# Genetic noise in the cyanobacterial circadian oscillator

by

Jeffrey R. Chabot

B.S. Physics
Rochester Institute of Technology, 1999

SUBMITTED TO THE DEPARTMENT OF PHYSICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN PHYSICS
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2005





Signature of author: _____
*Department of Physics*
April 8, 2005

Certified by: _____
*Alexander van Oudenaarden*
Associate Professor of Physics
Thesis Supervisor

Accepted by: _____
*Thomas Greytak*
Professor of Physics
Associate Department Head for Education

# Genetic noise in the cyanobacterial circadian oscillator

by

Jeffrey R. Chabot

Submitted to the Department of Physics on April 29, 2005
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in Physics

ABSTRACT

Clocks are generally considered as quintessential examples of accurate and precise devices. Biological clocks however, are continually subjected to intracellular and extracellular fluctuations that might reduce the fidelity of this periodic timer. One fundamental limitation that might set an upper bound on the fidelity is the stochastic nature of gene expression creating a noisy intracellular environment. Circadian rhythms, driven by oscillators which provide cells with an internal clock that controls the gene expression program, have been observed in a wide range of organisms from cyanobacteria to mammals. To explore the impact of stochastic expression fluctuations on the performance of the circadian clock, it is necessary to monitor single cells, since population experiments will average out these fluctuations.

The cyanobacterium *Synechococcus elongatus* PCC7942 is an excellent candidate for this study since its core circadian oscillator is well explored. We therefore measured, in single cells, the expression fluctuations of a fluorescent reporter driven by the cyanobacterial circadian oscillator. Repeated microscopic imaging of individual cells and their progeny revealed a robust circadian rhythm, and experiments with cells lacking the proposed central clock proteins confirm the crucial role they play in *Synechococcus*. Experiments conducted by microscopy and flow cytometry establish that the majority of genetic noise in *Synechococcus* arises from fluctuations uncorrelated between multiple genes (and therefore does not originate with a global clock noise). We observed via flow cytometry measurements a significant variability in cell-to-cell gene expression which is most pronounced during periods of net protein synthesis. An analytic model reveals that this nonequilibrium effect arises when protein lifetimes are comparable to the circadian period and the mean number of proteins translated per mRNA transcript is large. Stochastic Monte Carlo simulations verify this interpretation by providing a window into genetic behavior on an individual cell level. Estimates of the genetic noise supplied by intrinsic (uncorrelated), extrinsic (correlated within individual cells), and "clock" (unexplained sources with a 24 hour period) noise are developed. A quantitative description of noisy oscillators will be necessary for ultimately understanding the fidelity of circadian timekeeping and might inspire the design of robust synthetic oscillators.

Thesis supervisor: Alexander van Oudenaarden
Title: Associate Professor of Physics

# Table of contents.

# Table of figures.

# Table of tables.

# Preface.

In my experience as a scientist, I have found that (in retrospect) a direct path from the beginning to the end of a given project might be rather short; however, in practice, a much longer course is followed. I started as an undergraduate at the Rochester Institute of Technology with the notion of becoming a theoretical physicist. In fairly short order, I found that playing with toys was much more fun than playing with chalk. I then decided that experimental condensed matter physics (a catch-all phrase for "physics generally considered useful") was my path in life.

After performing several years of research in condensed matter labs at RIT and the National Institute of Standards and Technology, my final project before graduation utilized laser scattering to explore motility in a strain of *Pseudomonas* found in groundwater in Canada. I was fascinated by the challenges posed in doing work with biological organisms. Often, a physicist's instinctive reaction to a new system is "let's dunk it in liquid helium and see if it's superconducting" (or a similar notion); however, organisms in general object to such treatment, and any experimental probes need to be sufficiently gentle as to not destroy the system of interest.

Shortly after arriving at MIT, Alexander van Oudenaarden was accepted as a member of the physics faculty and was looking for a group of students to start his biophysics group. I joined at the beginning, and am proud to have helped design, build, outfit, and maintain the lab, and shape its goals over the past six years. Alexander stated in our very first group meeting that his expectation was nothing less than to be one of the top groups in the world in the fields we explore, and to hear the reactions of other scientists at conferences, I think we're well on our way.

My path to the cyanobacteria was also circuitous. My first project involved exploring the physics of individual biomolecules, and although it did not progress much past the construction and calibration of an optical tweezer, the techniques I learned in that short period of time are invaluable parts of my toolbox. I then joined in the exploration of the generation of force by the polymerization of actin proteins. For this project I did not only experimentation but also developed a theoretical model of the force balance and dynamics involved in polymerization-driven motion of lipid vesicles. This work was published as Upadhyaya, A., Chabot, J. R., *et al.*, Probing polymerization forces by using actin-propelled lipid vesicles, *Proc Natl Acad Sci U S A, 100*, 4521 (2003).

After this paper was published, I became aware of the study of circadian rhythms in cyanobacteria, which seemed to me a most remarkable prokaryotic system. As an anecdote, last fall I found myself at a marine biology field station on a small island off the coast of Belize, assisting my girlfriend Randi Rotjan (an excellent marine biologist). I was on the island for a couple weeks, during which time the Red Sox made it to the World Series (perhaps I should go more often). On the island, there are minimal cues as to the particular day of the week (only the resupply boat, which comes on Wednesdays). Within a couple days, nobody knows what day of the week it is. By comparison, these tiny prokaryotic cells, with only their femtoliter of possessions, can tell what time it is

with remarkable accuracy for months at a stretch with no assistance from external sources. Alexander was amenable to changing my study to this system, which was completely new to our lab. I have truly enjoyed the experience of developing the tools to probe this fascinating organism.

A few technical comments are in order. First, all temperatures in this thesis are given in centigrade, regardless of whether or not this is explicitly stated. Second, in diagrams of genetic regulation (such as Figure 2.2), a pointed arrow indicates a positive interaction (an enhancement of a rate or an increase in the levels of a species), while a blunted arrow indicates a negative interaction (repression or reduction). Third, when a particular biochemical species is indicated with italics (such as *kaiA*), this refers to the gene coding for the product. The protein product itself is in upright text and begins with a capital letter (KaiA). Promoters are indicated by a capital P followed by the gene they normally promote in subscripted italics ($P_{kaiA}$). Chromosomal insertions into neutral site I or neutral site II are indicated before the promoter by (I) or (II) respectively. Thus, an entire genetic construct may be listed as

$$(I)\ P_{kaiBC}::yfpLVA$$

meaning the gene *yfpLVA* coding for the YFPLVA protein, driven by the promoter of the *kaiBC* genes, and inserted in neutral site I of the chromosome. Fourth, when I refer to "noise" in the course of this work, I am referring to the standard deviation of the quantity in question normalized to its mean (a dimensionless ratio), unless explicitly stated otherwise. Finally, as the first member of my lab to study *Synechococcus*, I have attempted to compile my findings with respect to methodology into a single document (more readable than a lab notebook); therefore, the methods sections contain a significant amount of detail. I hope that future generations of graduate students may find this thesis a useful tool.

Finally, this work could not have been completed without the assistance of many individuals. I would like to thank my parents and sister for their love and encouragement of my interests in science. I also thank the members of my extended family, including Big Ron Morse (who can now honestly claim that his name appears in the archives of MIT). I thank the gentlemen of 291 Market Street, and the campers and staff of Camp Fatima Exceptional Citizens' Week, for constant support and friendship. I thank Mrs. (now Dr.) Swati Sharma of Manchester Memorial High School for driving my interest in mathematics, for following up on my activities after high school, and for motivating me to be a better person in many ways. I thank Professors Linda Meichle and Michael Kotlarchyk of RIT for teaching me to be an experimentalist and encouraging me to attend MIT, from which they are both alums. I thank Professor Alexander van Oudenaarden for giving me the freedom to explore different topics and entrusting hundreds of thousands of dollars worth of equipment to my care, which I feel has given me some of my most valuable training. I acknowledge and thank Michael J. T. O'Kelly and Inna Lipchina for assistance with data collection and cloning. I thank the members of my lab, both past and present, for support and wonderful discussions (both of science and life in general); in particular, Mukund Thattai, Juan Pedraza, Ertugrul Ozbudak, Han Lim, and Ben Kaufmann. Finally, I thank Randi Rotjan, whose support and curiosity for my work has exceeded even my own at times. To Randi I offer this one piece of advice: if ever you are uncertain of anything, always remember--google.

To dedicate this work to any one person would unjustly diminish the contributions of all the others. Therefore, I shall simply dedicate this thesis to the heroes of this story, the tiny workhorses who evolved chlorophyll and made our aerobic lives possible: the cyanobacteria. Their self-reliance, tirelessness, and determination to exploit every moment of the day to the maximum are models to us all. (Also, their circadian rhythm means that they are immune to the "hitting the snooze bar" habit.) So long, and thanks for all the oxygen.

Jeff Chabot
Cambridge, MA
April, 2005

# 1. Introduction.

This study is centrally concerned with the phenomenon of the circadian rhythm, an intrinsic clock giving organisms a way to monitor the passage of time in the absence of external cues, such as light and dark. These clocks are present in many organisms, ranging from simple fungi to mammals. Within the past fifteen years, it has been shown that cyanobacteria, some of the most abundant yet most simple organisms on the earth, also possess circadian rhythms capable of maintaining a remarkably accurate sense of time over extended periods.

To probe the genetic basis of these rhythms, the useful tool of bioluminescence has traditionally been employed. The bioluminescent luciferase protein causes the cells to emit light in proportional to the amount of bioluminescent protein in the cell. This reporter enables quick measurements of genetic expression in populations of cells, and has led to many discoveries related to the cyanobacterial clock mechanism (Chapter 2). However, it is not amenable to studying large numbers of individual cells.

In this study, fluorescent reporter proteins are used to probe genetic activity in the cyanobacterium *Synechococcus*. These reporters had not previously been utilized in this organism. To establish the properties of these reporters, the expression and decay kinetics of the proteins are determined. A method for shortening the half-life of the protein is employed and its effect quantified. The fluorescent reporter protein yields similar results as the luciferase reporter for bulk samples (Chapter 4).

Individual cells can be visualized with excellent spatial and temporal resolution and signal levels using the fluorescent reporter. Time-lapse "movies" reveal individual *Synechococcus* cells growing, dividing, and forming microcolonies under constant

conditions. The oscillation of the circadian clock, as reported by the fluorescence of the cells, is very clear in these time series. The importance of certain clock-related proteins at the individual cell level is verified by examination of cells with individual genes coding for these proteins disabled (Chapter 5).

One of the great advantages of the fluorescent reporter genes is the ability to construct histograms of large numbers of cells, enabling quantification of not only the mean level of expression (which can also easily be done using bioluminescence), but also the other moments of the distribution. In particular, the relative "noise" in the population (i.e., the standard deviation of gene expression normalized by its mean level) can readily be quantified using fluorescence (Chapter 6).

The variability across a population of individual, genetically identical cells can arise from several sources. These include differences in levels of various proteins important for gene expression (such as ribosomes, RNA polymerases, and so on) from cell to cell, and stochastic fluctuations in individual genes. Variability may also arise from noise in the circadian mechanism itself. A limit on the amount of noise injected into the system by the circadian clock is established (Chapter 8).

While numerous studies address noise in gene expression in other organisms, both prokaryotic and eukaryotic, these works have all explored steady state systems. The circadian clock of *Synechococcus* presents an opportunity to examine noise in a dynamic system, with variations over time governed by a robust oscillator which assures synchronicity between individuals, enabling the measurement of large numbers of cells with minimal noise introduced by decoherence. The cyanobacterial system is therefore used to probe nonequilibrium gene expression. Novel resulting effects, not predicted by existing steady-state theories, are modeled both analytically and numerically. These

models reveal time-dependent consequences of stochastic gene expression (Chapters 9 and 10).

The results of this work help explain how the circadian clock of *Synechococcus* can remain accurate for extended periods of time, and may prove useful in understanding the larger phenomenon of oscillatory regulation in general. In addition, knowledge of the methods employed by *Synechoccocus* to control genetic noise introduced by the clock mechanism may inspire the design of robust regulation by synthetic biological oscillators.

# 2. History.

## 2.1. General history.

One of the earliest endeavors of humanity has been the development of reliable means of delineating time and the passage thereof. Ancient sundials, hourglasses, and water clocks attest to our ancestors' desire to organize their lives according to a temporal schedule. Indeed, even nowadays our existence would be much more chaotic in the absence of wristwatches, alarm clocks, and personal organizers, which enable us to plan our days and nights to maximize the quantity and quality of labors we undertake.

Just as we have developed external devices for reporting time, humans and other organisms have evolved internal mechanisms to help us keep track of time and anticipate important events (such as the rising or setting of the sun). These mechanisms tell us, for example, when it is time to go to sleep (even if we are not particularly tired, or still have several hours of experiments to perform). Our body temperatures and endocrine systems are regulated with a 24 hour period (1). This phenomenon also manifests itself as "jet lag", as our internal clocks require some time to reset to a new environment.[1]

Humans are not alone in the requirement of organizing biological tasks according to the time of day. The reason for preparing for the arrival of daybreak is obvious: if this time can be anticipated, mechanisms can be in place in order to take advantage of those first rays of light, instead of wasting valuable time in activating systems to begin the preparation in response to those rays. The knowledge of internal clocks has existed for a long time. In 1729, the French astronomer J. J. d'Ortous deMairan performed

[1] The need to rapidly reset our internal clocks has not been evolutionarily selected, as the ability to rapidly change the portion of the day for which the sun is out has only recently arisen.

experiments (2) showing that a heliotropical plant would, even in a continually dark environment (lacking external cues of light and darkness), continue to open and close its leaves with roughly a 24 hour period. This behavior and other similar behaviors with an intrinsic, approximately 24 hour clock have been labeled "circadian rhythms" (from *circa* "around" + *diem* "day").[2]

## 2.2. Circadian rhythms.

In order to be considered a true circadian rhythm, three conditions must be satisfied by the mechanism being studied (3, 4). First, it must demonstrate a periodicity of approximately 24 hours, *even under constant conditions*. Typically, this means exposure to constant light levels and at a steady temperature and humidity; for studies performed in higher organisms, an even stricter definition of "constant" must be applied (including restricting body position and limiting visual cues in the environment) (5). Next, the mechanism must be temperature compensated such that a very minimal change in periodicity is observed with a change in thermal conditions. This condition can be restrictive in trying to custom design a synthetic "circadian" system from biological materials; the rates of many biological reactions vary strongly across physiological temperature ranges. However, just because the temperature has dropped by 10 degrees doesn't mean that the earth is rotating more or less quickly, and organisms mustn't be "fooled" by temperature changes. Finally, the phase of the rhythm must be resettable by

[2] Interestingly, most "circadian" rhythms have periods either shorter or longer than 24 hours. For example, the circadian rhythm of the fungus *Neurospora* displays a period of approximately 21.5 hours, whereas the clock of the squirrel monkey has a period of 25 hours. One possible reason relates to the facility of resetting the phase "forward" rather than "backwards"; another explanation considers the costs of "jumping the gun" versus getting a late start.

external cues (typically, light and dark). Since circadian rhythms are seldom exactly 24 hours, without the ability to resynchronize periodically, organisms would soon anticipate sunrise in the early evening, and the comparative advantages conferred by the circadian rhythm would quickly become a tremendous disadvantage.

Circadian rhythms appear to be ubiquitous across many species. Other mammals, ranging from primates like the squirrel monkey to rodents such as hamsters, have clear circadian rhythms, which help define times of greater mental activity. A blinded monkey will demonstrate extremely regular cycles of wakefulness/resting (6, 7), while hamsters will run on "exercise wheels" for only a certain part of the day, even in constant conditions. The circadian system of the fruitfly *Drosophila melanogaster* has been well studied, including to the genetic level (8, 9). The American lobster *Homarus americanus* displays circadian patterns of locomotor activity (10). The fungus *Neurospora crassa* has found a different use for its circadian clock: the reproductive cycle of the fungus is linked to a 24 hour clock, whereby the fungus will sporulate in the "late night" hours, and the spores will hatch and grow later (11).

All the organisms discussed thus far are eukaryotic and multicellular[3]. In some ways, it is not surprising that they may have evolved a circadian rhythm; the level of complexity present in these organisms leads to so many interactions at the genetic, cellular, tissue, and systemic levels that evolution surely has had many opportunities to select a time-keeping mechanism. However, there is another class of organisms that possess remarkably robust circadian rhythms, but are prokaryotic and (largely) unicellular: the cyanobacteria.

[3] *Neurospora* is multicellular for most of its life (as a filamentous fungus); the exception is the spore stage.

## 2.3. Cyanobacteria.

The cyanobacteria (also known as blue-green algae) are an ancient group of prokaryotic organisms. Cyanobacteria are a division (*Cyanophyta*) of the eubacteria, and are easily distinguished from the other bacteria as they contain chlorophyll.[4] There is evidence that cyanobacteria are the organisms who originally evolved chlorophyll, and became symbionts with larger eukaryotic cells some time later, becoming the chloroplasts found in plant cells today. Although bacteria do not generally contain much internal structure, cyanobacteria have a series of internal layers stacked like an onion (the thylakoid membranes) which contain the chlorophyll. Cyanobacteria can be unicellular, or may form filaments or larger colonial structures. Cyanobacteria occur ubiquitously across the world. They may be found in freshwater and saltwater environments (the organism *Prochlorococcus* is the most abundant autotrophic organism in the oceans, and is responsible for maintaining a good percentage of the total oxygen in the atmosphere (12)), and have been found in exotic environments including Antarctica (13). Most are nontoxic, and some may even be purchased freeze-dried from health food suppliers.

### 2.3.1. *Synechococcus elongatus* PCC7942.

One of the species of cyanobacteria often utilized as a model organism is a freshwater strain of *Synechococcus elongatus*, Pasteur Culture Collection (PCC) 7942. This unicellular cyanobacterium dwells in freshwater and is an obligate photoautotroph. The chromosome has recently been completely sequenced and contains approximately 2.7 megabases; approximately 2600 candidate protein-encoding genes have been

[4] Other bacteria are also photosynthetic, but do not use chlorophyll as the light-absorbing pigment for food synthesis. Also, only cyanobacteria use the full C3 Calvin cycle as the photosynthetic pathway.

identified. The organism also contains two endogenous plasmids, of sizes 8 and 46 kilobases. In 1970, it was shown that this organism (then known as *Anacystis nidulans* R2) could be easily transformed with exogenous DNA (14). *Synechococcus* PCC7942 will also grow readily on solid media, which further facilitates genetic study. Finally, the doubling time of *Synechococcus* (seven to eight hours under optimum conditions) is among the fastest for cyanobacteria, which makes its study somewhat easier than other strains, which may have doubling times of up to two days.

### 2.3.2. Circadian rhythms in cyanobacteria.

It was suspected for some time that cyanobacteria may be candidates for circadian rhythms. Many species of blue-green algae do not uptake nitrate from their environment, but rather prepare it themselves from diatomic nitrogen. The fixation of nitrogen is severely hindered by the presence of oxygen. Oxygen is a byproduct of photosynthesis, which is necessarily done during daylight hours. Since cyanobacteria do not have internal organelles to separate these two processes spatially, it is naturally suggestive that they might be separated temporally. Photosynthesis can occur during the day; nitrogen fixation will take place at night (15). As these two mechanisms occur using their own cellular machinery, the ability to anticipate the change of light to dark and vice versa can raise the overall efficiency of the cell. In a factory, a smart boss (with a good clock) will make sure the next shift has arrived before sending the first shift home. Cyanobacteria were believed to be very good managers. In 1990, evidence was found for circadian basis to these processes in a strain of *Synechococcus*. Later, some evidence was found for a circadian nature to both amino acid uptake (16) and cell division (17) in other

strains of *Synechococcus*. The next step was to find a genetically tractable system for investigating the basis of circadian rhythms in blue-green algae.

## 2.4. Circadian rhythms in *Synechococcus*.

In 1994, a group of researchers (primarily from Texas A&M, Vanderbilt University, and the National Institute for Basic Biology in Okazaki, Japan) identified *Synechococcus* PCC7942 as containing a robust circadian rhythm[5] (18). Northern blots of mRNA from the photosystem II gene *psbAI* were prepared from samples collected every three hours from *Synechococcus* growing originally in an alternating 12 hours light:12 hours dark environment, then transferred to constant light conditions. (The alternating light/dark cycling is indicated as LD; in contrast, a constant light environment is labeled LL.) The Northern blots appeared to indicate that the mRNA of the gene was not produced at a constant rate, but rather seemed to cycle with a 24 hour frequency, with the peak to trough ratio of the oscillation appearing to diminish over time. However, Northern blots are not the most accurate means of testing the levels of gene expression in a population, and a technique that allows for repeated *in vivo* measurements of cells was sought. The answer to this problem came like a flash from the blue. (Actually, approximately $10^9$ single-photon flashes from the blue-greens.)

[5] Interestingly, *Synechococcus* PCC7942 does not fix nitrogen (the original reason thought to lead to cyanobacterial circadian rhythms), but rather requires it as an externally supplied nutrient. Other, more primitive cyanobacterial species such as *Prochlorococcus*, as well as other species of *Synechococcus*, do fix nitrogen. It is possible that PCC7942 has lost the ability to fix nitrogen, while retaining the circadian rhythm.

### 2.4.1. Probing rhythms with bioluminescence.

These researchers incorporated the bacterial luciferase genes *luxAB* (originally isolated from *Vibrio harveyi*) under the control of the *psbAI* promoter into the cyanobacterium's chromosome. A site was chosen for the integration that did not lead to any apparent change in phenotype when disturbed (and is called a "neutral site" for this reason). After entrainment by several LD cycles, the cells were transferred to LL. LuxAB requires an aldehyde substrate to produce light, and this was introduced into the cell-containing media. To collect data, the cells were placed in the dark for a total of five minutes, during which time light from the population of cells was collected by a photomultiplier tube. The cells were then returned to the LL environment until the next measurement. When the bioluminescent signal was plotted versus time, a remarkable behavior emerged: the luminescence displayed a very clear oscillation with a period of very nearly 24 hours that persisted for at least a week (see Figure 2.1).

### 2.4.2. Identification of clock genes.

Once this simple organism was conclusively shown to possess a circadian rhythm, the next logical question became: what is the root cause of this rhythm? Many of the potential systems in higher eukaryotic organisms are not present in cyanobacteria. A logical approach to finding the key players in this drama involved a "shotgun" type approach. Large libraries of mutants with the bioluminescence reporter described above were randomly mutated, and the resulting microcolonies were synchronized in an LD environment and screened in an LL environment. Colonies that showed clear 24-hour rhythms implied that their mutation affected either a nonsense piece of the chromosome,

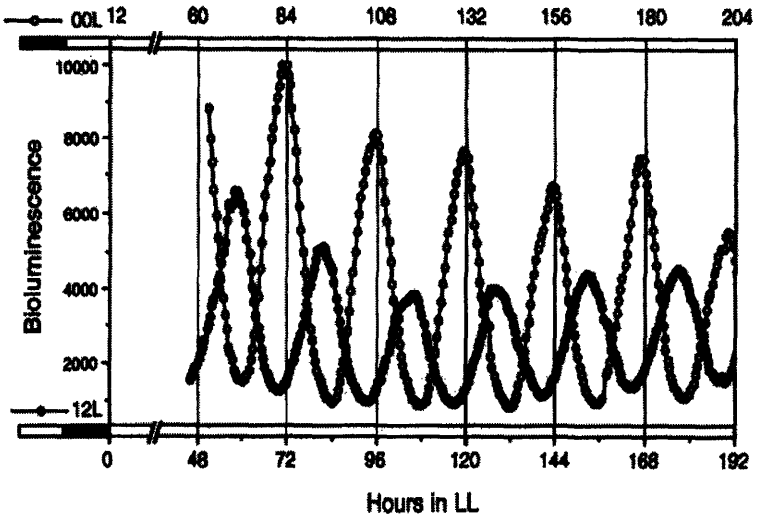


**Figure 2.1.** Circadian oscillation of bioluminescence levels in *Synechococcus* PCC7942 (Kondo *et al.*). The *luxAB* genes were expressed under the control of the *psbAI* promoter. Cells were synchronized in a 12:12 LD cycle. Open and closed circles represent cultures synchronized 12 hours out of phase, as shown on the left. (From (18); copyright 1993 National Academy of Sciences, U.S.A.)

---

or a gene whose loss was nonlethal but not part of the clock. On the other hand, a number of mutants showed either a loss of the rhythm from the bioluminescence signal,[6] or a significantly altered rhythm (with a longer or shorter period). The mutated DNA sequences were identified in the cells with altered clocks. All the mutations were found to lie within three genes in an operon. The genes were named *kaiA*, *kaiB*, and *kaiC* (from the Japanese word *kaiten* meaning "turning of the heavens", referring to the daily cycle the *kai* proteins cause).

KaiA is a 284 residue protein expressed under its own promoter ($P_{kaiA}$). KaiB (102 residues) and KaiC (519 residues) are expressed bicistronically under another

[6] These mutants continued to express the *luxAB* gene; the amplitude of the oscillating part of the signal appeared to be completely abolished.

promoter ($P_{kaiBC}$) immediately after *kaiA* in the chromosome. There is evidence that KaiA and KaiB can bind to KaiC under certain conditions. KaiC appears to form hexamers and has motifs for binding ATP similar to members of the RecA/DnaB superfamily of proteins (19, 20). As a hexamer, KaiC can bind to DNA under certain conditions (19). Mutation of *kaiC* may lead to clocks with different periods, either longer or shorter (21, 22). The particular placement of *kaiABC* in the chromosome does not appear to matter; in a strain with the wild type cluster removed and then re-inserted in a neutral site, full rhythmic ability appears to be restored (23). Close homologs of *kaiB* and *kaiC* appear in other cyanobacterial strains (24), as well as some other bacteria and archaebacteria (25). On the other hand, *kaiA* appears to be more unique to *Synechococcus*: other cyanobacteria contain fragments of the *kaiA* gene, while others contain no clear homologs (26, 27).

### 2.4.3. Additional components of circadian system.

Over time, additional proteins were identified as playing roles in the circadian dance of *Synechococcus*. These proteins play roles in the environmental input to the clock, or affect the operation of the core oscillator.

Two proteins that help *Synechococcus* synchronize its functioning to the environment are CikA and LdpA. CikA (*c*ircadian *i*nput *k*inase) plays a crucial role in resetting the phase of the circadian clock. This protein is a bacteriophytochrome (meaning that it interacts with light), forming an important bridge between the environmental condition and the core of the circadian clock. When its gene is inactivated, the rhythm is shortened, the phase of the oscillation is shifted, and the cells lose much of the ability to synchronize to external darkness cues (28). LdpA (*l*ight

*d*ependent *p*eriod), on the other hand, appears to help the cyanobacterium adjust its clock according to the light levels present in the environment (29). According to Aschoff's rule[7], the period of many circadian oscillators depends on the intensity of light in the environment[8]. Cells with *ldpA* knocked out lose this ability.

Other proteins that affect the functioning of the clock are CpmA, SasA, and Pex. CpmA (*c*ircadian *p*hase *m*odifier) plays a role in the way the clock regulates certain promoters. When the protein's function was modified by mutation, the phase and amplitude of the *psbAI* promoter was significantly altered (35). SasA (S*ynechococcus* *a*daptive *s*ensor) apparently interacts with KaiC and, when disrupted, resulted in lower *kaiB* and *kaiC* expression and shortened the period of the circadian clock (36). The role of Pex (*p*eriod *ex*tender) is less clear; what is known is that overexpression of Pex causes a longer oscillation period, and knocking *pex* out causes both a marginally shorter period and significantly greater *kaiA* expression (37).

### 2.4.4. Extent of circadian influence in *Synechococcus*.

How pervasive is the circadian rhythm in the cyanobacterial way of life? To test the scope of the control exerted by the clock, promoterless luciferase genes were randomly inserted at locations throughout the chromosome. The resulting colonies were screened for detectable bioluminescence levels. If a colony displayed a significant

[7] After Jurgen Aschoff, who did early studies of circadian rhythms in wide ranging systems, including humans (30), chaffinches (31) , chickens (32), and hamsters (33), as well as a very interesting study on the importance of naps (34).

[8] The period of diurnal species (including primates) goes down with increasing light levels, while nocturnal species (such as rodents) see their period go up (7). The cyanobacterial period varies inversely with light level, consistent with organisms that value the light.

bioluminescence, this result implied that the luciferase genes were inserted downstream of some promoter element, and therefore were measuring the gene expression of that promoter.[9] After screening 800 mutants exhibiting bioluminescence, nearly all showed very clear circadian variation in expression level[10]. Apparently, the clock mechanism exerts a global control over the genome in *Synechococcus* (38).

