the A-loop by way of the P-loop. The side chains of residues that may be part of this coupling are shown.

Putative interactions between the adenine base and residues I472 and D474, as inferred from the structure of KaiC (50), may help prevent ATP from approaching the sites of phosphorylation when A-loops are buried (Figure 2.9*A*). A D474A substitution attenuated the effect of KaiA and KaiB on KaiC phosphorylation kinetics (Figure 2.4, black ▲). The variant I472A (Figure 2.4, blue ■), however, was constitutively hyperphosphorylated, suggesting that ATP could more closely approach T432 and S431. We think that exposure of A-loops weakens the interaction between I472 and ATP. The ATP seems to also be stabilized by W331 because a W331A substitution also created a constitutively hyperphosphorylated KaiC variant (Figure 2.4, green ◆).

E444 apparently forms hydrogen bonds with A-loop residues and is part of a short segment (residues 438–444; blue ribbon, Figure 2.9*B*) that seems to interact with residues at or adjacent to the P-loop of the CII domain (residues 288–295; gold ribbon, Figure 2.9*B*). We anticipate that this chain of interactions couples the A-loop and γ-phosphate positions. A-loop exposure may disrupt its interaction with E444, causing the 438–444 segment to shift and thereby adjust the position of the P-loop, which, in turn, would move the γ-phosphate closer to the hydroxyl groups of T432 and S431. As a test of this hypothesis, we introduced an E444D substitution to disrupt the interactions with the A-loop. We found that this KaiC variant was constitutively hyperphosphorylated at ≈100% in the presence or absence of KaiA and KaiB (Figure 2.4, black ●). L297A and L440A substitutions, which would be a good test of the proposed coupling between the A-loop and ATP position, yielded insoluble KaiC variants.

Discussion

The A-loop Is the Master Switch That Determines KaiC Activities: KaiC is both an autokinase and an autophosphatase. The balance between these two activities is modulated by phosphoformdependent interactions between KaiC and KaiA and KaiB, and by the phosphoforms themselves, irrespective of KaiA and KaiB (57, 58). Here, we propose how KaiA and KaiB manipulate KaiC activity: the A-loop is the master switch and its dynamic equilibrium between buried and exposed states determines the levels of autokinase and autophosphatase activities (Figure 2.10); KaiA directly binds to and stabilizes the exposed state of the A-loop. KaiB does not directly interact with the A-loop; we predict that KaiB stabilizes the buried state of the A-loop indirectly by hindering its interaction with KaiA. Indeed, existing data suggest that KaiB does not affect KaiC activity directly but, rather, blocks KaiA stimulation of KaiC activity (Figure 2.2*C*, pink ●) (49, 51, 58).

A-Loop Burial/Exposure May Be Cooperative: Each buried A-loop in the KaiC hexamer appears to make several hydrogen bonds with the adjacent A-loops (Figure 2.3). The implication is that exposure of one A-loop destabilizes the buried states of the adjacent ones and raises the possibility that the burial and exposure of A-loops are cooperative processes. However, the rate of ST-KaiC → SpT-KaiC has a hyperbolic, rather than sigmoidal, dependence on KaiA concentration (58), which argues against cooperativity.

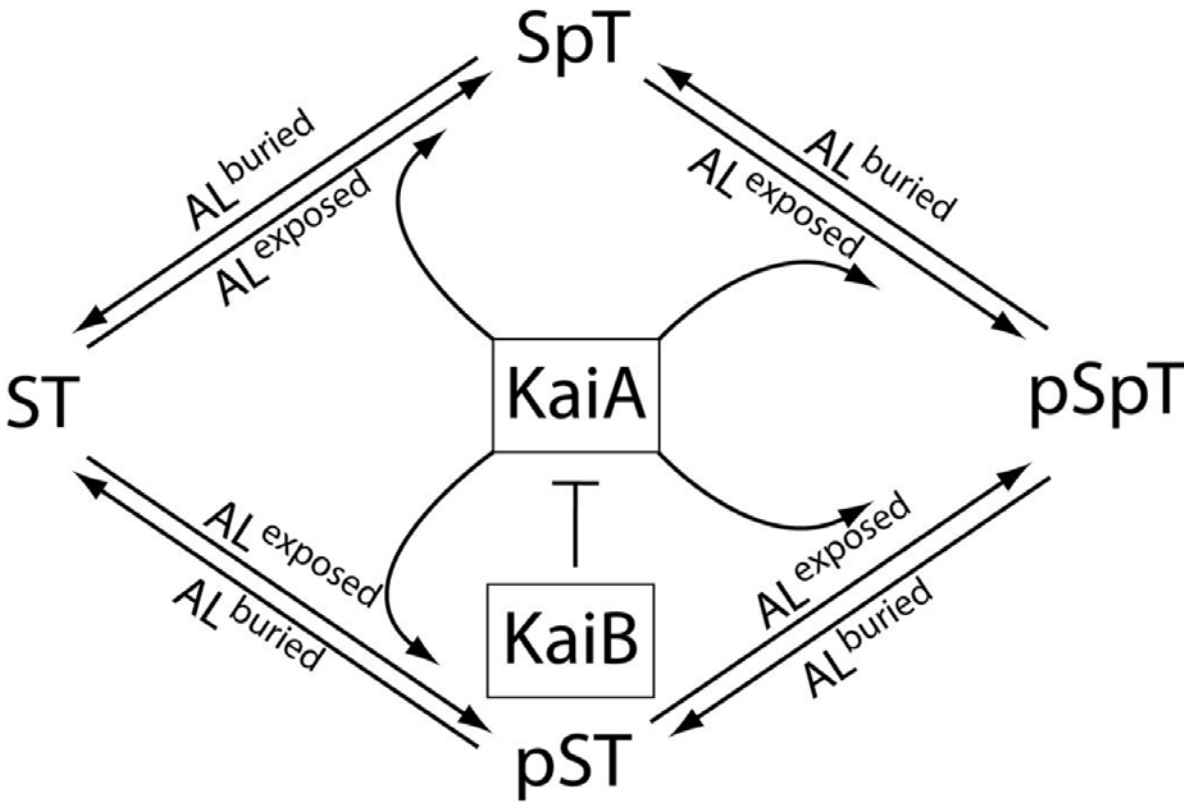

Figure 2.10. A-loop model. The steady-state level of phosphorylation of KaiC is determined by the dynamic equilibrium of the A-loops (AL). If the buried state of the A-loops (AL^{buried}) is favored, then the steady-state phosphorylation level of KaiC is lowered; if the exposed state of the A-loops ($AL^{exposed}$) is favored, the steady-state phosphorylation level is increased. KaiA stabilizes $AL^{exposed}$, whereas KaiB prevents KaiA from doing so by immobilizing it on the pST-KaiC phosphoform.

We suggest that when a single KaiA binds to a KaiC hexamer (53, 71), it stabilizes the exposed state of the two A-loops to which it binds (Figure 2.1*A*), shifting the dynamic equilibrium of the four remaining A-loops toward the exposed state. This scenario could produce the hyperbolic dependence for ST-KaiC → SpT-KaiC, as demonstrated by Rust *et al.* (58).

KaiB Inactivates KaiA by Hindering Its Interaction with A-Loops: Our data suggest that KaiB induces KaiC autodephosphorylation by hindering the interaction between KaiA and A-loops. There are two possible ways by which KaiB inactivates KaiA: either KaiB interacts directly with KaiA and thereby immobilizes it on pST-KaiC or the KaiB–pST-KaiC interaction affects the pSTKaiC–KaiA interaction, such that KaiA is trapped by pST-KaiC. Our observation that $KaiA^{C}$-induced KaiC autophosphorylation was immune to KaiB suggests that $KaiA^{N}$ is necessary for KaiB activity. Whether the KaiA–KaiB interaction is direct or indirect is still unclear.

Downstream Effects of A-Loop Exposure: The ATP bound in the CII domain is too far from S431 and T432 for phosphoryl transfer (Figure 2.9) and explains why KaiC by itself had a greater autophosphatase than autokinase activity (Figure 2.2*A*, pink ●). Our data suggest that I472 helps prevent ATP from approaching the sites of phosphorylation (Figure 2.4), probably through direct interactions with ATP. A-loop exposure likely repositions the ATP closer to the sites of phosphorylation through weakening the I472–ATP interaction, and disrupting the E444–A-loop interaction, thereby shifting the P-loop.

Differences Between in Vitro and in Vivo Oscillations: The dominant negative effects of KaiC487 and KaiC497 expression in a WT background suggest that these KaiC variants, which are unable to restore rhythmicity to a *kaiC*-null strain, either participate in hexamer formation with WT KaiC monomers, rendering them nonfunctional, or compete with WT subunits for interaction with the circadian output pathways. The former possibility, supported experimentally in Figure 2.11, contrasts with *in vitro* data in which oscillation of WT KaiC phosphorylation is unaffected by addition of nonshuffling KaiC variants, KaiC-AA (62). These data suggest that nascent proteins form mixed hexamers when both WT and mutant KaiC variants are coexpressed in the cell, such that protein dynamics of synthesis and degradation may play important roles *in vivo* that are not required for the basic oscillation *in vitro*.

Suggested Mechanism for the KaiA-Lack Species: Some cyanobacteria such as *Prochlorococcus marinus* PCC 9511 possess *kaiB* and *kaiC* genes, but lack *kaiA* homologs (29, 118), indicating that a potential KaiC-based circadian oscillator would operate by a different mechanism than that described here. *P. marinus* PCC 9511 displays robust 24-h rhythms of DNA replication and gene expression when it is incubated in alternating 12 h light:12 h dark periods, but the rhythms rapidly damp under continuous light (28).

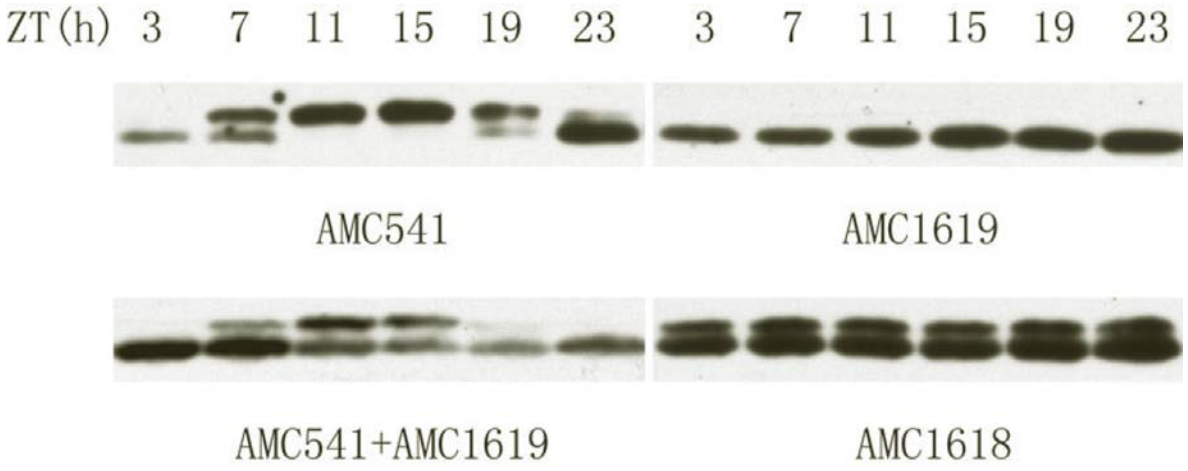

Figure 2.11. Expression of *kaiC-AA* in a WT background stops the phosphorylation cycle of KaiC. A WT strain (AMC541) and strains that carry *kaiC-AA* expressed in either the WT background (AMC1618) or Δ*kaiC* background (AMC1619) were subjected to two synchronizing light/dark cycles before being released in constant light. Samples were taken every 4 h and analyzed by immunoblot by using a KaiC antiserum. The KaiC phosphorylation state oscillates in AMC541, and no rhythm is observed with AMC1619, which expresses only the nonoscillating KaiC-AA. Oscillation of KaiC phosphorylation is still observable when AMC541 and AMC1619 samples are equally mixed *in vitro*, even though nonphosphorylated KaiC and KaiC-AA run as a single band. This figure demonstrates that oscillating phospho-KaiC would be observable above the background of nonphosphorylated KaiC in AMC1618 if WT KaiC continued to oscillate in the presence of KaiC-AA. AMC1618 shows no rhythm in KaiC phosphorylation, indicating that expression of KaiC-AA in the WT background has a dominant negative effect on the central oscillator.

These data suggest that *P. marinus* PCC 9511 uses an hourglass clock mechanism that requires daily resetting rather than a self-sustained circadian oscillator. We predict that KaiC phosphorylation is part of the timing mechanism in *P. marinus* PCC 9511 but that KaiC ensembles have a high steady-state level of phosphorylation as a result of a dynamic equilibrium that favors the exposed state of the A-loop. KaiB may stabilize the buried state of the A-loops for hyperphosphorylated KaiC and thereby stimulate autodephosphorylation. In this scenario, some physiological consequence of the diurnal cycle such as protein synthesis and degradation would provide a daily resetting of the phosphorylation state of KaiC.

The Discrepancy Between KaiA Stimulation of WT pST-KaiC and DT-KaiC: It was recently proposed by Nishiwaki *et al.* (57) that the autokinase and autophosphatase activities of KaiC are strongly influenced by its state of phosphorylation, irrespective of KaiA and KaiB. Particularly, it was suggested that KaiC switches from an autokinase to an autophosphatase when phosphorylated at S431. Indeed, a KaiC variant containing the S431D phosphomimetic substitution, DT-KaiC, was mostly dephosphorylated at T432 even in the presence of KaiA. In our hands, DT-KaiC is ≈40% phosphorylated in the presence of KaiA (Figure 2.12*B*, ○), which is qualitatively consistent with the previous study (57). However, the DT-KaiC487 variant is 90% phosphorylated regardless of KaiA (Figure 2.12*A* and *B*, ∆), indicating that more than S431 phosphorylation is involved in the switch.

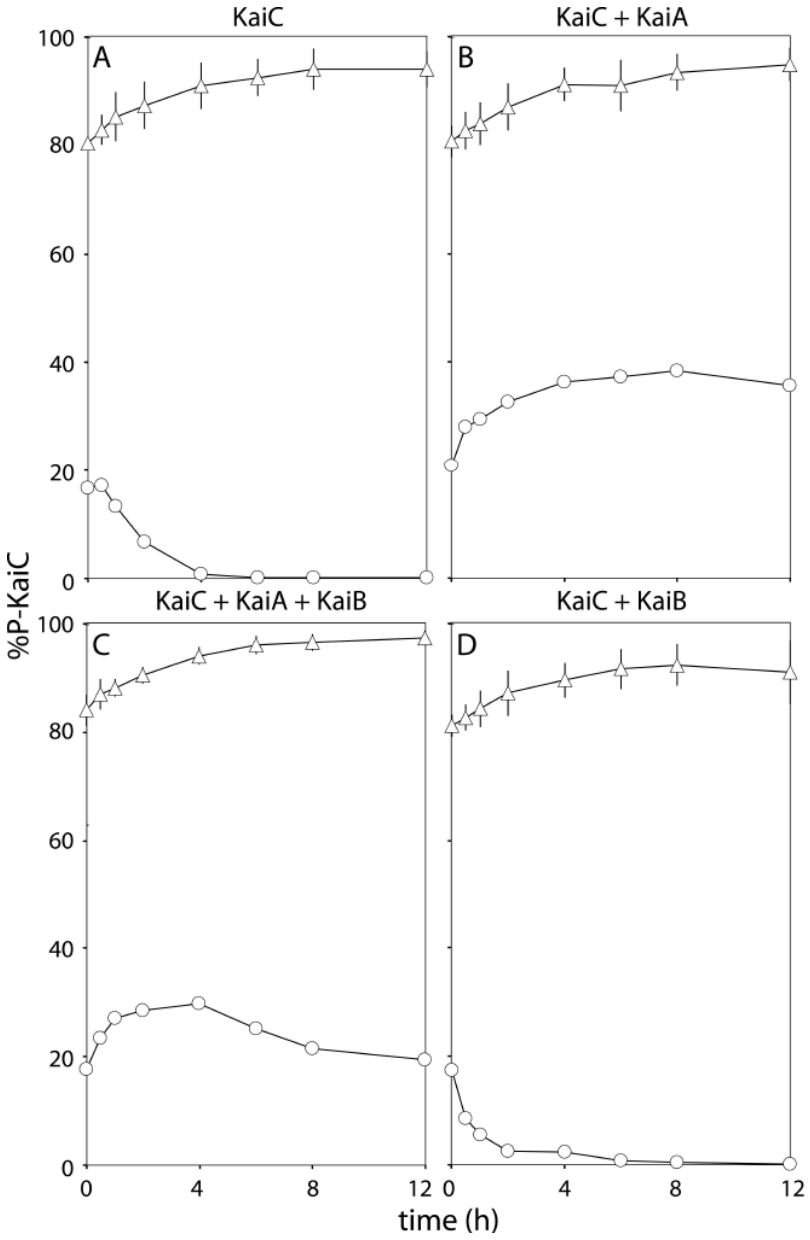

Figure 2.12. Phosphorylation of DT-KaiC (○) and DT-KaiC487 (△) variants for (A) KaiC alone, (B) KaiC + KaiA, (C) KaiC + KaiA + KaiB, and (D) KaiC + KaiB. For DT-KaiC487 each data point is the mean of two independent experiments; for DT-KaiC experiments were carried out only once. Solid lines connect data points for visual clarity. Images of the stained polyacrylamide gels are shown in Appendix A. Assignment of phosphorylated and unphosphorylated KaiC bands resolved by PAGE was determined from lambda phosphatase assays (Appendix A). All KaiC samples were fresh or from freshly frozen samples.

These observations in the context of our model suggest that the S431D substitution, and presumably pSpT and pST phosphoforms, make it more difficult for KaiA to stabilize the exposed state of the A-loop. However, Rust *et al.* (58) have evidence by using WT KaiC that the rate of pST-KaiC → pSpT-KaiC can be the fastest of the autophosphorylation reactions in the presence of KaiA and that the reaction is necessary for the generation of a self-sustained rhythm of KaiC phosphorylation. This observation by Rust *et al.* implies that stabilization of the exposed state of the A-loop by active KaiA is not inhibited by the pST-KaiC phosphoform. The discrepancy between KaiA stimulation of WT pST-KaiC and DT-KaiC could be attributable to an imperfect mimic of phosphoserine by aspartate at position 431.

Defects on thePphospho-Mimicking Mutants of KaiC: It was also observed by Nishiwaki *et al.* (57) that the T432E phosphomimetic variant of KaiC, SE-KaiC, incubated in isolation from the other Kai proteins, remains hyperphosphorylated at S431 with no sign of dephosphorylation after 20 h. Our model states that the dynamic equilibrium of the A-loop favors the buried state unless KaiA stabilizes the exposed state. Hence, pSE-KaiC should, according to our model, autodephosphorylate. Recently, Rust *et al.* (58) have presented evidence that WT KaiC indeed does autodephosphorylate from pSpT → SpT at a finite rate, which is in contrast to the work by Nishiwaki *et al.* (57). Indeed, the model of the oscillator presented by Rust *et al.* (58) fails to generate a sustained rhythm of KaiC phosphorylation when the rate of pSpT → SpT is set to zero. These discrepancies with the

nondephosphorylating pSE-KaiC variant suggest that glutamate is not a perfect mimic of phosphothreonine in this instance.