The precise mechanics of this control are not exactly known, but a clever experiment sheds some light on a candidate. As previously described, KaiC is known to form hexamers when bound to ATP, and these hexamers may bind to DNA using a motif similar to other proteins which can modify the topology of the chromosome. Some of these proteins, called helicases, can modify the supercoiling state of the DNA. Attempting to extract the chromosome and testing its level of supercoiling presents significant experimental difficulty. To get around this problem, one group of researchers extracted and purified one of the endogenous plasmids from *Synechococcus* at various times during the circadian cycle (a much more tractable proposition). The resulting samples were pushed through an agarose gel by electrophoresis. Even though the plasmids were all of the same length, if they happened to be more tightly supercoiled at one time than another, they would migrate through the gel more quickly under a given applied electric field. Their results showed a 24 hour pattern to the migratory speed of the plasmids (C. H. Johnson, unpublished results). If, in fact, KaiC affects the chromosome in the same way the plasmids are affected, the global mechanism for genetic

[9] The actual strength of the promoter may not have been accurately reflected, as the luciferase genes may have been inserted at different distances from the active elements of the promoters, potentially modifying their transcriptional efficiency. Regardless, variations in promoter activity throughout the day should have been faithfully reproduced in the time pattern of luciferase protein expressed.

[10] Phase variations exist between different promoters. This is a logical result if the clock serves to separate various processes temporally.

regulation may have been found.

The utility of the circadian rhythm is evident from studies comparing the performance of strains of *Synechococcus* with clocks of differing periods (made by mutation of the *kai* genes, particularly *kaiC*) (39). When two strains were seeded in equal numbers in LD environments where the period of light was "matched" to the natural frequency of one of the strains, that strain would massively outcompete the other[11]. This also held for wild type cells in short (11:11 LD) or long (15:15 LD) environments when competing against cells with appropriately mutated periods; this removes the possibility that the mutations themselves may cause an impediment to cell growth.

## 2.5. Models of circadian oscillation in *Synechococcus*.

### 2.5.1. Activator-repressor model (Barkai & Leibler).

Several models of the core molecular mechanism of the cyanobacterial circadian clock have been proposed. An early supposition was that the circadian clock of *Synechococcus* mimicked circadian mechanisms in other organisms. A number of papers have been written which attempt to model biological networks which can give rise to stable oscillations. Two such papers are (40) and the follow-up study (41). The model described in these papers model drew heavily from comparisons with other, better-known eukaryotic clocks (see Figure 2.2). A common theme in several circadian clocks is the interplay between an activator (which will increase the expression of genes targeted by the clock, including itself and the second gene) and a repressor, which can inhibit the

[11] This is to say, the composition of cells in the sample some time later (in (39), about seven generations) would be biased largely towards the strain matched to the LD period.

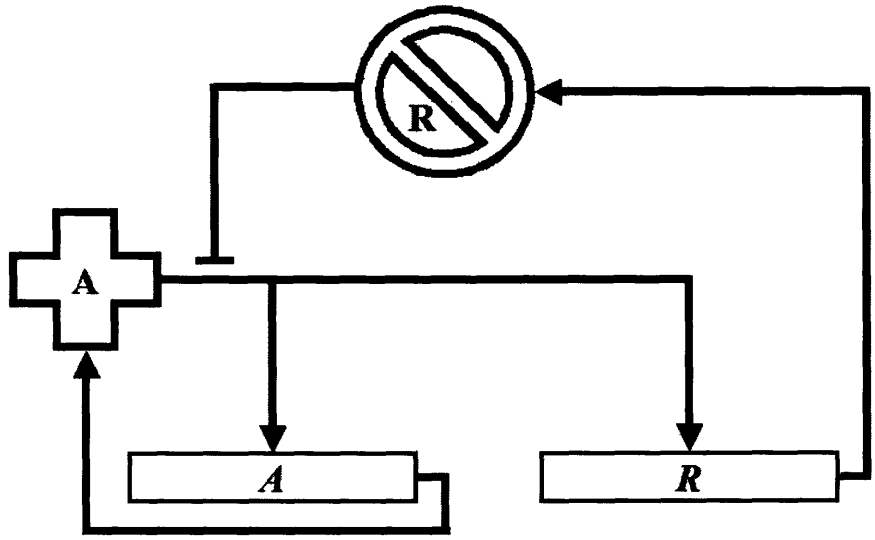


**Figure 2.2.** Scheme of activator-repressor model (Barkai & Leibler). In this model, genes (*A*, *R*) for an activator (A) and a repressor (R) are positively regulated by the activator protein. The repressor can bind to the activator, effectively eliminating its regulatory activity. When rate constants for the various steps are appropriately chosen, the levels of the A and R undergo oscillations. KaiA is assumed to be the activator; KaiB and KaiC are assumed to be repressors. Modified from (40) and (41).

---

activity of the activator[12]. A list of such proteins in several organisms is given in Table 2.1.

**Table 2.1.** Clock genes of several organisms.

| Organism | Activator | Repressor |
|---|---|---|
| *Neurospora* (fungus) | Wc1-2 | Frq |
| *Drosophila* (fruitfly) | Clc, Cyc | Tim, Per |
| Mouse | Clock, Bmal | Tim, Perl, Per2, Per3 |
| *Synechococcus* (proposed in (40)): | KaiA | KaiB, KaiC |

[12] This action may occur by complexing between the repressor and activator, which subsequently sequesters the activator and prevents it from enhancing the expression of the targeted genes.

The proposed molecular scheme results in oscillations in the following way. Initially, there is a low concentration of both activator (A) and repressor (R). The free activator rapidly causes an increase in A, concurrent with a slower increase in R (the production rates are chosen such that A's gene is transcribed ten times more than R, and each mRNA for A is translated five times more than mRNAs of R). Eventually, the population of R "catches up" to the population of A (owing to the much greater decay rate of A than R); at this time, the repressor complexes with nearly all of the available activator. Over time, the activator degrades, first leaving a large pool of repressor, which then slowly degrades. Eventually, we return to the conditions at the beginning of the cycle, with low levels of both R and A, at which time the activator level spikes again and the cycle continues.

This model of circadian rhythms, at least as it applies to *Synechococcus*, has some notable drawbacks. First, it was proposed at a time when the specific functions of the Kai proteins were not completely (or, verily, even partially) known. The roles of the Kai proteins are, at best, not as clear-cut as a simple "activator" and "repressor": KaiC is only sometimes a repressor; KaiA helps activate the repression; KaiB reduces the repression ability of KaiC; there is no known circadian activator that binds to specific sites on the chromosome (particularly not the promoters of the circadian genes). Second, as will be discussed later, there is some evidence that the role of transcription and translation is not essential to the core of the clock (42). The model also critically depends upon fifteen parameters, and involves assumptions that (for example) one type of mRNA must be twenty times less stable than another, while being translated at ten times the rate. In general, such "fine-tuned" models cannot often be lifted from one system and readily

applied to another; this is a problem with attempting to paint all circadian systems with a single brush stroke. What this model does achieve is a mathematical description of a two component system including a single unstable fixed point leading to oscillations. What it fails to do is describe reality, at least in the case of *Synechococcus*. It was clear that a more correct model would emerge from the sea of potential candidates only with further experimental study to better quantify and qualify the specifics of the cyanobacterial circadian mechanism.

### 2.5.2. KaiC phosphorylation and genetic repression (Xu *et al.*).

Two more experimentally based models emerged some time later. The first model ((43); see Figure 2.3) incorporates information showing that the particular promoters of the *kai* genes are not required for circadian oscillation. The *kaiBC* promoter was replaced with an inducible promoter, and expression was regulated to produce protein levels comparable with native quantities in wild-type cells. The resulting strains showed oscillating bioluminescence signals stable for over a week[13]. Additionally, the roles of some of the Kai proteins were better elucidated. First, in cells lacking the *kaiA* gene, nearly the entire population of KaiC was unphosphorylated (44). This implies that KaiA plays a critical role in either promoting or stabilizing the phosphorylation of KaiC (or both). In *kaiB* knockout strains, the population of KaiC shifts to being nearly completely phosphorylated. The interpretation of these results is that KaiB antagonizes the activity of KaiA. The end result of this "tug of war" is that KaiC moves between active (phosphorylated) and inactive (unphosphorylated) states, modulated by the relative

[13] This stands in stark contrast to eukaryotic clocks, where the particular promoters of the circadian genes are required to establish proper feedback in the system and maintain the oscillation.

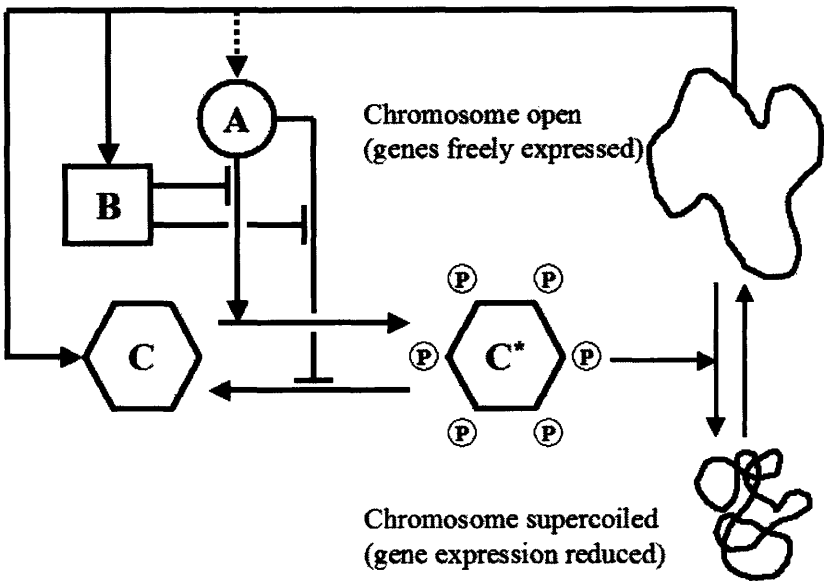


**Figure 2.3.** Scheme for KaiC phosphorylation and genetic repression model (Xu *et al.*). In this model, KaiC moves between unphosphorylated (C) and phosphorylated (C*) states. KaiA (A) both promotes the phosphorylation and reduces dephosphorylation. KaiB (B) antagonizes this action of KaiA. Phosphorylated KaiC causes the chromosome to become supercoiled; in this state, gene expression is limited. As the levels of the Kai proteins diminish, the chromosome returns to its open state; at this time, the levels of the Kai proteins rise again and the cycle repeats. The arrow from the chromosome to KaiA is dotted to indicate that KaiA seems to be less affected by the state of the chromosome, as experiments show that the level of KaiA seems to be more constant than KaiB and KaiC throughout the cycle. Modified from (43).

---

levels of KaiA and KaiB. A final piece of information is that the level of KaiA protein appears to be relatively constant throughout the circadian cycle (45). The model incorporates these facts in the following scheme. As KaiB and KaiC are produced from an initial state with relatively low levels, KaiA is readily able to convert KaiC to its active form; as the levels of KaiB and KaiC increase, this ability is modulated by the antagonistic effects of KaiB. The relative amounts of KaiC in each phosphorylation state

determine its impact on the supercoiling state of the chromosome, which in turn affects the expression of the *kai* cluster (as well as all other genes in the cyanobacterium).

### 2.5.3. KaiB spatiotemporal localization (Kitayama *et al.*).

The second model (46) focuses on another piece of experimental evidence. In protein purification, cells are lysed and fractions are collected which contain mainly cytosol or mainly membrane components. While KaiA and KaiC (in both phosphorylation states) are found predominantly in the cytosol, KaiB seems to oscillate (with a period matching the clock) between being found primarily in the membrane fraction to being present in significant amounts in the cytosol. Using this spatiotemporal oscillation of KaiB as a critical process in the rhythm, these authors suggested the following model (see Figure 2.4). During the "day", KaiB is located in the membrane, which allows KaiA to drive the KaiC population into its phosphorylated state. As "night" comes, KaiB emerges from the membrane, causing KaiA-KaiC complexes to break and returning KaiC to its unphosphorylated condition. As day arrives, the KaiB returns to the membrane and the cycle repeats. This model has the advantage of incorporating the localization of KaiB as a function of time, as well as not requiring a transcriptional feedback to maintain the clock. However, the mechanism by which KaiB moves between the membrane and cytosol is not explained, and so this model may have just changed the question instead of answering it.

Other "cartoon" models have been proposed (see, for example, (47)) that attempt to include other proteins such as SasA and CikA, but the mechanisms through which they interact with the core clock proteins are largely speculative. Additional

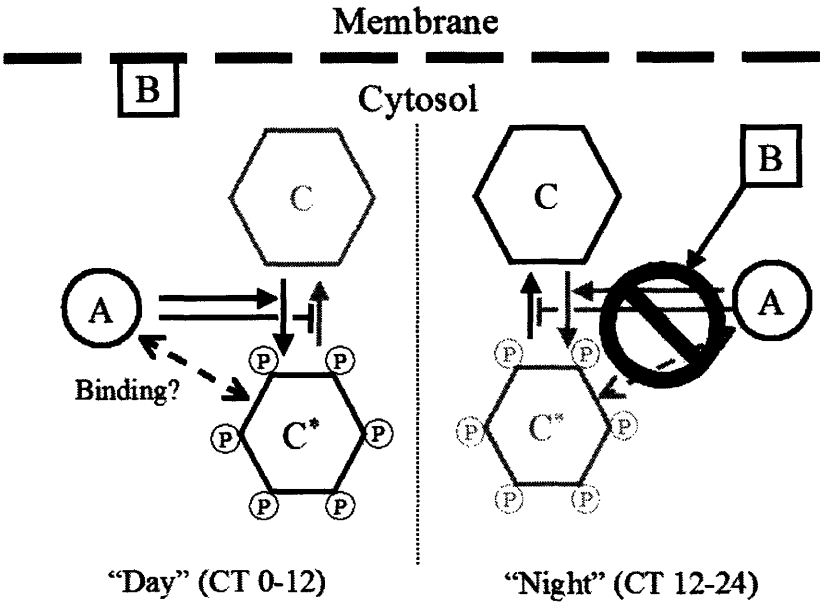


**Figure 2.4.** Scheme for KaiB spatiotemporal localization model (Kitayama *et al.*). In this model, during the "day" (either the time in light in an LD environment, or the first 12 hours of the circadian day in constant light), KaiB (B) localizes in the membrane. This allows KaiA (A) to drive the KaiC population into the phosphorylated state ($C^*$; C represents the unphosphorylated state), and complexes may form between these two proteins. As "night" arrives, the KaiB emerges from the membrane. KaiB interacts with the other two Kai proteins, resulting in the dissociation of complexes, and the antagonism of the KaiA activity. This causes the KaiC protein to return to its unphosphorylated state. When "day" arrives again, KaiB returns to the membrane and the cycle repeats. Modified from (46).

---

biochemistry experiments in the future will undoubtedly lead to a refined understanding of the cast of characters, and our knowledge of the plot of this circadian story will continue to grow.

## 2.6. Recent discoveries.

### 2.6.1. Single cell bioluminescence.

Recently, several studies have revealed information about previously uninvestigated aspects of the circadian oscillation of *Synechococcus*. A study published in July 2004 (48) utilized bioluminescence not as a reporter of bulk population state, but rather as a probe of genetic activity on an individual cell basis. Using an ultra-low noise CCD detector, cells growing on a slide were monitored for up to two weeks. Data was collected every 95-125 minutes, with exposures up to 30 minutes in order to integrate sufficient signal above background. By doing these experiments, the individual cells and the resulting microcolonies were shown to display circadian rhythms under "constant" conditions[14]. This study explored phase noise in *Synechococcus* individuals and microcolonies, and determined that the clock was stable for long periods of time, with strong temporal correlations extending for several months[15].

### 2.6.2. Cell-cell communication.

Another aspect of cyanobacterial circadian rhythms investigated in this study concerned communication between cells in the population. In some eukaryotic circadian systems, communication between various cells appears to play an important role in the stability of the clock (49). In contrast, *Synechococcus* was believed to be a unicellular oscillator, without communication between members of the bulk population. Theoretical studies (and at least one experimental work (50)) have shown that individual noisy

[14] The quotation marks will be explained shortly.

[15] A correlation time of 166 ± 100 days was calculated from the observed phase diffusion in the microcolonies studied; however, attempting to perform fits of curves with characteristic times so much longer than the duration of the experiment introduces the large uncertainty reported.

oscillators, if they are coupled, can collectively produce an oscillation with a stable, uniform period and amplitude, with the robustness of the rhythm increasing with the degree of coupling. To test if *Synechococcus* individuals were communicating to minimize the population phase noise, differently synchronized cells were mixed on the microscope and data was taken as before. Microcolonies resulting from individual cells synchronized 12 hours out of phase with each other would occasionally grow into contact with each other. By studying individuals growing along the interface between the two microcolonies, the tendency of an individual to change its phase to match its neighbor could be investigated. In this study, the average coupling strength observed was zero.

A few concerns may be legitimately raised by the experimental techniques used in the study described above. First, the cells were supposedly maintained under "constant" conditions (e.g., constant light levels and temperature). However, in order to integrate the level of signal required to exceed the noise from the detector, the cells were placed in the dark for 30 minutes every one to one and a half hours[16]. Ideally, a "constant light" environment would not have such a dramatic variation for 25 to 33 percent of the time. Second, although evidence exists that the circadian period is independent of the cell cycle in cyanobacteria (51, 52), artifacts may remain from cell division that affect the results of the experiment. For example, one of the primary mechanisms for lowering the number of protein molecules in a cell is dilution by cell division; if the cells elongate at different rates as they approach cell division, this can lead to an oscillating signal, even if the gene output is actually constant. In another example, at some point before the cells double, the

[16] Even after this integration time, the signal levels collected were less than those obtained by cosmic rays. However, due to the comparatively high signal levels and tight localization on the detector, these could be easily removed from the resulting images.

chromosome is replicated, resulting in duplicate copies of the reporter genes; this may cause an expression profile that oscillates with the same period as cell doubling (53). Finally, the measurement of phase variation in cells synchronized exactly out of phase but growing in close proximity has two drawbacks. First, the cells being targeted for analysis are in contact with roughly equal numbers of in-phase and out-of-phase cells; depending on the system of communication, they may not adjust their phase with so many "agreeing" cells nearby[17]. Second, in nature, cells are likely to not be suddenly 12 hours out of phase and mixed; the mechanics of phase adjustment may only function well for minor corrections. Regardless, this study presents several new and interesting results to build on.

### 2.6.3. Transcriptional/translational independence of circadian rhythm.

A final (and potentially most interesting) experiment has been very recently published in January 2005 (42). A key assumption since cyanobacterial circadian rhythms were discovered has been that clock gene regulation and feedback at the genetic level was required for producing stable oscillations (54). However, the processes of mRNA production and translation are severely impaired when the cells are moved to a dark environment, and yet *Synechococcus* remains on track with its clock. The authors of this study therefore questioned the importance of genetic regulation in the functioning of the circadian clock. To test this conjecture, the authors eliminated the processes of genetic transcription and translation. This was achieved in two ways: first, by placing the

[17] Some experiments have since been performed (but not yet published) by the same group in which low numbers of luminescent "reporter" cells were mixed with larger numbers of differently synchronized nonluminescent cells; no significant communication was found.

cells in the dark continuously (DD); second, by treating cells with rifampicin and/or chloramphenicol, which dramatically reduce transcription and translation respectively. Cells were collected at various times during these treatments; the KaiC protein was isolated and separated using polyacrylamide gel electrophoresis, yielding the dual bands representing phosphorylated and nonphosphorylated KaiC. The population of KaiC oscillated between these two bands with a periodicity of 24 hours for at least 56 hours *even in the absence of genetic feedback.* The period of this oscillation was only minimally affected by changing the temperature between 25-37 degrees Celsius (typical culture temperature for *Synechococcus* PCC7942 is 30 degrees). The persistence of this behavior under constant conditions and the temperature compensation (along with the prior synchronization of the cells to the LD cycles before the experimental treatment) indicate that this phosphorylation oscillation is truly a circadian behavior. Very recently, this group was able to reconstitute circadian patterns of KaiC phosphorylation using purified proteins *in vitro* (55).

This astonishing result calls into question the earlier models of the circadian oscillator of *Synechococcus*. The authors of this study proposed a more complicated model, using a transcription-independent "minimal" oscillator that feeds into an "extended timing module"[18] (see Figure 2.5). This, perhaps, represents the most complete model of the circadian clock mechanism of *Synechococcus* at this time. As biological systems are investigated and "custom" organisms are designed, knowledge of the basis of this remarkably robust cycle may assist in the inclusion of oscillatory

[18] This model does not explicitly include the spatiotemporal segregation of KaiB (which the authors of this study did not investigate under conditions of zero genetic activity); this perhaps is key to understanding the precise mechanism of the transcription-independent module and its role in stabilizing the oscillation.

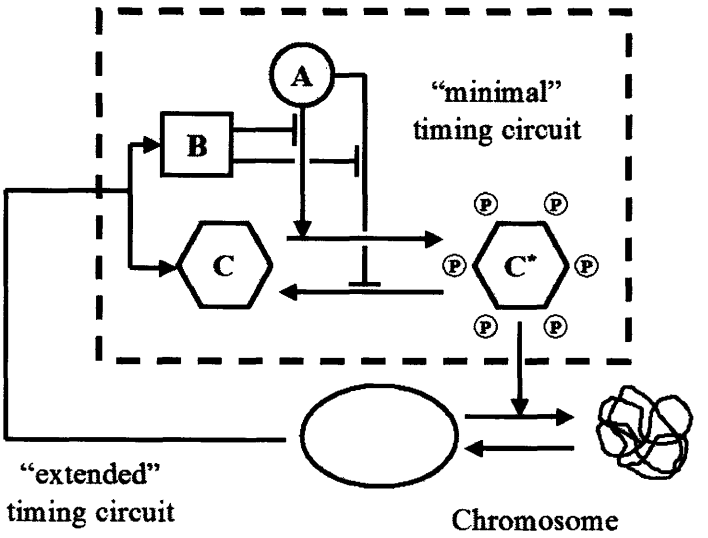


**Figure 2.5.** Proposed superstructure of circadian timing mechanism. This scheme incorporates a transcription- and translation-independent "minimal" timing unit, consisting of the Kai proteins, in which KaiC is rhythmically driven between phosporylated and nonphosphorylated forms. KaiC controls the state of the chromosome, which through a genetic pathway leads to regulation of the levels of the Kai proteins (specifically, KaiB and KaiC). Thus, the genetic feedback provides an "extended" timing unit, which may ensure stability over long times (greater than one or two circadian periods, for which the minimal timing circuit is sufficient). Modified from (42).

---

elements in these systems.

# 3. Methods.

## 3.1. Culturing cyanobacteria.

Growing and maintaining healthy photoautotrophic organisms involves a number of protocols different than those used in the culturing of other common lab organisms, such as the bacterium *Escherichia coli* or the budding yeast *Saccharomyces cerevisiae*. Although these cells are very self-sufficient and will make their own food, amino acids, and so on, they do need the appropriate building blocks for these compounds to be supplied in the correct amounts. The slow doubling time of cyanobacteria (compared to *E. coli* and *S. cerevisiae*) dictates that adherence to optimal conditions of light level and temperature should be as strict as possible. Whereas putting *E. coli* into minimal media at suboptimal temperatures may raise the generation time to one hour, with cyanobacteria the doubling time may exceed a day or more, making the culturing of large numbers of cells for cloning or data collection a time-consuming endeavor.

## 3.2. Light.

One of the most critical factors for growing *Synechococcus* is the level and quality of light to which the cells will be exposed during growth. For data collection, this light level was standardized at 4200 lux as measured using a light meter (VWR 62344-944) with a total dynamic range of 0 to 50,000 lux, and a resolution of 10 lux in the 2000-20,000 lux range. All devices used for preparation or observation of *Synechococcus* were maintained at this level to prevent any resynchronization owing to transfer between these devices[19]. Adjusting the proximity of the cells to the light source can modulate light

[19] For culturing cells during cloning, the light level would sometimes be higher, to give a faster growth rate.

levels; alternately, placing one or more pieces of window screening between the sample and illumination source will lower the light levels at the sample. Using pieces of screen made with different gauge wire allows light levels to be selected with fair resolution.

The specific source of light is also important to the growth rate of the population. Sources such as incandescent bulbs produce a significant amount of their light at longer wavelengths, which are not useful for photosynthesis by a majority of the light-absorbing pigments of cyanobacteria. In this study, cool white fluorescent bulbs (Sylvania; both linear and circular tubes) were utilized. These lights produce a high amount of photons appropriate for photosynthesis. 4200 lux from these bulbs corresponds to approximately 51 microeinsteins[20]; optimum conditions for culturing blue-green algae range from 20 to 200 microeinsteins (56, 57).

## 3.3. Temperature.

For all experiments, *Synechococcus* were maintained at 30 degrees Celsius. Although these cells may be effectively cultured over a range of temperatures[21], circadian experiments are usually conducted at this temperature. There is a weak temperature dependence to the period of the cyanobacterial oscillator (18); at 30 degrees, this period is nearly exactly 24 hours.

A custom growth enclosure was built for culturing *Synechococcus* (see Figure 3.1). A wooden box encloses five fluorescent light fixtures below a sheet of clear polycarbonate. An alternating current fan pulls air from outside the box and circulates it

[20] A microeinstein represents $6x10^{13}$ photons $m^{-2}$ $s^{-1}$. A very nice calculator to convert between light units for various illumination sources may be found at http://www.fb.u-tokai.ac.jp/WWW/hoshi/env/light.html.
[21] For example, the American Type Culture Collection (http://www.atcc.org) recommends growing PCC7942 at 26 degrees.

between the lights and clear surface. The heat produced by the lights serves to warm the box near the polycarbonate to 30 degrees; changing the size of ventilation holes on the sides below the surface can change this temperature level. The light levels are approximately 4000 to 6000 lux across the surface; in the brighter sections, they can be reduced to 4200 lux (for cells being used in experiments) by the use of window screen as described above.

---

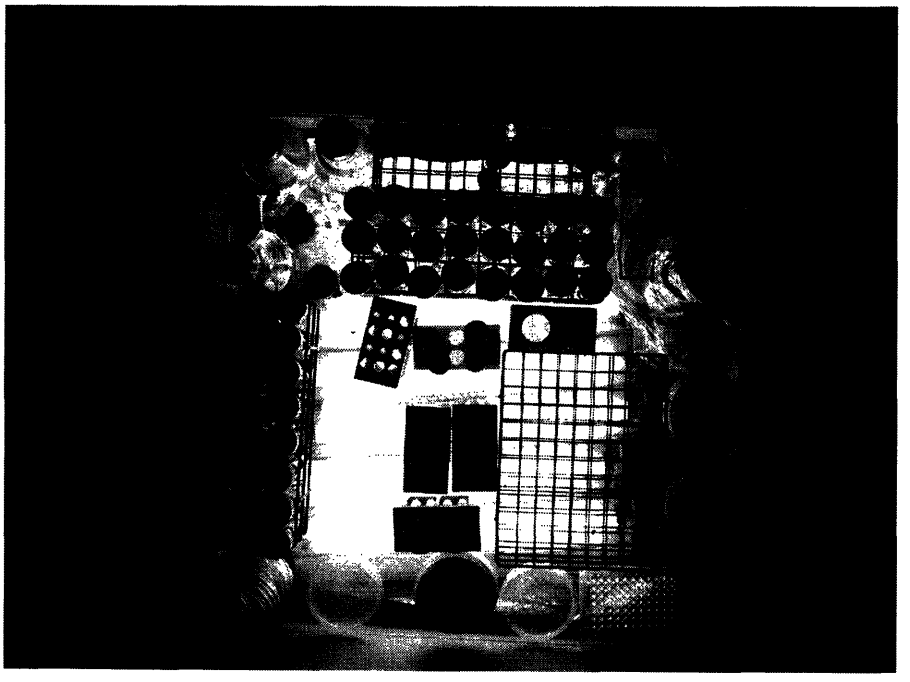

**Figure 3.1.** Constant light enclosure for culturing *Synechococcus*. The samples pictured here (including petri dishes, flasks, plastic culture tubes, and a 96 well plate) rest on a clear sheet of polycarbonate, lit from below by five cool fluorescent tube lights. The light level may be adjusted by placing sheets of window screen below the samples. A fan in the side of the box circulates air below the polycarbonate sheet, maintaining the temperature at the surface at 30 degrees.

## 3.4. Aeration.

*Synechococcus* are obligate photoautotrophs. As such, they do not metabolize exogenously supplied sugars such as glucose that are often used to culture other laboratory strains. The carbon source that these cyanobacteria utilize is dissolved carbon dioxide or carbonate in the media[22]. In order to ensure that the supply of carbon is replenished, it is important that the growing cultures are contained in such a way as to allow for air exchange with the greater environment[23]. This may be accomplished in different ways. The simplest is to leave caps or lids less than completely tightened; this method works well with 50 mL polystyrene culture tubes or 5 mL tubes used for flow cytometry. However, the rate of air exchange is not well regulated using this approach, and thus growth conditions may vary significantly from tube to tube.

To standardize gas exchange, holes were drilled in the lids of 15 and 50 mL polystyrene tubes, and 0.2 μm HT Tuffryn Acrodisc syringe filters (Pall Life Sciences) were epoxied in the holes. These filters allow air to freely exchange between the tube and the laboratory, while preventing contamination from airborne bacteria, yeast spores, and so on. The caps can be reused, re-sterilizing the plastic lids using 70% ethanol if contamination occurs. These tubes were used whenever collections of samples were prepared simultaneously as part of a large run, when sample-to-sample variation in growth conditions was to be minimized.

[22] Attempting to add large amounts of carbonate into the media (significantly beyond the amounts in the BG11 media, presented below) can cause a dramatic shift in pH, which is to be avoided.
[23] A completely airtight tube of cells will rapidly blanch and the cells will die; a sample with ample air exchange can be continuously grown for months.

## 3.5. Media.

The media for *Synechococcus* culture does not include any amino acids, proteins, or sugars; however, as PCC7942 does not fix nitrogen, it does contain nitrate. BG11 media (58) is made by combining the solutions described in Table 3.1 in the appropriate proportions. The media is then autoclaved for up to 40 minutes[24]. Often a precipitate will form in the autoclave; this will typically dissolve once the media cools and does not appear to be detrimental to the growth of *Synechococcus*. For strains with the appropriate antibiotic resistances, either spectinomycin (to a concentration of 40 μg/mL) or kanamycin (to a concentration of 50 μg/mL) are added to the media. For strains with both resistances, a final concentration of 2 μg/mL spectinomycin and 2.5 μg/mL kanamycin was used[25].

To make solid media for use in cloning, BG11 media was prepared at twice the concentration described in Table 2.1. Separately, agar was autoclaved in deionized water at a concentration of 31.25 g/L, then allowed to cool (while on a magnetic stirrer) to less than 55 degrees (when a hand can be held against the flask for at least a second without discomfort)[26]. At this time equal volumes of the liquid agar and the 2x BG11 were combined while stirring, and the mixture was dispensed into sterile 100x25 mm petri dishes (40 mL per plate)[27].

---

[24] Autoclaving for 40 minutes is not appropriate for media for culturing bacteria or yeast, as sugars and proteins will be caramelized or denatured by exposure to high temperatures for such long times. As BG11 media does not contain either of these, 40 minutes ensures a complete sterilization without harming the media.

[25] Either of these lower concentrations will kill non-resistant *Synechococcus*; some anecdotal evidence suggests that combining these two antibiotics at the higher concentrations may slow growth of the culture.

[26] Autoclaving agar with the various minerals in BG11 results in the evolution of toxic species (56); therefore, these must be sterilized separately and combined after allowing some cooling.

[27] A recipe for solid media provided by Susan Golden (Texas A&M University) also included the addition of 5 mL of 100 mM Na2S2O3 to the final mixture of agar and media; I found no difference in cells grown on plates with or without this component.

**Table 3.1.** Components of BG11 media.

| Component | Concentration in stock | Amount of stock in BG11 (1 L) |
|---|---|---|
| Deionized water | --- | 983 mL |
| $NaNO_3$ | 150 g/L | 10 mL |
| $CaCl_2 \cdot 2H_2O$ | 36 g/L | 1 mL |
| $FeNH_4$ | 12 g/L | 1 mL |
| Disodium EDTA | 1 g/L | 1 mL |
| $K_2HPO_4$ | 40 g/L | 1 mL |
| $MgSO_4 \cdot 7H_2O$ | 75 g/L | 1 mL |
| $Na_2CO_3$† | 20 g/L | 1 mL |
| Trace metals† | *see below | 1 mL |

*Trace metal solution consists of 2.86 g $H_3BO_3$, 1.81 g $MnCl_2 \cdot 4H_2O$, 0.222 g $ZnSO_4 \cdot 7H_2O$, 0.390 g $Na_2MoO_4 \cdot 2H_2O$, 0.079 g $CuSO_4 \cdot 5H_2O$, and 0.049 g $Co(NO_3)_2 \cdot 6H_2O$ dissolved in 1 L of deionized water.
† While the other solutions may safely be autoclaved to guarantee sterility, these solutions must be filter sterilized.

## 3.6. Genetics and cloning.

*Synechococcus* PCC7942 is the model strain of cyanobacteria largely because of the facility with which genetic modifications can be made. When foreign DNA is introduced into the cell, it will readily recombine with any parts of the chromosome bearing a significant homology. To genetically prepare strains of *Synechococcus*, a two step approach is taken. First, a plasmid containing the desired construct for insertion is prepared in *E. coli*, then the construct-bearing plasmid is introduced into the cyanobacterium, where the desired DNA is incorporated in the chromosome.

### 3.6.1. Preparation of constructs and plasmids.

Two plasmids are used as backbones for inserting the constructs to be incorporated in the cyanobacterial chromosome (Figure 3.2). These plasmids contain stretches of DNA from parts of the *Synechococcus* chromosome. These particular DNA segments were included because when they are disrupted, there is no apparent phenotypic difference between the resulting mutants and wild-type cells. These chromosomal sites are referred to as "neutral sites" for this reason. Plasmid pAM1303 contains DNA from neutral site I (NSI) and a spectinomycin resistance gene, and plasmid pAM1579 contains DNA from neutral site II (NSII) and a kanamycin resistance gene (plasmids kind gifts of Susan Golden, Texas A&M University). Constructs for chromosomal insertion are placed in these plasmids with the resistance genes between the neutral site DNA.

### 3.6.2. Transformation of *Synechococcus*.

The resulting plasmids are transformed into *Synechococcus* by the following protocol. First, mid to late exponential cells are collected from a master culture[28]. Cells are concentrated via centrifugation at 13,000 rpm for 2-3 minutes[29]; typically 3 mL of cells are used for transformation. The pelleted cells are resuspended in 1500 μL of 10 mM sodium chloride, and re-pelleted. The cells do not need to remain in the salt solution for any more time than required for resuspension and spin down. The cells are then resuspended in 300 μL BG11 without antibiotics. To this solution, 1-2 μL of plasmid

[28] Cells used for transformation included wild type PCC7942, cells with insertions already in the other neutral site, or cells with insertions in the targeted neutral site that will be "overwritten" by the new insert. This last type can be particularly difficult, as the cells will typically contain the same antibiotic resistance being inserted, making transformants difficult to select. Generally, the antibiotic resistance is changed in this case. Ampicillin and chloramphenicol are additional antibiotics useful for this purpose.

[29] *Synechococcus* do not pellet as readily as *E. coli*; consequently, longer spin times are required to get a firm pellet.

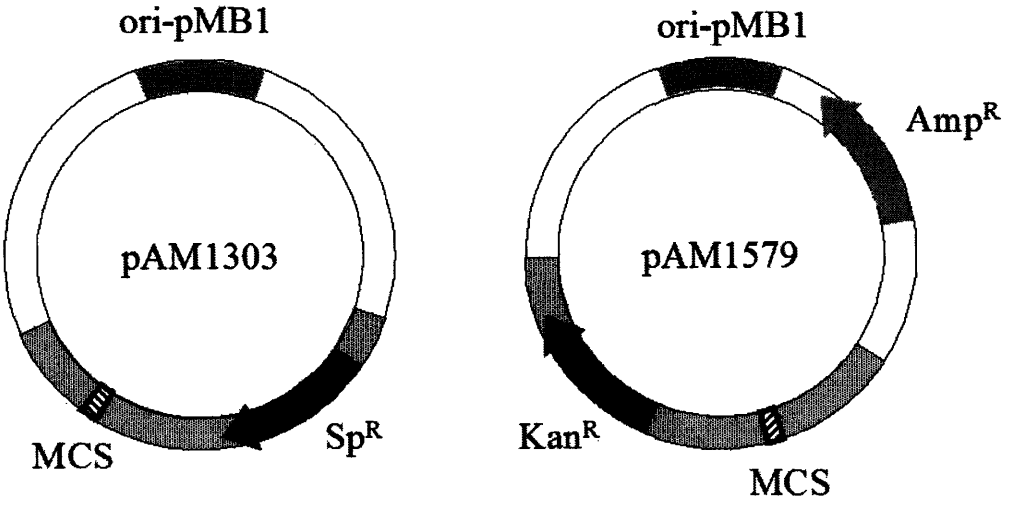


**Figure 3.2.** Plasmid maps for basic backbones. Both plasmids are based on pBR322. Plasmid pAM1303 contains DNA from neutral site I, and pAM1579 contains DNA from neutral site 2 (shaded gray for both plasmids). Sequences inserted into the multiple cloning site (MCS), as well as the resistance gene in the gray region, are incorporated into the chromosome at the appropriate neutral site. The multiple cloning sites contain the following unique restriction sites in clockwise order: pAM1303 contains NotI, BamHI, SmaI, and SacI; pAM1579 contains EcoRV, SalI, StuI, XbaI, and NheI.

---

from a miniprep are added. For simultaneous transformation with both NSI and NSII plasmids, 8-10 μL of each plasmid are added. The tube of cells with plasmid is wrapped in foil to exclude light, and placed on a rotator in a 30 degree incubator.

After four to sixteen hours (usually overnight), the cells are removed from the incubator and 150 μL of the solution are streaked on a 40 mL BG11 agar plate (prepared as described above), which should be warmed to 30 degrees prior to plating[30]. The plate is wrapped with parafilm and placed in a light environment at 30 degrees for three to five

[30] Two plates may be prepared from each transformation. This is useful, as often a plate may dry out before microcolonies appear. Having a backup plate can save a couple of weeks if this happens.

hours. After this time, the plate is removed and unwrapped. A metal spatula (preferably with a rounded edge) is heated in a bunsen burner, cooled in sterile water, and used to lift the agar slab. (This takes some skill and practice in order to avoid tearing the agar.) An appropriate amount of concentrated antibiotic (usually prepared in a 1000x stock) is introduced under the agar slab[31]. Introducing the antibiotic at the surface where the cells are located would present a lethal concentration, even if the cells carry the resistance gene; instead, the antibiotic diffuses through the agar, thus protecting the cells from a lethal dosage. The spatula is moved under the agar in such a way as to spread the antibiotic under the entire slab. The spatula is removed; if a large bubble remains under the agar, squeezing the plate on the sides most distant to the bubble can often force it to the edge and out. The plate is resealed with parafilm and returned to the light environment.

After seven to fourteen days, the plates should have small colonies (visible as small but distinct green dots) present on it. Often, up to one third of the plate will be covered with a continuous green mat, while the other end of the plate will have no colonies at all; individual colonies will be found between these two regions. This results from uneven distribution of antibiotic, which is often unavoidable. However, if individual colonies can be isolated at any point, the local concentration of antibiotic should be appropriate for selecting only the correct transformants. The colony is then collected with a sterile wooden applicator, and transferred to a tube or flask containing BG11. Antibiotics may be added immediately, or after allowing some time for the cells

[31] For transformants with two resistances, the full antibiotic dose (namely, 40 μg/mL for spectinomycin and 50 μg/mL for kanamycin) is given for each antibiotic. As the plate will typically have a gradient of antibiotic across it, at some location the dosage will effectively select colonies that can still grow at a decent rate.

grow, but no later than 24 hours after inoculation into the liquid media.

Colonies selected after appropriate antibiotic screening always (in this study) contained the appropriate construct. To confirm an insertion, the chromosome can be extracted (performed in this study using a kit from Amersham Biosciences for chromosomal DNA extraction from gram positive bacteria[32]), and the presence of the inserted constructs checked by PCR or sent for sequencing.

## 3.7. Maintenance of strains.

### 3.7.1. Continuous liquid cultures.

Strains can be maintained for several months by dilution after the cells reach a steady state density, visible as a rich green color which no longer appears even partially transparent. At this density, a majority of cells do not receive sufficient light to continue to grow (even if nutrient depletion is not an issue); often, cells will blanch and fall out of solution, forming a pale deposit at the bottom of the culture tube or flask. If such a deposit exists, cells are removed from the top of the media and introduced into a new container with fresh media; if this deposit has not formed, approximately ninety percent of the volume of the media is removed and replaced with fresh media.

### 3.7.2. Long term storage (freezing).

For long term storage of *Synechococcus* strains, 3 mL of cells from mid to late exponential cultures are collected and pelleted by centrifugation. The cells are resuspended in 450 μL of BG11 (without antibiotics). 50 μL of 50% DMSO in deionized

[32] This kit was also used to prepare PCC7942 DNA for use as a template for PCR to isolate native *Synechococcus* promoters.

water is added; inverting the tube several times to mix the resulting solution. The cells are then placed in a –80 degree freezer. To rejuvenate frozen cultures, the tube is removed from the freezer and allowed to thaw until 50 μL can be withdrawn with a pipettor (the remaining cells can be returned to the freezer for future rejuvenation). These cells are introduced into BG11 media without antibiotics and placed in a light environment at 30 degrees. Antibiotics are added later (typically after an overnight period); this allows the cells to express the antibiotic genes and more fully recover from the freezing process before being presented with antibiotics. After a day, these cells grow at rates matching cells at the same density that have not been recently frozen.

## 3.8. Synchronization of strains.

The synchronization of *Synechococcus* strains involves placing the cells in an environment where the light cycles from "high" levels (in this study, 4200 lux from soft white fluorescent lights) and "low" levels (ideally, 0 lux[33]). In this study, this was achieved in one of two ways. First, an Innova 4200 incubator/shaker (New Brunswick Scientific)[34] was fitted with soft white fluorescent tube lights connected to a 24 hour timer (Radio Shack) which would allow power to flow to the lights for 12 hours of each day. This technique was preferred when a single synchronization was required (for example, in "plate reader" experiments; see Chapter 4).

The other method for synchronization is useful for preparing a number of samples to be synchronized with different phases. A custom device, called the "circadian

[33] In this study, "low" levels ranged from 0 lux to 50 lux; no difference was observed between cells synchronized over this range of darkness.
[34] The cells were grown on a shelf, which did not move appreciably even when the rotating platform was engaged.

spinner" (see Figure 3.3), was constructed to synchronize up to 24 tubes with phases varying in one hour increments. The spinner is housed within a wooden fiberboard box. An AC synchronous gear motor rotating with a frequency of one revolution per hour is connected to a four start worm with a 96 tooth worm gear (all components from McMaster-Carr), causing the worm gear axis to rotate with a period of 24 hours. This construct is attached to a piece of fiberboard, which rests on the ground independently of the sides of the box. The axis of the worm gear was connected to a clear circular polycarbonate disk, with holes drilled every 15 degrees around the perimeter. These holes accommodate custom holders (made from 18 gauge copper wire) for either 15 mL or 50 mL tubes (Corning). As the disk rotates, tubes move through a pair of holes cut in a divider. One half of the chamber is illuminated from below by three soft fluorescent tube light fixtures (giving a measured light level within 10% of 4200 lux along the path the tubes travel), while the other half is shielded in darkness by a piece of fiberboard beneath the disk[35]. A length of heating tape runs around the top of the box, and is connected to a variable transformer (Staco Energy Products company). By changing the voltage on the transformer, and adjusting the length of tape on either side of the box, the temperatures on each side can be independently controlled[36]; in this study, both sides were set to 30 degrees. Typically, 24 tubes are prepared with one strain (or two strains with twelve tubes each, or three strains of eight tubes each) and placed on the spinner. After each tube has encountered at least two complete 12:12 LD cycles, the tubes in the

[35] The light level in the dark compartment registers a maximum of 50 lux, with most of this light leaking through the holes in the divider.

[36] The lights cause the temperature on the light side to be naturally higher than the dark side without any adjustment; therefore, both sides need an independent adjustment to make the temperature constant throughout.

light are removed to the custom constant light box described above, with a piece of screen matching the light levels. As tubes rotate into the light, they are moved to this box; after 11 hours, all tubes have been removed from the spinner.

---

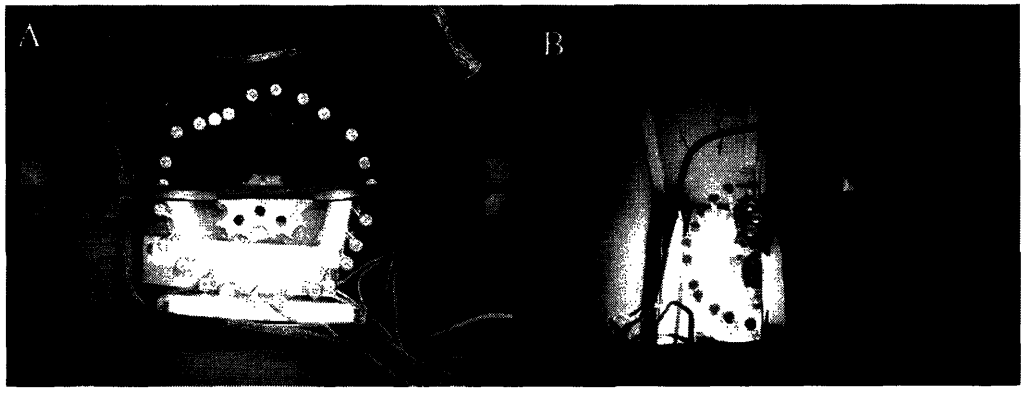


**Figure 3.3.** Circadian spinner. Figure A shows a top view (image taken with a flash). The disc containing 24 tubes (plus a control tube situated between two samples at approximately eleven o'clock) rotates counterclockwise; as cells pass through the divider, they move from a D environment (<50 lux) to an L environment (4200 lux ± 10% along the entire path). The rack in the center of the image can hold stationary samples in the L environment. Temperature probes are visible to the left in both the L and D sides. Heating tape (traveling across the top, left, and bottom of the image), powered through a variable voltage source, maintains the temperature at 30 degrees; pulling the tape through the hole in the divider allows the sides to be independently adjusted. Figure B shows a side view of the spinner, taken without a flash to show the difference in light levels.

# 4. Establishing Fluorescence as a Tool for *Synechococcus* Circadian Studies.

## 4.1. Choosing a reporter.

### 4.1.1. Bioluminescence.

Until this time, bioluminescence has been the sole reporter used in probing genetic activity in *Synechococcus* PCC7942 (see for example (18, 48, 51)). There are some significant advantages to this reporter (see (59) for a nice discussion of bioluminescent benefits). First, the background signal can be made as low as is possible from a given detector since these cyanobacterial cells have no native bioluminescence. Using quiet photomultiplier tubes or ultra low noise cooled CCD cameras, the background signal can approach zero compared to the level of signal emitted by a population of cells. Second, bioluminescence is apparently a "fast" reporter in cyanobacteria. By this, I mean that the bioluminescence oscillations seen in a sample of *Synechococcus* very closely agree in phase (i.e., with minimal lag) with Northern blots which monitor the mRNA levels for the promoter being probed[37] (18).

However, there are a number of disadvantages to bioluminescence studies, as were listed in Chapter 2 concerning the single cell bioluminescence studies. To reiterate briefly: for *single* cells, long exposure times and low signal levels make bioluminescence a tenuous experimental technique at the least. The nature of the observations also precludes examination of large numbers of cells (i.e., $10^3$-$10^5$) needed

[37] The mRNA levels measured by Northern blots are the most faithful indicators of genetic activity (as the creation of protein from a gene uses mRNA as its intermediary). However, this assay can only be performed 1) in bulk and 2) by collecting cells and lysing them, and is more time intensive than collecting a bioluminescence image. Therefore, to get data from a continuous population of cells with good time resolution, a technique such as bioluminescence or fluorescence is more useful.

to build confident statistics for describing population variations, etc. In addition to these concerns, bioluminescence results from a reaction involving an enzyme (which is actually expressed by two genes, which combine post-translationally to form the functional luciferase protein), a substrate, and ATP. The substrate may be produced genetically (by expressing another three genes), or by addition of a chemical (n-decanal). Each of these approaches has potential concerns. If the substrate is internally produced, the resulting signal is a convolution of the activities of the promoter driving the enzyme and the promoter controlling the substrate. When the substrate is supplied exogenously, care must be taken with the level of n-decanal used, as it has been shown to be toxic in other organisms (60)[38]. For these reasons, fluorescence was explored as a potential alternative to bioluminescence as a reporter in *Synechococcus* PCC7942.

### 4.1.2. Fluorescence.

Fluorescent proteins have been used in many studies to probe genetic activity in a variety of organisms (see for example (61-63)). The amount of fluorescent protein present can be tested in a number of ways, including using a spectrofluorimeter to monitor bulk fluorescence, or microscopy and flow cytometry to measure individual cells. To collect the measurement, the cells are exposed to light of a particular wavelength, which is absorbed and re-emitted by the fluorophore at a longer wavelength. Very rapid measurements are possible when the excitation light is strong (typically provided by a laser or arc lamp). Because this process is actively driven by an external

[38]This study describes toxicity in eukaryotic organisms including yeast and *C. elegans*, at levels that minimally impact *E. coli*. However, potential toxicity should still be considered when applying this reporter system to a new organism.

stimulus rather than a passive reaction in the cells, higher signals are possible in shorter times. This enables the measurements of large numbers of cells that are impractical with bioluminescence. Fluorescence also avoids the enzyme-substrate dilemma.

Fluorescence has disadvantages, the impact of which must be tested as well. First, cells may experience a reduced growth rate when repeatedly exposed to strong fluorescence excitation light (especially at short wavelengths). Second, as described above, high levels of expression may lead to some toxicity. Third, repeated measurements of fluorescence will cause a loss of signal, as the fluorophores are photobleached, whereby the fluorescence is irreversibly lost by a non-radiative release of the excitation energy. Fourth, cells may have a significant autofluorescence matching the absorption and emission spectrum of the desired fluorescent protein. Finally, especially in cells such as cyanobacteria that absorb light for various reasons, the impact of the bright light exposures on the life and metabolism of the cells needs to be examined. Nonetheless, under certain conditions[39], the use of fluorescence as a reporter of single cell genetic activity presents a significant advance over bioluminescent techniques.

## 4.2. Choice of particular fluorescent reporter.

Several studies have been published in which GFP (a green fluorescent protein originally isolated from the jellyfish *Aequorea victoria*) has been used as a reporter in cyanobacterial strains (64-66). It should be noted that the only unicellular nonfilamentous cyanobacteria in which GFP have been expressed are all marine strains, and, as such, typically have a different osmolarity and pH than the freshwater

[39] These conditions include the use of a reporter with a long wavelength excitation, moderate levels of expression, and either single exposures or the use of low levels of excitation light.

*Synechococcus* PCC7942[40]. When GFP[41] is expressed under a very strong promoter (the *trc* promoter ($P_{trc}$), a fusion of the *lac* and *trp* promoters of *E. coli*), and with a very strong ribosome binding site[42], no signal above background is detected in PCC7942[43]. Other fluorescent probes exist, including mutants of GFP with altered spectra and fluorophores derived from *Discosoma* (coral) proteins. Table 4.1 lists the various fluorophores considered in this study.

**Table 4.1.** Fluorescent proteins considered for this study.

| Fluorophore | Original source | Plasmid (Clontech) | Peak excitation (nm) | Peak emission (nm) |
|---|---|---|---|---|
| GFPmut3.1 | *Aequorea* | PGFPmut3.1 | 501 | 511 |
| CFP | *Aequorea* | PECFP | 433 | 475 |
| YFP | *Aequorea* | PEYFP | 513 | 527 |
| GFPuv | *Aequorea* | PGFPuv | 395 | 509 |
| DsRed | *Discosoma* | pDsRed-Express | 557 | 579 |
| HcRed | *Discosoma* | pHcRed1 | 588 | 618 |

Immediately, DsRed and HcRed were discarded as candidates; the chlorophyll fluorescence of cyanobacteria provides such a huge background for the excitation/emission pairs of these proteins that detection of modest levels of protein in *Synechococcus* is practically impossible. Once the remaining proteins were expressed

[40] GFP has been expressed with detectable signals in freshwater filamentous cyanobacteria, but fundamental differences exist between filamentous and unicellular blue-green algae, such as differentiation of cells (known as heterocysts) in filaments.

[41] Specifically, a mutant of GFP from a plasmid pGFPmut3.1 (Clontech), which has been optimized to give a larger signal for a given amount of excitation than the wild type GFP.

[42] In *E. coli*, strains with the ribosome binding site AGGAGGAAAAAAA preceding the ATG start codon produced the highest levels of protein compared to other strains utilizing the same promoter but different ribosome binding sites (67). This ribosome binding site was used in for all fluorescence reporters in this study.

[43] This was reported in a personal communication with Susan Golden (Texas A&M), and independently confirmed in our lab (see Figure 3.1).

under $P_{trc}$[44] in *Synechococcus* PCC7942, fluorescence comparisons were made between fluorophore-bearing mutants and wild type cells. The results are shown in Figure 4.1. From this figure, it can be noted that both CFP and YFP give clear signals much above background for expression driven by $P_{trc}$. GFP gives a minimal signal above background, rendering it unusable, especially for weaker promoters (all native promoters of *Synechococcus* are at least an order of magnitude less strong than $P_{trc}$). GFPuv was

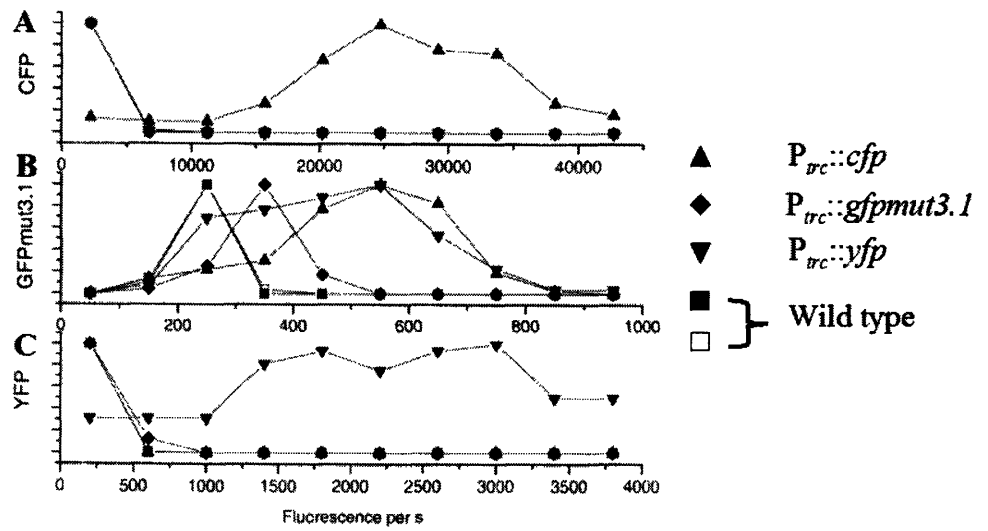


**Figure 4.1.** Testing signal from various fluorescent proteins in *Synechococcus*. The curves represent normalized histograms of fluorescences of individual cells measured by microscopy. Three filter sets were used: one optimized for CFP (A), one for GFP (B), and one for YFP (C). For CFP and YFP, the appropriate strain clearly gave significant signal above background. For the GFP filter set, the GFP strain gave insignificant signal above background. Fluorescence "leak through" from overlaps between YFP and CFP excitation/emission spectra with the wavelengths selected by the GFP cube is responsible for the signals seen from the CFP and YFP strains. Note that the absolute values on the abscissa should not be compared from filter set to filter set, as the transmissions of the filter cubes, the light level supplied by the lamp, and the quantum efficiency of the camera can and do vary as a function of wavelength.

[44] In *E. coli*, $P_{trc}$ is repressed by the *lac* repressor protein LacIQ; however, *Synechococcus* lack this repressor, and thus maximal expression is automatic.

tested by examining bulk fluorescence of fluorophore-containing cells compared to wild type using an ultraviolet transilluminator; minimal differences were seen. In addition, although cyanobacteria are more resistant to ultraviolet radiation than regular bacteria (56), repeated measurements using ultraviolet excitation may lead to possible mutations; this reporter is best avoided when another possible fluorophore is available. Thus, CFP and YFP emerged as the leading candidates for this study.