Materials and methods

Cloning and Protein Purification of KaiA and KaiB: The genes encoding KaiA and KaiB from *S. elongatus* and *T. elongatus* were cloned into the pET32a+ vector (Novagen) between BamHI and NcoI sites, the resulting plasmids were used to transform *Escherichia coli* BL21(DE3), and sequences were confirmed (Gene Technologies Laboratory, Texas A&M University) (see Table 2.1). Transformed *E. coli* cultures in log phase in LB at 37°C were induced to overexpress recombinant KaiA or KaiB with 1 mM isopropyl β-D-thiogalactopyranoside (Calbiochem). Cells were harvested after 6 h, and pellets were resuspended in 50 mM NaCl, 20 mM Tris-HCl, pH 7.0. Cell suspensions were passed twice through a chilled French press cell, and lysates were clarified by centrifugation at 20,000 × *g* for 60 min at 4°C. Tagged proteins were isolated on a Ni-charged chelating column. Proteases and ATPases were removed by anion-exchange chromatography (buffer A: 20 mM NaCl, 20 mM Tris-HCl, pH 7.0; buffer B: 1 M NaCl, 20 mM Tris-HCl, pH 7.0; gradient: 0–80% buffer B over 16 times of 5-mL column volumes). The tagged proteins were dialyzed against enterokinase (EK) cleavage buffer (Novagen), and reactions with 1 unit/ml EK were carried out at room temperature in 15–20 ml of ≈10 μM KaiA or ≈40 μM KaiB. Complete removal of the thioredoxin–His_6 tag from KaiA took 36 h; KaiB cleavage reactions were terminated after 16 h to prevent protein precipitation.

Table 2.1. Details on Recombinant Kai Proteins.

Protein	Vector	Insertion Site	Non-native residues[a]	*E. coli strain*
S.e. KaiA	pET32a+	EcoRV – HindIII	AMADI	BL21(DE3)
S.e. KaiA^{C}	pET32a+	NcoI – EcoRV	AMA	BL21(DE3)
S.e. KaiA^{N}	pET32a+	BglII – EcoRV	M	BL21(DE3)
S.e. KaiB	pET32a+	NcoI – HindIII	A[b]	BL21(DE3)
S.e. KaiC	pGEX-6P-2	BamHI – XhoI	GPLGS	DH5α
S.e. KaiC$^{AL+tail}$	pET32a+	NcoI – BamHI	AMC[c]	BL21(DE3)
T.e. KaiA	pET32a+	NcoI – HindIII	AMA	BL21(DE3)
T.e. KaiA^{C}	pET32a+	NcoI – EcoRI	AMA	BL21(DE3)
T.e. KaiA^{N}	pET32a+	NcoI – HindIII	AMA	BL21(DE3)
T.e. KaiB	pET32a+	NcoI – BamHI	AMA	BL21(DE3)
T.e. KaiC$^{AL+tail}$	pET32a+	NcoI – BamHI	AMC[c]	BL21(DE3)

[a]Non-native residues are at the N-terminus of the recombinant proteins. The C-termini all end with native residues.
[b]Originally AMA, but changed to AMS using site-directed mutagenesis. The first two residues of KaiB are MS, and the only additional non-native residue here is A.
[c]Originally AMA, but changed to AMC using site-directed mutagenesis.

KaiA and KaiB were separated from thioredoxin–His_6 tags and uncut proteins by a second passage through a Ni-charged chelating column. All proteins were analyzed for purity by SDS/PAGE and dialyzed against autophosphorylation assay buffer (20 mM Tris-HCl, 150 mM NaCl, 0.5 mM EDTA, 5 mM $MgCl_2$, 1 mM ATP, pH 8.0). Protein solutions were concentrated to ≈20 μM for KaiA and at least 60 μM for KaiB. Above ≈20 μM, KaiA started to precipitate from solution. The proteins were then passed through a sterile 0.2-μm filter and stored at -80°C. KaiA and KaiB did not lose function after repeated cycles of freezing and thawing. One liter of induced *E. coli* culture yields ≈220 nmol of KaiA and 830 nmol of KaiB after purification. Protein concentrations for KaiA, KaiB, and KaiC were determined by using Coomassie Plus–The Better Method Assay Reagent (Pierce).

Cloning and Protein Purification of KaiC: Genes encoding KaiC from *S. elongatus* and *T. elongatus* were cloned into the pGEX-6P-2 vector (GE Healthcare) between BamHI and XhoI sites, and the resulting plasmids were used to transform *E. coli* DH5α, which produced more soluble recombinant KaiC than did *E. coli* BL21(DE3). KaiC truncation variants KaiC487, KaiC494, KaiC495, KaiC496, and KaiC497 were constructed by inserting stop codons to terminate translation after residues 487, 494, 495, 496, and 497, respectively, using the QuikChange method from Stratagene. Protein purification, analysis, and storage were performed as for KaiA and KaiB with the following exceptions: transformed cultures of *E. coli* DH5α in LB were continuously shaken in 1-liter flasks at 30°C for 2.5 days. Cell pellets were resuspended in 50 mM Tris-HCl at pH 7.3 with 150 mM NaCl, 5 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT and 5 mM ATP. Tagged KaiC proteins

were separated from the supernatant fraction on a GSTrap HP column (GE Healthcare). Prescission Protease (GE Healthcare) was used (1 unit/ml in 12 ml) to cut the GST tag from 5 μM tagged KaiC. KaiC was separated from the cleaved GST tag, tagged KaiC, and protease by passage a second time through a GSTrap HP column. Cycles of thawing and freezing did not affect KaiC activity. On average, 2 liters of *E. coli* culture yielded 75 nmol of purified KaiC and its variants. Table 2.1 lists the recombinant Kai proteins, their nonnative residues, vectors, cloning insertion sites, and strains of *E. coli* used for their overexpression.

Preparation of the KaiC$^{AL+tail}$ Peptide Labeled with Fluorescein: The gene fragments encoding KaiC$^{AL+tail}$ peptides from *S. elongatus* (residues 488–519) and *T. elongatus* (488–518) and an added N terminal Cys codon were cloned into the pET32a+ vector between NcoI and BamHI sites, verified by sequencing, and introduced into *E. coli* BL21(DE3). Protein overexpression and purification were similar to methods for KaiA and KaiB. The tagged peptide was labeled with 6-iodoacetamidofluorescein according to the protocol of the manufacturer (Molecular Probes, Inc.). Samples were buffer exchanged into 50 mM NaCl, 20 mM Tris-HCl, pH 7.4, and the thioredoxin-His_6 tag was removed by using enterokinase. Cleavage by this enzyme results in three additional non-KaiC-derived residues (AMC) at the N terminus of the KaiC$^{AL+tail}$ peptides. Thus, the final sequences of *S. elongatus* and *T. elongatus* KaiC$^{AL+tail}$ peptides were ‘‘AMCRIISGSPTRITVDEKSELSRIVRGVQEKGPES’’ and ‘‘AMCGIISGTPTRISVDEKTELARIAKGMQDLESE’’, respectively. Labeled peptides

were resolved from unlabeled peptides and free fluorophore by reversed phase chromatography, and quantification was achieved by measuring the UV absorbance of fluorescein. Peptide identity and purity were confirmed by matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) spectroscopy. Lyophilized fluorescein-labeled peptide was dissolved into autophosphorylation assay buffer without ATP.

KaiC Phosphorylation Kinetics: Phosphorylation assays in sterile 1 ml tubes in a 30°C water bath included KaiA and KaiB at 1.5 μM and 4.5 μM final concentrationsin the autophosphorylation assay buffer. A ‘‘time zero’’ sample was taken immediately after KaiC (3.4 μM) was added. Periodically, 38-μl aliquots were removed and denatured at 60°C for 15 min with 6 μl of SDS/PAGE gelloading dye (100 mM Tris-HCl at pH 6.8 with 4% SDS, 0.2% bromophenol blue, 20% glycerol, and 400 mM β-mercaptoethanol). A sample (10 μl) of each was loaded onto 9 × 10 cm SDS polyacrylamide gels (4% stacking, 6.5% running) with 15 wells (10 × 3 × 0.75 mm). The stacking gel extended only 2 mm from the bottom of the wells to the running gel. Maximum resolution of the KaiC phosphoforms by PAGE by using our setup was achieved by (*i*) immersing the gel box in ice, (*ii*) filling the gel box with buffer (700 ml), (*iii*) removing all bubbles along the bottom of the gel plates, (*iv*) loading the first and last two lanes with loading buffer only, and (*iv*) running the experiments at constant voltage (140 V). Gels were stained with Coomassie brilliant blue, and the percentage of KaiC that was phosphorylated in each lane was determined by densitometric analysis by using Image J (National Institutes of Health) and PeakFit (SeaSolve Software, Inc.). Images of the stained gels are provided in Appendeix A.

Fluorescence Anisotropy Experiments: Fluorescence anisotropy experiments on the fluorescein-labeled KaiC$^{AL+tail}$ peptide were carried out with a PC1 photon counting spectrofluorometer (ISS) with a sample temperature of 25°C. The excitation wavelength was set to 487 nm, and orthogonal emission intensities, $I_{\parallel}$ and $I_{\perp}$, were measured at 523 nm; fluorescence anisotropy, r, was determined using the equation: $r = (I_{\parallel} - I_{\perp})/(I_{\parallel} + I_{\perp})$. The KaiC$^{AL+tail}$ peptide concentration in an initial volume of 1.8 ml was 100 nM. Fluorescence anisotropies were measured as a function of the concentrations of KaiA, KaiA variants, and KaiB. Dissociation constants were calculated by fitting anisotropy data using DYNAFIT (119) with a simple 1:1 binding model.

Cyanobacterial Strains, Culture Conditions, and Bioluminescence Assays: WT *S. elongatus* PCC 7942 and its derivatives were propagated in BG-11 medium with appropriate antibiotics at 30°C, as described previously (21, 120). Ectopic alleles of various KaiC constructs, described in Table 2.2, were introduced into neutral site I of the *S. elongatus* chromosome (21, 120). Bioluminescence assays of these strains were performed on a Packard TopCount scintillation counter (PerkinElmer Life Sciences) according to a previous protocol (21).

Table 2.2. Constructs and cyanobacterial strains used in this study.

Plasmid	Characteristics		Source or reference
pAM2302	WT *kaiC* driven by *kaiBC* promoter		(115)
pAM2596	WT *kaiC* driven by an IPTG-inducible *trc* promoter		(49)
pAM2969	Km^R-Ω cassette in the BamHI site of *kaiA*		(115)
pAM3868	kaiC487 (based on pAM2302)		This study
pAM3871	kaiC487 (based on pAM2596)		This study
pAM3910	kaiC497 (based on pAM2302)		This study
pAM4047	kaiC497 (based on pAM2596)		This study
S. elongatus strain*	Genetic background	Ectopic kai plasmid	Source or reference
AMC541	WT	none	(9)
AMC704	*kaiC* deletion	none	(115)
AMC705	*kaiB* and *kaiC* deletion	none	(115)
AMC1617	*kaiA* insertion and *kaiC* deletion	none	This study
AMC1276	WT	pAM2302	(115)
AMC1274	*kaiC* deletion	pAM2302	(115)
AMC1620	WT	pAM3868	This study
AMC1621	*kaiC deletion*	pAM3868	This study
AMC1622	WT	pAM3910	This study
AMC1623	*kaiC* deletion	pAM3910	This study
AMC1624	*kaiB* and *kaiC* deletion	pAM2302	This study
AMC1625	*kaiB* and *kaiC* deletion	pAM4047	This study
AMC1626	*kaiA* insertion and *kaiC* deletion	pAM3871	This study

* All *S. elongatus* strains in this study carry a firefly luciferase gene (*luc*) driven by the *kaiBC* promoter.

CHAPTER III

COMPUTATIONAL AND EMPIRICAL TRANS-HYDROGEN BOND DEUTERIUM ISOTOPE SHIFTS SUGGEST THAT N1–N3 A:U HYDROGEN BONDS OF RNA ARE SHORTER THAN THOSE OF A:T HYDROGEN BONDS OF DNA*

Introduction

It is well established that there are significant structural differences between RNA and DNA (121). However, it is still arguable whether there are any measurable differences in their hydrogen-bond lengths or strengths. Even a small difference in hydrogen-bond strength can have a significant cumulative effect. Comparisons of the highest resolved X-ray crystal structures of RNA and DNA, however, revealed no differences within the experimental scatter of N1–N3 hydrogen-bond lengths of Watson–Crick A:U and A:T base pairs (106). It was shown in non-base paired mononucleotides that the difference in pK_a values of rA and rU is less than that for dA and dT, from which it was inferred that RNA hydrogen bonds can be stronger than those of DNA (122). However, it has been shown that pK_a values of nucleobases can shift significantly upon base pairing (123-125).

---

* Reprinted with permission from "Computational and empirical trans-hydrogen bond deuterium isotope shifts suggest that N1-N3 A:U hydrogen bonds of RNA are shorter than those of A:T hydrogen bonds of DNA." by Kim, Y.-I., Manalo M.N., Perez L.M., LiWang, A.C. (2006) *J Biomol NMR* **34**, 229-236. Copyright 2006 Kluwer Academic Publishers. All computational calculations were performed by Dr. Marlon N. Manalo in the laboratory of Dr. Andy C. LiWang (Texas A&M University, Department of Biochemistry and Biophysics).

Substitution of a proton for a deuteron changes the average bond lengths at and near the site of substitution, which in turn perturbs the nuclear shielding of nearby spins (126). The resulting change in chemical shift is called the deuterium isotope shift and is defined for ^{13}C as $^{n}\Delta^{13}C = \delta^{13}C(^{1}H) - \delta^{13}C(^{2}H)$, where n is the number of bonds separating the ^{13}C and the site of substitution, and $\delta^{13}C(^{1}H)$ and $\delta^{13}C(^{2}H)$ are the ^{13}C isotropic chemical shifts of the two isotopomers. Studies of small molecules have shown that $^{2}\Delta^{13}C$ are sensitive to hydrogen-bond strength (104). For weak hydrogen bonds, the normal-mode stretching vibrational potentials for N–H (or O–H) bonds become more anharmonic with increasing hydrogen-bond strength, which are reflected in larger $^{2}\Delta^{13}C$ magnitudes.

In DNA, deuterium substitution of the imino H3 of thymine produces a downfield shift of the $^{13}C2$ resonance of the Watson–Crick base-paired adenine (103), which is a through-space or trans-hydrogen bond isotope effect. In this case, the deuterium isotope shift is $^{2h}\Delta^{13}C2 = \delta^{13}C2(^{1}H3) - \delta^{13}C2(^{2}H3)$. Recently, $^{2h}\Delta^{13}C2$ values of RNA Watson–Crick A:U base pairs were found to be more negative than those of A:T base pairs of DNA, which suggests that RNA N1–N3 hydrogen bonds are shorter than those of DNA (106). However, investigators of a density functional theory (DFT) calculation study proposed that $^{2h}\Delta^{13}C2$ of DNA are significantly affected by the electron-donating character of the C7 methyl group of thymine (109), and that, importantly, the experimental differences observed between $^{2h}\Delta^{13}C2$ values in RNA and DNA A:U and A:T base pairs (106) do not reflect differences in hydrogen bond lengths but merely the absence or presence of the C7 methyl group. Indeed, the utility of $^{2h}\Delta^{13}C2$ as a gauge of hydrogen-bond length is inferred

from small molecule studies and is an untested presumption for nucleic acids. Independent lines of evidence will be needed to establish $^{2h}\Delta^{13}C2$ as a measure of hydrogen-bond length for RNA and DNA.

First, we will present DFT calculations of isolated A:U and A:T base pairs that suggest that $^{2h}\Delta^{13}C2$ is sensitive to hydrogen-bond length. Then, we will show using both computation and experimentation, that the C7 methyl group of thymine does not measurably influence $^{2h}\Delta^{13}C2$. We will finally show that $^{2h}\Delta^{13}C2$ has a context dependence that is similar in RNA and DNA.

Results

Studies have shown that $^{2}\Delta^{13}C2$ is sensitive to hydrogen-bond strength in small molecules (104, 126). We showed previously that deuterium substitution at the uracil or thymine imino H3 site results in trans-hydrogen bond isotope shift of $^{13}C2$ of the Watson–Crick base-paired adenine, $^{2h}\Delta^{13}C2$ (103). The notion that $^{2h}\Delta^{13}C2$ is measure of hydrogen-bond length has been inferred from small molecule studies, but was never supported by independent lines of evidence for RNA and DNA. Thus, in an effort to estimate the sensitivity of $^{2h}\Delta^{13}C2$ to hydrogen-bond length in A:U/T Watson–Crick base pairs, we carried out a series of DFT calculations using the approach described by Abilgaard *et al.* (104). The $^{2h}\Delta^{13}C2$ was calculated using Equation 1: $^{2h}\Delta^{13}C2 = -d\sigma/dR_{NH} \times \Delta R$, where $d\sigma/dR_{NH}$ characterizes the conveyance of the isotope effect through the hydrogen bond.

Shown in Figure 3.1 is plot of $^{2h}\Delta^{13}C2$ calculated for isolated A:U and A:T base pairs as function of the hydrogen-bond distance between the N1 of adenine and the N3 of uracil or thymine. As the N1–N3 hydrogen-bond distance decreases, the calculated $^{2h}\Delta^{13}C2$ becomes increasingly negative. Thus, our calculations support the notion that $^{2h}\Delta^{13}C2$ is sensitive to hydrogen-bond length. It can also be seen that isolated A:U and A:T base pairs have very similar $^{2h}\Delta^{13}C2$ values (within 1 ppb) at given N1–N3 distance, which suggests that the empirical difference between RNA and DNA $^{2h}\Delta^{13}C2$ values (106) is not due to the chemical difference between uracil and thymine, but shorter hydrogen bonds in double-stranded RNA.

In order to further assess the effect of the chemical difference between uracil and thymine on $^{2h}\Delta^{13}C2$ values, we measured $^{2h}\Delta^{13}C2$ on chemically modified RNA and DNA, in which the uridine residues of RNA were substituted with 5-methyl uridine and the deoxythymidine residues of DNA were replaced with deoxyuridine. We will refer to the chemically modified RNA and DNA as RNA^{5mU} and DNA^{dU}. Shown in Figure 3.2 are small regions of ^{1}H, ^{13}C TROSY-HSQC spectra of the RNA^{5mU} dodecamer $r(CGAAAAU^{5m}U^{5m}U^{5m}U^{5m}CG)_2$ and the isosequential DNA^{dU} dodecamer $d(CGAAAAUUUUCG)_2$ dissolved in 50% H_2O, 50% D_2O buffer. Due to slow exchange of the pyrimidine H3 with the solvent, there are two peaks for every adenine $^{1}H2$, $^{13}C2$ pair. The peak with the higher $^{13}C2$ frequency corresponds to the $^{2}H3$ isotopomer, and the lower frequency $^{13}C2$ peak arises from the $^{1}H3$ isotopomer. The $^{2h}\Delta^{13}C2$ values were determined from the difference in the peak positions: $^{2h}\Delta^{13}C2 = \delta^{13}C2(^{1}H3) - \delta^{13}C2(^{2}H3)$.