Which protein should be used to study *Synechococcus*? Three factors come into play. First, cyanobacteria absorb light across most of the visible spectrum, due to their array of photosynthetic pigments, which include both chlorophyll and phycobilins (cyanobacterial pigments which extend the cells' "diet" of photons). Fluorescence excitation light that overlaps with the more naturally absorbed wavelengths tends to lead to higher autofluorescence. The one appreciable gap in the absorption spectrum of cyanobacteria in general includes the excitation wavelengths for GFP and YFP. *Synechococcus* have approximately the same fluorescence background at these wavelengths as *E. coli* cells (which are comparable in size and morphology); by comparison, the CFP background is over ten times that of *E. coli*. The fact that *Synechococcus* do not appreciably absorb these wavelengths also makes it unlikely that the flashes of excitation light will trigger some receptor leading to altered circadian rhythms. Second, because YFP and GFP have similar absorption and emission spectra, common tools such as flow cytometry that have been designed to study the more commonly used GFP can also be used for experimentation with YFP as the probe. Third, repeated exposures to CFP excitation light have been shown to lead to some phototoxicity in organisms such as *S. cerevisiae* (Ben Kaufmann, unpublished observation). To avoid any possible detrimental photoinduced effects in the

cyanobacteria, as well as for the first two reasons listed above, YFP was the preferred fluorophore for the experiments performed in this study.

## 4.3. Effect of fluorescent probes on measured oscillation amplitude.

A common concern when using fluorescent proteins to study dynamic processes concerns the lifetime of the reporter proteins. Most prokaryotic cells lack machinery specifically designed to degrade these proteins derived from other organisms. Consequently, the stability of proteins such as GFP and its mutants leads to half lives of 24 hours or longer. The primary means of lowering the concentration of protein becomes the growth and division of the cell:

$$\frac{1}{\tau_{\text{observed}}} = \frac{1}{\tau_{\text{dilution}}} + \frac{1}{\tau_{\text{decay}}}$$

where $\tau_{\text{decay}}$ is the chemical half life of the reporter species, $\tau_{\text{dilution}}$ is the doubling time of the cell, and $\tau_{\text{observed}}$ is the time to reduce the amount of protein in a cell by fifty percent. In cells such as *E. coli*, where doubling times can be as low as 20 minutes, the use of stable fluorescent reporters may be amenable for most processes; in *Synechococcus*, where doubling times are (at least) seven to eight hours, this lifetime may be more restrictive.

Consider a species $y$ produced with a time dependent rate $x(t)$, and undergoing first order decay with a rate constant $\gamma$. The rate of change can be written as

$$\dot{y}(t) = x(t) - \gamma\, y(t)$$

The Laplace transform (using $y=e^{st}$) of this function is

$$s\hat{y} = \hat{x} - \gamma\, \hat{y}$$

The transform of the output can then be described as

$$\hat{y} = \frac{\hat{x}}{s+\gamma} = H(s)\hat{x}$$

where $H(s)$ is the transfer function from $\hat{x}$ to $\hat{y}$. When $x$ is driven by a frequency $\omega$, the amplitude of $y$ is affected by this transfer function in a frequency dependent way. If the amplitude of the oscillation in $x$ is $A$, and the DC component of the input is $C$ (necessarily greater than $A$ to make $x$ and $y$ positive definite), the peak-to-trough ratio ($PTR$) of the oscillation in $y$ is

$$PTR_y = \frac{C + A\dfrac{|H(i\omega)|}{|H(0)|}}{C - A\dfrac{|H(i\omega)|}{|H(0)|}}$$

For the transfer function in this case, the $PTR$ is[45]

$$PTR_y = \frac{C + A\dfrac{\gamma}{\sqrt{\omega^2+\gamma^2}}}{C - A\dfrac{\gamma}{\sqrt{\omega^2+\gamma^2}}}$$

Therefore, as the decay time approaches the period of the oscillation, the peak to trough ratio of the reporter will begin to be significantly diminished[46]. For YFP molecules with an infinite lifetime, and a cell doubling time of 8 hours, and with $C$=$A$ (infinite peak to trough ratio of genetic transcriptional activity), this gives a $PTR$ of ~1.9 (see Figure 4.2A). This ratio can be significantly improved by shortening the lifetime of the reporter.

[45] This derivation has been performed in other sources, such as (68).

[46] Conceptually, this can be imagined as if a huge number of molecules have built up over time, such that the few molecules produced in a single cycle matches the few number of molecules destroyed over that same time period. In this case, a nearly imperceptible wiggle will occur around that huge number of molecules.

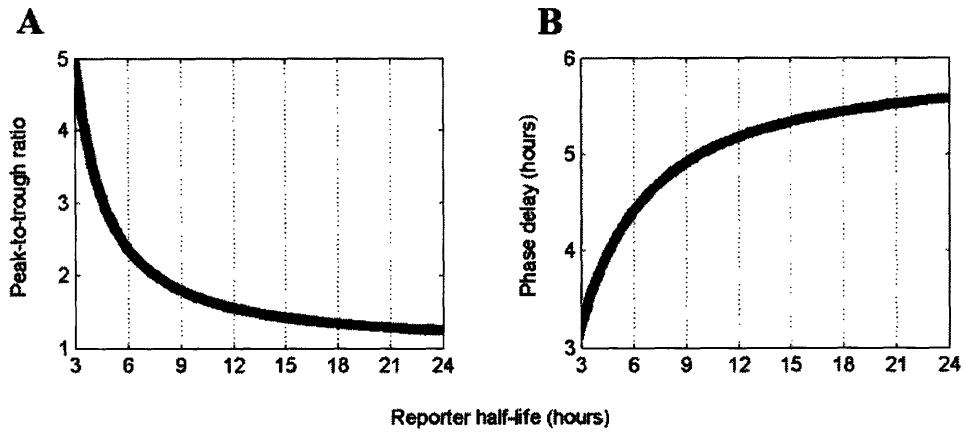


**Figure 4.2.** Effects of long-lived reporter proteins on oscillation levels. In both figures, an input signal with infinite peak-to-trough ratio and a 24 hour period is assumed. (A) Output peak-to-trough ratio for reporters with indicated half-lives. As the reporter lifetime is increased, the amplitude of the oscillation is diminished as the level of protein integrated from previous cycles increases. The function describing the curve is derived in the Section 4.3. (B) Phase delay of reporter signal. As the reporter half-life increases, the output becomes an effective integration of the input, and the delay asymptotically approaches 6 hours (corresponding to a $\pi/2$ phase shift). The shape of the curve is an inverse tangent, and is derived in Section 4.6.

---

In order to do this, a *yfp* mutant was constructed that would hopefully enable proteases in *Synechococcus* to recognize the YFP and mark it for a more speedy degradation.

## 4.4. *LVA* degradation tag.

Many organisms, both prokaryotic and eukaryotic, possess proteases belonging to the Clp family. These proteases specifically attack proteins with distinct recognition sequences, called *ssrA* tags. A study was performed comparing the efficiency of Clp protease response in two prokaryotic species to the *E. coli ssrA* tag and some mutations thereof. In both *E. coli* and *Pseudomonas putida*, a tag consisting of amino acid residues AANDENYALVA was shown to lead to the fastest degradation times (69). I therefore

inserted the appropriate DNA sequence before the stop codon of *yfp*, forming the gene *yfpLVA* (and its protein product, YFPLVA)[47].

### 4.4.1. Construction of tagged YFP test strains.

*Synechococcus* strains IL001 and JRCS48 (see Appendices for complete lists of plasmids (Appendix 1) and strains (Appendix 2) constructed as part of this study) were prepared by introducing *yfp* and *yfpLVA* respectively under the control of the *trc* promoter into neutral site I. A construct expressing the *lac* repressor LacIQ under its native *E. coli* promoter was placed in neutral site II. This combination causes $P_{trc}$ to be repressed and consequently the expression of *yfp*(*LVA*) is minimal. Addition of IPTG (isopropyl-$\beta$-D-thiogalactopyranoside) causes the LacIQ to reversibly bind with the IPTG, rendering it less effective as a repressor and increasing the amount of $P_{trc}$-driven transcription[48].

### 4.4.2. Quantifying effect of tag.

To test and compare the lifetimes of YFP and YFPLVA, master cultures of IL001 and JRCS48 were grown for several days in LL in BG11 containing 2 mM IPTG (at which concentration the promoter is fully induced, giving signal levels comparable to strains lacking the *lacIQ* gene). Once per hour, an aliquot of cells from each strain were "de-induced" by pelleting via centrifugation and subsequent resuspension in non-IPTG-containing media; these aliquots were then returned to the LL environment. After thirty

[47] In (69), the last three amino acids of the tag were varied to test the impact on the effectiveness of the tag. They indicated the particular tag being studied by using the last three amino acids as a label; I have chosen to do the same in this study.
[48] For this reason, IPTG is called an *inducer* and $P_{trc}$ an *inducible promoter*.

hours of iterating this process, the de-induced cells were measured using flow cytometry (details of flow cytometry measurements are given in Chapter 6). The data, along with first order exponential decays, are shown in Figure 4.3A. These results indicate a shortening of the effective half-life of the reporter from 12.8 hours without the tag to 5.6 hours with the tag. The doubling time of cells in this experiment was consistent with the half-life of YFP observed (implying a decay half-life much longer than 13 hours), whereas YFPLVA displays a calculated decay half-life of 10 hours using this same doubling time. Using the formula for peak to trough ratio, this decay half-life and doubling times from 8-13 hours imply a *PTR* ranging from 2.5 to 3.1 (assuming infinite *PTR* of the driving signal).

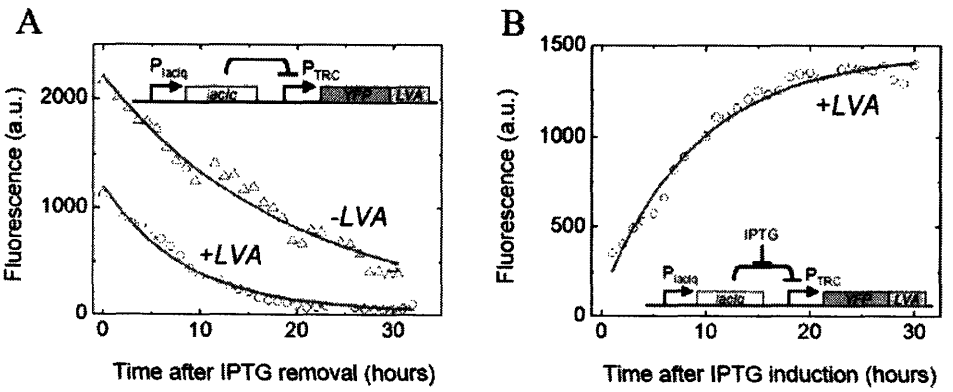


**Figure 4.3.** Decay and induction kinetics of YFP and YFPLVA. (A) Decay of YFP (triangles) and YFPLVA (circles). Strains expressing these reporters under the repressible *trc* promoter are maximally induced with 2mM IPTG until $t = 0$, at which time they are removed from inducer-containing media to regular media (restoring the maximal repression of the reporter gene). This procedure is repeated once per hour, then all samples are measured using flow cytometry. The fits are simple exponential decays ($y = Ae^{-\alpha t}$), and give half lives of 12.8 (YFP) and 5.6 (YFPLVA) hours. (B) Induction kinetics of YFPLVA. Cells are uninduced (i.e., maximally repressed) until $t = 0$, at which time they are maximally induced by 2mM IPTG. This is repeated once per hour, then all samples were measured using flow cytometry. The fit represents a simple first order approach to a limit ($y = A_1 - A_2 e^{-\alpha t}$), and gives a calculated half life of 6.3 hours.

### 4.4.3. Induction kinetics of *yfpLVA*.

The induction kinetics of $P_{trc}$::*yfpLVA* were also tested using a similar experiment. A master culture of JRCS48 was prepared in LL, in BG11 media without IPTG (and thus maximally repressed expression of YFPLVA). Once per hour, IPTG was added to an aliquot from this stock to a final concentration of 2 mM, immediately raising expression to its maximal level. After 30 hours, these tubes were collected and measured using flow cytometry. The data with an exponentially asymptotic approach to a constant limit is shown in Figure 4.3B. This represents production which experiences a step increase at $t$=0 and a first order decay. If, for example, there was a lag period associated with induction and expression[49], this would evidence itself as an inflection point before the steep increase; however, such a lag is not seen within the one hour resolution of the experiment. The calculated half-life of the YFPLVA in this experiment is 6.3 hours, consistent with the previous experiment[50].

## 4.5. "Plate reader" population-level measurements.

The majority of measurements performed *in vivo* on *Synechococcus* have involved observing a population of cells (either in liquid culture or as microcolonies on a solid surface) with luciferase driven by a particular native promoter. The resulting oscillating signal is then analyzed to compare periods, amplitudes, and so on to test some aspect of the circadian system. To further establish fluorescence as a well-understood

[49] This lag could come from a slow release of the repressor, or a delay in folding the protein, for example.
[50] The slight increase could come from a longer doubling time. These cells are producing YFPLVA at a much larger rate than they normally produce other proteins, which could potentially tax their metabolic resources. This problem is avoided in the rest of this study by the use of weaker native promoters.

tool in *Synechococcus*, I performed similar experiments utilizing YFPLVA as the reporter.

### 4.5.1. Plate reader apparatus.

A Nikon Eclipse E800 upright microscope was prepared for long term observation of *Synechococcus* (see Figure 4.4). An enclosure was made from 1/16" aluminum sheets; the components were designed with overlapping edges and could be made light-tight by covering these seams with electrical tape for bioluminescence measurements. The front panel was easily removable to allow samples to be inserted or removed and allow direct visual observation via the eyepieces with ready access to the coarse focus knob. A ProScan controller (Prior) controlled the stage and fine focus. The camera used for all observations on this microscope was a 1392 x 1000 pixel, 12 bit, thermoelectrically cooled CoolSnap HQ camera (Photometrics). Uniblitz shutters (Vincent Associates) were placed in the transmitted and fluorescence excitation light pathways. Fluorescence excitation light was provided by a xenon arc lamp (Sutter Instruments); this source provides the temporal stability needed for observations over the time scale of several days. Metamorph software (Universal Imaging) controlled all of these components. A circular fluorescent light bulb (the "feeder bulb", to avoid confusion with the fluorescence excitation lamp) was positioned around the objective; with the use of a piece of window screen, this bulb (when powered) provided 4200 lux at the sample location. The feeder bulb was connected to an extension cord with a relay along its length; the state of the relay was regulated by a TTL signal from a serial port on the computer running Metamorph. In addition to providing the light needed to sustain

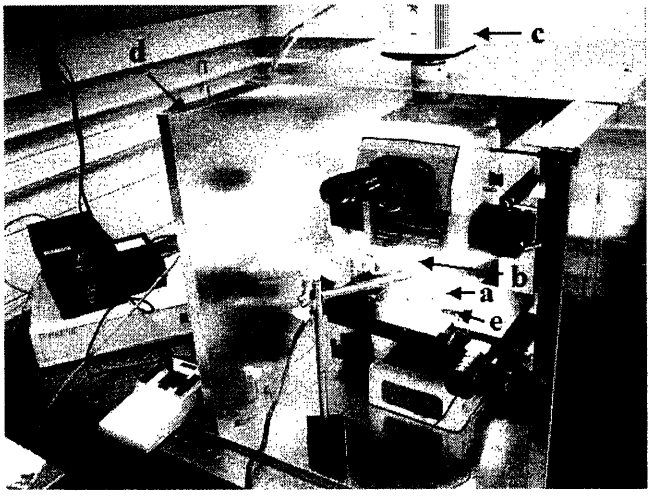


**Figure 4.4.** "Plate reader" microscopy setup. A sample (a) is positioned below a low power objective. A circular cool fluorescent light bulb (b) provides the light needed to sustain growth; the state of this light is controlled by a computer (not shown). At programmed intervals, this light turns off, the sample is exposed to fluorescence excitation light, and the emission is detected by a CCD camera (c). The entire setup is enclosed in a custom aluminum box (d, shown with the front access piece removed). This enclosure isolates the sample from the ambient light level in the lab, and (coupled with the light bulb) maintains the temperature at the desired level. The temperature can be checked by a temperature probe attached to the stage near the sample (e). To the left in the image various shutter and stage controllers can be seen, as well as a joystick to manually control the stage.

---

growth of *Synechococcus*, the feeder bulb maintained the temperature of the enclosure at 30 degrees[51]. A 10x objective was used for the observations of populations (to sample from a large volume of cells); YFP and Texas Red filter cubes (Omega) were used for selecting fluorescence excitation and emission wavelengths.

[51] The temperature could be controlled by adjusting the tightness of the contact between the front panel and the rest of the enclosure.

### 4.5.2. Experimental protocol.

For a typical plate readerexperiment, several wells of a 96 well plate (Corning) were filled with 250 μL of early to mid exponential cells from various strains or preparations[52]. The remaining wells were filled with 300 μL of BG11 media, and the entire plate was sealed with Parafilm. During a run, some evaporation would occur quickly, but the wells were sufficiently isolated that it generally accounted for approximately less than 5% of the well volume. For runs with less than 96 samples to be examined, the targeted wells would be located within a rectangular area including the center of the plate.

A typical data collection run would occur as follows. The plate would be loaded onto the stage. The center axis of a well was found using transmitted light. Using the Texas Red filter cube, the focus was adjusted until the camera detected the largest average red fluorescence signal (from the chlorophyll) across the field of view. The stage was moved so the center of the rectangle to be probed was directly under the objective. At this time, a journal script in Metamorph was executed every time interval (usually twenty minutes), iterating the following process. The feeder light was turned off. The stage would translate so each well would be positioned under the objective for measurement. The fluorescence excitation lamp would illuminate the well (through the YFP filter cube) for five seconds, during which time the camera would integrate the signal from the cells. The resulting image was recorded and the stage moved to the next well. After all wells were sampled, the resulting stack of images was automatically saved

[52] These may include cells with fluorescent reporters driven by different promoters, with different synchronizations, or wild type cells or blank BG11 media (for background correction). Additionally, multiple wells would be filled with identical samples to test reproducibility.

to a hard disk and the feeder light was turned on again. The typical length of an experiment was three to five days[53].

### 4.5.3. Analysis of plate reader data.

The data was analyzed by examining the average fluorescence over the entire collected image[54]. Typical results from these "plate reader" experiments for the native *kaiBC* (strain JRCS32) and *psbAI* (JRCS33) promoters driving *yfpLVA* are shown in Figure 4.5. The fit represents an exponentially growing culture with a sinusoidal oscillation of arbitrary phase and amplitude, along with a constant level of signal[55]

$$y = O + 2^{-t/\tau}\{A[\sin(2\pi t/T + 2\pi I/24) + 1] + C\}$$

Here, $A$ represents the amplitude of the oscillatory part of the signal from the culture, while $C$ describes the non-oscillating part of the growing signal; $(A+C)/C$ can be considered as a peak to trough ratio[56]. $\tau$ represents the time to double the signal. A previous study (51) used this equation (without the $O$) to describe the behavior they observed using luciferase reporters. The interpretation of their parameters differs from those in this study, and comparisons should be made carefully. For example, bioluminescence does not seem to build up over time; therefore, $C$ for that study truly is a

[53] A concern with longer experiments was the depletion of available nutrients (namely, nitrogen and carbon dioxide) in the relatively small wells.

[54] Analyses of smaller regions within the image may yield different mean signals, as the fluorescence excitation lamp creates light gradients across the field of view. The behavior of the signal over time was exactly reproduced regardless of the size of the region considered. Therefore, the entire image was used to smooth any pixel to pixel noise.

[55] This equation is modified from (51) to include the constant offset term $O$, which represents fluorescence that does not increase over time. The camera provides a certain background (even when the shutter is not opened), the plastic in the wells has some intrinsic fluorescence, and the media has a (very weak) autofluorescence. The data plotted has not had any of these corrected, and so they appear in $O$.

[56] If the constant offset $O$ was zero (as is the case for bioluminescence), the *PTR* of the signal is fixed; however, with fluorescence, the peak-to-trough ratio does increase over time.

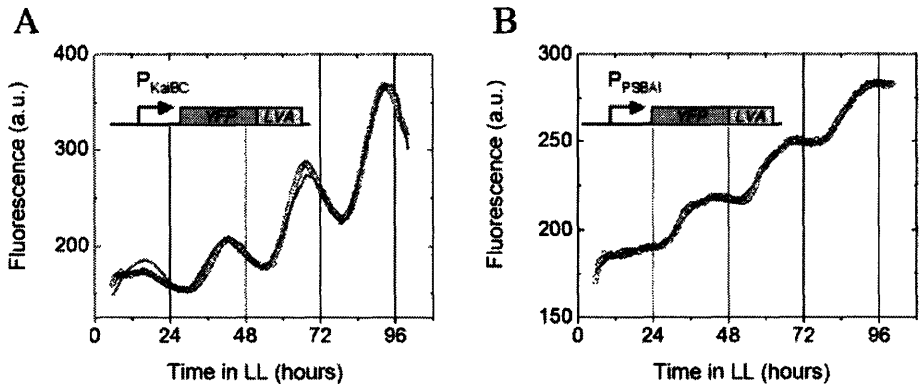


**Figure 4.5.** Plate reader results. YFPLVA was expressed under the control of the *kaiBC* (A) or *psbAI* (B) promoters. The bulk fluorescence was measured every twenty minutes using the plate reader microscopy setup. The line represents the best fit to the function $y = O + 2^{-t/\tau}\{A[\sin(2\pi t/T + 2\pi I/24) + 1] + C\}$; fitting parameters are given in Table 4.2.

---

basal expression (or "leakiness"), whereas in this study it is both leakiness and integrated protein from previous oscillations. Another difference is the interpretation of $\tau$. For bioluminescence, the signal will double when the cells have doubled (at the same circadian phase). With fluorescence, as cells are exposed to the excitation light and YFP absorbs the light, there is a chance that a fluorescent protein may be rendered nonfunctional by the absorption of the quantum of energy and subsequent nonradiative relaxation. This photobleaching process means that even once the cells have doubled, some of the accumulated protein has been destroyed from previous exposures and thus the signal level will be less than twice the amount before the cell doubling. The fitting parameters for the data in Figure 4.5 are shown in Table 4.2.

**Table 4.2.** Fitting parameters for data in Figure 4.5.

| Strain | $\tau$ (hours) | $T$ (hours) | $I$ (hours) | $(A+C)/C$ |
|---|---|---|---|---|
| JRCS32 | 36.7 | 25.4 | 9.19 | 1.407 |
| JRCS33 | 106.6 | 26.4 | 13.4 | 1.03 |

## 4.6. Effect of fluorescent reporter on phase of oscillation.

The phase of the oscillations in JRCS32 ($P_{kaiBC}$::*yfpLVA*) can be readily compared to results in a previous study (51), showing oscillations in bioluminescence driven by the same promoter. The bioluminescence results indicate that the peak signal occurs around hour 14-15 of the circadian day (with hour 0 defined as the time the cells transition from darkness to light). The fluorescence, however, peaks around hour 18. Early results (18) showed an excellent correspondence between luciferase signal and mRNA levels (as was discussed in Chapter 2). Since no lag phase was seen for YFPLVA expression, the source of this delay must be explained. Once again, the answer can be found in the lifetime of the reporter molecule.

Consider a driving signal with frequency $\omega$ and arbitrary phase $\phi = 0$. A reporter being produced with a production rate $k$ modulated by such a signal, with an constant production offset of $A$ and a first order decay with constant $\gamma$, is described by

$$\dot{y} = k[\sin(\omega t) + A] - \gamma y$$

The solution to this takes the form

$$y = y_0 \sin(\omega t + \phi) + c$$

Inserting this solution yields

$$y_0 \omega \cos(\omega t + \phi) = k \sin(\omega t) - \gamma y_0 \sin(\omega t + \phi)$$

Trigonometric substitution gives

$$y_0\omega[\cos(\omega t)\cos(\phi)-\sin(\omega t)\sin(\phi)] = k\sin(\omega t)-\gamma y_0[\sin(\omega t)\cos\phi+\cos(\omega t)\sin(\phi)]$$

Dividing by $\cos(\omega t)$ and grouping terms independent of $t$ reveals

$$\tan^{-1}(\phi) = -\frac{\omega}{\gamma}$$

This term represents the phase difference observed between the reporter and the driving signal.

In this system, with a measured decay rate of $\gamma$= ln(2)/(5.6 hours) = 0.124 $hr^{-1}$ and $\omega = 2\pi$ / (24 hours), a calculated phase delay of 4.3 hours would be expected (see Figure 4.2B), in good agreement with the experimental evidence.

## 4.7. Summary of effects introduced by fluorescence.

In summary, making the transition from bioluminescence to fluorescence as a reporter introduces the following effects:

1) The fluorescence signal displays a diminished peak to trough ratio compared to the input signal (level of genetic activity), owing to integration of previously-produced protein. The use of an *ssrA* tag attached to the fluorescent protein can shorten its lifetime to improve the *PTR* of the signal.
2) Repeated measurements on the same sample can cause photobleaching, which may be misinterpreted in bulk studies as a lengthened doubling time.
3) A phase shift owing to the long lifetime of fluorescent proteins appears.

The impact of each of these factors is well studied, and can easily be accounted in cyanobacterial circadian studies using fluorescent proteins to probe genetic activity.

# 5. Single cell dynamics.

## 5.1. Use of bioluminescence for single cell time courses.

Once fluorescence has been established as a technique for probing circadian rhythms on the bulk population level, the natural extension of this method is to examine the behavior of individual cells as they progress through the circadian cycle. As discussed in Chapter 2, a study (48) has shown circadian rhythms in individual *Synechococcus* but with major concerns regarding a) long exposure times in a dark environment; b) low signal to noise; and c) division times nearly matching the observed period of oscillation. The use of fluorescent reporters (along with appropriate growth conditions) for single cell microscopy addresses each of these concerns.

## 5.2. Experimental apparatus.

### 5.2.1. Requirements.

Keeping any organism alive and in good condition in a microscopic observation setup presents certain experimental challenges, and *Synechococcus* is no exception. The cells require light to produce food, yet data acquisition must take place in the dark. The cells themselves must be maintained at a fixed temperature. Growth for extended periods of time requires a significant amount of culture media to buffer the nutrient concentration around the cells. The cells themselves must be immobilized within a plane to permit the observations, yet must maintain access to the media. Finally, air should be present in the culture chamber to act as a reservoir of carbon dioxide, yet if the sample is open to the environment, evaporation may cause the sample to dry out over the long stretches of time needed to collect several complete oscillatory periods.

### 5.2.2. Growth conditions.

The light required for growth was provided as described in Chapter 4, by a circular fluorescent bulb controlled by Metamorph. Loops constructed of window screening surrounding the sample adjust the light to the desired level. The light level for these single cell experiments was maintained at 4200 lux for two reasons. First, this level is the same as that used for all other experiments in this study. Second, this light level leads to doubling times approaching ten to twelve hours; this eliminates artifacts of cell division leading to observed fluctuations with the same frequency as the circadian rhythm. The light also maintains the enclosure surrounding the microscope at 30 degrees; this temperature is uniform across the entire sample chamber.

### 5.2.3. Growth chamber.

The requirements of media, air, and sample immobilization were solved using a custom sample chamber (see Figure 5.1). This chamber was constructed by cutting a 1 cm hole (using a sheet metal drill bit) in the lid of a 60x15 mm petri dish (Corning). A 15 mm #1 coverslip was epoxied over this hole. A 4 cm hole was bored (using a lathe) into the lid of a 100x15 mm petri dish (Corning). The smaller lid was inverted and glued with silicone RTV inside the larger lid, such that the coverslip was centered under the large hole. To prepare a sample, this construction was placed with the large hole facing downward. A 100 mL droplet of mid-exponential cells was placed on the coverslip, and allowed to settle for 30 minutes (at 30 degrees and 4200 lux) and naturally adhere to the

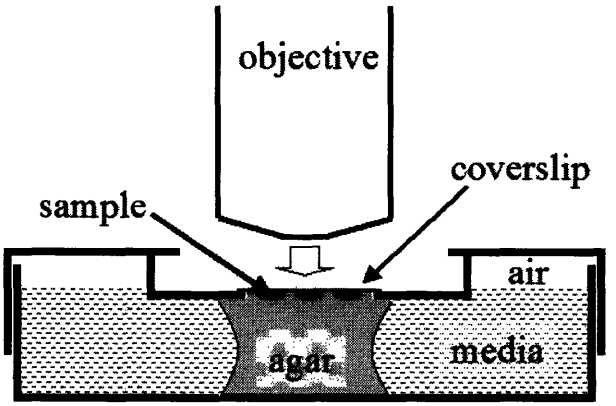


**Figure 5.1.** Single cell microscopy growth chamber. The chamber itself is assembled from two petri dishes. The larger has a hole in the lid to admit a microscope objective. The smaller (of which only the lid is used) has a hole which is covered with a coverslip. Cells adhere naturally to the glass of the coverslip, and are held against the glass as they grow and divide by a column of low-melting point agar. The cells and agar are submerged within media; the chamber also contains air, which acts as a reservoir of carbon dioxide. Cells may be maintained within such a chamber for over a week without diminishing growth rates.