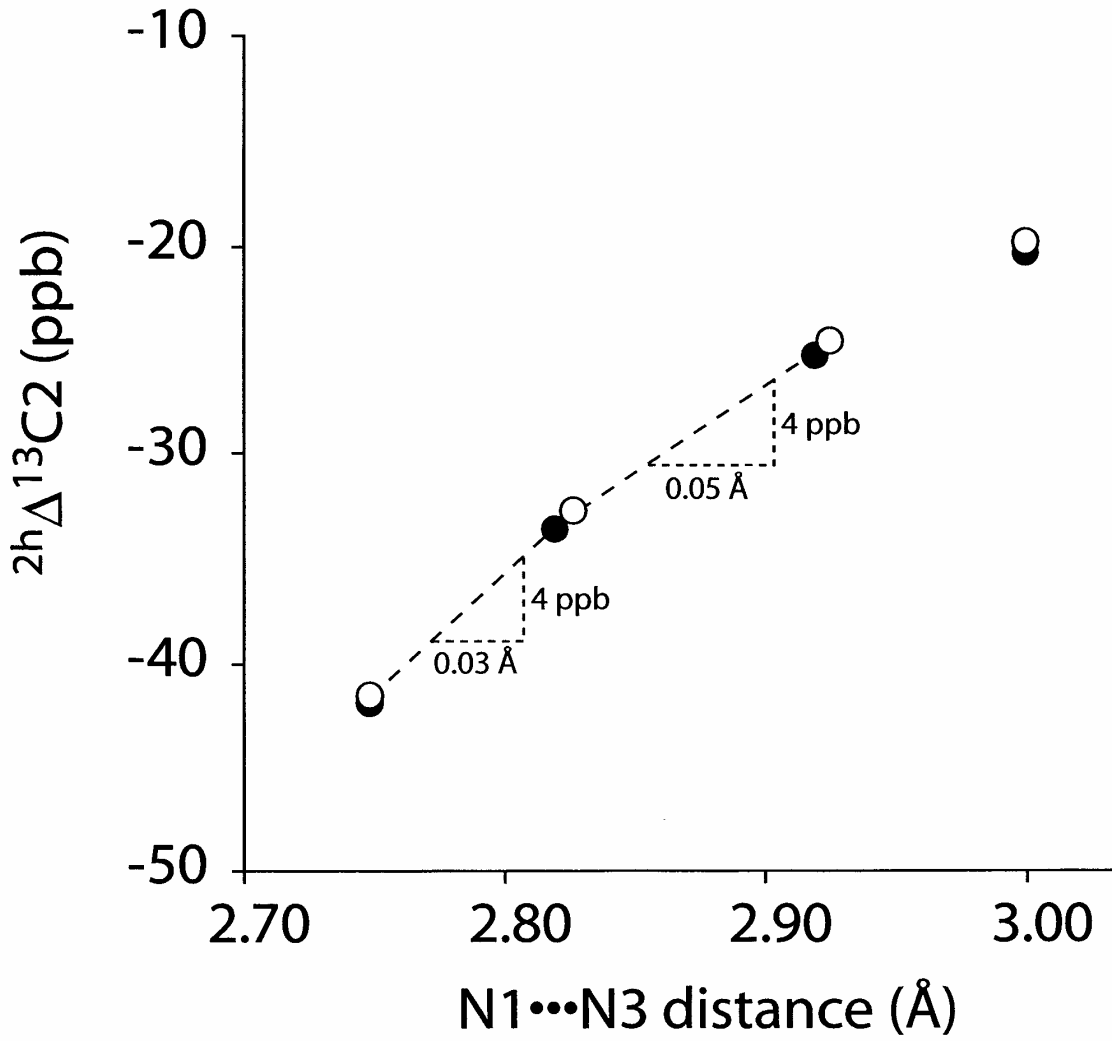

Figure 3.1. The calculated dependence of $^{2h}\Delta^{13}C2$ on the distance between the N1 of adenine and N3 of uracil (solid circle) and thymine (open circle). From the points calculated near the fully optimized N1–N3 distance of 2.82 Å, a change in $^{2h}\Delta^{13}C2$ by 4 ± 3 ppb corresponds to a 0.04 ± 0.03 Å change in N1–N3.

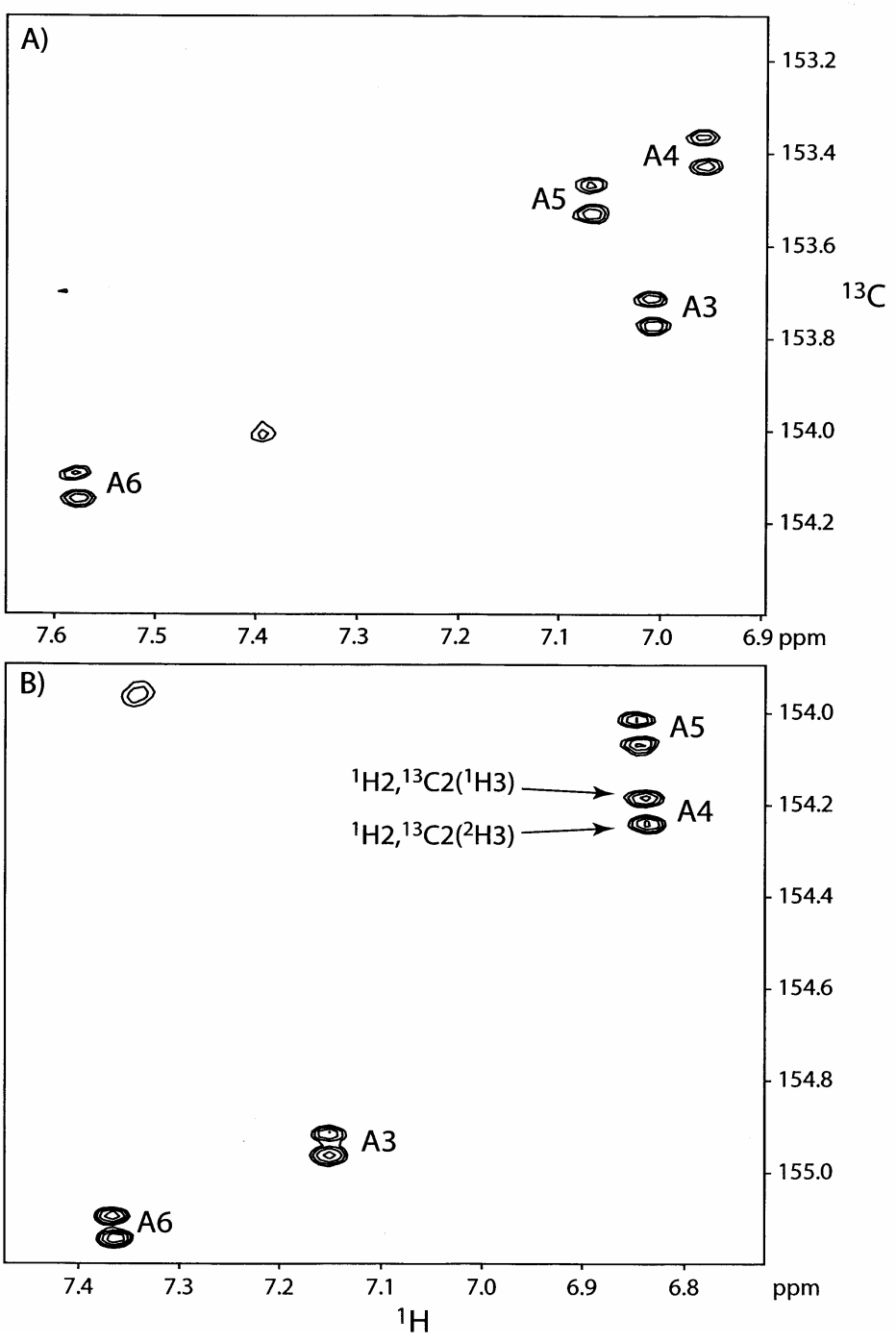

Figure 3.2. Small regions of ^{1}H, ^{13}C TROSY-HSQC spectra of (a) r(CGAAAAU^{5m}U^{5m}U^{5m}U^{5m}CG)$_2$ and (b) d(CGAAAAUUUUCG)$_2$ dissolved in 50% H_2O, 50% D_2O at 25 ºC at 500 MHz proton frequency. The adenine ^{1}H2, ^{13}C2 correlations are labeled. Each spectrum took about 20 h to acquire.

The five pairs of chemically modified RNA^{5mU} and DNA^{dU} duplexes studied here are isosequential to the unmodified RNA and DNA duplexes whose $^{2h}\Delta^{13}$C2 values were reported earlier (103, 106). As can be seen in Figure 3.3, there is no systematic difference in $^{2h}\Delta^{13}$C2 values between RNA and RNA^{5mU} or between DNA and DNA^{dU}. Therefore, empirically there is no discernable effect of the C7 methyl group on $^{2h}\Delta^{13}$C2 values. The chemically modified RNA and DNA remained A-and B-form, respectively, as determined by circular dichroism (Figure 3.4). The scatter in Figure 3.3 is larger than the experimental uncertainty and reflects the perturbing effects of the chemical modifications, which can also be seen from the moderate differences in the CD spectra between the modified and unmodified duplexes.

Shown in Figure 3.5 is a plot of $^{2h}\Delta^{13}$C2 values of DNA and DNA^{dU} vs. those of RNA and RNA^{5mU}. Here it can be seen that $^{2h}\Delta^{13}$C2 values of RNA and RNA^{5mU} are 4 ± 3 ppb more negative than those of DNA and DNA^{dU}. On the basis of our DFT calculations near 2.82 Å of isolated base pairs (Figure 3.1), this 4 ± 3 ppb difference is consistent with an N1–N3 hydrogen bond that is 0.04 ± 0.03 Å shorter in RNA than in DNA. It can also be seen from Figure that $^{2h}\Delta^{13}$C2 values have range of about 17 ppb, depend on sequence context, and that this dependence is the same for RNA and DNA. The $^{2h}\Delta^{13}$C2 values are most negative when the adenine intrastrand nearest neighbors are purines and least negative when the adenine is flanked by pyrimidines.

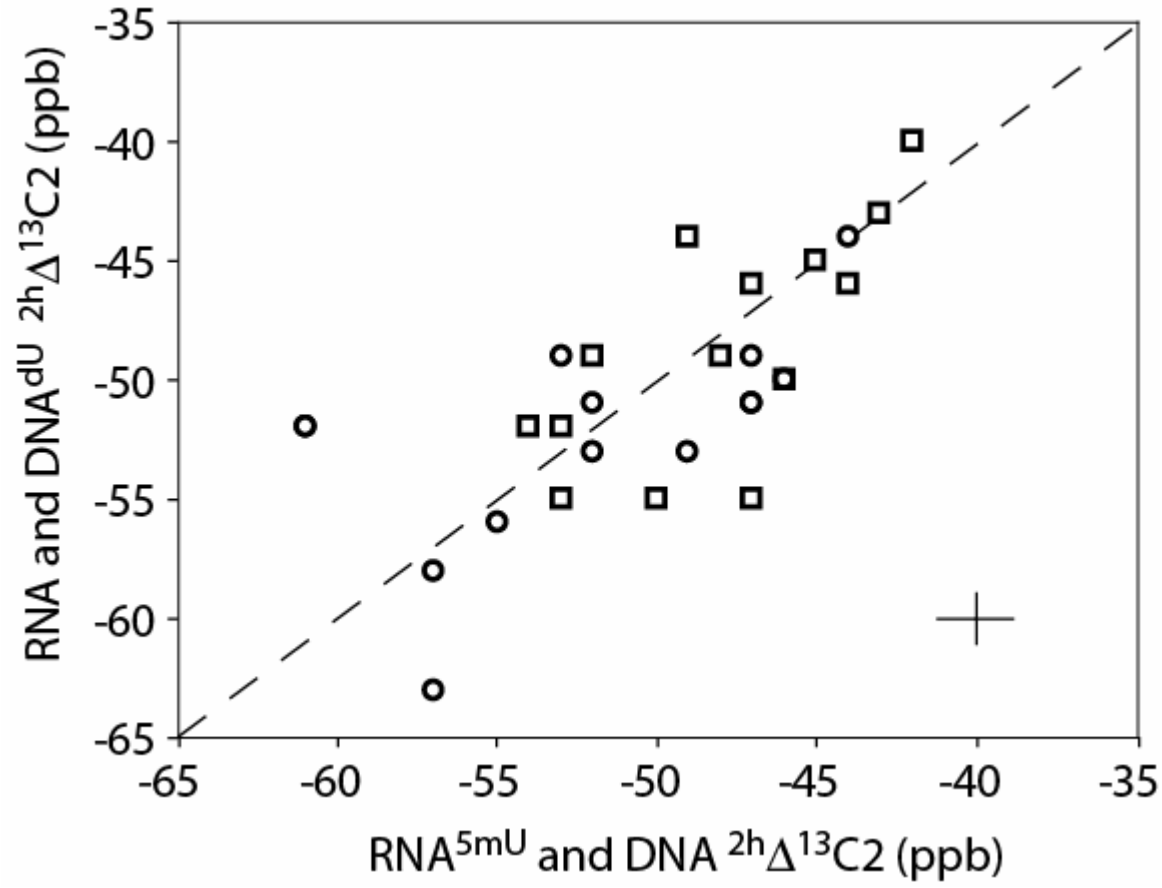

Figure 3.3. Effect of the C7 methyl group on empirical $^{2h}\Delta^{13}$C2 values. Correlations between RNA and RNA^{5mU} and DNA^{dU} and DNA values are shown using circles and squares, respectively. The dashed line is along the diagonal. A paired Student's t-test yields a probability of $p>>0.05$, which shows that there is no statistically significant effect of the C7 methyl group on $^{2h}\Delta^{13}$C2 in the data presented here. Shown in the lower right is the average uncertainty in the measurements. The values used here are listed in the Appendix B.

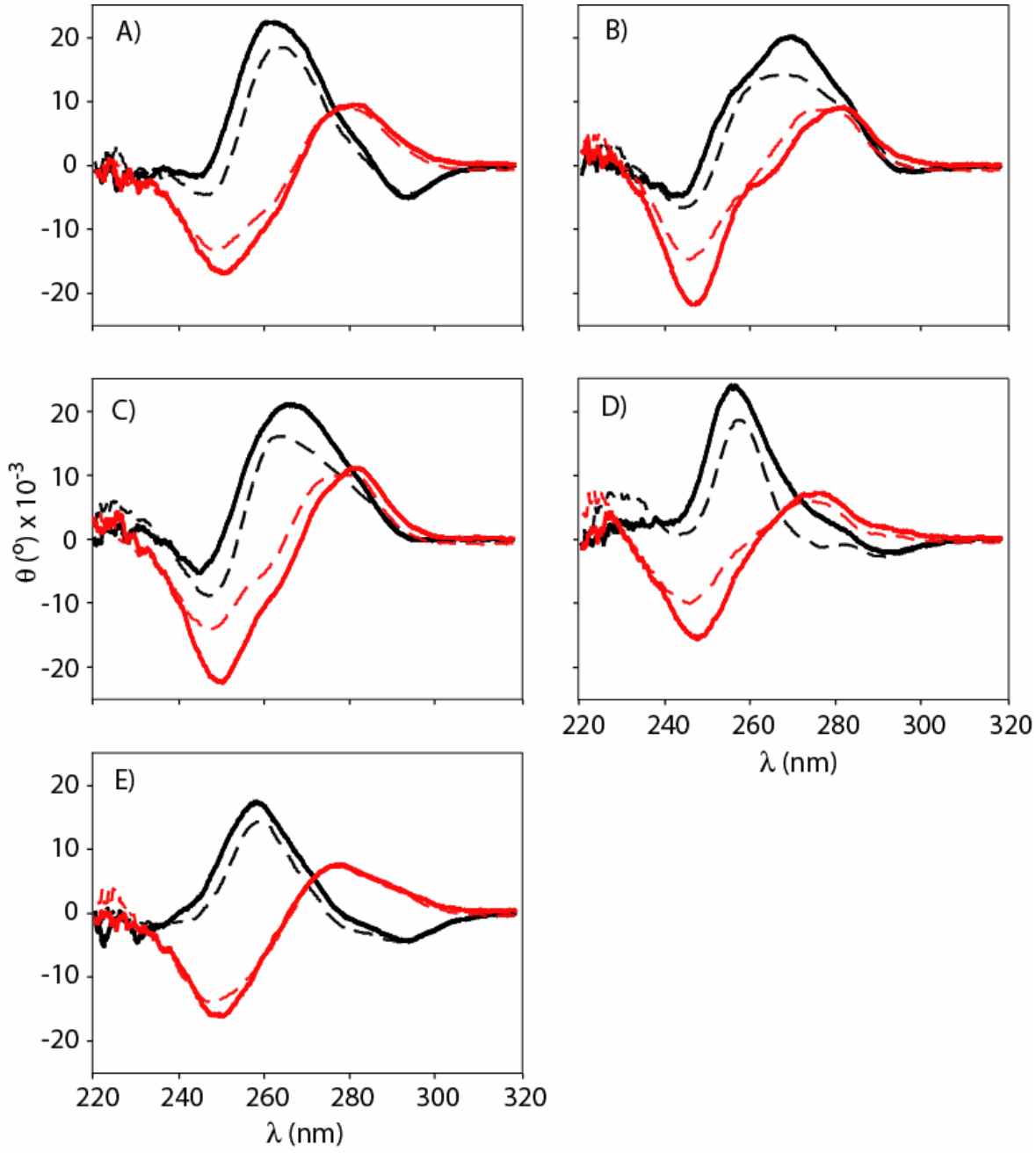

Figure 3.4. Circular dichroism spectra. Spectra for RNA and RNA^{5mU} are shown using solid and dashed black lines, respectively. Spectra for DNA and DNA^{dU} are shown using solid and dashed red lines, respectively. Shown in each panel are overlays of CD spectra of RNA, RNA^{5mU}, DNA, and DNA^{dU}. Listing only the RNA sequences, the panels correspond to A) $r(CGCGAAUUCGCG)_2$, B) $r(CGUUUUAAAACG)_2$, C) $r(CGAAAAUUUUCG)_2$, D) $r(CGUAUAUAUACG)_2$, and E) $r(CGCGUAUACGCG)_2$. The buffer condition was the same as that used for the NMR experiments.

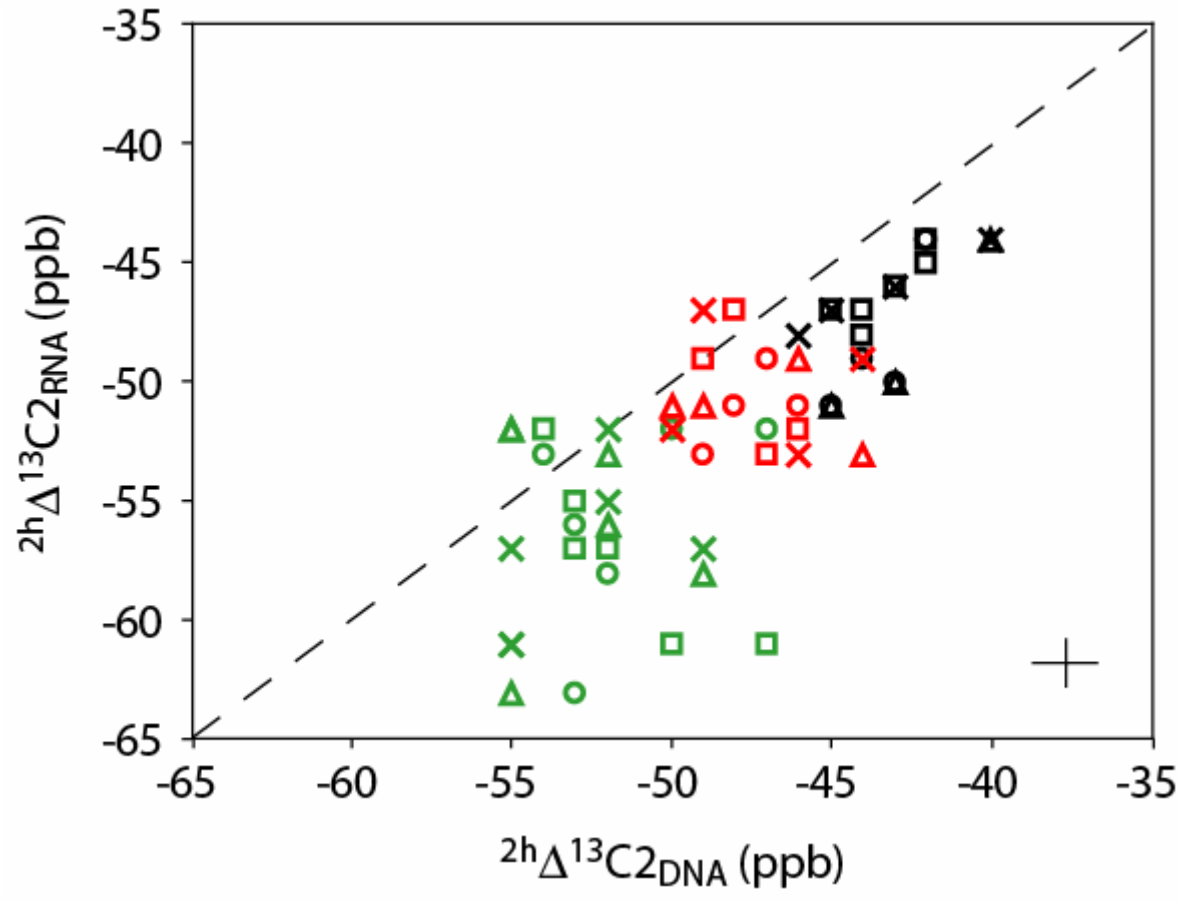

Figure 3.5. Plot of $^{2h}\Delta^{13}C2$ values of DNA vs. those of RNA. Circles, squares, triangles, and "x"s are used for DNA vs. RNA, DNA vs. RNA^{5mU}, DNA^{dU} vs. RNA, and DNA^{dU} vs. RNA^{5mU}, respectively. Green, red, and black colors are used to denote adenines with intra-strand nearest neighbors that are both purines, a purine and a pyrimidine, and both pyrimidines, respectively. The dashed line is along the diagonal. The average uncertainty in the data is shown in the lower right corner of the plot.

Discussion

The work presented here set out to test $^{2h}\Delta^{13}C2$ as measure of hydrogen-bond length and thereby test our hypothesis that RNA hydrogen bonds are shorter than those of DNA (106). Our DFT calculations predict that $^{2h}\Delta^{13}C2$ is indeed sensitive to the N1–N3 hydrogen-bond length of A:U and A:T base pairs. Furthermore, they predict that $^{2h}\Delta^{13}C2$ is insensitive to the chemical difference between uracil and thymine, which is verified by our empirical results.

Recently, investigators using DFT calculations suggested that $^{2h}\Delta^{13}C2$ values do not reflect hydrogen-bond length, but merely the presence or absence of the C7 methyl group (109). Swart et al. investigated the relationship between the NMR shielding of adenine $^{13}C2$ and total hydrogen-bond energies of isolated A:U and A:T base pairs using a technique they call "crosscoupling". They found that, unlike $^{13}C2$ shielding, hydrogen-bond energies are not completely recovered through cross-coupling and therefore, concluded that these two effects are separate and not correlated. As with Swart et al., we also find that the calculated $^{13}C2$ chemical shifts of uracil and thymine differ by about 0.1 ppm as result of the C7 methyl group of thymine (Figure 3.6). However, Swart et al. use this chemical shift difference as measure of $^{2h}\Delta^{13}C2$, whereas we find that using more rigorous approach $^{2h}\Delta^{13}C2$ calculated for A:U and A:T base pairs differ by <1 ppb at given N1–N3 distance. Importantly, our calculations are supported by our empirical observation that the C7 methyl group has no impact on $^{2h}\Delta^{13}C2$ values.

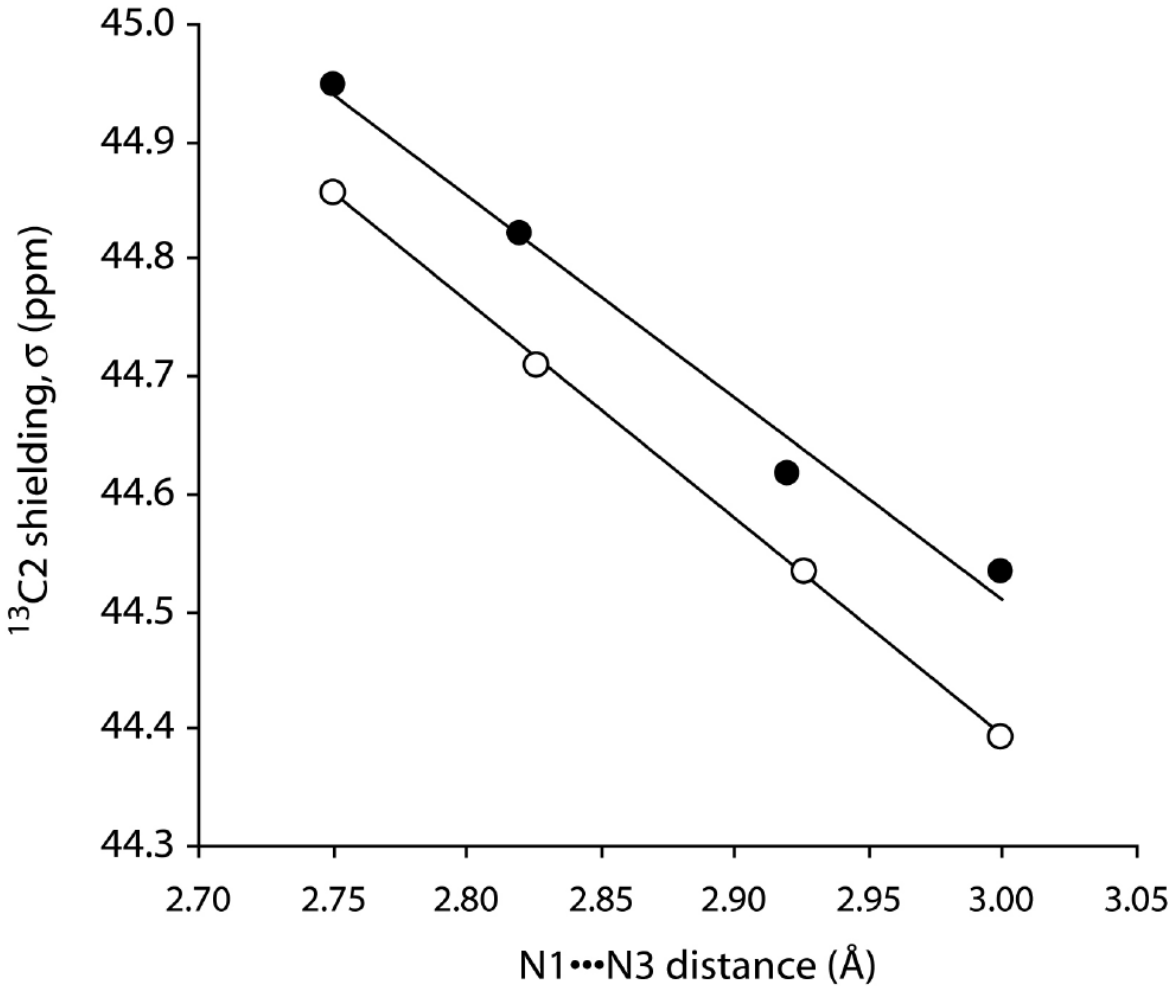

Figure 3.6. Calculated dependence of the adenine ^{13}C2 shielding constant, σ, as a function of the N1···N3 distance of isolated A:U (solid symbols) and A:T (open symbols) base pairs. Linear fits of the points yields $\sigma_{AT} = -1.84\ r + 49.92$ ppm and $\sigma_{AU} = -1.72\ r + 49.67$ ppm with $R^2 = 1.00$ and $R^2 = 0.99$, respectively.

Instead, we think that the 4 ppb difference in $^{2h}\Delta^{13}C2$ values of RNA and DNA is because of stronger RNA N1–N3 hydrogen bond, which can arise from shorter distance (or differences in hydrogen-bond angles). Perhaps the shorter RNA hydrogen bonds are inherent to the A-form secondary structure. Alternatively, differences in hydration between RNA and DNA may play role in their different hydrogen-bond lengths. Although RNA has been shown to have more structured water molecules in the major and minor grooves and around the ribose O2′ (127), DNA is more hydrated overall (108, 128). In small molecules, hydrogen-bond strengths have been demonstrated to increase in non-aqueous solvents (107). Thus, the lower hydration of RNA may promote stronger hydrogen bonds. Our results are consistent with only 0.04 Å shorter N1–N3 distance in RNA, which is difficult to discern from the larger scatter in the RNA and DNA X-ray crystal structures in the Protein Data Bank. A thermodynamic comparison of isosequential 12 base-pair duplexes found that RNA was more stable than DNA^{dU} by 3.8 kcal/ mol and that RNA^{5mU} was more stable than DNA by 4.7 kcal/mol, which gives an average of 0.4 kcal/mol more stabilization per base pair for RNA (129). Our calculated energies suggest that ~25% of the 0.4 kcal/mol arises from the shorter separation between the RNA A:U base pairs (data not shown).

For both RNA and DNA, $^{2h}\Delta^{13}C2$ values are most negative when the adenine is flanked by purines and least negative when pyrimidines are the intrastrand nearest neighbors. It is tempting to postulate that hydrogen-bond strengths have sequence dependence such that they are strongest in polypurine:polypyrimidine tracts. This hypothesis predicts that hydrogen-bonding and base-stacking interactions are coupled, which would have implications for cooperativity and long-range structure and function.

Indeed, recent calculations have suggested that π–π interactions between aromatic heterocycles play significant role in the hydrogen-bonding potential of an aromatic nitrogen base (130). However, we recognize the distinct possibility that base-stacking interactions may modulate $^{2h}\Delta^{13}C2$ without affecting hydrogen-bond strengths. Thus, we are initiating extensive DFT calculations of $^{2h}\Delta^{13}C2$ of A:U base pairs in different sequence contexts in an effort to determine the effect of base-stacking interactions on $^{2h}\Delta^{13}C2$ and hydrogen-bond strength.

Calculations and empirical data on DNA have shown that not only are $^{2h}J_{NN}$ sensitive to the hydrogen-bond length, but $^{1}J_{NH}$ coupling constants of imino groups are as well (131). Specifically, as the N1–N3 distance decreases J_{NH} is calculated to become less negative. The utility of one-bond scalar couplings as reporters of hydrogen-bond strength have also been demonstrated for amide $^{1}J_{NH}$ couplings in proteins (132, 133). Recent study found that in RNA $^{1}J_{NH}$ values were 0.4 ± 0.6[1] Hz less negative than those of DNA (134), which according to calculations on DNA triplex (131) corresponds to an N1–N3 hydrogen bond that is 0.02 ± 0.03 Å shorter (at 2.80 Å). Thus, the $^{2h}\Delta^{13}C2$- and $^{1}J_{NH}$-predicted differences between RNA and DNA hydrogen-bond lengths are similar. Furthermore, the $^{1}J_{NH}$ values showed statistically significant correlation with the $^{2h}\Delta^{13}C2$ values (134).

It should be noted that this study did not consider G:C base pairs. Furthermore, the N6–O4 hydrogen bond of A:U/T base pairs was not observed in our experiments and any

[1] The 0.4 ± 0.6 Hz is reported in the paper by Manalo *et al*. (116) as 0.4 ± 0.4 Hz, which is an error. Student's t-test calculation, which shows relevant difference between RNA and DNA $^{1}J_{NH}$ values, is reported correctly.

conclusions regarding differences in overall hydrogen-bond strengths require consideration of both A:U/T hydrogen bonds. However, a recent study has suggested that cooperativity between the N1–N3 and N6–O4 hydrogen bonds of an A:T base-pair contribute 31% to the overall stability(135), which raises the possibility that N6– O4 hydrogen bonds in RNA are stronger than those of DNA as well

Materials and methods

NMR Sample Preparation and Experiments: The DNA samples were purchased from Integrated DNA Technologies (Coralville, IA, USA) and the RNA samples were purchased from Dharmacon (Lafayette, CO, USA). The DNA sequences are $d(CGCGAATTCGCG)_2$, $d(CGTTTTAAAACG)_2$, $d(CGAAAATTTTCG)_2$, $d(CGTATATATACG)_2$, and $d(CGCGTATACGCG)_2$. For each DNA sequence there are corresponding RNA, modified DNA, and modified RNA sequences. The modified DNAs contain deoxyuridine in place of deoxythymidine. The modified RNAs contain 5-methyl uridine instead of uridine. All NMR experiments described here were performed on natural-abundance samples equilibrated in a solvent mixture of 50% H_2O and 50% D_2O as described previously (106). Proton resonance assignments were determined from WATERGATE NOESY spectra (136). All chemical shifts were referenced to internal DSS (137).

To measure $^{2h}\Delta^{13}C2$, we acquired ^{1}H, ^{13}C gradient-enhanced TROSY–HSQC spectra correlating adenine $^{1}H2$ with $^{13}C2$, as described previously (103). All experiments

were conducted at either 11.7 or 14.1 T (500 or 600 MHz ^{1}H frequency) and a sample temperature of 25 °C on Varian Inova NMR spectrometers. Chemical shift evolutions for the two-dimensional spectra were 50 ms along t2 and 181 ms along t1. An exponential 1.0 Hz linenarrowing function was applied along the t1 dimension during processing. Final digital resolutions were 5.9 and 1.8 Hz along F2 and F1, respectively. Each spectrum was collected in approximately 20 h, used an interscan delay of 2.5 s, and a delay of 3.3 ms for each INEPT dephasing period. Sample concentrations varied between 1.4 and 3.2 mM duplex for the oligonucleotides. Buffer conditions were 125 mM NaCl, 50 mM NaH_2PO_4, 0.75 mM EDTA, 0.2 mM DSS, 0.02% NaN_3, pH 7, 50% H_2O, and 50% D_2O. Peak positions were determined using polynomial interpolation with the program PIPP (138). As exchange rates of RNA and DNA imino protons with the 50% H_2O, 50% D_2O solvent mixture are slow under the conditions used here, each adenine ^{1}H2, ^{13}C2 pair presents two peaks, one for each pyrimidine H3 isotopomer. It was determined earlier that ^{13}C2(^{1}H3) resonates upfield relative to ^{13}C2(^{2}H3) (103). The trans-hydrogen bond deuterium isotope effect is defined here as $^{2h}\Delta^{13}C2 = \delta^{13}C2(^{1}H3) - \delta^{13}C2(^{2}H3)$ and is therefore, negative.

It should be noted that $^{2h}\Delta^{13}$C2 values were previously published for the unmodified RNA and DNA sequences. However, only one spectrum was collected for each unmodified DNA sample. As such, one or two additional data sets were acquired on the unmodified DNA samples and the resulting $^{2h}\Delta^{13}$C2 values averaged with the original data set. Similarly, one additional data set was acquired on the unmodified RNA samples and the new $^{2h}\Delta^{13}$C2 values were averaged with the two original data sets. For the modified RNA

and DNA samples, reported $^{2h}\Delta^{13}C2$ values are the averages from two or three spectra (Appendix B).

Density Functional Theory Calculations: A ^{13}C nucleus two bonds from the site of deuterium substitution (C–A–H) experiences a frequency shift called the deuterium isotope shift, $^{2}\Delta^{13}C = \delta^{13}C(^{1}H) - \delta^{13}C(^{2}H)$, and can be approximated by Equation 1 (104, 126):

$$^{2}\Delta^{13}C = -d\sigma/R_{AH} \times \Delta R \quad (1)$$

where σ is the ^{13}C NMR shielding constant, and $\Delta R = R_{AH} - R_{AD}$ is the difference in the mean bond lengths of A─^{1}H and A─^{2}H, respectively.

The DFT calculations are used to calculate $d\sigma/R_{AH}$ and ΔR. Specifically, in order to calculate $^{2h}\Delta^{13}C2$ using Equation 1, we need to calculate $d\sigma/R_{AH}$, which is the first derivative of the NMR shielding constant of $^{13}C2$ of adenine with respect to the pyrimidine N3–H3 bond length, and $\Delta R = R_{AH} - R_{AD}$, where R_{AH} and R_{AD} are the mean N3─^{1}H3 and N3─^{2}H3 bond lengths of the base-paired uracil or thymine. The DFT calculations of $^{2h}\Delta^{13}C2$ were carried out on isolated A:U and A:T base pairs (Figure 3.7) according to the method of Abildgaard et al. (104), using Becke's exchange (139) and Perdew and Wang's correlation functional (140) (BPW91), as implemented in (141). Calculations were carried out on an SGI Altix 3700 supercomputer at the Texas A&M University Supercomputing Facility. Full geometry optimization of A:U and A:T base pairs were performed with the

Pople basis sets (142) 6-31G(d) on all heavy atoms (C, N, O) and hydrogen atoms bound to carbon, but with 6-31G(d, p) on hydrogen atoms bound to nitrogen (104).

From Equation 1, calculation of $^{2h}\Delta^{13}C2$ requires the separate calculations of $d\sigma/R_{AH}$ and ΔR. The $d\sigma/R_{AH}$ is determined from the slope of calculated adenine $^{13}C2$ shielding values, σ, as a function of the U/T N3–H3 bond length (Figure 3.8a): σ at the N3–H3 bond length of fully optimized base pairs is calculated with the 6-31G(d) basis set using the Gauge-Independent Atomic Orbital (GIAO) method (143, 144). Then, the fully optimized N3–H3 bond length is shortened by 0.01 Å and the r is recalculated. These two points are used to calculate $d\sigma/R_{AH}$ as shown in Figure 3.8a for an isolated A:U base pair. It should be noted that r is linear within a 0.05 Å range (data not shown) and that DR is only ~0.01 Å.

The ΔR was determined from a potential energy surface scan along the pyrimidine N3–H3 bond in 0.05 Å increments around the fully optimized U/T N3–H3 bond lengths to produce a total of 18 points. The basis set used in these calculations was the same as that for geometry optimizations. Nine points surrounding the energy minimum were fit to the Morse potential function (145). Zeropoint energies were calculated using the reduced masses from frequency calculations of protonated (N3–^{1}H3) and deuterated (N3–^{2}H3) isotopomers of A:U and A:T base pairs. The mean pyrimidine N3–^{1}H3 and N3–^{2}H3 bond lengths R_{AH} and R_{AD} were then obtained from the Morse potential equation by using the fit parameters and zero-point energies (Figure 3.8b).