---

coverslip[57]. After this time, the liquid media was removed with a pipet, and the deposited cells were immediately covered with approximately 1 mL of 4% low melting point agarose (Shelton Scientific Inc.) in BG11 media, which had been melted in a microwave and allowed to cool to nearly 30 degrees[58] while being rotated to remove bubbles. The bottom of the large petri dish was placed over this sample forming a column of agarose between the bottom half of the dish and the coverslip. The two halves of the dish were sealed with parafilm and allowed to sit for 30 minutes to allow the agarose to harden.

[57] In order to form a stronger connection between the cells and coverslip, some coverslips were pretreated with 1% poly-l-lysine. However, *Synechococcus* never grew on these coverslips, possibly because the poly-l-lysine interferes with the membranes used for photosynthesis; this is purely speculative. In any case, only untreated coverslips were used for the experiments discussed in this chapter.

[58] The melting temperature of this agarose is 65 degrees, with a gelling point of 28 degrees. It will eventually become firm at 30 degrees, but it remains fluid long enough to cover the cells.

After this time, the dish was returned to its upright position, and BG11 media was introduced through a small hole in the lid. Enough media was supplied (usually 10-15 mL) to cover the agarose column and the cells while still leaving a significant amount (10 mL) of air above the liquid. The small hole was covered with electrical tape to cut evaporative losses into the lab environment. The entire sample holder was glued (with silicone RTV[59]) onto a microscope stage insert; after allowing the adhesive to set (usually overnight), the sample was placed onto the Nikon E800 microscope.

## 5.3. Observation of growing cells.

Once the sample was on the microscope, a field of cells was selected for repeated observation. An ideal field contained approximately ten individual cells[60] spaced by five to ten cell lengths to allow for an increasingly crowded field as the cells divide. The complicated series of optical interfaces in the sample often prevented the acquisition of sharp phase contrast images using the normal annular rings and condenser alignment. Therefore, various combinations of these settings (phase rings; condenser alignment) were tested to give the best phase contrast image. An autofocusing program (part of Metamorph) used to find the ideal plane of focus. Once this plane was located, the autofocusing routine was iterated to verify that the same plane would be identified each time; if the plane would vary appreciably, another field of cells was selected.

Another consequence of the optically complex pathway followed by the

[59] Gluing the holder serves to prevent translation of the sample during the course of the experiment. The RTV can be cut away with a razor blade after the conclusion of the experiment, and the stage insert can be reused.

[60] Over time, the majority of these cells would not grow and divide (often as many as eighty percent); these cells perhaps did not survive the deposition onto the coverslip.

transmitted light through this sample is that the fluorescence focal plane is displaced from the plane where the phase contrast is most sharply in focus. However, the fluorescence cannot be used for focusing, as photobleaching would rapidly destroy the reporter molecules and the signal-to-noise ratio would diminish to an unusable level. To accommodate these concerns, the fluorescence focal plane was located during the setup for each sample using a Texas Red filter cube (Omega) which effectively selects chlorophyll fluorescence. During the data collection, the phase contrast image would be used to adjust the focus, then the height of the stage was automatically raised or lowered to bring the sample to the fluorescence focal plane prior to collecting a fluorescence image.

A typical run would involve executing the autofocus routine every three minutes. After a total of forty-five minutes, a one-second fluorescence exposure was taken. The images would be automatically saved after collecting the fluorescence image. Runs would proceed for up to five days. After approximately one week, the agar column would often start to pull away from the coverslip, and the cells would escape into the media. To prevent this in future studies, a permeable membrane might be used to hold the agar firmly against the surface (as was done in (48), in which data was collected for up to two weeks). However, with doubling times of ten to twelve hours, the density of cells becomes so high after one week that further analysis becomes difficult; cells grow out of the plane and microcolonies start to be in contact with one another.

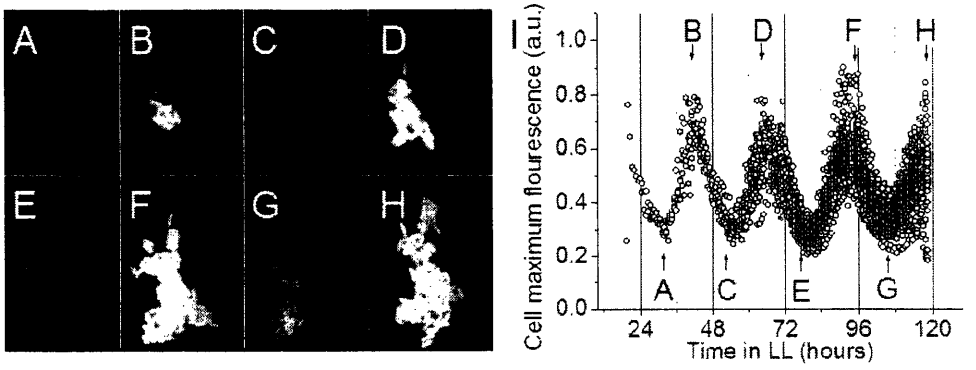


**Figure 5.2.** Growing microcolony of strain JRCS32 ($P_{kaiBC}$::*yfpLVA*) growing on microscope. (A-H) Selected fluorescence images from various times throughout the experiment. The images are shown with identical scaling. (I) Quantitative plot of data from images in this run. Each circle represents the maximum fluorescence pixel value from an individual cell at the time indicated on the abscissa. The data in (I) is corrected for uneven illumination from the xenon arc lamp.

## 5.4. Results.

Figure 5.2(A-H) shows selected images from a time series of exposures of strain JRCS32, which expresses YFPLVA under the control of the *kaiBC* promoter. The gray level indicates the fluorescence of the cells; the scaling in all images is identical. Figure 5.2(I) shows the maximum fluorescence pixel value of each cell in the image as a function of time[61]. The phase and period of these measurements agrees with the results for this same strain as shown in Figure 4.5.

[61] Because the cells in a microcolony contain many edges that are in contact, regions defining the borders of each cell were defined by hand. The maximum was used to avoid averaging issues if the region borders extended beyond the actual size of the cell. Watershed algorithms may be used in the future to automate this process. For larger microcolonies, only cells within 2 cells of the colony edge were analyzed, to ensure that a second layer of cells was not below the plane of focus (and would contribute to the fluorescence measured).

## 5.5. Observations of *kai* knockout strains.

One of my original motivations to study circadian rhythms in individual cells concerned the causes of the loss of rhythmicity in strains with inactivated clock proteins. As discussed in Chapter 2, the Kai clock proteins were originally identified in strains constructed with random transposons, which disrupt genes into which they are inserted. When bulk samples of these strains were screened for bioluminescence, the signal remained constant, in contrast to the clear oscillations seen in samples with a wild-type genetic background. What is unclear from these results, however, is whether or not the Kai proteins are responsible for oscillations on a single-cell level, or if they might (for example) be responsible for entrainment to external cues. If the cells constituting a population lack the ability to align their oscillations to external cues, then even small amounts of phase noise will cause these independent oscillators to rapidly desynchronize from one another. The resulting observation of bulk bioluminescence would then give a flat signal, owing to averaging over millions of effectively randomly-phased oscillators. Performing single-cell observations of these knockouts can eliminate this concern and truly verify that the Kai proteins are necessary to produce oscillations in individual cells.

To this end, knockout strains AMC702, AMC703, and AMC704 (kind gifts of Susan Golden, Texas A&M University), which have *kaiA*, *kaiB*, and *kaiC* respectively inactivated, were transformed using pJRC27, a NSI reporter plasmid with YFPLVA expressed under the *kaiBC* promoter[62]. Cultures were synchronized for at least two full 12:12 LD cycles, and samples were prepared as above. Long-term observations (at least three full cycles) were performed on each strain. The mean fluorescence of the members

[62] This is the same reporter construct used for the strain shown in Figures 4.4B and 5.2.

of a growing microcolony was determined from the images with a time resolution of 40 minutes. The resulting data (although clearly non-oscillating) was fit with the best possible 24 hour sinusoidal oscillation, with phase, amplitude, and DC offset allowed as fitting parameters. The peak to trough ratio was computed, and compared to the data shown in Figure 5.2. The results are given in Table 5.1.

**Table 5.1.** Results of single-cell observations of *kai* deletion strains.

| Strain | Peak-to-trough ratio | $(PTR_{strain\text{-}x}-1)/(PTR_{JRCS32}-1)$ |
|---|---|---|
| JRCS32 (wild-type background) | 2.085 | 1 |
| JRCS62 (Δ*kaiA*) | 1.031 | 0.028 |
| JRCS63 (Δ*kaiB*) | 1.074 | 0.068 |
| JRCS64 (Δ*kaiC*) | 1.084 | 0.077 |

These results clearly show that the loss of any one of the Kai proteins abolishes the circadian rhythm in *Synechococcus*. These data justify the development of models (such as those presented in Chapter 2) in which these proteins form the core of the circadian oscillator[63]. This study is the first to conclusively show the critical, fundamental role that these proteins play in the circadian oscillation of this cyanobacterium on an individual cell basis.

[63] Note that, although these proteins are shown here to be essential for the individual cell oscillation, it has not yet been demonstrated that *only* these proteins form the core oscillatory mechanism.

# 6. Flow Cytometry.

## 6.1. Why cytometry?

The plate reader is very useful for obtaining data of the type that has been previously collected using bioluminescence, and the microscopy discussed in Chapter 5 permits repeated measurements of low numbers of cells. However, to build the statistics needed to report genetic noise in the circadian oscillator, many cells (i.e., thousands or more) need to be measured. One solution is to make a series of slides representing various circadian times and use the microscope to take images of many fields of view on each slide. A single slide can produce data on ~1000 cells in fifteen to twenty minutes of time. This means that in order to collect 24 points over a single circadian period, eight hours of data collection is needed (and typical runs will involve collection for several periods). This is impractical, given the passage of time during the experiment (during which samples will advance through an appreciable part of the circadian cycle), inconsistencies in instrument performance over long times, finite access to microscopy equipment, and so on. Another method for collecting data rapidly on large numbers of individual cells is flow cytometry. Using this approach, 24 hours of data can be collected in a single hour of sample measurement, addressing the various concerns just mentioned.

## 6.2. Instrument setup.

The discussion of flow cytometry that follows is specific for the FACScan cytometer (BD Biosciences), with which the data presented was collected; the software controlling the cytometer is CellQuest (also BD Biosciences), running on a Macintosh

computer. The data was collected in the Cell Sorting Facility in the Center for Cancer Research at MIT.

Flow cytometry involves flowing cells in a narrow capillary past a laser beam focused in the path of the cells. As cells (or other particulate matter in the solution) pass through the laser beam, light is scattered in all directions; a detector located orthogonally to the flow and laser measures side scattered light, and a detector just off the straight path of the beam measures forward scattered light. The forward scatter detector is a photodiode; all other detectors in this system are photomultiplier tubes. For large eukaryotic cells, these scattering properties are related to the size of the cell and its internal complexity; for small prokaryotic cells, they are both approximately related to the size of the cells. The laser also excites fluorescence in the cell. Along the path leading to the side scattering detector, a series of dichroic filters select emitted light with different wavelengths; a total of three fluorescence emission wavelengths are measured. In the FACScan, the argon excitation laser has a 488 nm wavelength, and the three fluorescence emission wavelengths monitored are at 530 ± 15, 585 ± 21, and ≥650 nm. The laser significantly excites YFP, and the 530 nm filter effectively measures fluorescence of this protein. The 650 nm filter is also useful for measuring the chlorophyll fluorescence of the cells, which can be a useful tool for distinguishing live cyanobacteria from dead cells[64]. The biases on each detector can be independently set to give the optimum dynamic range for each parameter observed. For all experiments in this study, the settings used are listed in Table 6.1.

[64] Dead cyanobacteria have dramatically less red fluorescence than live cells; this may result from a compromised cell membrane or denaturing of the chlorophyll after cell death.

**Table 6.1.** Settings used for flow cytometry studies of *Synechococcus*.

| Parameter | Voltage | Amp gain |
|---|---|---|
| FSC (forward scatter) | E01 | 4.26 |
| SSC (side scatter) | 443 | 1 |
| FL1 (530 nm fluorescence emission) | 900 | 1 |
| FL2 (585 nm fluorescence emission) | 801 | 1 |
| FL3 (650 nm fluorescence emission) | 505 | 1 |

When an object passes through the beam, the forward and side scattering properties are measured. If the cell lies within a certain region (or "gate") in forward scattering versus side scattering space, these scattering parameters and its fluorescence at the three wavelengths are recorded to a file. This region is selected to contain the majority of the population of cells (readily evident in samples of moderate density, $OD_{700} \approx 0.1$)[65]. Data for each sample is collected until either 300,000 scattering events in the gate are logged, or two minutes have transpired since beginning the run[66]. This time limit was enforced to allow an entire run (~25 samples) to take place in one hour or less.

## 6.3. Sample preparation.

In order to collect data on populations of individual cells with a time resolution of one hour, a sample with a single synchronization must be measured every hour for the length of the run, or many samples differently synchronized can be measured several times. For obvious reasons, the latter strategy was employed, using the circadian spinner

[65] Later, this gate can be restricted in post-collection analysis.
[66] The FACScan can measure a maximum of 10,000 events per second (regardless of their scattering properties). For the fastest samples, approximately 5,000 events per second would lie within the gate; therefore, the shortest runs were around one minute in length.

described in Chapter 3. Cells were loaded into 15 mL tubes with 0.2 μm filters epoxied into the lids to allow gas exchange, in 3 to 4 mL of BG11 with antibiotics as appropriate. A tube of wild-type cells was also prepared, and traveled around the spinner between the tubes of cells moving into the light at 12AM and 1AM. All tubes were placed on the spinner simulataneously and removed two complete LD cycles later as described previously. The cells were then divided into 0.5-0.75 mL aliquots in 5 mL polystyrene culture tubes (Falcon) with loose caps that allowed gas exchange and stored in the constant light box. For a measurement, a complete set of tubes would be taken to the cytometer and measured in a random order to cause any systematic variation in the cytometer to appear as high frequency noise compared to the expected slow circadian variation. Measurements were performed less than 24 hours apart; this leads to overlap in the data sets and ensures that the relative phase of each tube would vary from run to run. In this way, the oscillation of each tube could be checked by observing its behavior relative to neighboring tubes, eliminating concern that the observed oscillation was an artifact of variations in growth conditions from tube to tube.

## 6.4. Data analysis.

The data collected from the cytometer was analyzed using a combination of software packages. The raw data files (in a binary format) were converted to ASCII using MFI (Median Fluorescence Intensity, Eric Martz, University of Massachusetts; available at http://www.umass.edu/microbio/mfi/). Once in ASCII format, the sample could be subsampled using a smaller forward/side scattering gate using a custom script in

Perl. Usually, the larger population was restricted to a gate consisting of 25 bins by 25 bins[67]. This data was then imported into Origin (Microcal); the logarithmic bin number was converted to a linear fluorescence value using the formula

$$f_{\text{linear}} = 10^{\left(\frac{4*(\text{bin \#})}{1024}\right)}$$

distributing the fluorescence values over a range of 1 to $10^4$. These linear fluorescence values were binned to make a histogram, with bin sizes chosen such that all linear bins would be occupied by at least two logarithmic bins for the range of data values selected[68].

Once this histogram was constructed, a gaussian fit was applied, with a baseline of zero. I chose to fit a gaussian rather than use the mean and standard deviation of the population for two reasons. First, occasionally a secondary (much smaller) peak would appear at fluorescences consistent with non-cyanobacterial particles in the solution. The quality of the fit is highest by very closely following the curve of the peak corresponding to the cells, thereby effectively excluding these background events without changing the mean and standard deviation calculated for the target cells. The second reason to fit a gaussian is to minimize the impact of substantial outlying data points. Discretion must be employed in using this approach, as many biological process produce distributions which are log normal, rather than normally distributed in a linear space; in this case, points which appear as outliers may actually be part of the distribution. All samples measured appeared to be normally distributed in a linear space, but to probe the nature of the distribution experimentally, over 100 distributions were fit to both linear gaussian and log

[67] Each detector's output is scaled into 1024 bins covering the detection range. These bins can be either linearly or logarithmically distributed; for all runs in this study, a logarithmic range was used.

[68] As the data is falls into logarithmically distrubited discrete values, as the fluorescence increases, the spacing between possible fluorescence values increases. If the linear bins are too narrow, they may be either unoccupied or underpopulated if they just "miss" including a bin. By including at least two logarithmic bins, the data appears much smoother and the gaussian fits are better.

normal curves and the quality of fits were compared. The results indicate that the curves are in fact gaussian in a linear space.

To correct for autofluorescence of the cells (independent of YFP), wild type *Synechococcus* PCC7942 were measured as part of each run. The mean fluorescence of these cells is subtracted from the mean of the samples. In addition, the variance of the wild type cells is subtracted from the variance of the samples; this should correct for both noise introduced by the cytometer as well as the background spread[69]. To describe the genetic "noise" in the system, the parameter chosen is the background- and instrument-corrected standard deviation normalized by the mean expression at a given time; this dimensionless quantity is also known as the coefficient of variation (CV).

## 6.5. Results.

Figure 6.1 shows data for a typical experimental series (3 measurements of 24 tubes with phases varying by 1 hour) for JRCS32 ( (I)$P_{kaiBC}$::*yfpLVA* ). The mean fluorescence oscillates in agreement with the data shown in Chapter 4; however, the signal is now normalized to be per cell. The CV oscillates as well, roughly out of phase with the mean. Existing gene noise theories (which will be discussed further in Chapter 7) predict that in a system where different mean values can be achieved (for example, by using an inducible promoter), the noise should vary inversely. Simply put, systems with higher means should have lower noises; systems with lower means should have higher

[69] The noise from the cytometer may affect signals of different strength by varying amounts. For the bulk of these experiments, the signals are typically 3-5 times larger than background, so the correction is only approximate (but certainly better than either ignoring it or if the signal was 100 times greater than background).

noises[70]. However, when we plot the CV as a function of the mean expression (Figure 6.2), an interesting behavior emerges. The resulting curve is not a line that follows the same path from low to high mean, but rather traces a loop in CV versus mean space. The interpretation of these "noise loops", and understanding their causes, will form the goals for the next 4 chapters.

---

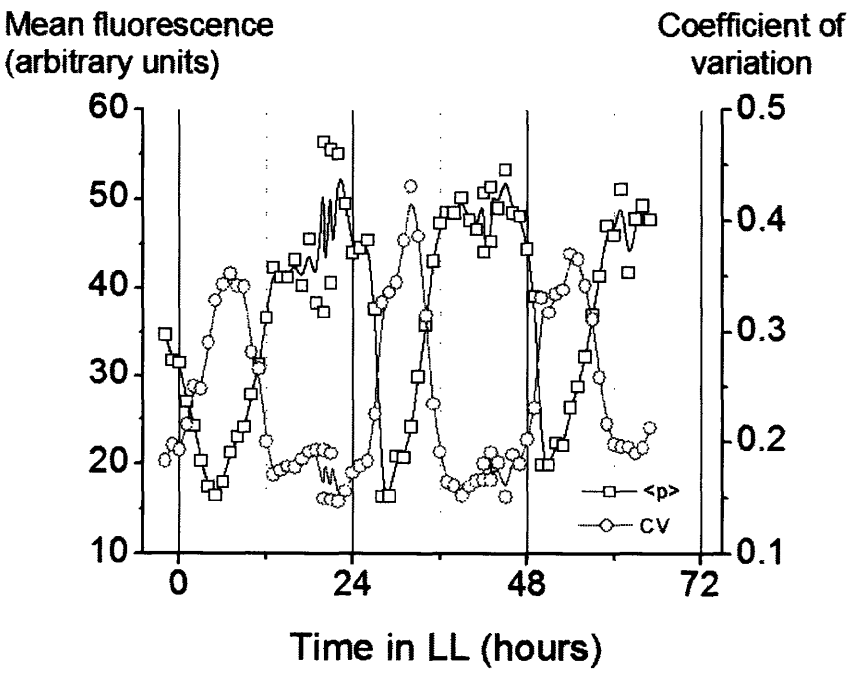


**Figure 6.1.** Results from flow cytometry. Data is from strain JRCS32, expressing YFPLVA under the *kaiBC* promoter. The background-corrected mean fluorescence (squares) and coefficient of variation (or standard deviation normalized by the mean, circles) are plotted against time.

[70] This originates with the central limit theorem, and will be discussed further in the next chapter.

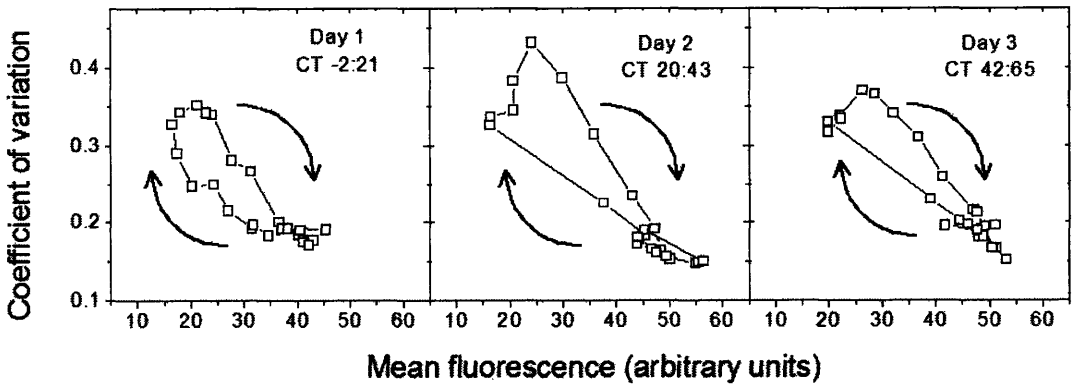


**Figure 6.2.** Noise loops. The data from JRCS32 ($P_{kaiBC}$::*yfpLVA*) for the three days shown in Figure 6.1 are reproduced here, with the coefficient of variation plotted against the mean fluorescence. The direction of rotation of the loops in each case is clockwise, with the noise lower as the mean decreases and higher as it increases.

# 7. Introduction to genetic noise.

## 7.1. Introduction.

The understanding of noise loops in the *Synechococcus* circadian system requires an examination of noise in other well-studied genetic systems. The study of the sources and description of noise of genetic origin has been developed since 1978 (70). However, the tools to investigate this variability in large populations of individual cells have only emerged recently and thus a significant number of studies (both theoretical and experimental) have emerged in recent years. Here, I will discuss two theoretical studies and two experimental studies, which will develop the basis for the understanding and explanation of noise loops in the oscillating circadian system.

## 7.2. Noise arising from mRNA stochasticity.

### 7.2.1. Theory.

One source of variation in cells arises naturally from number fluctuations in chemical species with low absolute numbers of molecules. Bacterial cells will typically contain a single copy of a gene, tens of repressor or activator molecules (71) and available RNA polymerase molecules (72), $10^2$-$10^4$ ribosomes (of which only a few percent are free for translation at any time (73)), and so on. The result of these low numbers is a significant stochasticity in the production and decay of individual mRNA and protein molecules. One study (74) considered number fluctuations in the product of a single constantly expressed gene, by considering the production or destruction of

mRNAs and proteins as Markov processes[71]. In an increment of time in this system, an mRNA may be produced from the DNA encoding it with a fixed probability. Proteins are produced from the current pool of mRNAs with a probability proportional to the number of encoding mRNAs in the cell. Both mRNAs and proteins may disappear with probabilities proportional to their respective numbers. This scheme is shown in Figure 7.1. The system takes discrete unit steps in a two dimensional mRNA/protein space. The system will approach a steady state probability distribution in this space. A cell's protein number will fluctuate, and the probability of observing it having a given number of proteins will be described by the distribution obtained by summing the probabilities of all states with the same protein number (but any mRNA numbers). If cells are independent, observing a large number of cells is equivalent to observing a single cell many times[72].

In order to describe the static protein probability distribution, Monte Carlo programs may be used to simulate the population of cells. However, an analytical approach confers greater insight by transparently revealing the final distribution's dependence on the various parameters describing the core system. Using a master equation approach (which will be discussed in more detail in Chapter 9), it was shown that the standard deviation of a population of cells $\sigma_p$ is related to its mean value $\langle p \rangle$ by the following relation (with variables as in Figure 7.1)

$$\frac{\sigma_p^2}{\langle p \rangle} = \frac{k_p/\gamma_r}{1+\gamma_p/\gamma_r} + 1$$

[71] A Markov process is a random process where the future behavior is determined only by the current state of the system, lacking a "memory" of previous values.

[72] This technique may be a more rapid way to build the probability distribution, because knowledge of the static probability distribution does not reveal how long it will take a cell to traverse the range of protein values. If the protein level changes very slowly, the observations need to occur over exceptionally long times in order to build the distribution in this way.

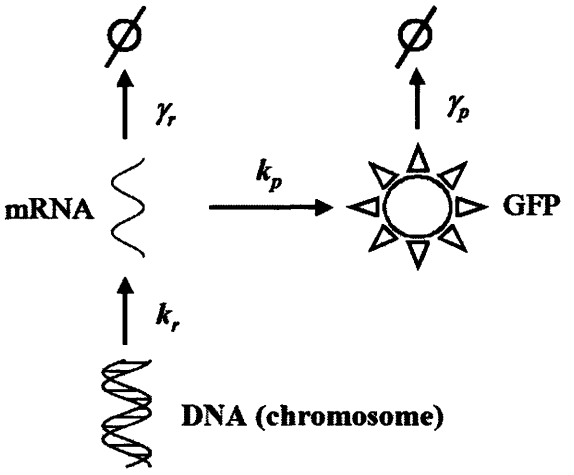


**Figure 7.1.** Basic scheme for modeling steady-state expression of reporter protein. The DNA is transcribed into mRNA at a rate $k_r$ (this rate may be modified *in vivo* by the use of inducible promoters). The mRNA is, in turn, translated into proteins by ribosomes with rate constant $k_p$ (this rate constant may be modified by the use of ribosome binding sites before the start codon with different affinities for ribosomes). The mRNA and protein undergo first order decay with rates $\gamma_r$ and $\gamma_p$ respectively.

---

The term $\gamma_p/\gamma_r$ represents the ratio of protein to mRNA decay rates, or equivalently, the ratio of mRNA to protein lifetimes. mRNA molecules tend to be very short lived (half lives of under ten minutes are very common), whereas proteins are more long lived (with effective half lives often equal to the division time of the cell); thus, this ratio goes to zero. The term $k_p/\gamma_r$ can be thought of as the average number of proteins produced by an individual mRNA during its lifetime. If mRNAs are produced infrequently and live for a short time, the proteins will appear in "bursts" of average size $k_p/\gamma_r$ ; this ratio is therefore called the "burst size" (indicated by the letter $b$). The above

equation can then be simplified to

$$\frac{\sigma_p^2}{\langle p \rangle} \approx b+1$$

A Poisson process leads to a distribution where $\sigma_p^2 = \langle p \rangle$. Therefore, the ratio $\sigma_p^2 / \langle p \rangle$ in this system is increased above its "basal" level by the burst size $b$. Consider two systems with identical mean expression $\langle p \rangle$, but different burst sizes $b_1 >> b_2$. The first system will have protein levels that drop to a low level, then a large burst overshoots the average, and the process repeats (at stochastic intervals). The second case will remain much closer to the average level at all times, with small bursts occurring frequently to maintain the average. An alternative view of this solution is that, to get a given level of protein, either large bursts occur infrequently, or small bursts happen with a high frequency. Infrequent random events imply a large role for stochasticity, whereas frequent random events take on a certain regularity. An analogy is popping corn: during the height of the popping, individual pops come at short, almost regular intervals; at the end, with few remaining unpopped kernels, the time between pops varies significantly. This analogy will aid in understanding the source of the *Synechococcus* noise loops.

### 7.2.2. Experiments.

The predictions of this model were tested experimentally in another study (75) examining the expression of *gfp* driven by an inducible promoter in the bacterium *Bacillus subtilis*. The parameter $k_r$ was varied by changing the amount of inducer used; the parameter $k_p$ was changed by mutating the ribosome binding site of the reporter

gene[73]. This study revealed that noise (as measured by $\sigma_p^2/\langle p \rangle$) was relatively constant for different values of $k_r$ (leaving $b$ unchanged), whereas the noise varied linearly with changes in $k_p$ (and therefore $b$; see Figure 7.2). This result confirms the predictions of (74) in *B. subtilis.*

## 7.3. Intrinsic and extrinsic noise.

### 7.3.1. Theory.

This description of genetic noise includes only number fluctuations arising from random production and destruction events. However, other sources of noise in genetic systems do exist. For example, the number of ribosomes may vary from cell to cell; this leads to differences in the rate "constant" $k_p$ between individual cells. Cells may have different numbers of RNA polymerase proteins, changing $k_r$. Similarly, the decay rates may differ from variations in proteases, RNases, or division times. The overall environment of the cells may change (for example, the availability of nutrients, or dissolved gases). Whereas the number fluctuations are *intrinsic* to the gene being expressed, these other sources of noise are described as *extrinsic* to the gene (as they affect more than the single gene being discussed). Another excellent theoretical study (76) pieced together the contributions of intrinsic and extrinsic noise in the expression of a gene. The noises $\eta_{\text{intrinsic (extrinsic)}}$ from intrinsic and extrinsic sources add in quadrature:

$$\eta_{\text{total}}^2 = \frac{\sigma_p^2}{\langle p \rangle^2} = \eta_{\text{intrinsic}}^2 + \eta_{\text{extrinsic}}^2$$

[73] The ribosome binding site is part of the DNA transcribed by RNA polymerase before the coding section of the gene. Mutating this sequence modulates the affinity of ribosomes for the mRNA produced by the gene, changing the amount of protein translation per mRNA.