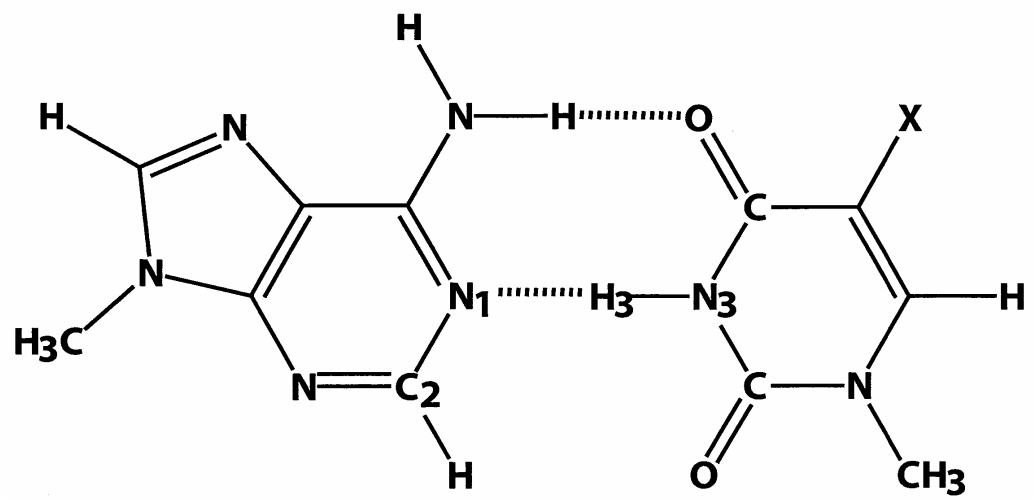

Figure 3.7. Model structures of A:U (X=H) and A:T (X=CH_3) base pairs used in the DFT calculations. The ribose groups were replaced by methyl groups to reduce computational costs.

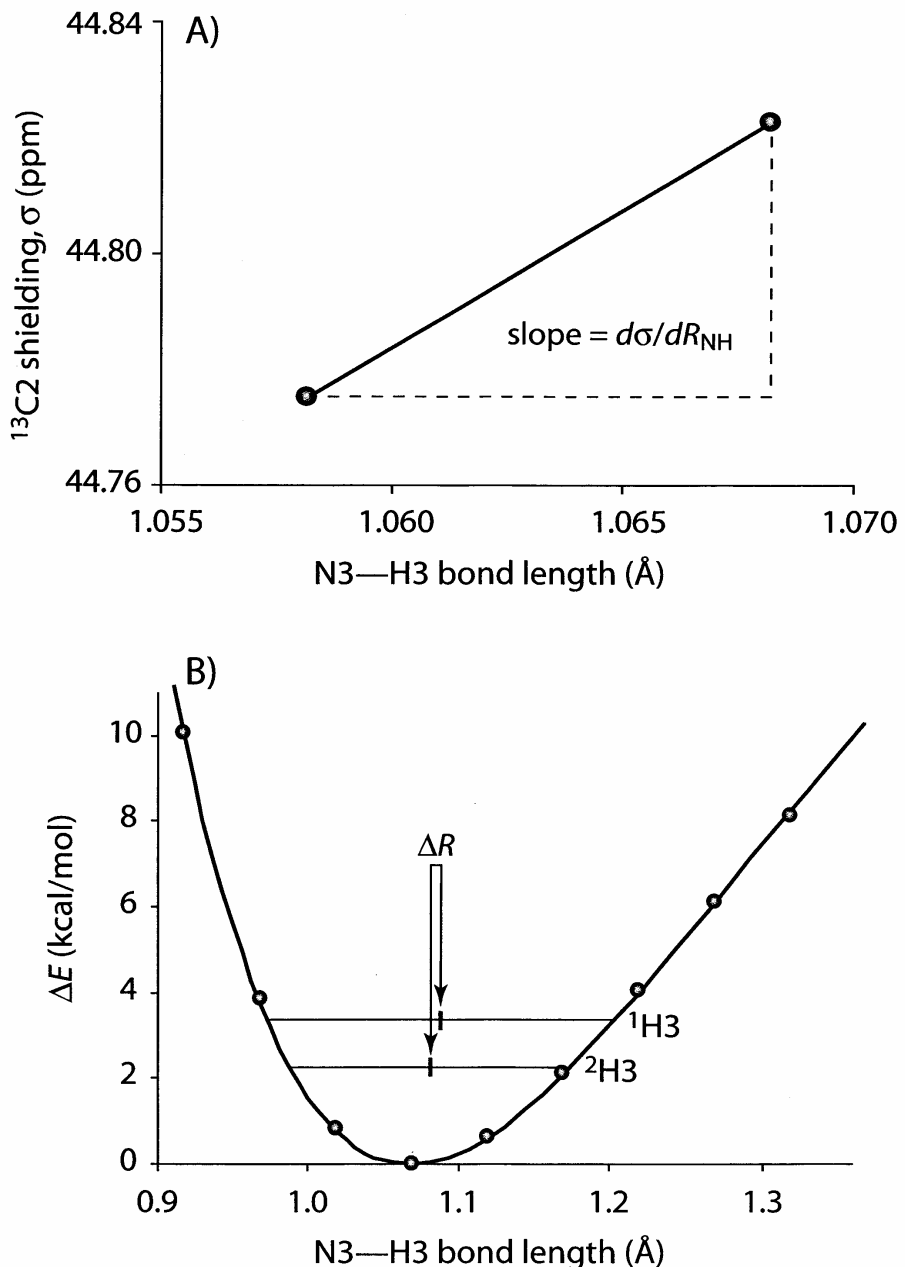

Figure 3.8. DFT calculations of (a) the ^{13}C2 shielding of adenine and (b) the total energy change of a Watson–Crick A:U base pair as functions of the N3–H3 bond length of uracil. The two points in (a) correspond to N3–H3 bond lengths of 1.0682 and 1.0582 Å, which were used to determine that $d\sigma/d\mathrm{R}_{NH}$ = 4.74 ppm/ Å. The nine calculated points in (b) were fit to the Morse potential function (Equation 2). The zero-point energies of the proton and deuteron are 3.3606 and 2.2572 kcal/mol, respectively, with mean N3^{1}H3 and N3^{2}H3 bond lengths of 1.0880 and 1.0809 Å , which yield ΔR=0.0071 Å . From (a) and (b), we get $^{2}\mathrm{h}\Delta^{13}$C2 = $-d\sigma/d\mathrm{R}_{NH} \times \Delta R$ = –4.74 ppm/Å × 0.0071 Å = –33.7 ppb.

The Morse potential has the form given in Equation 2 (145):

$$\Delta E = D\{1 - \mathrm{Exp}[-a(R - R_e)]\}^2, \qquad (2)$$

where R is the pyrimidine N3–H3 bond length, R_e is the N3–H3 bond length at the potential minimum, D is the 'depth' of the potential energy function, $a = [k/(2D)]^{1/2}$ describes the 'width' of the potential, and k is the force constant of the bond. The permitted energy levels are given in Equation 3:

$$\Delta \mathrm{E} = (\upsilon + 1/2)hc\tilde{\upsilon} - (\upsilon + 1/2)^2\chi hc\tilde{\upsilon}\,, \qquad (3)$$

where υ is the vibrational quantum number and can take the integer values 0, 1, 2, etc., χ is called the anharmonicity constant and is equal to $hc\tilde{\upsilon}/(4D)$, $\tilde{\upsilon}$ is the wave number, h is the Planck constant, and c is the speed of light. The wavenumber is defined as $\tilde{\upsilon} = \{1/(2\pi c)\}(k/m)^{1/2}$, where m is the reduced mass. The zero-point energies, E_{ZP}, for the two isotopomers are calculated at $\upsilon = 0$: $E_{ZP} = (hc\tilde{\upsilon}/2)(1 - \chi/2)$. Fitting points such as those shown in Figure 3.8b to Equation 2 yields D, a, and R_e, from which χ and $\tilde{\upsilon}$ can be derived and used to determine E_{ZP}. The R_{NH} is the average of the two N3─^{1}H3 bond length solutions to the Morse potential function at E_{ZP} for the ^{1}H3 isotopomer. The R_{ND} is determined analogously.

The $^{2h}\Delta^{13}$C2 is calculated as the product of $-d\sigma/dR_{NH}$ and ΔR (104). Calculations of $^{2h}\Delta^{13}$C2 were performed at the fully optimized A:U and A:T geometries and at different

(constrained) N1─N3 distances. In deriving ΔR for the structures with constrained N1─N3 distances the reduced masses obtained from frequency calculations of the fully optimized base pairs were used since frequency calculations are valid only at stationary points on the potential energy surface (146). Energy corrections for basis set superposition error (BSSE) are very small (less than 0.0007% of the counterpoise-corrected energy of fully optimized A:U and A:T); hence all calculations were performed without BSSE correction). Several input files used in our Gaussian calculations are provided in the Appendix B.

CHAPTER IV

CONCLUSION AND DISCUSSION

The day/night switch of the circadian clock of *Synechococcus Elongatus*

The work accomplished here shows that rhythmic phosphorylation of KaiC is determined in large part by the dynamic equilibrium of the A-loops. KaiA shifts the equilibrium towards the exposed state through direct interactions with the A-loop, thereby enhancing the autokinase activity over that of the autophosphatase of KaiC. In contrast, KaiB shifts the dynamic equilibrium of the A-loop towards the buried state not by directly interacting with it but rather through hindering KaiA. Work here also showed that the equilibrium position of the A-loop likely determines the position of ATP. When the A-loops favor the buried state, the ATP is held far away from the sites of phosphorylation. However, when the exposed state is favored, ATP approaches T432 and S431 for phosphoryl transfer.

The work presented here opens the following several important new questions that should be pursued with all due haste.

1. How does the conformation of KaiC depend on A-loop position?

   The work presented here shows how the position of the dynamic equilibrium of the A-loop determines the relative activities of KaiC's autokinase and autophosphatase activities. It is unknown how the position of the A-loop changes the conformation of KaiC to influence these activities. It is

proposed here that the A-loop position determines the conformation of the ATP binding pocket, but structural details are lacking. A primary future objective should be to elucidate such details.

2. Does KaiA interact with more than just the 30 C-terminal residues of KaiC?

   Structural details are available only for a complex between the C-terminal domain of KaiA and a peptide of the 30 C-terminal residues of KaiC. KaiA interacts much more tightly with full-length KaiC than with just free C-terminal peptides. The higher local concentration of C-terminal segments in the KaiC hexamer relative to free peptides in solution used in this study likely accounts for part of this discrepancy. It is unknown whether part of the tighter binding between full-length proteins arises from KaiA–KaiC interactions that do not include the 30 residue C-terminal segment. So far, there is only a partial accounting of the interactions between KaiA and KaiC. Structural details of the interactions between the full-length proteins is required to completely understand how KaiA induces KaiC autophosphorylation.

3. How does the N-terminal domain of KaiA regulate the KaiA–KaiC interaction?

   The affinity of full-length KaiA for the C-terminal 30 residues of KaiC is several fold weaker than that of the truncated C-terminal domain of KaiA. Thus, the N-terminal domain of KaiA somehow weakens the affinity of KaiA for KaiC. Yet, the N-terminal domain of KaiA does not have any

detectable affinity for the C-terminal residues of KaiC. So, how is this attenuation achieved? Does the presence of the N-terminal domain of KaiA affect the structure of its C-terminal domain? A comparison of the NMR structure of the C-terminal domain of KaiA with the X-ray crystal structure of full-length KaiA hints at such an affection. What is missing in this comparison is a structure of full-length KaiA with the C-terminal KaiC peptide bound. Structural details of this complex can be obtained using state-of-the art NMR experiments.

4. What are the different phosphorylation-dependent conformational states of KaiC?

   KaiA–KaiC, KaiB–KaiC, and KaiA-KaiB–KaiC interactions are driven by the phosphorylation states of KaiC. Reversible phosphorylation is a common strategy among proteins for switching conformational states (147). As the sites of phosphorylation of KaiC are at the hexameric subunit interfaces of the CII domains, it is proposed here that phosphorylation causes dramatic remodeling of KaiC. Structural details of KaiC in each of its phosphorylation states are needed to understand its phosphorylation-dependent interactions with KaiA and KaiB.

5. How do pST-KaiC and KaiB interact?

   The KaiB–KaiC interaction is important as it determines the reversal from KaiC autophosphorylation to autodephosphorylation. Yet high resolution structures of a KaiB-KaiC complex are lacking. KaiB interacts with a

specific phosphoform of KaiC, pST-KaiC, where only S431 is phosphorylated (57, 58). It was shown here that KaiB does not directly interact with the C-terminal tail of KaiC, and does not affect KaiC activity in the absence of KaiA. So KaiB's affect on the dynamic equilibrium of the A-loop is indirect. Electron microscopy experiments show that KaiB binds as a dimer to the top of the CII dome (72), but the resolution is 25Å, much too low to resolve the interactions important for stabilizing the KaiB–pST-KaiC complex. Structural details of this complex in comparison with the structures of the other phosphoforms of KaiC will reveal why KaiB as a particular affinity for pST-KaiC over KaiC, SpT-KaiC, and pSpT-KaiC.

6. How does KaiB hinder KaiA?

   KaiB does not directly affect KaiC activity but exerts itself by blocking KaiA stimulation of KaiC, resulting in the autodephosphorylation half of the cycle of KaiC phosphorylation. Currently, how KaiB accomplishes this task is completely unknown. Any interactions between KaiA and KaiB have not yet been detectable in the absence of KaiC. However, KaiA and KaiB form large complexes with pST-KaiC (57, 58). The KaiA–KaiB interactions on pST-KaiC are unknown, but are critically important to the function of this circadian oscillator.

7. Why does KaiB require the presence of the N-terminal domain of KaiA for its function?

It was shown here that the truncated C-terminal domain of KaiA is immune to KaiB. In other words, KaiB cannot stop the C-terminal domain of KaiA from stimulating KaiC in the absence of the N-terminal domain of KaiA. Mysteriously, the N-terminal domain plays a fundamental role in KaiB function. Do KaiB and the N-terminal domain of KaiA physically contact or is their interaction indirect?

All of the questions posed above can be approached using NMR, even though the molecular weights of some of the complexes will be several hundreds of kilodaltons. Recent advances overcome the problems of fast signal relaxation and of overly complex spectra in large molecules, by combining selective labeling of proteins with a new type of NMR experiment (148, 149). The proteins of interest are expressed such that hydrogen sites are uniformly deuterated, except for a single methyl group in valine, leucine, and isoleucine residues (150). This dilution of hydrogen atoms with deuterium atoms reduces spectra complexity and increases signal strength by taking advantage of certain NMR relaxation phenomena (149). These methyl groups are uniformly spaced in proteins and serve as reporters of protein dynamics or of structural perturbations due to phosphorylation or protein-protein interactions.

Even in the absence of specific assignments of these methyl resonances, spectra report on changes in the overall protein structure and dynamics under different sample conditions (151). For example, NMR can be used to determine the distinct spectral signatures for labeled KaiC in its different phosphoforms or, more likely, stable

phosphomimics (152) and in the presence and absence of unlabeled KaiB and/or KaiA. Indeed, several methyl groups accessible by this method are located along the subunit interfaces of KaiC and near the sites of phosphorylation, S431 and T432. This would allow the direct observation of changes as these proteins interact. Spectral assignments allow the mapping of observed structural and dynamic changes onto the available crystal structures. Recently, another large system, the 670 kD 20S proteasome of Archaea was studied (153) using methods that can probably be applied to the 350 kD KaiC hexamer and its complexes.

Hydrogen bonds of DNA and RNA

The work presented here shows for the first time that RNA hydrogen bonds are inherently shorter than those of DNA by a few hundredths of an Angstrom. The lateness of this discovery is rather remarkable in light of the extensive history of biophysical studies on nucleic acids and the 1150 DNA and 669 RNA structures deposited to date in the Protein Data Bank. The reason this difference was not discovered earlier was because previous NMR and X-ray experiments were not sensitive enough. $^{2h}\Delta^{13}C2$ is clearly the most sensitive gauge of DNA and RNA hydrogen bonds. It is demonstrated here that the chemical difference between uracil in RNA and thymine in DNA have no effect on hydrogen bond lengths.

However, $^{2h}\Delta^{13}C2$ can in principle be affected by interactions other than hydrogen bonding, raising the possibility that the difference in DNA and RNA $^{2h}\Delta^{13}C2$ values is just a consequence of their respective B- and A-form conformations and not the result of

hydrogen bond differences. An independent line of evidence was needed. Calculations show that the one-bond scalar coupling constant of an $^{15}N–^{1}H$ pair, $^{1}J_{NH}$, decreases in magnitude as the length of a $N\cdots^{1}H–^{15}N$ hydrogen bond decreases (131). $^{1}J_{NH}$ arises from the magnetic polarization of electrons by nuclear spins (154) whereas $^{2h}\Delta^{13}C2$ is vibrational in origin. Thus, they are mutually independent observables. It was shown by others in the LiWang laboratory that $^{1}J_{NH}$ values of RNA were indeed smaller in magnitude than those of DNA (134), enforcing our original conclusion that RNA hydrogen bonds are shorter.

It was suggested here that the sequence dependence of $^{2h}\Delta^{13}C2$ reflects the sequence dependence of hydrogen-bond lengths in DNA and RNA. However, the vibrational and shielding factors of $^{2h}\Delta^{13}C2$ can in principle be affected by interactions other than hydrogen bonding. Sequence-dependent base stacking may alter the vibrational and shielding factors to different extents independently of hydrogen bonding. Thus, the sequence dependence of $^{2h}\Delta^{13}C2$ may be due to the sequence dependence of base stacking rather than to differences in hydrogen bond lengths. Others in the LiWang laboratory calculated that indeed sequence-dependent base stacking produces a sequence dependence in $^{2h}\Delta^{13}C2$ (155). These calculations predicted that base stacking contributes to hydrogen-bond strength and that the experimental sequence dependence of $^{2h}\Delta^{13}C2$ does in fact reflect sequence dependence in hydrogen-bond lengths and strengths. These calculations predict that the cooperativity of DNA and RNA structure arises at least in part from the coupling between hydrogen-bond and base stacking interactions. Experimentally weakening base stacking interactions was found to remove this coupling (155). Future work should test the

contribution of this coupling in DNA and RNA motifs that have been previously shown to have high or low cooperativity in their folding/unfolding behavior.

The physicochemical basis for shorter hydrogen bonds in RNA with respect to DNA was proposed in the work presented to arise from differences in hydration (156). It was known already that 1) hydrogen bonds become stronger as water activity is reduced (107), and 2) RNA is less hydrated than DNA (108, 157). Indeed, the A-form conformation adopted by RNA is the conformation of dehydrated DNA (74). Others in the LiWang laboratory used dilute ethanol-water mixtures to reduce water activity and thus the water available for hydration of solutions of DNA and RNA (158). $^{1}J_{NH}$ magnitudes of DNA and RNA became smaller in 8 mol% ethanol, supporting the notion that dehydration decreases hydrogen-bond lengths. $^{1}J_{NH}$ values of DNA in 8 mol% ethanol were statistically equal to those of RNA in water, which is consistent with our original hypothesis that the reduced hydration of RNA with respect to DNA was a major factor in their hydrogen-bond length differences. A direct observation of hydrogen-bond length changes in a protein as water activity is decreased would show that this effect is general.

REFERENCES

1. Dunlap, JC (1999) Molecular basis for circadian clocks. *Cell* 96:271-290.
2. Bell-Pedersen, D, Cassone, VM, Earnest, DJ, Golden, SS, Hardin, PE*, et al.* (2005) Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nat Rev Genet* 6:544-556.
3. Woelfle, MA, Ouyang, Y, Phanvijhitsiri, K, Johnson, CH (2004) The adaptive value of circadian clocks: an experimental assessment in cyanobacteria. *Curr Biol* 14:1481-1486.
4. Green, RM, Tingay, S, Wang, ZY, Tobin, EM (2002) Circadian rhythms confer a higher level of fitness to *Arabidopsis* plants. *Plant Physiol* 129:576-584.
5. Michael, TP, Salome, PA, Yu, HJ, Spencer, TR, Sharp, EL*, et al.* (2003) Enhanced fitness conferred by naturally occurring variation in the circadian clock. *Science* 302:1049-1053.
6. Barkai, N, Leibler, S (2000) Circadian clocks limited by noise. *Nature* 403:267-268.
7. Edery, I (2000) Circadian rhythms in a nutshell. *Physiol Genomics* 3:59-74.
8. Pittendrigh, CS (1954) On temperature independence in the clock system controlling emergence time in *Drosophila*. *Proc Natl Acad Sci USA* 40:1018-1029.
9. Ditty, JL, Williams, SB, Golden, SS (2003) A cyanobacterial circadian timing mechanism. *Annu Rev Genet* 37:513-543.
10. Golden, SS, Ishiura, M, Johnson, CH, Kondo, T (1997) Cyanobacterial circadian rhythms. *Annu Rev Plant Physiol Plant Mol Biol* 48:327-354.

11. Golden, SS, Johnson, CH, Kondo, T (1998) The cyanobacterial circadian system: a clock apart. *Curr Opin Microbiol* 1:669-673.

12. Lorne, J, Scheffer, J, Lee, A, Painter, M, Miao, VP (2000) Genes controlling circadian rhythm are widely distributed in cyanobacteria. *FEMS Microbiol Lett* 189:129-133.

13. Kondo, T, Ishiura, M (2000) The circadian clock of cyanobacteria. *Bioessays* 22:10-15.

14. Iwasaki, H, Williams, SB, Kitayama, Y, Ishiura, M, Golden, SS*, et al.* (2000) A KaiC-interacting sensory histidine kinase, SasA, necessary to sustain robust circadian oscillation in cyanobacteria. *Cell* 101:223-233.

15. Grobbelaar, N, Huang, TC, Lin, HY, Chow, TJ (1986) Dinitrogen-fixing endogenous rhythm in *Synechococcus* Rf-1. *Fems Microbiology Letters* 37:173-177.

16. Onai, K, Morishita, M, Itoh, S, Okamoto, K, Ishiura, M (2004) Circadian rhythms in the thermophilic cyanobacterium Thermosynechococcus elongatus: compensation of period length over a wide temperature range. *J Bacteriol* 186:4972-4977.

17. Aoki, S, Kondo, T, Wada, H, Ishiura, M (1997) Circadian rhythm of the cyanobacterium *Synechocystis* sp. strain PCC 6803 in the dark. *J Bacteriol* 179:5751-5755.

18. Kondo, T, Strayer, CA, Kulkarni, RD, Taylor, W, Ishiura, M*, et al.* (1993) Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria. *Proc Natl Acad Sci USA* 90:5672-5676.

19. Liu, Y, Tsinoremas, NF, Johnson, CH, Lebedeva, NV, Golden, SS*, et al.* (1995) Circadian orchestration of gene expression in cyanobacteria. *Genes Dev* 9:1469-1478.

20. Smith, RM, Williams, SB (2006) Circadian rhythms in gene transcription imparted by chromosome compaction in the cyanobacterium *Synechococcus elongatus*. *Proc Natl Acad Sci USA* 103:8564-8569.

21. Mackey, SR, Ditty, JL, Clerico, EM, Golden, SS (2007) Detection of rhythmic bioluminescence from luciferase reporters in cyanobacteria. *Methods Mol Biol* 362:115-129.

22. Kondo, T, Mori, T, Lebedeva, NV, Aoki, S, Ishiura, M*, et al.* (1997) Circadian rhythms in rapidly dividing cyanobacteria. *Science* 275:224-227.

23. Kondo, T, Tsinoremas, NF, Golden, SS, Johnson, CH, Kutsuna, S*, et al.* (1994) Circadian clock mutants of cyanobacteria. *Science* 266:1233-1236.

24. Ishiura, M, Kutsuna, S, Aoki, S, Iwasaki, H, Andersson, CR*, et al.* (1998) Expression of a gene cluster kaiABC as a circadian feedback process in cyanobacteria. *Science* 281:1519-1523.

25. Nishimura, H, Nakahira, Y, Imai, K, Tsuruhara, A, Kondo, H*, et al.* (2002) Mutations in KaiA, a clock protein, extend the period of circadian rhythm in the cyanobacterium Synechococcus elongatus PCC 7942. *Microbiology* 148:2903-2909.

26. Taniguchi, Y, Yamaguchi, A, Hijikata, A, Iwasaki, H, Kamagata, K*, et al.* (2001) Two KaiA-binding domains of cyanobacterial circadian clock protein KaiC. *FEBS Lett* 496:86-90.

27. Kageyama, H, Kondo, T, Iwasaki, H (2003) Circadian formation of clock protein complexes by KaiA, KaiB, KaiC, and SasA in cyanobacteria. *J Biol Chem* 278:2388-2395.

28. Holtzendorff, J, Partensky, F, Mella, D, Lennon, JF, Hess, WR*, et al.* (2008) Genome streamlining results in loss of robustness of the circadian clock in the marine cyanobacterium *Prochlorococcus marinus* PCC 9511. *J Biol Rhythms* 23:187-199.

29. Rocap, G, Larimer, FW, Lamerdin, J, Malfatti, S, Chain, P*, et al.* (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424:1042-1047.

30. Johnson, CH, Golden, SS, Kondo, T (1998) Adaptive significance of circadian programs in cyanobacteria. *Trends Microbiol* 6:407-410.

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

32. Dunlap, JC (1999) Molecular bases for circadian clocks. *Cell* 96:271-290.

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

34. Hardin, PE, Hall, JC, Rosbash, M (1990) Feedback of the *Drosophila* period gene product on circadian cycling of its messenger RNA levels. *Nature* 343:536-540.

35. Nakahira, Y, Katayama, M, Miyashita, H, Kutsuna, S, Iwasaki, H*, et al.* (2004) Global gene repression by KaiC as a master process of prokaryotic circadian system. *Proc Natl Acad Sci USA* 101:881-885.

36. Johnson, CH (2004) Precise circadian clocks in prokaryotic cyanobacteria. *Curr Issues Mol Biol* 6:103-110.

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

38. Takigawa-Imamura, H, Mochizuki, A (2006) Transcriptional autoregulation by phosphorylated and non-phosphorylated KaiC in cyanobacterial circadian rhythms. *J Theor Biol* 241:178-192.

39. Iwasaki, H, Taniguchi, Y, Ishiura, M, Kondo, T (1999) Physical interactions among circadian clock proteins KaiA, KaiB and KaiC in cyanobacteria. *EMBO J* 18:1137-1145.

40. Garces, RG, Wu, N, Gillon, W, Pai, EF (2004) Anabaena circadian clock proteins KaiA and KaiB reveal a potential common binding site to their partner KaiC. *EMBO J* 23:1688-1698.

41. Mori, T, Saveliev, SV, Xu, Y, Stafford, WF, Cox, MM*, et al.* (2002) Circadian clock protein KaiC forms ATP-dependent hexameric rings and binds DNA. *Proc Natl Acad Sci USA* 99:17203-17208.

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

43. Hayashi, F, Suzuki, H, Iwase, R, Uzumaki, T, Miyake, A*, et al.* (2003) ATP-induced hexameric ring structure of the cyanobacterial circadian clock protein KaiC. *Genes Cells* 8:287-296.

44. Nishiwaki, T, Iwasaki, H, Ishiura, M, Kondo, T (2000) Nucleotide binding and autophosphorylation of the clock protein KaiC as a circadian timing process of cyanobacteria. *Proc Natl Acad Sci USA* 97:495-499.

45. Hayashi, F, Itoh, N, Uzumaki, T, Iwase, R, Tsuchiya, Y*, et al.* (2004) Roles of two ATPase-motif-containing domains in cyanobacterial circadian clock protein KaiC. *J Biol Chem* 279:52331-52337.

46. 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. *Biochem Biophys Res Commun* 348:864-872.

47. Nishiwaki, T, Satomi, Y, Nakajima, M, Lee, C, Kiyohara, R*, et al.* (2004) Role of KaiC phosphorylation in the circadian clock system of *Synechococcus elongatus* PCC 7942. *Proc Natl Acad Sci USA* 101:13927-13932.

48. Xu, Y, Mori, T, Pattanayek, R, Pattanayek, S, Egli, M*, et al.* (2004) Identification of key phosphorylation sites in the circadian clock protein KaiC by crystallographic and mutagenetic analyses. *Proc Natl Acad Sci USA* 101:13933-13938.

49. Xu, Y, Mori, T, Johnson, CH (2003) Cyanobacterial circadian clockwork: roles of KaiA, KaiB and the kaiBC promoter in regulating KaiC. *EMBO J* 22:2117-2126.

50. Pattanayek, R, Wang, J, Mori, T, Xu, Y, Johnson, CH*, et al.* (2004) Visualizing a circadian clock protein: crystal structure of KaiC and functional insights. *Mol Cell* 15:375-388.

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

52. 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. *Proc Natl Acad Sci USA* 99:15357-15362.

53. Hayashi, F, Ito, H, Fujita, M, Iwase, R, Uzumaki, T*, et al.* (2004) Stoichiometric interactions between cyanobacterial clock proteins KaiA and KaiC. *Biochem Biophys Res Commun* 316:195-202.

54. Kageyama, H, Nishiwaki, T, Nakajima, M, Iwasaki, H, Oyama, T*, et al.* (2006) Cyanobacterial circadian pacemaker: Kai protein complex dynamics in the KaiC phosphorylation cycle in vitro. *Mol Cell* 23:161-171.

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

56. Nakajima, M, Imai, K, Ito, H, Nishiwaki, T, Murayama, Y*, et al.* (2005) Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science* 308:414-415.

57. Nishiwaki, T, Satomi, Y, Kitayama, Y, Terauchi, K, Kiyohara, R*, et al.* (2007) A sequential program of dual phosphorylation of KaiC as a basis for circadian rhythm in cyanobacteria. *EMBO J* 26:4029-4037.

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

59. Terauchi, K, Kitayama, Y, Nishiwaki, T, Miwa, K, Murayama, Y*, et al.* (2007) ATPase activity of KaiC determines the basic timing for circadian clock of cyanobacteria. *Proc Natl Acad Sci USA* 104:16377-16381.

60. Mori, T, Williams, DR, Byrne, MO, Qin, X, Egli, M*, et al.* (2007) Elucidating the ticking of an in vitro circadian clockwork. *PLoS Biol* 5:e93.

61. Yoda, M, Eguchi, K, Terada, TP, Sasai, M (2007) Monomer-shuffling and allosteric transition in KaiC circadian oscillation. *PLoS ONE* 2:e408.

62. Ito, H, Kageyama, H, Mutsuda, M, Nakajima, M, Oyama, T*, et al.* (2007) Autonomous synchronization of the circadian KaiC phosphorylation rhythm. *Nat Struct Mol Biol* 14:1084-1088.

63. Uzumaki, T, Fujita, M, Nakatsu, T, Hayashi, F, Shibata, H*, et al.* (2004) Crystal structure of the C-terminal clock-oscillator domain of the cyanobacterial KaiA protein. *Nat Struct Mol Biol* 11:623-631.

64. Vakonakis, I, LiWang, AC (2004) Structure of the C-terminal domain of the clock protein KaiA in complex with a KaiC-derived peptide: implications for KaiC regulation. *Proc Natl Acad Sci USA* 101:10925-10930.

65. Ye, S, Vakonakis, I, Ioerger, TR, LiWang, AC, Sacchettini, JC (2004) Crystal structure of circadian clock protein KaiA from *Synechococcus elongatus*. *J Biol Chem* 279:20511-20518.

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

67. Iwase, R, Imada, K, Hayashi, F, Uzumaki, T, Morishita, M*, et al.* (2005) Functionally important substructures of circadian clock protein KaiB in a unique tetramer complex. *J Biol Chem* 280:43141-43149.

68. Vakonakis, I, Sun, J, Wu, T, Holzenburg, A, Golden, SS*, et al.* (2004) NMR structure of the KaiC-interacting C-terminal domain of KaiA, a circadian clock protein: implications for KaiA-KaiC interaction. *Proc Natl Acad Sci USA* 101:1479-1484.

69. Leipe, DD, Aravind, L, Grishin, NV, Koonin, EV (2000) The bacterial replicative helicase DnaB evolved from a RecA duplication. *Genome Res* 10:5-16.

70. Murakami, R, Miyake, A, Iwase, R, Hayashi, F, Uzumaki, T*, et al.* (2008) ATPase activity and its temperature compensation of the cyanobacterial clock protein KaiC. *Genes Cells* 13:387-395.

71. Pattanayek, R, Williams, DR, Pattanayek, S, Xu, Y, Mori, T*, et al.* (2006) Analysis of KaiA-KaiC protein interactions in the cyano-bacterial circadian clock using hybrid structural methods. *EMBO J* 25:2017-2028.

72. Pattanayek, R, Williams, DR, Pattanayek, S, Mori, T, Johnson, CH*, et al.* (2008) Structural model of the circadian clock KaiB-KaiC complex and mechanism for modulation of KaiC phosphorylation. *EMBO J* 27:1767-1778.

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

74. Saenger, W (1984) *Principles of Nucleic Acid Structure* (Springer-Verlag, New York).

75. Puglisi, JD, Wyatt, JR (1995) Biochemical and NMR studies of RNA conformation with an emphasis on RNA pseudoknots. *Methods Enzymol* 261:323-350.

76. Gehring, K, Leroy, JL, Gueron, M (1993) A tetrameric DNA structure with protonated cytosine:cytosine base pairs. *Nature* 363:561-565.

77. Raghunathan, G, Miles, HT, Sasisekharan, V (1995) Symmetry and structure of RNA and DNA triple helices. *Biopolymers* 36:333-343.

78. Brown, T, Hunter, WN (1997) Non-Watson-Crick base associations in DNA and RNA revealed by single crystal x-ray diffraction methods: mismatches, modified bases, and nonduplex DNA. *Biopolymers* 44:91-103.

79. Fedoroff, OY, Ge, Y, Reid, BR (1997) Solution structure of r(GAGGACUG):d(CAGTCCTC) hybrid: implications for the initiation of HIV-1 (+)-strand synthesis. *J Mol Biol* 269:225-239.

80. Zimmermann, GR, Jenison, RD, Wick, CL, Simorre, JP, Pardi, A (1997) Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. *Nat Struct Biol* 4:644-649.

81. Tarasow, TM, Tarasow, SL, Eaton, BE (1997) RNA-catalysed carbon-carbon bond formation. *Nature* 389:54-57.

82. Konig, P, Giraldo, R, Chapman, L, Rhodes, D (1996) The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. *Cell* 85:125-136.

83. Omichinski, JG, Clore, GM, Schaad, O, Felsenfeld, G, Trainor, C*, et al.* (1993) NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* 261:438-446.

84. Cuenoud, B, Szostak, JW (1995) A DNA metalloenzyme with DNA ligase activity. *Nature* 375:611-614.

85. Keniry, MA, Shafer, RH (1995) NMR studies of drug-DNA complexes. *Methods Enzymol* 261:575-604.

86. Mao, B, Vyas, RR, Hingerty, BE, Broyde, S, Basu, AK*, et al.* (1996) Solution conformation of the N-(deoxyguanosin-8-yl)-1-aminopyrene ([AP]dG) adduct opposite dC in a DNA duplex. *Biochemistry* 35:12659-12670.

87. de Clairac, RPL, Geierstanger, BH, Mrksich, M, Dervan, PB, Wemmer, DE (1997) NMR characterization of hairpin polyamide complexes with the minor groove of DNA. *J Am Chem Soc* 119:7909-7916.

88. Lewin, B (1999) *Genes VII* (Oxford University Press, New York).

89. Cate, JH, Gooding, AR, Podell, E, Zhou, K, Golden, BL*, et al.* (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* 273:1678-1685.

90. Luger, K, Mader, AW, Richmond, RK, Sargent, DF, Richmond, TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389:251-260.

91. Dingley, AJ, Grzesiek, S (1998) Direct observation of hydrogen bonds in nucleic ncid base pairs by internucleotide $^{2}J_{NN}$ couplings. *J Am Chem Soc* 120:8293-8297.

92. Dingley, AJ, Masse, JE, Peterson, RD, Barfield, M, Feigon, J*, et al.* (1999) Internucleotide scalar couplings across hydrogen bonds in Watson-Crick and Hoogsteen base pairs of a DNA triplex. *J Am Chem Soc* 121:6019-6027.

93. Jeffrey, GA, Saenger, W (1991) *Hydrogen Bonding in Biological Structures* (Springer-Verlag, New York).

94. Hvidt, A, Nielsen, SO, Anfinsen, CBJ, Anson, ML, Edsall, JT*, et al.* (1966) Hydrogen exchange in proteins. *Adv Prot Chem* 21:287-386.

95. Wagner, G (1983) Characterization of the distribution of internal motions in the basic pancreatic trypsin inhibitor using a large number of internal NMR probes. *Q Rev Biophys* 16:1-57.

96. Englander, SW, Kallenbach, NR (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys* 16:521-655.

97. Udgaonkar, JB, Baldwin, RL (1990) Early folding intermediate of ribonuclease A. *Proc of the Natl Acad Sci USA* 87:8197-8201.

98. Guéron, M, Leroy, J-L (1995) Studies of base pair kinetics by NMR measurement of proton exchange. *Methods Enzymol.* 261:383-413.

99. Pardi, A, Wagner, G, Wüthrich, K (1983) Protein conformation and proton nuclear-magnetic-resonance chemical shifts. *Eur J Biochem* 137:445-454.

100. Ösapay, K, Case, DA (1991) A new analysis of proton chemical shifts in proteins. *J A Chem Soc* 113:9436-9444.

101. Pardi, A, Hare, DR, Wang, C (1988) Determination of DNA structures by NMR and distance geometry techniques: a computer simulation. *Proc Natl Acad Sci USA* 85:8785-8789.

102. Cantor, CR, Schimmel, PR (1980) *Biophysical Chemistry* (W. H. Freeman, San Francisco).

103. Vakonakis, I, LiWang, AC (2004) Trans-hydrogen bond deuterium isotope effects of A:T base pairs in DNA. *J Biomol NMR* 29:65-72.

104. Abildgaard, J, Bolvig, S, Hansen, PE (1998) Unraveling the electronic and vibrational contributions to deuterium isotope effects on ^{13}C chemical shifts using *ab initio* model calculations. Analysis of the observed isotope effects on sterically perturbed intramolecular hydrogen-bonded O-hydroxy acyl aromatics. *J Am Chem Soc* 120:9063-9069.

105. Wang, AC, S. -G, K, Flynn, PF, Chou, S-H, Orban, J*, et al.* (1992) Errors in RNA NOESY distance measurements in chimeric and hybrid duplexes: differences in RNA and DNA proton relaxation. *Biochemistry* 31:3940-3946.

106. Vakonakis, I, LiWang, AC (2004) N1···N3 hydrogen bonds of A:U base pairs of RNA are stronger than those of A:T base pairs of DNA. *J Am Chem Soc* 126:5688-5689.

107. Shan, SO, Herschlag, D (1996) The change in hydrogen bond strength accompanying charge rearrangement: implications for enzymatic catalysis. *Proc Natl Acad Sci USA* 93:14474-14479.

108. Kankia, BI, Marky, LA (1999) DNA, RNA, and DNA/RNA oligomer duplexes: a comparative study of their stability, heat, hydration, and Mg^{2+} binding properties. *J Phys Chem B* 103:8759-8767.