Comparison with the previous model for intrinsic noise gives

$$\eta^2_{\text{total}} = \frac{(b+1)}{\langle p \rangle} + \eta^2_{\text{extrinsic}}$$

Extrinsic noise is not a function of the level of protein expressed (i.e., it contributes equally to all expression levels). Therefore, if the quantity $\sigma_p^2/\langle p \rangle^2$ is plotted against $1/\langle p \rangle$ for various mean levels of expression, the resulting fit should be a line with slope related to internal noise[74], and intercept related to external noise. Re-examining the data of (75) in this light reveals consistency with this more complete description.

### 7.3.2. Experiments.

Several studies have attempted to quantify the relative contributions of intrinsic and extrinsic noise in genetic systems. One excellent example (77) used a system of two fluorescent reporters with non-overlapping excitation and emission spectra to allow separate measurement of the levels of each. The genes for these two proteins (CFP and YFP) were expressed under a *lac*-repressible (and IPTG-inducible) promoter and located in the chromosome equidistant from the chromosomal origin of replication[75]. Consider the two extreme cases of zero intrinsic noise, and zero extrinsic noise. In the first, the expression of each gene is fixed deterministically by the particular rate parameters in the cell, which (although allowed to fluctuate) affect both gene products equally. In this example, the amount of CFP and YFP expression measured should be identical; the range

[74] In general, experimental techniques for measuring noise do not directly quantify the number of proteins. A scaling factor is therefore introduced in the dimensionful quantity 1/<p>. The slope is not actually ($b$+1) but rather has this scaling factor tied into it.

[75] This positioning minimizes the time when only one of the two genes has been replicated as part of the overall chromosome replication preceding cell division. This differing gene copy number can introduce another source of noise which is not truly "global" to the cell, as it affects separate genes differently.

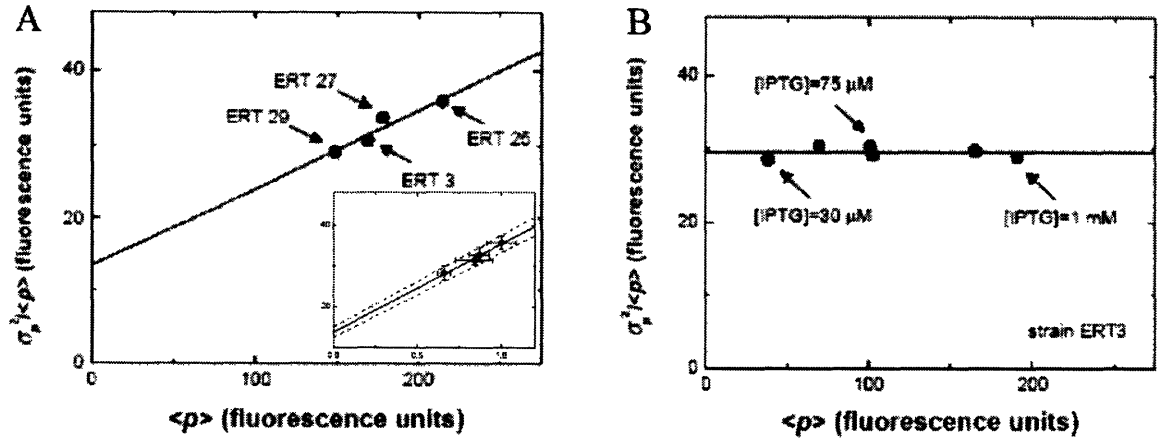


**Figure 7.2.** Steady state measurements of fluorescent reporter noise in *Bacillus subtilis* (Ozbudak *et al.*). (A) Noise (quantified as $\sigma_p^2/\langle p \rangle$) as a function of average fluorescence $\langle p \rangle$ for strains with identical promoters at the same induction level (and therefore the same $k_r$), but with different ribosome binding sites (and therefore different burst sizes and total amounts of protein). The noise is a linearly increasing function of the mean. When the noise is plotted as a function of the translational efficiency of the ribosome binding site (inset), the linear relation between the noise and $b$ is clearly shown. (B) Noise (again, $\sigma_p^2/\langle p \rangle$) as a function of average fluorescence for a single strain; in all measurements, the ribosome binding site (and therefore $b$) is the same. The level of inducer is changed, which changes the mean expression. The noise remains the same. These results indicate that an expression describing the noise in the output of a single gene should certainly include translational efficiency, but may be independent of transcription (as (74) predicts). Figures copyright E. M. Ozbudak *et al.*, 2002, and appear in (75).

---

of CFP (and YFP) values represents the noise introduced by the varying of the rate parameters from cell to cell (which defines the extrinsic noise). In the second case, the rate parameters would be fixed for all cells at identical values, but intrinsic fluctuations would affect CFP and YFP in non-correlated ways. In a plot of CFP versus YFP values, the cells would produce a circular distribution[76]; the width of the circle is the noise

[76] When measuring CFP and YFP, the fluorescence values detected do not scale equally with protein number in each case; therefore, some axis rescaling is needed to make the distribution look like a circle. The scaling can be determined by matching the background-subtracted mean signal levels for each gene.

introduced from intrinsic number fluctuations. When both of these noise sources are combined, the CFP/YFP distribution looks like an ellipse. Using this technique to quantify internal and external noise, the expected $1/\langle p \rangle$ behavior was seen for the internal noise. The external noise actually was not constant over the entire range of induction tested. The interpretation of this result was that at zero or maximal induction, the levels of *lac* repressor or IPTG were respectively buffered, and therefore fluctuations in the mRNA production rate were less than at intermediate concentrations, in agreement with the calculated extrinsic noise. This two color technique has become a standard experimental technique for the study of intrinsic versus extrinsic noise.

## 7.4. Significance of noise in the circadian clock of *Synechococcus.*

The circadian oscillator of *Synechococcus* is remarkably robust against noise. This is evident in many ways. Many experiments (for example, (18, 48)) show clear oscillations after at least 1 week in a constant environment. *Synechococcus* can divide up to three to four times per cycle. As a cell divides, the information about the circadian state needs to be accurately transmitted to the daughter cells, despite errors in partitioning proteins between the daughters. Cells that have not been synchronized for very long periods of time continue to display a 24 hour rhythm. "Movies" of cells growing into microcolonies which have been in constant light environments for over two months show that the cells continue to oscillate in phase with one another[77]. Estimates of the phase noise reported in (48) indicate correlation times of over four months. Examination of samples as part of this study which had not been synchronized for over three months are

[77] Although the number of cells examined was low, the individuals examined did have the same phase to within the error bars of the experiment.

consistent with these correlation times (Figure 7.3). All of this evidence leads to the conclusion that the clock itself is very robust against noise.

Another important factor to consider is how noise in the clock itself can add to the noise of other genetic systems in *Synechococcus*. A recent study (78) has explored how variability in a gene product that regulates other genes is transmitted to the output of the downstream gene products. In the cyanobacteria, the circadian clock appears to control the expression of *every* gene in the cell. Therefore, an understanding of the noise in the clock is needed to explain the variability measured in the output of other genes in the cell.

---

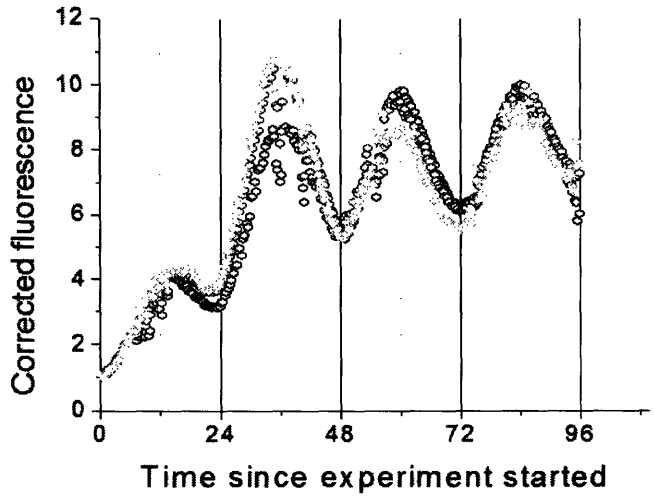


**Figure 7.3.** Bulk fluorescence measurements of *Synechococcus* after long continuous growth in constant conditions. Strain JRCS32 ( (I)$P_{kaiBC}$::*yfpLVA*) was continuously cultured under constant conditions in the LL box (shown in Figure 3.1) for over three months, and measured using the "plate reader" setup described in Chapter 4. Oscillations are still very clear.

# 8. Identifying sources of noise in *Synechococcus*.

## 8.1. Two-color microscopy.

In order to characterize internal and external noise in *Synechococcus*, I examined samples using both approaches described in Chapter 7 (two color microscopy and multiple reporter gene copies). First, strains were prepared with *cfpLVA* and *yfpLVA* driven by identical promoters[78]. Strains JRCS14, JRCS22, JRCS26, and JRCS30 used the *trc*, *kaiA*, *kaiBC*, and *psbAI* promoters respectively. Samples were synchronized to remove variability arising from differently phased cells. Slides were prepared with these cells, and examined on a Nikon Eclipe TE2000 inverted microscope with a MicroMax 512BFT cooled CCD camera (Princeton Instruments). Images of the CFP, YFP, and chlorophyll (using a Texas Red filter cube) fluorescences were recorded, as well as a phase contrast image. Cells were identified in the phase contrast images using a morphological analysis package in Metamorph (Universal Imaging). Regions defining each candidate cell were transferred to the various fluorescence images, and the maximum fluorescence pixel value was recorded. Cells exhibiting fluorescence within one standard deviation of the mean of a population of wild-type cells were excluded, as well as any clear outliers. The remaining cells were normalized to adjust for inhomogeneities resulting from uneven illumination from the lamp.

The intrinsic noise is calculated by the following formula (77)

$$CV_{\text{int}} = \sqrt{\frac{\langle (c-y)^2 \rangle}{2\langle c \rangle \langle y \rangle}}$$

[78] Unlike the experiments in (77), the relative positions of the sites for *cfp* and *yfp* integration (NSI and NSII respectively) are not known with respect to the origin of replication. I will neglect the contribution of gene copy number in my analysis.

where $c$ and $y$ represent CFP and YFP measurements respectively. Fluorescence measurements for CFP and YFP were normalized to their respective background-subtracted mean levels. The background correction to the noise is obtained by substituting $c$-$c_b$ and $y$-$y_b$ into the above formula and taking the appropriate population averages. Extrinsic noise is calculated by the formula

$$CV_{\text{ext}} = \sqrt{\frac{\langle cy \rangle - \langle c \rangle \langle y \rangle}{\langle c \rangle \langle y \rangle}}$$

and performing background corrections as for intrinsic noise.

These measurements were performed at two times during the circadian cycle (representing circadian times of 2 and 6 hours). The results of this analysis are listed in Table 8.1.

**Table 8.1.** Intrinsic and extrinsic noise for various promoters measured using two-color technique.

| Strain (promoter) | Circadian time (hours) | Intrinsic CV | Extrinsic CV |
|---|---|---|---|
| JRCS14 ($P_{trc}$) | 2 | 0.1406 | 0.4360 |
| | 6 | 0.1350 | 0.2981 |
| JRCS22 ($P_{kaiA}$) | 2 | 1.8423 | 2.0594 |
| | 6 | 0.7418 | 1.2066 |
| JRCS26 ($P_{kaiBC}$) | 2 | 0.2162 | 0.4826 |
| | 6 | 0.2707 | 0.5419 |
| JRCS30 ($P_{psbAI}$) | 2 | 0.5438 | 0.7316 |
| | 6 | 1.9130 | 1.3040 |

The results for JRCS26 can be compared to the flow cytometry data shown in Figure 7.1. The total noise measured by flow cytometry is similar to the intrinsic noise

calculated by the two-color microscopy method. This implies that for the *kaiBC* promoter, intrinsic noise apparently dominates the total noise. The considerable external noise measured may represent an artifact of the microscopy process itself[79]. In any case, as noise can only be added by "noisy" measurement techniques, the lowest total noise measured (here, from flow cytometry) should be regarded as the best measure of actual gene noise.

## 8.2. Variation of gene copy number.

Another technique for measuring the relative contribution of intrinsic noise to a system was recently developed by Becskei *et al.* (manuscript in preparation). The basic premise involves placing multiple genes coding for the same reporter protein within a single strain. If the number of copies of the gene is doubled, it is trivial to show that the mean protein level should also double. However, the behavior of the noise in the system does not automatically scale with the number of gene copies, but rather depends on the nature of the noise.

Extrinsic noise, originating in cell-to-cell variations in ribosome number, polymerase protein levels, and so on, affects all genes equally (in principle). If the primary source of noise in the population of cells originates from these fluctuations, the output of the two reporter genes will vary in a correlated way in each cell. Thus, the total variation will also double, and the dimensionless noise remains identical to the single reporter gene result. If, on the other hand, the primary source of noise is number

[79] For example, slow temporal fluctuations in lamp intensity lead to correlated variations in CFP and YFP signals, which is interpreted as extrinsic noise. The spatial variation of the lamp also causes "extrinsic noise", but the normalization applied to the data should resolve this effect.

fluctuations from stochastic production and decay of individual molecules, the output of each gene will fluctuate in an uncorrelated way. Thus, if the nominal contribution of each gene remains roughly equal, the noise should be reduced by a factor of $1/\sqrt{2}$. In practice, the total noise within the cell is composed of both intrinsic and extrinsic (i.e., uncorrelated and correlated) noise, and the total noise reduction will be less than this factor. Measuring the actual drop will give an estimate of the relative contribution of each type of noise.[80]

To collect this data, strain JRCS35 (containing (II) $P_{kaiBC}$::*yfpLVA*) was transformed with pJRC27, a neutral site I plasmid containing *yfpLVA* under the *kaiBC* promoter. The resulting strain, JRCS70, contained functional copies of *yfpLVA* at two different locations in the chromosome, under control of the same promoter. To measure the impact on the noise profiles of the population, tubes containing JRCS70 and JRCS32 (with the construct only in neutral site I) were synchronized on the circadian spinner as described in Chapter 6, and data was collected on each tube. Figure 8.1 shows the results of the measurements. Indeed, the signal detected from the two-copy strain doubles, while the noise is reduced. A linear fit of the square of the coefficient of variation of the two-copy strain to the same quantity for the one-copy strain (with a tube of each strain at similar circadian times providing the coordinates for each point) reveals a slope of slightly more than 0.5. As just mentioned, if the noise in the system comes almost entirely from intrinsic sources, a value of 0.5 would be expected; therefore, the noise in *Synechococcus* seems to be dominated by intrinsic sources, reaffirming the observations from comparing two-color microscopy with flow cytometry results.

[80] In *S. cerevisiae*, Becskei *et al.* found that the noise remained mostly unchanged upon introduction of additional gene copies, and therefore concluded that gene noise in this yeast is mainly extrinsic in origin.

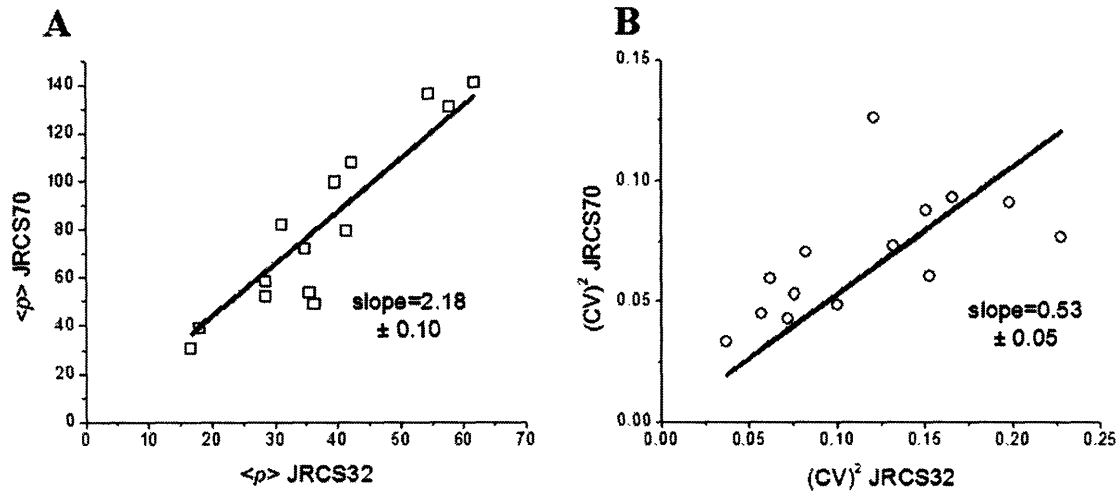


**Figure 8.1.** Results of doubling gene copy number. (A) Plot of mean fluorescence of tubes of JRCS70 (containing two copies of *yfpLVA*) against tubes of JRCS32 (containing only one copy of *yfpLVA*) phased one circadian hour away. A linear fit through the origin is shown with slope approximately 2. This is consistent with doubling the total amount of protein by doubling the gene number. (B) Plot of the square of the coefficient of variation of JRCS70 against the square of the CV for JRCS32, again comparing tubes phased one hour apart. This time, a linear fit through the origin reveals a slope slightly greater than ½. This implies that the noise in *Synechococcus* is dominated by intrinsic factors; if the noise were primarily extrinsic, the slope would be closer to unity.

---

## 8.3. Discussion.

The results presented above demonstrate that the majority of the noise in *Synechococcus* arises from uncorrelated fluctuations between individual genes within a cell. In principle, the circadian clock should affect the various genes in a cell in similar ways; for example, if the clock is running "fast", all genes should respond accordingly. Since the fluorescent proteins do not evince such correlated fluctuations, the implication is that the contribution of the clock to the total noise is very low. How can the clock of

*Synechococcus* get around the significant noises associated with highly stochastic protein and mRNA production and decay? The answer perhaps lies in the studies presented in Section 2.6.3 (42, 55), showing that the Kai proteins can produce an oscillation with a 24 hour period independent of the noisy processes of transcription and translation. The noise associated with the phosphorylation state of thousands of molecules of KaiC will be much less than the noise associated with infrequent production of mRNA from the *kai* genes. Simply put, *Synechococcus* may have circumvented the problem of noise in its regulatory mechanism by constructing a clock rooted in biochemical processes, rather than genetic ones. The results shown in this chapter show that using this approach for periodic regulation introduces minimal noise above the levels intrinsic in the gene being regulated, and thus may provide a direction for understanding of other natural oscillators and assist in the design of biologically-inspired synthetic oscillators.

# 9. Modeling noise loops (master equation).

## 9.1. Modeling approaches.

Understanding the origin of the noise loops requires a quantitative, analytic approach to modeling circadian-regulated genes. However, the use of simple differential equations to describe the creation and destruction of various relevant species, such as are used in deterministic chemistry systems involving molar quantities of particles, are not suitable to describe this system where noise is an integral factor. Two approaches to modeling systems with noise in very different ways are the Langevin and the master equation approaches. I will discuss both briefly before developing a circadian noise model[81].

### 9.1.1. Langevin method.

The Langevin technique involves explicitly adding a time-varying noise term to the deterministic chemical reaction equations. The noise is a randomly fluctuating variable, with an arbitrary power spectrum in frequency space[82]. The chemical equations are linearized about the steady state (e.g., noise-free) solution, defining $\delta x = x - x_{ss}$ with $x_{ss}$ the steady state value. The resulting differential equations for the fluctuations of the various species are Fourier transformed. After multiplication with the complex conjugate of this transform, they are averaged over the random variable (in Fourier space) to give a mean fluctuation $\left\langle \left| \delta\tilde{x}(\omega) \right|^2 \right\rangle$ as a function of frequency. The Wiener-Khintchine theorem

[81] For discussions and derivations with greater depth, an excellent reference is (79).
[82] Practically, a white noise spectrum is often used, where the average contribution of the noise term is zero, and the correlation time of the noise vanishes when averaged over all time.

(80) then allows the calculation of the root mean square fluctuation in species $x$ by taking the inverse Fourier transform of this function with $t = 0$.

The Langevin method provides a fairly straightforward means of calculating the expected width of the distribution of each species being considered, if the noise strength is known. The noise term may include both intrinsic (number fluctuation) and extrinsic (global parameter fluctuation) sources. In the case of internal noise, the parameter describing the mRNA noise can be calculated by applying the assumption that the production and decay of mRNA are Poisson processes. The parameter describing the noise in protein dynamics can also be calculated by the assumption that for a fixed population of mRNA the protein production is also poissonian. Thus, an expression for the variance of the protein population can be written as a function of the various reaction rate constants in the genetic system.

A major drawback of the Langevin approach for circadian modeling comes about in the final step of the calculation, where the average fluctuation size is calculated from the inverse transform of the root mean square fluctuations in Fourier space. The Wiener-Khintchine theorem allows the ready calculation of this fluctuation size by evaluating the transform at $t = 0$, which greatly simplifies the approach for the steady state problem. However, for systems with an intrinsic oscillation, there is no "steady state", and thus the full solution needs to be considered. This greatly complicates matters. In addition, as the circadian system does have a natural frequency, describing the noise in the system with a white noise profile (with no characteristic frequency) is a questionable method. Therefore, another approach for describing noise in a circadian-regulated gene was sought.

### 9.1.2. Master equation method.

The master equation approach takes a different view of the gene expression system[83]. The master equation itself is a description of the time rate of change of the probability of finding the system in (for the case of a single gene, its mRNA and protein product) a state $\{r, p\}$, with discrete numbers $r$ of mRNA and $p$ of protein[84]. The probability $P(r, p, t)$ (allowing explicitly for a time dependence) evolves according to its own value and the probabilities of the neighboring states with rates dependent on the reaction constants in the system, as previously shown in Figure 7.1.

The end goal of the master equation approach is to describe the probability distribution of finding the system in any of its possible states. This distribution is characterized by its various moments. To find these properties of the distribution, the master equation is transformed into a moment-generating space, in which the $N$ system variables $x_1 \ldots x_N$ (for a single gene, $N = 2$ with $x_1 = r$ and $x_2 = p$) are replaced by terms $z_1 \ldots z_N$. The equation describing the moment generating function for a system of $N$ variables is

$$F(z_1, z_2, \ldots, z_N, t) = \sum_{x_1, x_2, \ldots, x_N} z_1^{x_1} z_2^{x_2} \ldots z_N^{x_N} p(x_1, x_2, \ldots, x_N, t)$$

The sum is performed over all possible values for each variable; for system variables corresponding to chemical species, the occupancies are integers from zero to infinity[85].

Various derivatives of the moment generating function with respect to transformed variables yield properties of the distributions of the original quantities. In

---

[83] This is the basis of the study described in the Chapter 8 (74).

[84] This approach is easily extended to arbitrary numbers of genes, in which case the state of the system is described by $\{r_1, p_1, r_2, p_2, \ldots r_n, p_n\}$.

[85] All genetic systems studied have at the least first order decays of the various species; therefore, the system is naturally bounded and species are not driven to infinity.

particular, the following relations will be important in finding the moments of protein distributions

$$F\big|_1 = 1$$
$$\left.\frac{\partial F}{\partial z_i}\right|_1 = \langle x_i \rangle$$
$$\left.\frac{\partial^2 F}{\partial z_i^2}\right|_1 = \langle x_i (x_i - 1) \rangle$$
$$\left.\frac{\partial^2 F}{\partial z_i \partial z_j}\right|_1 = \langle x_i x_j \rangle$$

The notation $f\big|_1$ indicates that $f$ is to be evaluated with all $z_i = 1$.

The equation $\dot{P}(r,p,t) = f\left(\sum_{r',t'=0}^{\infty} P(r',p',t)\right)$ is transformed into an equation $\dot{F}(z_r,z_p,t) = g(z_r,z_p,t,F(z_r,z_p,t))$. Taking the derivatives as above sets up differential equations describing time derivatives of the various moments of the distributions of $r$ and $p$ with respect to the moments themselves (for example, $\langle \dot{p} \rangle = f(\langle r \rangle, \langle p \rangle)$). For a steady state system, the average rate of change of any quantity is zero. Consequently, the time varying quantities disappear, leaving a system of linear equations. These are solved to give the steady state distribution (and the single-gene result given in the previous chapter).

The advantage of the master equation approach is that the Wiener-Khintchine theorem used at the end of the Langevin approach to obtain a time-averaged steady state result is not applicable in the dynamic circadian system. The extension from the steady state solution to a dynamic solution is realized by simply not setting the right side of the differential equations equal to zero. The time-dependent moments of the predicted protein distribution can then be calculated, and the dynamic noise behavior of non-

steady-state systems can be studied. This work represents the first use of the master equation to describe noise in a nonequilibrium genetic system.

## 9.2. Modeling circadian production of YFP.

### 9.2.1. Basic scheme.

To model the production of fluorescent reporter protein under the control of the circadian clock, the genetic activity related to YFP is assumed to proceed according to the following model (Figure 9.1A). The gene is transcribed at a rate $k_r(\sin\omega t + A)$. Here, $\omega = 2\pi/T$, with $T$ the circadian period. The value $A$ is constrained to be $\geq 1$ (to maintain a nonnegative production rate), with inequality implying a "leakiness" of the promoter even under conditions of maximum circadian repression. The resulting mRNA is translated with a rate $k_p$, assuming no circadian phase dependence on ribosome activity. Both mRNA and protein degrade with first order rate constants $\gamma_r$ and $\gamma_p$ respectively; again, assuming no circadian impact on factors such as RNAase or protease levels or activity. I also include no feedback from the reporter protein into the circadian network.

### 9.2.2. Master equation; probability flow; moment generating function.

This system leads to the following master equation describing probability flow (with $r$ and $p$ representing mRNA and protein numbers as before)

$$\begin{aligned}\dot{P}(r,p,t) = {} & [P(r-1,p,t) - P(r,p,t)]k_r(\sin(\omega t) + A) \\ & + [P(r+1,p,t)(r+1) - P(r,p,t)r]\gamma_r \\ & + [P(r,p-1,t) - P(r,p,t)]rk_p \\ & + [P(r,p+1,t)(p+1) - P(r,p,t)p]\gamma_p\end{aligned}$$

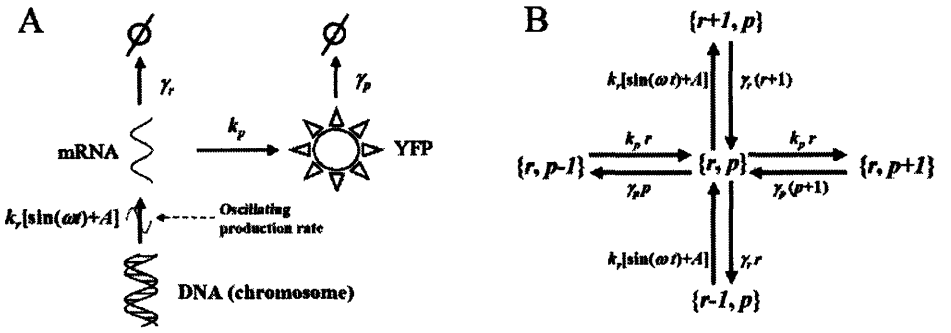


**Figure 9.1.** Scheme for modeling circadian production of YFP. (A) The rate of transcription of the reporter gene is assumed to follow a sinusoidal oscillation, with amplitude $k_r$ and basal expression $k_r(A-1)$. The transcribed mRNA is translated into YFP with rate constant $k_p$. First order decays of mRNA and protein occur with rate constants $\gamma_r$ and $\gamma_p$ respectively. (B) Probability flow between discrete states of mRNA and protein number $\{r, p\}$ is given in terms of these rate constants and the number of mRNA and protein in the various states.

---

This probability flow is represented in Figure 9.1B. The contributions of each reaction to the moment generating function will now be considered.

Certain motifs appear frequently in genetic systems. For example, if species 2 is created using species 1 as a template (while leaving species 1 unaltered), this would be described by the chemical reaction equation

$$\dot{x}_2 = kx_1$$

For discrete numbers of particles, this becomes

$$\dot{P}(x_1, x_2, t) = kx_1 P(x_1, x_2 - 1, t) - kx_1 P(x_1, x_2, t)$$

The moment generating differential equation for this process is

$$\dot{F}(z_1, z_2, t) = kz_1(z_2 - 1)\frac{\partial F}{\partial z_1}$$

This process describes the translation of protein from mRNAs, as well as the transcription of mRNA from the chromosome. In addition, the production of mRNA may be simplified (as the template can be assumed to have a fixed number of copies[86]); this reduces the moment generating function to

$$\dot{F}(z_2,t) = k(z_2 - 1)F$$

The other motif for consideration in this work is that of first order decay of a species $x$. In chemistry terms

$$\dot{x} = -\gamma x$$

For discrete particles

$$\dot{P}(x,t) = \gamma(x+1)P(x+1,t) - \gamma xP(x,t)$$

The resulting moment generating function is

$$\dot{F}(z,t) = -\gamma(z-1)\frac{\partial F}{\partial z}$$

This term will appear in the total moment generating function for both $r$ and $p$.

Using these motifs, it is evident that the moment generating function for the system is

$$\dot{F} = k_r(\sin(\omega t) + A)(z_r - 1)F - \gamma_r(z_r - 1)\frac{\partial F}{\partial z_r} + k_p z_r(z_p - 1)\frac{\partial F}{\partial z_r} - \gamma_p(z_p - 1)\frac{\partial F}{\partial z_p}$$

### 9.2.3. Calculation of distribution moments.

The derivatives of $\dot{F}$ are now taken to get the quantities of interest

[86] This number is nominally 1; but for any fixed number of copies, the parameter $k$ can absorb the constant.

$$\langle \dot{r} \rangle = \left. \frac{\partial \dot{F}}{\partial z_r} \right|_{z_{r,p}=1} = k_r(\sin \omega t + A) - \gamma_r \langle r \rangle$$

$$\langle \dot{r}^2 \rangle - \langle \dot{r} \rangle = \left. \frac{\partial^2 \dot{F}}{\partial z_r^2} \right|_{z_{r,p}=1} = 2k_r(\sin \omega t + A)\langle r \rangle - 2\gamma_r\left(\langle r^2 \rangle - \langle r \rangle\right)$$

$$\langle \dot{p} \rangle = \left. \frac{\partial \dot{F}}{\partial z_p} \right|_{z_{r,p}=1} = k_p \langle r \rangle - \gamma_p \langle p \rangle$$

$$\langle r\dot{p} \rangle = \left. \frac{\partial^2 \dot{F}}{\partial z_r \partial z_p} \right|_{z_{r,p}=1} = k_p\left(\langle r^2 \rangle + \langle r \rangle\right) + k_r(\sin \omega t + A)\langle p \rangle - (\gamma_r + \gamma_p)\langle rp \rangle$$

$$\langle \dot{p}^2 \rangle - \langle \dot{p} \rangle = \left. \frac{\partial^2 \dot{F}}{\partial z_p^2} \right|_{z_{r,p}=1} = 2k_p \langle rp \rangle - 2\gamma_p\left(\langle p^2 \rangle - \langle p \rangle\right)$$

These differential equations can be solved to yield time-dependent solutions in terms of $k_r$, $k_p$, $\gamma_r$, $\gamma_p$, and $A$.

The homogeneous solutions to these differential equations all take the form

$$\alpha \exp^{-\beta t}$$

and represent an approach to the limit cycle from initial conditions away from the natural oscillation. I ignore these terms, since the experiments are conducted in such a way as to avoid driving the system away from its natural cycle[87].

The solutions for the various quantities listed above are

$$\langle r \rangle = \frac{Ak_r}{\gamma_r} + \frac{\gamma_r k_r}{\gamma_r^2 + \omega^2}\sin(\omega t) - \frac{k_r \omega}{\gamma_r^2 + \omega^2}\cos(\omega t)$$

[87] Displacement from the "natural" oscillation could be achieved, for example, by leaving cells in the dark for long periods of time to slow growth and lower gene expression (42), or suddenly changing the mean expression level by use of an inducible promoter.

$$\langle r^2 \rangle = -\frac{\gamma_r k_r^2 \omega}{\left(\gamma_r^2+\omega^2\right)^2}\sin(2\omega t) - \frac{k_r^2\left(\gamma_r^2-\omega^2\right)}{2\left(\gamma_r^2+\omega^2\right)^2}\cos(2\omega t) + \frac{k_r\left(\gamma_r+2Ak_r\right)}{\gamma_r^2+\omega^2}\sin(\omega t)$$
$$-\frac{k_r\left(\gamma_r+2Ak_r\right)\omega}{\gamma_r\left(\gamma_r^2+\omega^2\right)^2}\cos(\omega t) + \frac{Ak_r\left(\gamma_r+Ak_r\right)}{\gamma_r^2} + \frac{k_r^2}{2\left(\gamma_r^2+\omega^2\right)}$$

$$\langle p \rangle = \frac{Ak_r k_p}{\gamma_r\gamma_p} + \frac{k_r k_p\left(\gamma_r\gamma_p-\omega^2\right)}{\left(\gamma_r{}^2+\omega^2\right)\left(\gamma_p{}^2+\omega^2\right)}\sin(\omega t) - \frac{\left(\gamma_r+\gamma_p\right)k_r k_p\omega}{\left(\gamma_r{}^2+\omega^2\right)\left(\gamma_p{}^2+\omega^2\right)}\cos(\omega t)$$

$$\langle rp \rangle = \frac{k_r k_p\left\{\gamma_r{}^2\gamma_p{}^2 k_r\left(\gamma_r+\gamma_p\right)+2A\gamma_r\gamma_p\left(\gamma_r{}^2+\omega^2\right)\left(\gamma_p{}^2+\omega^2\right)+2A^2k_r\left(\gamma_r+\gamma_p\right)\left(\gamma_r{}^2+\omega^2\right)\left(\gamma_p{}^2+\omega^2\right)\right\}}{2\gamma_r{}^2\gamma_p\left(\gamma_r+\gamma_p\right)\left(\gamma_r{}^2+\omega^2\right)\left(\gamma_p{}^2+\omega^2\right)}$$
$$+\frac{k_r k_p\left\{\begin{array}{l}\gamma_r\gamma_p{}^2\left(\gamma_r+\gamma_p\right)\left[\gamma_r\gamma_p+2Ak_r\left(\gamma_r+\gamma_p\right)\right]\\+\left[-Ak_r\gamma_p{}^3+\gamma_r\gamma_p{}^2\left(Ak_r-\gamma_p\right)+\gamma_r{}^2\left(\gamma_r+\gamma_p\right)\left(\gamma_p+Ak_r\right)\right]\omega^2\\-\left[\gamma_r\gamma_p+Ak_r\left(\gamma_p-\gamma_r\right)\right]\omega^4\end{array}\right\}}{\gamma_r\gamma_p\left(\gamma_r{}^2+\omega^2\right)\left(\gamma_p{}^2+\omega^2\right)\left[\left(\gamma_r+\gamma_p\right)^2+\omega^2\right]}\sin(\omega t)$$
$$+\frac{k_r k_p\omega\left\{\begin{array}{l}\gamma_r\gamma_p{}^3\left(2\gamma_r+\gamma_p\right)+Ak_r\gamma_p\left(\gamma_r+\gamma_p\right)^2\left(\gamma_r+2\gamma_p\right)\\+\left[\gamma_r\gamma_p\left(2\gamma_r+\gamma_p\right)+Ak_r\left(\gamma_r{}^2+3\gamma_r\gamma_p+3\gamma_p{}^2\right)\right]\omega^2+Ak_r\omega^4\end{array}\right\}}{\gamma_r\gamma_p\left(\gamma_r{}^2+\omega^2\right)\left(\gamma_p{}^2+\omega^2\right)\left[\left(\gamma_r+\gamma_p\right)^2+\omega^2\right]}\cos(\omega t)$$
$$-\frac{k_r^2 k_p\omega\left[\gamma_r\left(\gamma_r+2\gamma_p\right)+\omega^2\right]}{2\left(\gamma_r{}^2+\omega^2\right)^2\left(\gamma_p{}^2+\omega^2\right)}\sin(2\omega t) - \frac{k_r^2 k_p\left[\gamma_r{}^2\gamma_p-\left(2\gamma_r+\gamma_p\right)\omega^2\right]}{2\left(\gamma_r{}^2+\omega^2\right)^2\left(\gamma_p{}^2+\omega^2\right)}\cos(2\omega t)$$

$$\langle p^2\rangle = \frac{k_r k_p\left[\begin{array}{l} k_r k_p \gamma_r^2 \gamma_p{}^2(\gamma_r+\gamma_p)+2A^2 k_r k_p(\gamma_r+\gamma_p)(\gamma_r^2+\omega^2)(\gamma_p{}^2+\omega^2) \\ +2A\gamma_r\gamma_p(\gamma_r+\gamma_p+k_p)(\gamma_r^2+\omega^2)(\gamma_p{}^2+\omega^2) \end{array}\right]}{2\gamma_r^2\gamma_p{}^2(\gamma_r+\gamma_p)(\gamma_r^2+\omega^2)(\gamma_p{}^2+\omega^2)}$$

$$+\frac{k_r k_p\left\{\begin{array}{l} 4\gamma_r\gamma_p{}^3(\gamma_r+\gamma_p)[\gamma_r\gamma_p(\gamma_r+\gamma_p)+k_p(\gamma_r\gamma_p+2Ak_r(\gamma_r+\gamma_p))] \\ -\gamma_p\left[\begin{array}{l} -\gamma_r^4\gamma_p+4\gamma_r\gamma_p{}^4+8Ak_rk_p\gamma_p{}^3+6k_p\gamma_r\gamma_p{}^2(\gamma_p+Ak_r) \\ +2\gamma_r^3(\gamma_p{}^2-2\gamma_pk_p-Ak_rk_p)+\gamma_r^2(3\gamma_p{}^3+4Ak_rk_p\gamma_p) \end{array}\right]\omega^2 \\ -\left[\begin{array}{l} \gamma_r\gamma_p(\gamma_r^2+\gamma_r\gamma_p+5\gamma_p{}^2)+2k_p\left(\begin{array}{l} \gamma_r\gamma_p(2\gamma_r+3\gamma_p) \\ +Ak_r(\gamma_r^2+\gamma_r\gamma_p+5\gamma_p{}^2) \end{array}\right) \end{array}\right]\omega^4-(\gamma_r\gamma_p+2Ak_rk_p)\omega^6 \end{array}\right\}}{\gamma_r\gamma_p(\gamma_r^2+\omega^2)(\gamma_p{}^2+\omega^2)(4\gamma_p{}^2+\omega^2)[(\gamma_r+\gamma_p)^2+\omega^2]}\sin(\omega t)$$

$$-\frac{k_r k_p\omega\left\{\begin{array}{l} 2\gamma_p{}^2[2\gamma_r\gamma_p(\gamma_r+\gamma_p)^3+k_p(\gamma_r\gamma_p(\gamma_r^2+5\gamma_r\gamma_p+2\gamma_p{}^2)+4Ak_r(\gamma_r+\gamma_p)^3)] \\ +\left[\begin{array}{l} \gamma_r\gamma_p(\gamma_r+\gamma_p)(\gamma_r^2+2\gamma_r\gamma_p+5\gamma_p{}^2)+2k_p\left(\begin{array}{l} \gamma_r\gamma_p(\gamma_r^2+5\gamma_r\gamma_p+\gamma_p{}^2) \\ +Ak_r(\gamma_r+\gamma_p)(\gamma_r^2+2\gamma_r\gamma_p+5\gamma_p{}^2) \end{array}\right) \end{array}\right]\omega^2 \\ +[\gamma_r\gamma_p(\gamma_r+\gamma_p)+2k_p(-\gamma_r\gamma_p+Ak_r(\gamma_r+\gamma_p))]\omega^4 \end{array}\right\}}{\gamma_r\gamma_p(\gamma_r^2+\omega^2)(\gamma_p{}^2+\omega^2)(4\gamma_p{}^2+\omega^2)[(\gamma_r+\gamma_p)^2+\omega^2]}\cos(\omega t)$$

$$-\frac{k_r^2k_p^2\omega(\gamma_r+\gamma_p)(\gamma_r\gamma_p-\omega^2)}{(\gamma_r^2+\omega^2)^2(\gamma_p{}^2+\omega^2)^2}\sin(2\omega t)+\frac{k_r^2k_p^2[4\gamma_r\gamma_p\omega^2-(\gamma_r^2-\omega^2)(\gamma_p{}^2-\omega^2)]}{2(\gamma_r^2+\omega^2)^2(\gamma_p{}^2+\omega^2)^2}\cos(2\omega t)$$

The variances are given by the formula $\sigma_x^2 = \langle x^2\rangle - \langle x\rangle^2$, and are calculated from the above quantities.

$$\sigma_r^2 = \frac{\gamma_r k_r}{\gamma_r^2+\omega^2}\sin(\omega t)-\frac{k_r\omega}{\gamma_r^2+\omega^2}\cos(\omega t)+\frac{Ak_r}{\gamma_r}$$

$$\sigma_p^2 = \frac{k_r k_p \left\{ \begin{array}{l} 4\gamma_r\gamma_p^3(\gamma_r+\gamma_p)(\gamma_r+\gamma_p+k_p) - \gamma_p\left[-\gamma_r^3+3\gamma_r\gamma_p^2+2\gamma_r^2(\gamma_p-2k_p)+2\gamma_p^2(2\gamma_p+3k_p)\right]\omega^2 \\ -\left[\gamma_r^2+\gamma_r(\gamma_p+4k_p)+\gamma_p(5\gamma_p+6k_p)\right]\omega^4 - \omega^6 \end{array} \right\}}{(\gamma_r^2+\omega^2)(\gamma_p^2+\omega^2)(4\gamma_p^2+\omega^2)\left[(\gamma_r+\gamma_p)^2+\omega^2\right]}\sin(\omega t)$$
$$+\frac{k_r k_p \omega \left\{ \begin{array}{l} 2\gamma_p^2\left[-2(\gamma_r+\gamma_p)^3 - k_p(\gamma_r^2+5\gamma_r\gamma_p+2\gamma_p^2)\right] \\ -\left[\gamma_r^3+3\gamma_r^2\gamma_p+7\gamma_r\gamma_p^2+5\gamma_p^3+2k_p(\gamma_r^2+5\gamma_r\gamma_p+\gamma_p^2)\right]\omega^2 + (\gamma_r+\gamma_p-2k_p)\omega^4 \end{array} \right\}}{(\gamma_r^2+\omega^2)(\gamma_p^2+\omega^2)(4\gamma_p^2+\omega^2)\left[(\gamma_r+\gamma_p)^2+\omega^2\right]}\cos(\omega t)$$
$$+\frac{Ak_r k_p(\gamma_r+\gamma_p+k_p)}{\gamma_r\gamma_p(\gamma_r+\gamma_p)}$$

The coefficients of variation are then simply given by

$$CV_x = \frac{\sqrt{\sigma_x^2}}{\langle x \rangle}$$

## 9.3. Observations.

Based on these results, several observations can be made. All quantities (means and variances) oscillate with the same frequency. The phase of these oscillating moments may be readily checked by taking the inverse tangent of the ratio of the sine and cosine terms. For both $\langle r \rangle$ and $\sigma_r^2$, this ratio is $-\gamma_r/\omega$, corresponding to the delay associated with a species with decay constant $\gamma_r$ as calculated in Chapter 4. Since the mean and variance of the mRNA oscillate with the same frequency and phase, this means that mRNA should never produce noise loops[88].

This observation contradicts an early idea for the source of the noise loops. Consider that the noise fundamentally arises from stochastic events that occur

[88] Practically, this prediction is extremely hard (if not impossible) to test, as no techniques exist to test each cell quantitatively for the levels of particular mRNAs.

infrequently. The two types of events associated with mRNA levels are the transcription from the gene (assumed to oscillate with a frequency $\omega$ and zero arbitrary phase), and destruction of mRNA (which occurs at a maximal rate when the mean level of mRNA is the highest). If the mRNA has a long lifetime, these two processes will be out of phase with one another (by the $-\gamma_r / \omega$ factor); the time of highest noise comes when the frequency of events is lowest[89], and thus the noise should be out of phase with the mean (see Figure 9.2). This view of the noise loops is deficient in that it only considers the noise being introduced in a given time interval by the events during that interval. In reality, the population "inherits" noise from the history of production and decay. The

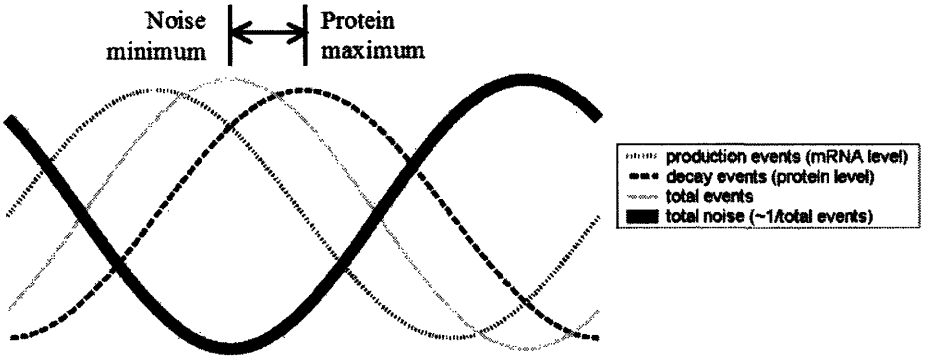


**Figure 9.2.** Possible explanation of noise loops. The total number of production events of YFP should be proportional to the amount of mRNA present, shown as a dotted line; the total number of decay events of YFP is proportional to the amount of YFP, the dashed line. The total number of events (light gray line) is the sum of these two curves (weighing each curve equally). Since the noise is highest when the frequency of events is lowest, the heavy black line represents the total noise. Indeed, a phase shift between noise minimum and protein maximum is clear.

[89] When I say "noise", I again refer to the coefficient of variation, a dimensionless quantity. Normalization by $(\langle p \rangle)^2$ makes the noise peak when $\langle p \rangle$ is lowest if $\sigma_p^2$ and $\langle p \rangle$ are equal.

master equation approach includes this history and thus reveals the flaw in this explanation of noise loop generation.

The ratio of the sine to cosine terms for the protein distribution moments are calculated next. For the mean protein, the result $-\frac{\gamma_r\gamma_p-\omega^2}{(\gamma_r+\gamma_p)\omega}$ is obtained. The expression for the phase shift for $\sigma_p^2$ is significantly more complicated and the result, while clearly different than the simple phase shift of $<p>$, does not immediately provide much insight into the origins of the loop. A simplification can deliver a striking clarity.

In most biological systems, especially where the protein utilized is a stable fluorescent reporter, mRNA half-lives are tremendously shorter than the protein half-lives[90], meaning $\gamma_r >> \gamma_p$. The parameter $\omega$, describing an oscillation with a period of 24 hours, is also extremely small compared to $\gamma_r$. Applying this approximation, the phase shift for $<p>$ reduces to $-\frac{\gamma_p}{\omega}$, which is the expected result if the RNA introduces no phase shift (which agrees with the assumption of $\gamma_r >> \omega$).

The approximation $\gamma_r >> (\gamma_p, \omega)$ is next applied to the phase shift for $\sigma_p^2$, giving the following result

$$-\frac{\gamma_p}{\omega}\left[\frac{4k_p\left(\gamma_p^2+\omega^2\right)+\gamma_r\left(4\gamma_p^2+\omega^2\right)}{2k_p\left(\gamma_p^2+\omega^2\right)+\gamma_r\left(4\gamma_p^2+\omega^2\right)}\right]$$

As discussed in Chapter 7, the parameter $k_p / \gamma_r$ can be considered as a "burst size" corresponding to the typical number of proteins produced by an individual mRNA during

[90] mRNA half-lives are typically on the order of five to ten minutes (81). In contrast, the YFPLVA protein half-life was measured at approximately six hours in Chapter 4.

its lifetime. Labeling this quantity $b$ and inserting into the above equation gives a simple expression for the phase of the standard deviation

$$-\frac{\gamma_p}{\omega}\left[\frac{4b\left(\gamma_p^2+\omega^2\right)+4\gamma_p^2+\omega^2}{2b\left(\gamma_p^2+\omega^2\right)+4\gamma_p^2+\omega^2}\right]$$

From the comparison of this result with the phase of $<p>$, three factors emerge which can cause the loops to vanish:

1) The ratio of the protein decay rate to the circadian frequency goes to zero. This corresponds to a near complete loss of peak-to-trough ratio (as discussed in Chapter 4), as the circadian oscillations are so fast that the reporter cannot respond quickly enough to track them.

2) The ratio of the protein decay rate to the circadian frequency goes to infinity. This implies that the system responds so quickly on the time scale of the oscillation that it may be considered to be in equilibrium with a slowly varying expression rate, and the predictions of steady-state theories (where the noise is a single-valued function of the mean) apply. (These first two factors are trivial to the understanding of the system.)

3) The burst size $b$ goes to zero: according to theoretical (74, 76) and experimental (75) studies, the noise associated with producing proteins by a two-step process (transcription and translation) causes noise to be increased above poissonian (or baseline) noise by a factor dependent on the burst size. Without this amplification, the loops vanish.

In light of this observation, a better interpretation of the source of the noise loops emerges. The dynamic master equation results predict the noise to be higher as the mean

protein level increases, and lower as it decreases. During the mean increase, production outstrips destruction, but examination of this stage in light of the protein-burst idea reveals that relatively few events are responsible (each mRNA produced is magnified by a factor of $b$). For the decreasing stage, each protein must individually be destroyed; this means many more events on the lower part of the loop, and thus these events become more regular in time, and the decay appears nearly deterministic. The Monte Carlo simulations of individual cell behavior presented in Chapter 10 will further illustrate this explanation of the noise loops.

## 9.4. Determination of parameters.

The model contains a total of five parameters: the decay times of both mRNA and the reporter protein ($\gamma_r$ and $\gamma_p$), the mRNA and protein production rate constants ($k_r$ and $k_p$), and the promoter "leakiness" factor $A$. Some of these can be directly measured, while others must be inferred. Connections with other quantities can be made, allowing the calculation of some constants from other quantities in the system. From the master equation solution for $<p>$, the mean number of proteins $N$ is given by

$$N = \frac{k_r k_p A}{\gamma_r \gamma_p}$$

Also, recall that the burst number was defined as

$$b = \frac{k_p}{\gamma_r}$$

Therefore, the set of parameters $\{k_r, k_p, \gamma_r, \gamma_p, A\}$ can be mapped to the set $\{N, b, \gamma_r, \gamma_p, A\}$.

### 9.4.1. Decay rates.

The parameter $\gamma_p$ has already been measured in Chapter 4, and is given by

$$\gamma_p = \frac{\ln(2)}{\tau_{p,1/2}}$$

with $\tau_{p,1/2}$ the effective half life of the protein (measured in this study at 5.6 hours). The parameter $\gamma_r$ can be estimated from data included in (42). In this study, all mRNA production was halted using different techniques. Northern blots reveal that the total mRNA level drops to less than 0.1% of its initial value when transcription is halted for four hours. The 0.1% value puts an upper bound on the half life of 24 minutes. I will use a 15 minute half life for the mRNA[91].

### 9.4.2. Basal activity of promoter ("leakiness").

The basal expression factor $A$ can be inferred in a couple different ways. First, bioluminescence measurements (such as those shown in Figure 2.1, but for expression driven by the *kaiBC* promoter (23)) indicate that the total signal does indeed drop to very nearly zero at the minimum, suggesting practically no leaking of the *kaiBC* promoter when maximally repressed by the circadian system. Second, as described in Chapter 4, if the mRNA levels have an infinite peak-to-trough ratio (i.e., zero leaking of the promoter), the measured protein half life implies a resulting *PTR* of 2.5 to 3.1 for the fluorescence signal; an actual value of approximately 2.7 is seen experimentally. As the only parameters involved in this calculation are the period of the oscillation and decay constant, the *PTR* of the mRNA may be stated to be extremely large with good certainty.

[91] This is consistent with measurements of mRNA lifetimes in other prokaryotes (for example, (81) measured a global mRNA half life of 6.8 minutes in *E. coli*).

Therefore, the value for $A$ will be nominally set to one, corresponding to zero basal expression[92].

### 9.4.3. Total number of proteins.

The total number of proteins $N$ may be calculated using a novel approach set forth in (53). As cells divide, the proteins may be assumed to partition randomly between the two daughter cells[93]. The resulting fluorescence pairs should obey a binomial distribution. Consequently, an ensemble of pairs of cells with protein number $n_1$ and $n_2$ adding up to total protein number $N$ can be shown to have the following property

$$\sqrt{\left\langle \left( \frac{n_1 - n_2}{2} \right)^2 \right\rangle} = \frac{\sqrt{N}}{2}$$

The fluorescence $f$ is linearly related to the number of proteins $n$ by a constant factor $y$; this means the equation above may be rewritten (with squaring)

$$\frac{\left\langle (f_1 - f_2)^2 \right\rangle}{4y^2} = \frac{N}{4}$$

Using $y = f_1/n_1 = f_2/n_2$, we obtain

$$\frac{\left\langle (f_1 - f_2)^2 \right\rangle}{\frac{\langle f_1 f_2 \rangle}{\langle n_1 n_2 \rangle}} = N$$

[92] The master equation results have no discontinuity as the basal expression goes to zero; therefore, we need not be concerned that some small but finite value need be used instead of zero.

[93] Some critical proteins may be divided using active mechanisms to ensure that nearly equal amounts are partitioned between the daughters, just as the chromosome copies are equally split. However, the non-native fluorescent proteins used most certainly are not part of such a system.

For a binomial distribution with equal probability of partitioning to either daughter,

$$\langle n_1 n_2 \rangle = \frac{N^2}{4}$$

and we obtain the final result

$$\frac{\langle (\Delta f)^2 \rangle}{\langle f_1 f_2 \rangle} = \frac{4}{N}$$

To perform this analysis, seven samples of *Synechococcus* strain JRCS32 (expressing *yfpLVA* under the *kaiBC* promoter) representing times throughout the circadian cycle were examined using microscopy. Cells with clearly complete septa were identified in phase contrast images. Regions were automatically created around these cells, and manually divided along the septa. The maximum fluorescence value of each daughter cell was recorded[94]. These data were corrected for inhomogeneities in the fluorescence illumination profile, and background fluorescence equal to the autofluorescence of wild type *Synechococcus* was subtracted. Daughter pairs were discarded if either had a maximum fluorescence less than background, if the maximum fluorescence was greater than five standard deviations above the population mean, or if the relative difference between the daughters was more than fifty percent[95]. For each pair of cells with maximum fluorescences $f_1$ and $f_2$ respectively, the quantities $(f_1-f_2)^2$ and $f_1 f_2$ were calculated. Cells were grouped by average maximum fluorescence into bins containing approximately 100 pairs of daughters, and for each bin, the mean number of

[94] Again, the maximum value was used to avoid issues with incorrectly identifying the boundaries of the cells.

[95] To obtain the signal levels seen, hundreds of YFP proteins were expected. Daughters with 50% relative difference should only occur less than 0.01% of the time. More likely, there was a "hot pixel" or other camera readout error in cells exhibiting such a difference.

proteins was calculated by

$$N = \frac{4}{\left[\frac{\langle (f_1 - f_2)^2 \rangle}{\langle f_1 f_2 \rangle}\right]}$$

The results are plotted in Figure 9.3 versus the mean maximum fluorescence of cells in the respective bins[96]. For the purposes of the model, I will use a value of 200 for the mean number of proteins during the cycle.

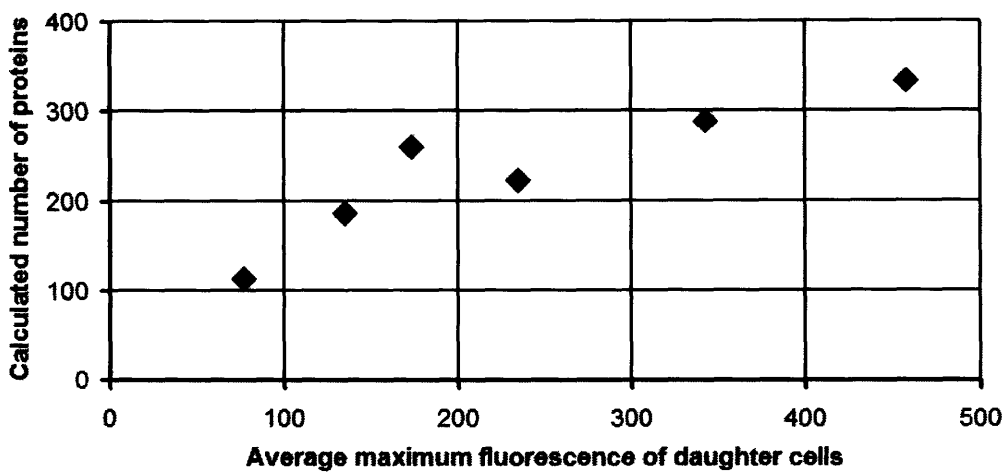


**Figure 9.3.** Calculation of average protein number. Pairs of dividing cells were identified from microscopy measurements on samples throughout the circadian cycle. The maximum fluorescence pixel value of each daughter was found, and the difference and product of these pairs were calculated. Cells were binned in groups of approximately 100 by the average maximum fluorescence of the two daughters. The mean number of proteins in each bin was calculated using the method described in the chapter. A nominal value of 200 proteins per cell was selected for the calculations in both this chapter and the next.