109. Swart, M, FonsecaGuerra, C, Bickelhaupt, FM (2004) Hydrogen bonds of RNA are stronger than those of DNA, but NMR monitors only presence of methyl substituent in uracil/thymine. *J Am Chem Soc* 126:16718-16719.

110. Dunlap, JC, Loros, JJ, DeCoursey, PJ (2004) *Chronobiology: Biological Timekeeping*. (Sinauer, Sunderland, MA).

111. Emberly, E, Wingreen, NS (2006) Hourglass model for a protein-based circadian oscillator. *Phys Rev Lett* 96:038303.

112. Pettersen, EF, Goddard, TD, Huang, CC, Couch, GS, Greenblatt, DM*, et al.* (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605-1612.

113. Akiyama, S, Nohara, A, Ito, K, Maeda, Y (2008) Assembly and disassembly dynamics of the cyanobacterial periodosome. *Mol Cell* 29:703-716.

114. Ivleva, NB, Golden, SS (2007) Protein extraction, fractionation, and purification from cyanobacteria. *Methods Mol Biol* 362:365-373.

115. Ditty, JL, Canales, SR, Anderson, BE, Williams, SB, Golden, SS (2005) Stability of the *Synechococcus elongatus* PCC 7942 circadian clock under directed anti-phase expression of the kai genes. *Microbiology* 151:2605-2613.

116. Granot, J, Mildvan, AS, Bramson, HN, Kaiser, ET (1980) Magnetic resonance measurements of intersubstrate distances at the active site of protein kinase using substitution-inert cobalt(III) and chromium(III) complexes of adenosine 5'-(β,γ-methylenetriphosphate). *Biochemistry* 19:3537-3543.

117. Mildvan, AS (1997) Mechanisms of signaling and related enzymes. *Protein Struct Funct Genet* 29:401-416.

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

119. Kuzmic, P (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal Biochem* 237:260-273.

120. Clerico, EM, Ditty, JL, Golden, SS (2007) Specialized techniques for site-directed mutagenesis in cyanobacteria. *Methods Mol Biol* 362:155-171.

121. Saenger, W (1984) *Principles of Nucleic Acid Structure* (Springer-Verlag, New York).

122. Acharya, P, Cheruku, P, Chatterjee, S, Acharya, S, Chattopadhyaya, J (2004) Measurement of nucleobase pKa values in model mononucleotides shows RNA-

RNA duplexes to be more stable than DNA-DNA duplexes. *J Am Chem Soc* 126:2862-2869.

123. Legault, P, Pardi, A (1997) Unusual dynamics and pKa shift at the active site of a lead-dependent ribozyme. *J Am Chem Soc* 119:6621-6628.

124. Narlikar, GJ, Herschlag, D (1997) Mechanistic aspects of enzymatic catalysis: lessons from comparison of RNA and protein enzymes. *Annu Rev Biochem* 66:19-59.

125. Moody, EM, Brown, TS, Bevilacqua, PC (2004) Simple method for determining nucleobase pKa values by indirect labeling and demonstration of a pKa of neutrality in dsDNA. *J Am Chem Soc* 126:10200-10201.

126. Dziembowska, T, Hansen, PE, Rozwadowski, Z (2004) Studies based on deuterium isotope effect on ^{13}C chemical shifts. *Prog Nucl Magn Reson Spectrosc* 45:1-29.

127. Egli, M, Portmann, S, Usman, N (1996) RNA hydration: A detailed look. *Biochemistry* 35:8489-8494.

128. Chalikian, TV, Völker, J, Srinivasan, AR, Olson, WK, Breslauer, KJ (1999) The hydration of nucleic acid duplexes as assessed by a combination of volumetric and structural techniques. *Biopolymers* 50:459-471.

129. Wang, S, Kool, ET (1995) Origins of the large differences in stability of DNA and RNA helixes: C-5 methyl and 2'-hydroxyl effects. *Biochemistry* 34:4125-4132.

130. Mignon, P, Loverix, S, Steyaert, J, Geerlings, P (2005) Influence of the pi-pi interaction on the hydrogen bonding capacity of stacked DNA/RNA bases. *Nucleic Acids Res* 33:1779-1789.

131. Barfield, M, Dingley, AJ, Feigon, J, Grzesiek, S (2001) A DFT study of the interresidue dependencies of scalar J-coupling and magnetic shielding in the hydrogen-bonding regions of a DNA triplex. *J Am Chem Soc* 123:4014-4022.

132. Juranic, N, Likic, VA, Prendergast, FG, Macura, S (1996) Protein-solvent hydrogen bonding studied by NMR $^{1}J_{NC'}$ coupling constant determination and molecular dynamics simulations. *J Am Chem Soc* 118:7859-7860.

133. Juranic, N, Ilich, PK, Macura, S (1995) Hydrogen bonding networks in proteins as revealed by the amide $^{1}J_{NC'}$ coupling constant. *J Am Chem Soc* 117:405-410.

134. Manalo, MN, Kong, X, LiWang, A (2005) $^{1}J_{NH}$ values show that N1···N3 hydrogen bonds are stronger in dsRNA A:U than dsDNA A:T base pairs. *J Am Chem Soc* 127:17974-17975.

135. Asensio, A, Kobko, N, Dannenberg, JJ (2003) Cooperative hydrogen-bonding in adenine-thymine and guanine-cytosine base pairs. Density functional theory and Møller-Plesset molecular orbital study. *J Phys Chem A* 107:6441-6443.

136. Piotto, M, Saudek, V, Sklenář, V (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 2:661-665.

137. Markley, JL, Bax, A, Arata, Y, Hilbers, CW, Kaptein, R, *et al.* (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB inter-union task group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. *J Biomol NMR* 12:1-23.

138. Garrett, DS, Powers, R, Gronenborn, AM, Clore, GM (1991) A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J Magn Reson* 95:214-220.

139. Becke, AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* 38:3098.

140. Perdew, JP, Wang, Y (1992) Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B* 45:13244.

141. Frisch, MJ, Trucks, GW, Schlegel, HB, Scuseria, GE, Robb, MA*, et al.* (2004) Gaussian 03, Revision C.02 (Gaussian, Inc., Wallingford CT).

142. Hariharan, PC, Pople, JA (1973) The influence of polarization functions on molecular orbital hydrogenation energies. *Theor Chem Accounts Theor Comput Model Theor Chim Acta* 28:213-222.

143. Ditchfield, R (1974) Self-consistent perturbation theory of diamagnetism. *Mol Phys* 27:789 - 807.

144. Wolinski, K, Hinton, JF, Pulay, P (1990) Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. *J Am Chem Soc* 112:8251-8260.

145. Atkins, P (1998) *Physical Chemistry* (W.H. Freeman and Company, New York).

146. Foresman, JB, Frisch, AE (1996) *Exploring Chemistry with Electronic Structure Methods* (Gaussian Inc, Pittsburgh).

147. Russo, AA, Jeffrey, PD, Pavletich, NP (1996) Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat Struct Biol* 3:696-700.

148. Ollerenshaw, JE, Tugarinov, V, Kay, LE (2003) Methyl TROSY: explanation and experimental verification. *Magn Reson Chem* 41:843-852.

149. Tugarinov, V, Hwang, PM, Ollerenshaw, JE, Kay, LE (2003) Cross-correlated relaxation enhanced ^{1}H-^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* 125:10420-10428.

150. Sprangers, R, Velyvis, A, Kay, LE (2007) Solution NMR of supramolecular complexes: providing new insights into function. *Nat Methods* 4:697-703.

151. Velyvis, A, Yang, YR, Schachman, HK, Kay, LE (2007) A solution NMR study showing that active site ligands and nucleotides directly perturb the allosteric equilibrium in aspartate transcarbamoylase. *Proc Nat Acad Sci USA* 104:8815-8820.

152. Nishiwaki, T, Satomi, Y, Kitayama, Y, Terauchi, K, Kiyohara, R*, et al.* (2007) A sequential program of dual phosphorylation of KaiC as a basis for circadian rhythm in cyanobacteria. *EMBO J.* 26:4029-4037.

153. Sprangers, R, Kay, LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* 445:618-622.

154. Grzesiek, S, Cordier, F, Jaravine, V, Barfield, M (2004) Insights into biomolecular hydrogen bonds from hydrogen bond scalar couplings. *Prog Nucl Magn Reson Spectrosc* 45:275-300.

155. Manalo, MN, Perez, LM, LiWang, A (2007) Hydrogen-bonding and pi-pi base-stacking interactions are coupled in DNA, as suggested by calculated and

experimental trans-Hbond deuterium isotope shifts. *J Am Chem Soc* 129:11298-11299.

156. Kim, YI, Manalo, MN, Perez, LM, LiWang, A (2006) Computational and empirical trans-hydrogen bond deuterium isotope shifts suggest that N1-N3 A:U hydrogen bonds of RNA are shorter than those of A:T hydrogen bonds of DNA. *J Biomol NMR* 34:229-236.

157. Chalikian, TV, Völker, J, Srinivasan, AR, Olson, WK, Breslauer, KJ (1999) The hydration of nucleic acid duplexes as assessed by a combination of volumetric and structural techniques. *Biopolymers* 50:459-471.

158. Manalo, MN, Kong, X, LiWang, A (2007) Sensitivity of hydrogen bonds of DNA and RNA to hydration, as gauged by $^{1}J_{NH}$ measurements in ethanol-water mixtures. *J Biomol NMR* 37:257-263.

159. Wishart, DS, Bigam, CG, Yao, J, Abildgaard, F, Dyson, HJ*, et al.* (1995) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J Biomol NMR* 6:135-140.

APPENDIX A

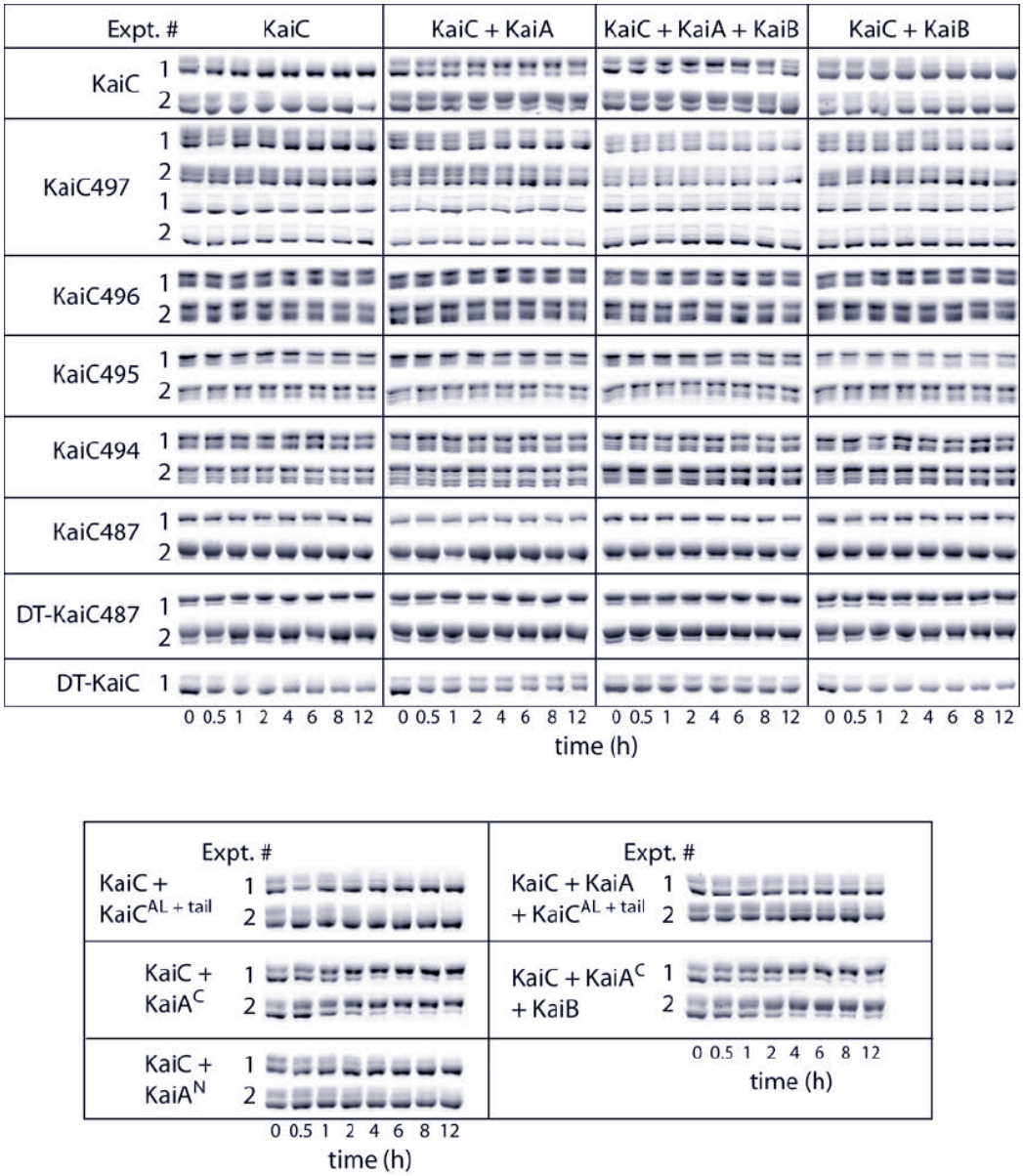

Figure A.1. SDS/PAGE gels of KaiC and KaiC variants ± KaiA ± KaiB as a function of time. These images were quantitated by densitometry for Figure 2.2. All KaiC samples were fresh, except for the second set of experiments for KaiC497, which were incubated 18 h at 30°C ± KaiA ± KaiB before data collection.

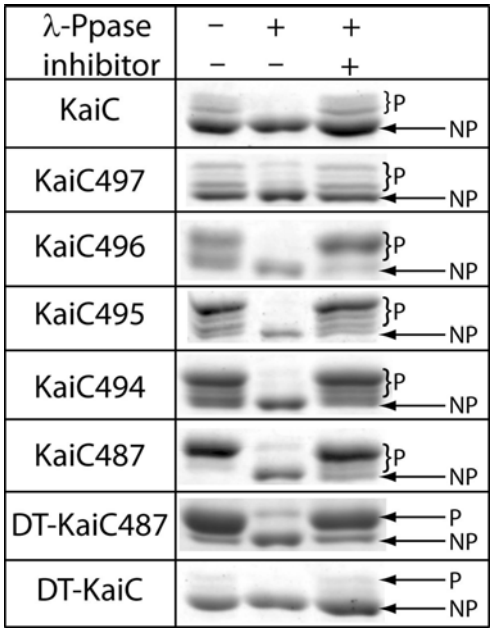

Figure A.2. Lambda phosphatase assays of KaiC and KaiC variants. KaiC protein (2.3 M) was incubated with λ-phosphatase (12000 U) in 50 L at 30 °C for six hours. λ−phosphatase inhibitor was a cocktail of 10 mM Na_3VO_4 + 50 mM NaF (final concentrations). Aliquots (14μL first and last lanes; 3μL middle lane) were loaded onto 6.5% SDS polyacrylamide gels. The gels were stained with Coomassie Brilliant Blue. The phosphorylated and unphosphorylated forms of KaiC are indicated as P and NP, respectively.

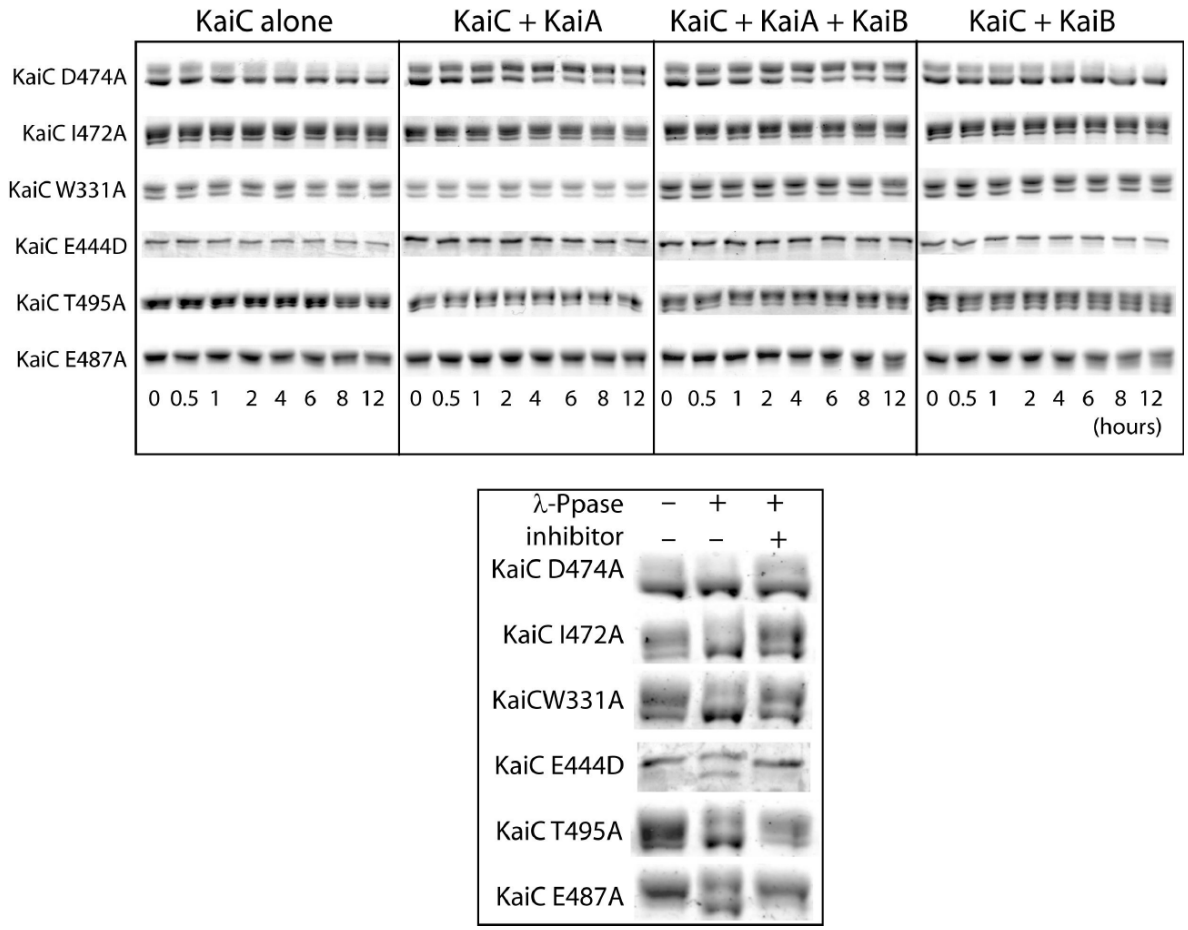

Figure A.3. SDS PAGE gels of KaiC D474A, KaiC I472A, KaiC W331A and KaiC E444D as a function of time ± KaiA ± KaiB. These images were quantitated by densitometry. Also shown are lambda phosphatase assays for KaiC D474A, KaiC I472A, KaiC W331A and KaiC E444D. The experiments and conditions used here are identical to those described earlier.

APPENDIX B

Table B.1. ^{1}H3 chemical shifts of RNA, DNA, RNA^{5mU}, and DNA^{dU}.[a]

Base Pair	RNA[b]	DNA[b]	RNA^{5mU}	DNA^{dU}
1, A5:U/T8	13.76	13.76	13.45	13.92
1, A6:U/T7	14.02	13.63	14.02	13.68
2, U/T3:A10	14.20	14.00	14.19	14.05
2, U/T4:A9	13.76	13.94	13.68	14.07
2, U/T5:A8	13.62	13.80	13.52	13.94
2, U/T6:A7	12.86	13.43	12.60	13.63
3, A3:U/T10	13.66	13.74	13.39	13.90
3, A4:U/T9	13.61	13.88	13.51	14.01
3, A5:U/T8	13.66	13.99	13.53	14.11
3, A6:U/T7	13.96	13.72	14.06	13.79
4, U/T3:A10	13.40	13.33[c]	13.14	13.45
4, A4:U/T9	13.12	13.26[c]	13.00	13.35
4, U/T5:A8	13.12	13.08[c]	12.92	13.24
4, A6:U/T7	13.11	13.11[c]	12.94	13.24
5, U/T5:A8	13.44	13.34	13.13	13.49
5, A6:U/T7	13.14	13.19	13.06	13.31

[a]Chemical shifts are given in parts per million relative to internal DSS. The five RNA, DNA, RNA^{5mU}, and DNA^{dU} sequences are **1** = r(CGCGAAUUCGCG)$_2$, **2** = r(CGUUUUAAAACG)$_2$, **3** = r(CGAAAAUUUUCG)$_2$, **4** = r(CGUAUAUAUACG)$_2$, and **5** = r(CGCGUAUACGCG)$_2$, where only the RNA sequences are given.

[b]Values are averages from additional data sets and those of Vakonakis *et al.* (103, 106).

[c]Values published earlier (106) were incorrectly assigned and are corrected here.

Table B.2. $^{13}C2$ chemical shifts of RNA, DNA, RNA^{5mU}, and DNA^{dU}.[a]

Base Pair	RNA[b]	DNA[b]	RNA^{5mU}	DNA^{dU}
1, A5:U/T8	153.65	154.75	153.83	154.79
1, A6:U/T7	154.42	155.32	154.30	155.23
2, U/T3:A10	154.71	155.61	154.74	155.45
2, U/T4:A9	153.54	154.52	153.61	154.26
2, U/T5:A8	153.36	154.41	153.46	154.23
2, U/T6:A7	152.62	154.42	152.81	154.57
3, A3:U/T10	153.59	155.04	153.77	154.96
3, A4:U/T9	153.38	154.38	153.45	154.23
3, A5:U/T8	153.52	154.22	153.53	154.06
3, A6:U/T7	154.44	155.29	154.14	155.14
4, U/T3:A10	154.13	155.35[d]	154.31	155.33
4, A4:U/T9	N.A.[c]	154.99[d]	153.71	155.11
4, U/T5:A8	N.A.[c]	154.76[d]	153.93	N.A.[c]
4, A6:U/T7	153.87	154.72[d]	153.83	N.A.[c]
5, U/T5:A8	154.03	155.14	154.23	155.16
5, A6:U/T7	153.84	154.89	153.80	155.03

[a]Chemical shifts are given in parts per million and indirectly referenced to internal DSS (159). The sequences are defined in the caption of Table B.1.

[b]Values are averages from additional data sets and those of Vakonakis *et al*. (103, 106).

[c]Not available due to resonance overlap.

[d]Values published earlier (106) were incorrectly assigned and are corrected here.

Table B.3. $^{2h}\Delta^{13}C2$ values of RNA, DNA, RNA^{5mU}, and DNA^{dU}.[a]

Base Pair	RNA[b]	DNA[b]	RNA^{5mU}	DNA^{dU}
1, A5:U/T8	-53 ± 1.2	-54 ± 1.2	-52 ± 2.2	-52 ± 0.9
1, A6:U/T7	-51 ± 1.5	-48 ± 1.4	-47 ± 0.8	-49 ± 0.1
2, U/T3:A10	-49 ± 2.8	-47 ± 1.3	-53 ± 0.3	-46 ± 0.2
2, U/T4:A9	-52 ± 2.2	-47 ± 2.0	-61 ± 1.6	-55 ± 0.8
2, U/T5:A8	-56 ± 1.6	-53 ± 1.0	-55 ± 1.8	-52 ± 1.2
2, U/T6:A7	-53 ± 1.3	-49 ± 1.5	-49 ± 1.1	-44 ± 1.2
3, A3:U/T10	-58 ± 1.2	-52 ± 1.3	-57 ± 1.2	-49 ± 1.9
3, A4:U/T9	-52 ± 1.0	-50 ± 0.7	-61 ± 0.8	-55 ± 0.9
3, A5:U/T8	-63 ± 0.6	-53 ± 0.6	-57 ± 2.1	-55 ± 1.2
3, A6:U/T7	-51 ± 0.6	-46 ± 1.2	-52 ± 1.2	-50 ± 0.4
4, U/T3:A10	-44 ± 1.3	-42 ± 0.9[d]	-44 ± 0.8	-40 ± 1.4
4, A4:U/T9	N.A.[c]	-44 ± 0.3[d]	-48 ± 0.6	-46 ± 0.7
4, U/T5:A8	N.A.[c]	-42 ± 0.8[d]	-45 ± 1.8	N.A.[c]
4, A6:U/T7	-49 ± 1.2	-44 ± 1.1[d]	-47 ± 1.5	N.A.[c]
5, U/T5:A8	-51 ± 0.9	-45 ± 1.3	-47 ± 1.8	-45 ± 1.4
5, A6:U/T7	-50 ± 0.5	-43 ± 0.8	-46 ± 1.4	-43 ± 1.1

[a] $^{2h}\Delta^{13}C2$ values are given in parts per billion. Each value is the average of 2–3 independent measurements. The sequences are defined in the caption of Table B.1.

[b] Values are averages from additional data sets and those of Vakonakis *et al.* (103, 106).

[c] Not available due to resonance overlap.

[d] Values published earlier (106) were incorrectly assigned and are corrected here.

For a general introduction to Gaussian and Gaussian input files, please see *Exploring Chemistry with Electronic Structure Methods* (Second edition) by James B. Foreman and Æleen Frisch; Copyright © 1993, 1995-96, Gaussian, Inc.: Pittsburgh, PA.

1. Gaussian input file for full geometry optimization of an AU base pair

```
#p bpw91/gen opt pop=full gfinput

AU opt

0 1
H       -0.000014       0.948603        -0.014271
N       -0.000027       1.996855        0.033740
C       -0.000011       2.638656        -1.186382
N       -0.000022       4.043277        -1.108695
C       -0.000029       4.685547        0.105936
C       -0.000042       4.031024        1.288922
C       -0.000037       2.578184        1.297593
O       -0.000043       1.884663        2.318912
H       -0.000057       4.557142        2.233148
H       -0.000027       5.769283        0.047786
O       0.000040        2.042769        -2.254344
C       0.000060        4.811446        -2.353756
H       -0.000332       4.099494        -3.176650
H       0.893086        5.440561        -2.415084
H       -0.892529       5.441207        -2.414759
N       0.000009        -0.847272       -0.081598
C       0.000032        -1.424063       -1.299529
N       0.000041        -2.722694       -1.594075
C       0.000049        -3.474845       -0.481708
C       0.000029        -3.025339       0.841228
C       0.000006        -1.623045       1.028449
N       0.000014        -4.083791       1.730558
C       0.000171        -5.143094       0.954999
N       0.000004        -4.850603       -0.395128
H       0.000214        -6.168162       1.302838
C       -0.000101       -5.773879       -1.518171
H       -0.892074       -6.406739       -1.501866
H       0.891803        -6.406835       -1.501920
H       -0.000100       -5.177638       -2.431470
N       -0.000019       -1.042878       2.238681
H       -0.000021       -1.631380       3.055732
H       -0.000033       -0.024444       2.328057
H       0.000018        -0.724561       -2.132376
```

C N O 0
6-31g(d)
****
1 31 32
6-31g(d,p)
****
9 10 13 14 15 25 27 28 29 33
6-31g(d)
****

2. Gaussian input file for geometry optimization of an AU base pair with constrained N1···N3 distance

#p bpw91/gen opt=modredund pop=full gfinput

AU opt with N--N = 2.75 A

0 1

H	0.000043	1.137983	0.036121
N	0.000053	2.169844	0.086076
C	0.000133	2.817623	-1.137834
N	-0.000001	4.235732	-1.051148
C	-0.000026	4.874057	0.168258
C	-0.000001	4.206089	1.356734
C	0.000011	2.751572	1.359992
O	-0.000013	2.049684	2.390515
H	-0.000037	4.733929	2.308610
H	-0.000084	5.965685	0.115479
O	-0.000010	2.228289	-2.220472
C	-0.000047	5.005501	-2.298036
H	-0.000283	4.282968	-3.122844
H	0.899711	5.638030	-2.361074
H	-0.899602	5.638342	-2.360796
N	0.000028	-0.577605	-0.032361
C	0.000065	-1.164452	-1.253224
N	0.000062	-2.473023	-1.549187
C	0.000029	-3.219100	-0.425431
C	-0.000009	-2.765675	0.906191
C	-0.000011	-1.355251	1.089016
N	-0.000012	-3.822333	1.804049
C	-0.000010	-4.891761	1.022838
N	-0.000034	-4.602526	-0.333884
H	-0.000028	-5.922453	1.376725
C	-0.000089	-5.526274	-1.461074

H	-0.898656	-6.162852	-1.445364
H	0.898378	-6.162992	-1.445322
H	-0.000023	-4.921954	-2.378497
N	-0.000027	-0.764088	2.299421
H	-0.000047	-1.352400	3.123967
H	-0.000025	0.266340	2.381587
H	0.000094	-0.464220	-2.094608

*B
1
2 F
3
4
5
6
7
8
9
10
11
12
13
14
15
16 F
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
C N O 0
6-31g(d)
****

1 31 32
6-31g(d,p)
****
9 10 13 14 15 25 27 28 29 33
6-31g(d)
****

3. Gaussian input file for a shielding calculation of the fully-optimized AU base pair

#p bpw91/6-31g(d) nmr pop=full gfinput
scf=(conver=8)

AU shielding at opt N-H

0 1

H				
N	1	B1		
C	2	B2 1	A1	
N	3	B3 2	A2 1	D1
C	4	B4 3	A3 2	D2
C	5	B5 4	A4 3	D3
C	2	B6 1	A5 3	D4
O	7	B7 2	A6 1	D5
H	6	B8 5	A7 4	D6
H	5	B9 4	A8 3	D7
O	3	B10 2	A9 1	D8
C	4	B11 3	A10 2	D9
H	12	B12 4	A11 3	D10
H	12	B13 4	A12 3	D11
H	12	B14 4	A13 3	D12
N	2	B15 3	A14 11	D13
C	16	B16 2	A15 3	D14
N	17	B17 16	A16 2	D15
C	18	B18 17	A17 16	D16
C	19	B19 18	A18 17	D17
C	16	B20 2	A19 3	D18
N	20	B21 19	A20 18	D19
C	22	B22 20	A21 19	D20
N	19	B23 18	A22 17	D21
H	23	B24 22	A23 20	D22
C	24	B25 19	A24 18	D23
H	26	B26 24	A25 19	D24
H	26	B27 24	A26 19	D25

H 26 B28 24 A27 19 D26
N 21 B29 16 A28 2 D27
H 30 B30 21 A29 16 D28
H 30 B31 21 A30 16 D29
H 17 B32 16 A31 2 D30
B1 1.068193
B2 1.384765
B3 1.420756
B4 1.376376
B5 1.363326
B6 1.400453
B7 1.246846
B8 1.088431
B9 1.092903
B10 1.232647
B11 1.465358
B12 1.096523
B13 1.101649
B14 1.101647
B15 2.820248
B16 1.354584
B17 1.341624
B18 1.348873
B19 1.406702
B20 1.364631
B21 1.386606
B22 1.324374
B23 1.386452
B24 1.089753
B25 1.457350
B26 1.101318
B27 1.101318
B28 1.098575
B29 1.347054
B30 1.012910
B31 1.033699
B32 1.094646
A1 115.129235
A2 114.392938
A3 121.128790
A4 123.031575
A5 117.305327
A6 121.197650
A7 121.652836

A8 114.862772
A9 123.547379
A10 118.191194
A11 107.092356
A12 110.504448
A13 110.503571
A14 115.422567
A15 118.141123
A16 128.417055
A17 110.836383
A18 127.615294
A19 122.271974
A20 111.551040
A21 103.497056
A22 127.366652
A23 124.902295
A24 125.549129
A25 110.816430
A26 110.815803
A27 107.291426
A28 119.228715
A29 118.461261
A30 120.590020
A31 114.558735
D1 179.993481
D2 0.008174
D3 -0.003750
D4 -179.998822
D5 0.001276
D6 -179.999041
D7 179.997160
D8 0.011187
D9 -179.994776
D10 0.017573
D11 -119.303662
D12 119.340205
D13 0.011236
D14 -0.001381
D15 179.999989
D16 -0.000690
D17 0.000566
D18 179.998159
D19 -179.998985
D20 -0.001534

D21 179.997364
D22 -179.998900
D23 -0.000486
D24 -119.208360
D25 119.213414
D26 0.002594
D27 -0.000502
D28 -179.999395
D29 0.000899
D30 -0.000494

4. Gaussian input file for an energy scan along the uracil imino N–H bond of the fully optimized AU base pair

#p bpw91/gen scan gfinput

AU scan

0 1

H				
N	1	B1		
C	2	B2 1	A1	
N	3	B3 2	A2 1	D1
C	4	B4 3	A3 2	D2
C	5	B5 4	A4 3	D3
C	2	B6 1	A5 3	D4
O	7	B7 2	A6 1	D5
H	6	B8 5	A7 4	D6
H	5	B9 4	A8 3	D7
O	3	B10 2	A9 1	D8
C	4	B11 3	A10 2	D9
H	12	B12 4	A11 3	D10
H	12	B13 4	A12 3	D11
H	12	B14 4	A13 3	D12
N	2	B15 3	A14 11	D13
C	16	B16 2	A15 3	D14
N	17	B17 16	A16 2	D15
C	18	B18 17	A17 16	D16
C	19	B19 18	A18 17	D17
C	16	B20 2	A19 3	D18
N	20	B21 19	A20 18	D19
C	22	B22 20	A21 19	D20
N	19	B23 18	A22 17	D21

H	23	B24 22	A23 20	D22
C	24	B25 19	A24 18	D23
H	26	B26 24	A25 19	D24
H	26	B27 24	A26 19	D25
H	26	B28 24	A27 19	D26
N	21	B29 16	A28 2	D27
H	30	B30 21	A29 16	D28
H	30	B31 21	A30 16	D29
H	17	B32 16	A31 2	D30

B1	1.068193	S	9 0.05
B2	1.384765		
B3	1.420756		
B4	1.376376		
B5	1.363326		
B6	1.400453		
B7	1.246846		
B8	1.088431		
B9	1.092903		
B10	1.232647		
B11	1.465358		
B12	1.096523		
B13	1.101649		
B14	1.101647		
B15	2.820248		
B16	1.354584		
B17	1.341624		
B18	1.348873		
B19	1.406702		
B20	1.364631		
B21	1.386606		
B22	1.324374		
B23	1.386452		
B24	1.089753		
B25	1.457350		
B26	1.101318		
B27	1.101318		
B28	1.098575		
B29	1.347054		
B30	1.012910		
B31	1.033699		
B32	1.094646		
A1	115.129235		
A2	114.392938		
A3	121.128790		

A4 123.031575
A5 117.305327
A6 121.197650
A7 121.652836
A8 114.862772
A9 123.547379
A10 118.191194
A11 107.092356
A12 110.504448
A13 110.503571
A14 115.422567
A15 118.141123
A16 128.417055
A17 110.836383
A18 127.615294
A19 122.271974
A20 111.551040
A21 103.497056
A22 127.366652
A23 124.902295
A24 125.549129
A25 110.816430
A26 110.815803
A27 107.291426
A28 119.228715
A29 118.461261
A30 120.590020
A31 114.558735
D1 179.993481
D2 0.008174
D3 -0.003750
D4 -179.998822
D5 0.001276
D6 -179.999041

D7 179.997160
D8 0.011187
D9 -179.994776
D10 0.017573
D11 -119.303662
D12 119.340205
D13 0.011236
D14 -0.001381
D15 179.999989
D16 -0.000690
D17 0.000566
D18 179.998159
D19 -179.998985
D20 -0.001534
D21 179.997364
D22 -179.998900
D23 -0.000486
D24 -119.208360
D25 119.213414
D26 0.002594
D27 -0.000502
D28 -179.999395
D29 0.000899
D30 -0.000494

C N O 0
6-31g(d)
****
1 31 32
6-31g(d,p)
****
9 10 13 14 15 25 27 28 29 33
6-31g(d)
****

5. Gaussian input file for frequency calculation of the fully-optimized AU base pair

#p freq=(noraman,readisotopes) bpw91/gen pop=full gfinput

AU freq

0 1
H 0.000043 1.137983 0.036121

N	0.000053	2.204935	0.087589
C	0.000134	2.852715	-1.136321
N	-0.000001	4.270824	-1.049636
C	-0.000026	4.909148	0.169771
C	-0.000000	4.241180	1.358247
C	0.000011	2.786664	1.361504
O	-0.000013	2.084776	2.392028
H	-0.000037	4.769020	2.310122
H	-0.000084	6.000776	0.116992
O	-0.000010	2.263380	-2.218959
C	-0.000046	5.040593	-2.296524
H	-0.000283	4.318059	-3.121331
H	0.899712	5.673121	-2.359561
H	-0.899601	5.673434	-2.359283
N	0.000028	-0.612696	-0.033873
C	0.000065	-1.199543	-1.254737
N	0.000061	-2.508115	-1.550700
C	0.000029	-3.254191	-0.426943
C	-0.000010	-2.800766	0.904678
C	-0.000011	-1.390343	1.087503
N	-0.000012	-3.857424	1.802536
C	-0.000010	-4.926853	1.021326
N	-0.000035	-4.637617	-0.335397
H	-0.000029	-5.957545	1.375212
C	-0.000089	-5.561365	-1.462587
H	-0.898657	-6.197944	-1.446877
H	0.898378	-6.198083	-1.446835
H	-0.000023	-4.957045	-2.380010
N	-0.000028	-0.799179	2.297908
H	-0.000047	-1.387491	3.122454
H	-0.000025	0.231249	2.380074
H	0.000093	-0.499312	-2.096121

298.15 1.0
1
14
12
14
12
12
12
16
1
1

16
12
1
1
1
14
12
14
12
12
12
14
12
14
1
12
1
1
1
14
1
1
1

C N O 0
6-31g(d)
****
1 31 32
6-31g(d,p)
****
9 10 13 14 15 25 27 28 29 33
6-31g(d)
****

VITA

Name: Yong-Ick Kim

Address: Dept. of Biochemistry/Biophysics
c/o Dr. Andy LiWang
2128 TAMU
College Station, TX 77843-2128

Email Address: yongick@tamu.edu

Education: B.S., Chemistry, Sung Kyun Kwan University, 1992
M.S., Organic Chemistry, Sung Kyun Kwan University, 1994
Ph.D., Biochemistry, Texas A&M University, 2008

Peer reviewed publications:

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. *Proc Natl Acad Sci USA* **105**:12825-12830.

Kim, YI, Manalo, MN, Perez, LM, LiWang, A (2006) Computational and empirical trans-hydrogen bond deuterium isotope shifts suggest that N1-N3 A:U hydrogen bonds of RNA are shorter than those of A:T hydrogen bonds of DNA. *J Biomol NMR* **34**:229-236.