[96] Ideally, these data points should lie on a line intersecting the origin. However, external sources of noise can cause pairs of cells with high fluorescence to appear to have larger deviations than would be expected by the binomial distribution, and thus underestimate the amount of protein in the bright pairs.

### 9.4.4. Burst size.

The only remaining parameter is $b$, the burst size. Typical bursts in other prokaryotic organisms are on the order of ten proteins per mRNA transcript (71). Various values for $b$ less than twenty were tested in the model. Values for $b \leq 10$ produce loops that do not exceed the experimental loops to within the error in the measurements. A nominal value of 10 is assigned to $b$, and it should be noted in the following discussion that this is the only parameter not experimentally deduced.

The complete list of nominal parameters is summarized in Table 9.1.

**Table 9.1.** Nominal parameters for model calculations.

| Parameter | Nominal value |
|---|---|
| $N$ (average total protein number) | 200 (no units) |
| $b$ (burst size) | 10 (no units) |
| $\gamma_r$ (mRNA decay rate) | [ln(2)/(15 min)] = 2.773 $hr^{-1}$ |
| $\gamma_p$ (protein decay rate) | [ln(2)/(5.6 hours)] = 0.124 $hr^{-1}$ |
| $A$ (basal production) | 1 (no units; implies zero "leakiness") |

## 9.5. Graphical results with nominal parameters.

The model results are plotted with the nominal parameters in Figure 9.4. The time axis is adjusted by six hours so the genetic activity agrees with the observations in (18) (the promoter activity now appears as $-\cos(\omega t)$ ). The mRNA closely tracks the promoter activity, as expected. The coefficient of variation for the mRNA is large (the average number of mRNAs during the cycle is approximately one, and therefore large

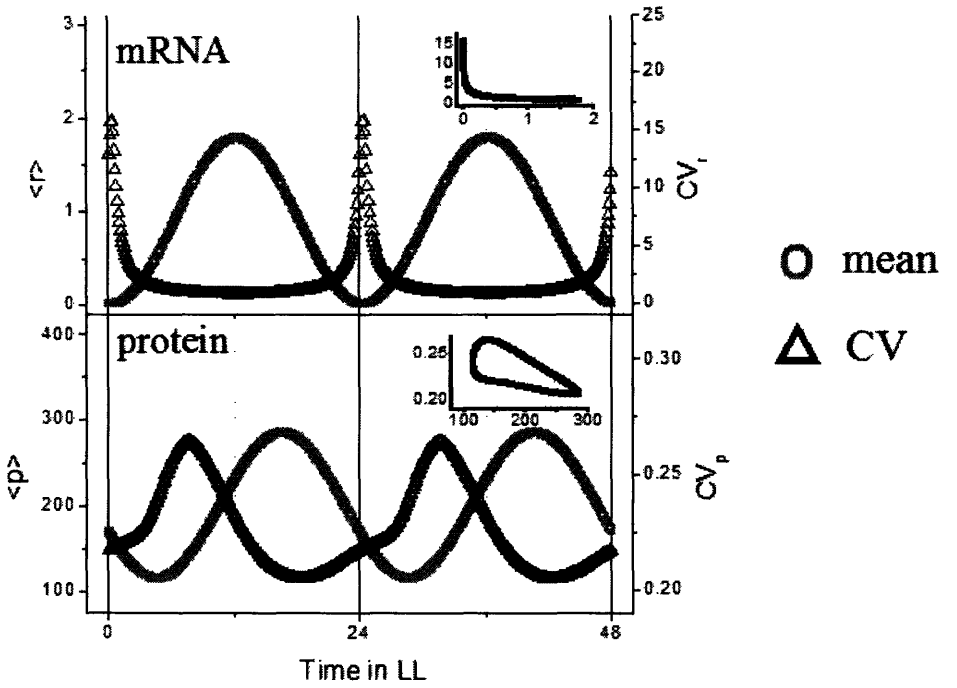


**Figure 9.4.** Master equation model results. Solutions are plotted with nominal parameters as described in the chapter. For mRNA and protein, both mean and coefficient of variation (CV) are plotted. The time axis for all quantities is adjusted to make the mRNA traces match the circadian time as reported in (18). The phase delay is clear for the protein, and agrees with experimental results (such as in Figure 4.4). The *PTR* of the protein also agrees with experimental measurements (such as in Figure 6.1). The insets show a plot of CV versus mean level for each species. The mRNA does not make loops, while the protein clearly does.

---

normalized noises are expected), and is exactly out of phase with the mRNA; therefore, no noise loop is seen. The protein mean oscillates with the delay and peak-to-trough ratio matching the experimental results, with the mean level of 200 proteins. The noise peaks after the minimum mean protein value, in agreement with experimental results (such as those shown in Figure 6.1); plotting the CV versus the mean clearly shows the loop.

Figure 9.5 shows the effect of varying the burst size from 0.1 to 10 (while maintaining the other parameters). The overall noise goes up with larger $b$ (as increasing the burst size increases the amplification of noise in the production stage), and the difference between the top and bottom of the loops becomes greater. This plot helps illuminate the importance of the burst size to the structure of the loops.

---

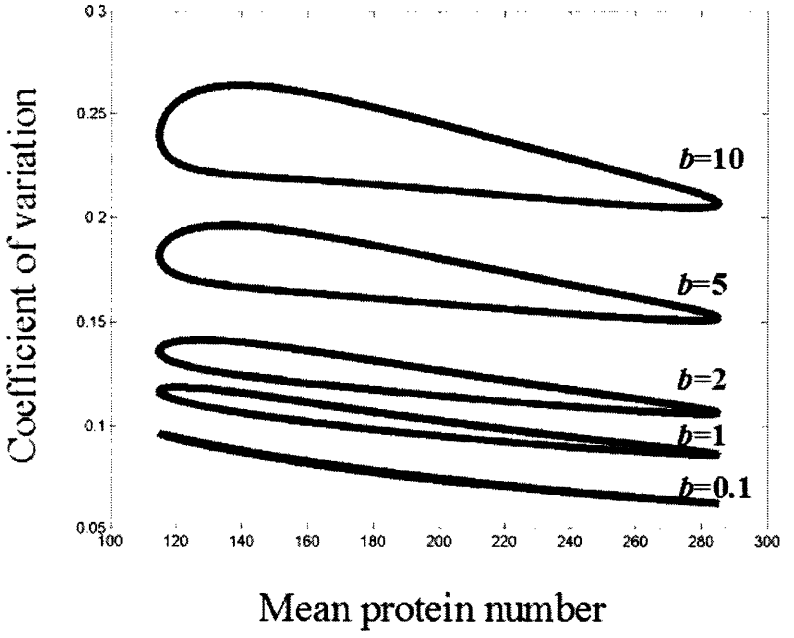


**Figure 9.5.** Effect of varying burst size. While keeping the other parameters at their nominal values, the burst size was varied from 0.1 to 10. As the burst size increases, the overall noise level in the system goes up, and the opening of the noise loops becomes much more pronounced.

## 9.6. Noise contributions from circadian clock.

### 9.6.1. Estimation of clock noise.

It should be noted that the master equation results predict noise loops resulting from purely intrinsic fluctuations in the cell. However, cells are also subject to extrinsic noise, and the clock mechanism itself may introduce noise with a 24-hour periodicity. These additional noise sources may be estimated from the intrinsic noise result from the master equation and the experimentally measured total noise.

Extrinsic noise that is time-independent would serve to translate the loop upward without changing its shape[97]. Therefore, a comparison of the baseline of the experimental and master equation loops gives an estimate of the extrinsic noise in the system

$$\eta_{\text{extrinsic}} = \sqrt{\eta^2_{\text{baseline,expt.}} - \eta^2_{\text{baseline, m.eq.}}}$$

On the other hand, noise with a 24-hour period will cause the loop to open further than predicted by the intrinsic noise alone. This may be calculated by taking the maximum value of the experimental noise, and subtracting the maximum noise in the intrinsic loop (and the extrinsic noise just calculated)

$$\eta_{\text{clock}} = \sqrt{\eta^2_{\text{max,expt}} - \eta^2_{\text{max, m.eq.}} - \eta^2_{\text{extrinsic}}}$$

Using the average of the loops in Figure 6.2 for the experimental noise, and the loop in Figure 9.4 for the intrinsic noise, gives the estimates for the other noises given in Table 9.2.

[97] As the noises add in quadrature, the best "space" to visualize this shifting is one employing the square of the coefficient of variation; otherwise, the loop will change its shape slightly upon addition of non-intrinsic noise. I will continue to use the CV as my definition of noise for the rest of this chapter.

**Table 9.2.** Estimates of various noises in system.

| Noise source | Value (coefficient of variation) |
|---|---|
| Total (baseline) | 0.22 |
| Total (maximum) | 0.35 |
| Intrinsic (baseline) | 0.21 |
| Intrinsic (maximum) | 0.265 |
| Extrinsic | 0.066 |
| "Clock" | 0.219 |

First, these results echo what has been measured in Chapter 8; namely, that intrinsic noise dominates over time-independent extrinsic noise in this system. In addition, for the first time we have an estimate of the noise contributed by the clock itself. This noise should in principle affect every gene under circadian regulation (which is to say, all the genes in *Synechococcus*). This noise is estimated to be slightly less than the contribution of the intrinsic fluctuations. However, the results in Chapter 8 suggest that nearly all of the noise in the cell is uncorrelated between the various genes of the cell. Noise in the clock mechanism should affect the genes in correlated ways. Therefore, the meaning of the "clock" noise calculated above must be considered.

### 9.6.2. Sources of clock noise.

What is the origin of this clock noise? There are several possibilities. First, the modeling performed in this chapter assumed no circadian variation in the rates of translation and protein or mRNA decay. If, however, the ribosome number or protease/RNase levels oscillate with a 24 hour period, this variation may evince itself as unexplained "clock noise" in this calculation. Another possibility is that the ribosome number may remain fixed, while the global amount of transcription of mRNA varies; this

would create competition for the available ribosomes and lead to an effective variation in the translation rate.

Evidence presented in Chapter 8 in fact suggests strongly that the "clock" noise just calculated is largely composed of effects that originate in intrinsic fluctuations. Presumably, the clock noise is a ubiquitous phenomenon within the cell, meaning that it should affect all genes in a particular cell in a correlated way. The experiments shown in Figure 8.1 demonstrate that the vast majority of the variability for each reporter gene is uncorrelated with the other copy. This is characteristic of intrinsic noise. Future studies may investigate the contributions of effects such as rhythmic variations in polymerases, (free) ribosomes, proteases, and RNases. The results of these studies can be incorporated into the master equation modeling to adjust the level of intrinsic noise, and further refine the estimation of the noise originating from the clock regulation itself.

A third possibility exists for the origin of this 24 hour periodic noise, and the one that most truly reflects the idea of "clock" noise. The regulation of the promoters in the cell may not oscillate in the smooth manner described in the modeling, but rather has some error that fluctuates over time. The amount of deviation from the ideal regulation, as well as the frequency spectrum of these fluctuations, are of central importance in understanding the output of the clock in *Synechococcus*. In addition, this measure may be used in the future with synthetic oscillatory regulatory genetic circuits to quantify the quality of the clocks as regulatory elements, to design robust genetic systems.

# 10. Modeling noise loops (Monte Carlo simulations).

## 10.1. Construction of Monte Carlo simulation.

In general, analytic solutions (such as the results of the master equation calculations) offer more insight than finite element or Monte Carlo simulations. However, in the case of the noise loops, the analytic answer provides information regarding collective properties of the entire population, without giving a description of typical behavior of individual cells. A Monte Carlo simulation of the processes in individual *Synechococcus* will continue to develop the understanding of this system.

To this end, I coded a Monte Carlo description of the mRNA and YFP protein fluctuations in *Synechococcus*. The program uses the same basic scheme shown in Figure 9.1. The program was written in MATLAB (The Mathworks) and employed a direct Gillespie algorithm (82). In this algorithm, the rates of all possible reactions are summed. A single random number is selected, which (coupled with the sum of the reaction rates) indicates the amount of time before an event (i.e., production or destruction of a single mRNA or protein) occurs. A second random number is chosen, which selects the particular type of event based on the probabilities of each possible event. This algorithm is much more efficient than other alternatives, as only two random numbers need be selected per reaction event.

The initial protein and mRNA values are seeded with their average levels, equal to $\bar{r} = \frac{Ak_r}{\gamma_r}$ for mRNA and $\bar{p} = \frac{Ak_r k_p}{\gamma_r \gamma_p}$ for the protein. In order to allow any effects from the choice of initial conditions to dissipate, three full periods elapse before the data is

saved. When the initial values are changed (within reasonable limits), the statistical results after three periods remain similar.

## 10.2. Results.

Figure 10.1 shows a typical track for a single simulation performed with the nominal parameters described in Chapter 9, with levels recorded after every event. The mRNA is present in low numbers during the part of the cycle with above average expression (circadian time 6-18), and virtually absent during the rest of the cycle. While the mRNA is being produced, a net increase in the protein level results, effectively integrating the genetic transcriptional activity. The jagged, noisy nature of the increase (originating from the burst phenomenon) is evident. During the period with no mRNA production (circadian time 18-6), the protein levels have a net loss, which appears much smoother than the production phase. The phase shift between the peak protein and mRNA levels is also clear.

Figure 10.2 shows a series of protein levels (recorded at one hour intervals) for 100 such simulations. This figure allows a more clear visualization of the origin of the noise loops. As the protein levels increase, the lines are very "jumbled", a consequence of uncorrelated bursts between individual cells. As the levels decrease, the curves are mostly smooth and travel together in a parallel bundle. This view reinforces the idea of bursts as the origin of noise.

Figure 10.3 shows the statistical results of 1000 independent runs. The mean and coefficient of variation oscillate with similar phase offset as the experimental results shown in Figure 6.1 and the analytic master equation results in Figure 9.4. These results

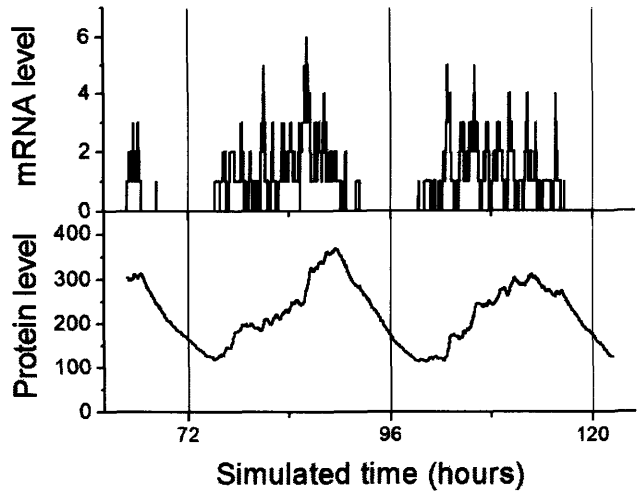


**Figure 10.1.** mRNA and protein levels in a simulated cell. The nominal parameters listed in Chapter 9 were used to generate this simulation. The protein experiences a net increase during the periods of higher mRNA production controlled by the circadian cycle. During this growth, the protein production is related to the mRNA production by the burst factor, and thus the increase is significantly noisy. The part of the cycle with greater repression sees nearly no mRNA production, and the protein level decays very smoothly.

---

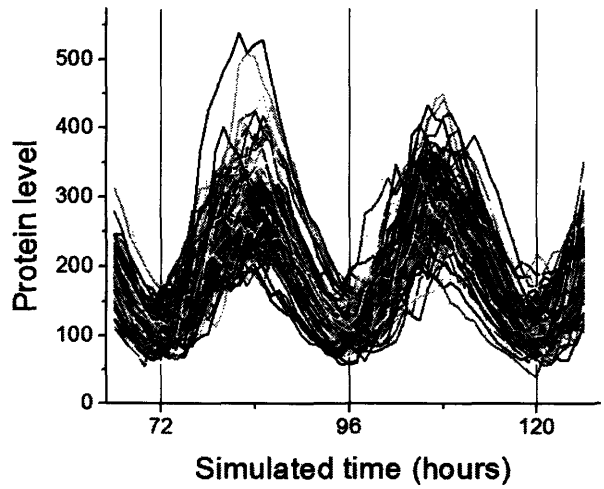


**Figure 10.2.** Protein level traces for 100 simulated cells. Data is plotted representing protein levels in one-hour increments; the nominal parameters were used. The uncorrelated bursts during the production phase are evident, as are the smooth, mostly parallel trajectories during the decay phase. These results suggest that the bursts are responsible for the higher noise during the net production stage.

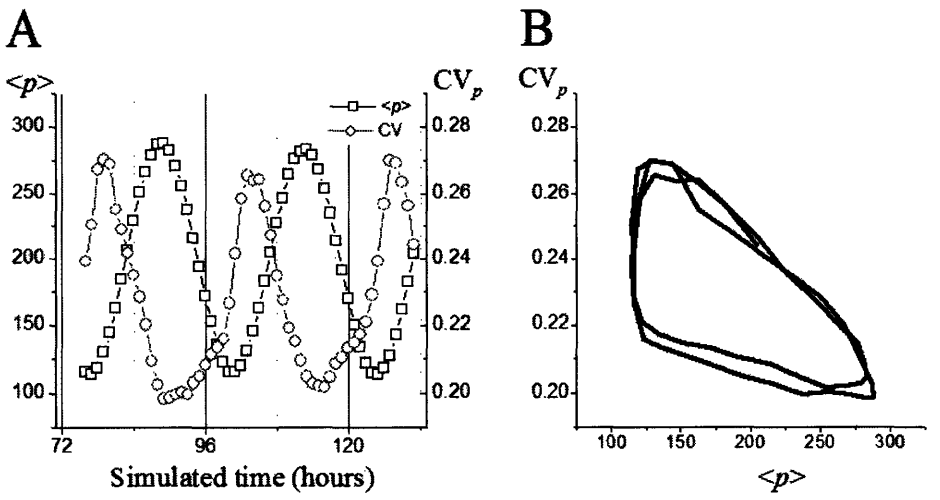


**Figure 10.3.** Statistical results of Monte Carlo simulations. (A) Plots of the mean and coefficient of variation of the protein level in 1000 separate simulations. (B) Noise loops from the results in (A). The results in both plots are identical to the analytic master equation results displayed in Figure 9.4 to within statistical error.

---

can be combined to display noise loops; these loops trace (to within statistical sampling error) the loops described by the master equation solution.

## 10.3. Monte Carlo simulations with deterministic protein production and decay.

Using the Monte Carlo simulations, it is possible to separate the contributions of noise originating in mRNA processes (i.e., transcription and decay) and the noise arising from the protein (translation and decay). This is accomplished by returning to the differential equation describing protein production and decay

$$\frac{dp}{dt} = k_p r - \gamma_p p$$

where both $r$ and $p$ are functions of time. Leaving $r$ as an as-yet-undetermined function, the protein level at time $t$ can be written as

$$p(t) = k_p e^{-\gamma_p t} \int_0^t r e^{\gamma_p \tau} d\tau$$

To remove the noise associated with protein production and decay, I used a Monte Carlo simulation to generate the mRNA levels as a function of time. At each time, the protein was calculated using the equation above, numerically integrating the mRNA history. The initial values were the same as in the full simulations; once again, only data after three full periods was utilized. The results of this reduced simulation are shown in Figure 10.4. The mean protein oscillates with the same phase and absolute values as in the full model employed in Figure 10.2; the noise keeps the same relative phase. However, examining the resulting noise loops shows that the overall noise has been reduced, although the "opening" of the loop is only minimally changed.

The separate noise contributions of the mRNA and protein events to the overall noise and loop opening can now be compared. The baseline of the mRNA loop gives the mRNA "constant" noise, and the quadrature difference of the maximum and minimum noise values gives the mRNA "period related" noise. The baseline of the full noise loop then allows the calculation of the protein "constant" noise (by subtracting in quadrature the mRNA constant noise), and maximum of the loop gives the protein period-related noise (by subtracting in quadrature the mRNA maximum and protein constant noises). The results are summarized in Table 10.1.

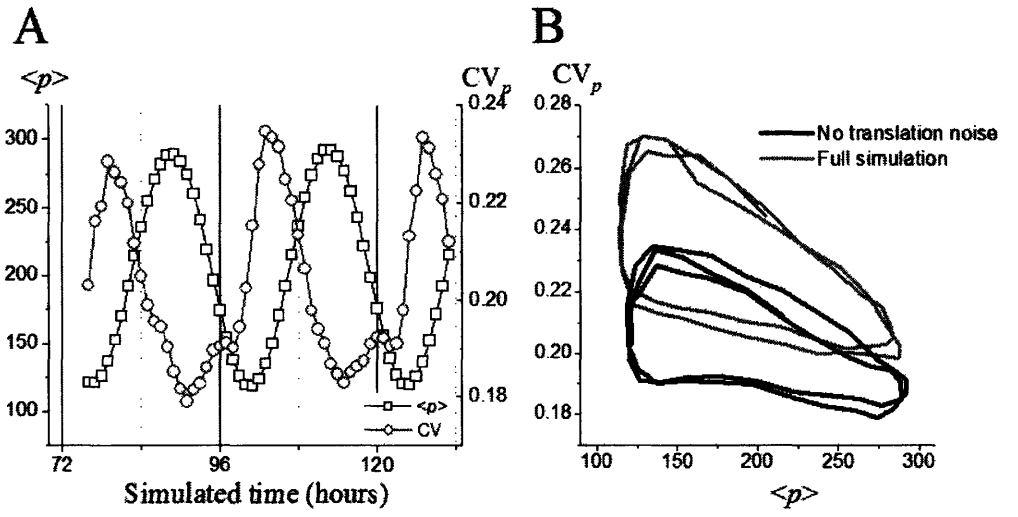


**Figure 10.4.** Monte Carlo simulations with deterministic protein production and decay. The mRNA levels over time are simulated using the stochastic Gillespie algorithm, and the protein levels are calculated from the resulting mRNA populations. (A) Protein mean and coefficient of variation for 1000 simulated cells. The mean levels are identical to the full Monte Carlo simulations in Figure 10.3, while the calculated noise is less than with the full random simulation. (B) Noise loops from the results in (A), compared to the fully stochastic simulation. Again, the range of mean values is the same, but the noise is less with deterministic protein kinetics. This allows for separate estimation of intrinsic noise arising from mRNA and protein fluctuations, on both a static and dynamic time scale.

---

**Table 10.1.** mRNA and protein noises calculated from Monte Carlo simulations.

| Noise source | Type | Value | Relative to mRNA constant noise |
|---|---|---|---|
| mRNA | Constant | 0.190 | 1 |
| | Period-related | 0.1349 | 0.71 |
| protein | Constant | 0.1006 | 0.52 |
| | Period-related | 0.0921 | 0.48 |

These results indicate that, for the nominal parameters, stochasticity associated with translation and protein decay contribute significantly to the total noise in the system. As the burst size increases, the relative contribution of the protein to the total noise is

reduced; eventually, for burst sizes of 20 or more, the mRNA-only loops become virtually identical to the loops from the full solution for that burst size.

# 11. Conclusions.

In recent years, tools have become available to explore the subject of variability (or "noise") in genetic systems. This noise can arise from intrinsic fluctuations originating in the stochastic behavior of low numbers of molecules, or extrinsic variation of rates from cell to cell. Numerous theoretical and experimental studies predict and measure the contributions of noise in various systems, both prokaryotic and eukaryotic. One common feature of these studies is that they all explore systems that operate in the steady state.

Depending on the system being probed, noise can be desirable (the $\lambda$ phage) or undesirable (embryonic development). A clock is an archetypal system in which noise should be minimized, given the important role that clocks play in coordinating many other behaviors. Many biological organisms, from prokaryotes to mammals, have built-in clocks called circadian rhythms. These clocks enable the tracking of the time of day, even under constant conditions, and regulate various important aspects of the day-to-day lives of these organisms. Because they are vital regulators of essential biological operations, any noise in the clock itself will affect these downstream processes. Therefore, characterization of the noise in the clock mechanism is fundamental to understanding the overall functioning of the organism.

The cyanobacterium *Synechococcus* has evolved a circadian rhythm mechanism, which effectively regulates all genes within the organism with an internal 24 hour clock. This clock remains exceptionally accurate over extended periods of time under constant conditions. Proteins expressed under the control of this robust system cycle over a range

of values, but somehow the population of cells remains well synchronized, despite evidence that multicellular communication is absent in this system.

Prior to this study, the tools available to probe noise in the cyanobacterial system were limited to bioluminescence and other bulk measurements, such as Northern and Western blots for mRNA and protein levels respectively. In this work, I developed the use of fluorescent proteins as a reporter for genetic activity in *Synechococcus*, including quantifying the phase delay and impact on peak-to-trough ratio of the detected signal compared to the driving function.

The use of the much brighter fluorescent proteins instead of bioluminescent proteins enabled imaging of individual cells growing, dividing, and oscillating with higher resolution and under more ideal conditions than previous work. This work enabled the visualization of the robust rhythm shared by cells and their progeny. These fluorescent "movies" also revealed for the first time the necessity of the Kai clock proteins for oscillation at the individual cell level.

The use of fluorescence also enabled the rapid measurement of the variability of protein levels in populations of synchronized cyanobacterial cells. These experiments revealed that, contrary to existing theories of genetic noise, the variability (defined as standard deviation normalized by the mean) was not a single-valued function inversely depending on the mean. Rather, the noise in the system was consistently higher as the mean level increased, and lower as the mean decreased. These "noise loops" formed the motivation for non-steady state modeling of genetic expression driven by an oscillating regulatory mechanism.

Further experiments were performed to probe the sources of the genetic noise in *Synechococcus*. The consistent result of independent experiments was that the majority

of the variability in these cells arises from events associated with stochastic reactions downstream of the clock regulation. This type of noise is uncorrelated between various genes within a given cell. The implication is that the noise introduced by the clock itself is low. This is consistent with recent discoveries that the core of the clock is rooted in biochemical reactions between tens of thousands of molecules, instead of the more noisy process of genetic expression, and helps explain how the clock can remain accurate for great lengths of time.

An analytic expression for the time-dependent moments of the resulting mRNA and protein probability distributions was obtained by use of a master equation approach. These results showed that while mRNA levels should never produce noise loops, the expressed protein will lead to noise loops as long as the protein lifetime and oscillatory period are comparable and the "burst size", or average number of proteins produced per mRNA transcript, is nonzero. Increasing the burst size serves to open the noise loops further. The interpretation of the loops is that bursts of protein production lead to high noise (relatively few events of mRNA transcription are involved), while the decay of individual proteins is a much lower noise process (many events of protein degradation take place). Stochastic Monte Carlo programs validate this explanation. Simulations using a random mRNA algorithm and deterministic protein production and decay permitted comparison of the relative contributions of each species to the total intrinsic fluctuation noise in the system.

The model, with only five parameters (of which four are experimentally determined, and the fifth takes on a value typical in prokaryotes), allowed for the calculation of intrinsic noise within the cyanobacterium. Comparison with experimental results led to estimations of the time-independent extrinsic noise, as well as the remaining

unexplained noise with a 24-hour periodicity. This oscillating noise is termed the "clock noise" and an estimation of the magnitude of this variability was determined.

This work extends the understanding of variability in gene expression beyond pre-existing theories and experiments, which were confined to the steady state. The identification, quantification, and modeling of genetic noise oscillations, presented here for the first time, will be extended to other natural oscillators, as well as the burgeoning field of synthetic genetic network design. Just as oscillatory elements are ubiquitous in electronic circuits for such functions as timing and regulation, so will they play a vital role in the synthesis and functioning of custom genetic-based systems. The knowledge

# Appendix 1: Plasmids

| Name | NS (I,II) | cut | Promoter | cut | Gene | cut |
|---|---|---|---|---|---|---|
| pJRC01 | I | NotI | $P_{trc}$ | SphI,BamHI | -- | SmaI,SacI |
| pJRC02 | I | NotI | $P_{trc}$ | SphI,BamHI | *cfp* | SacI |
| pJRC03 | I | NotI | $P_{trc}$ | SphI,BamHI | *gfpmut3.1* | SacI |
| pJRC04 | I | NotI | $P_{trc}$ | SphI | *gfpuv* | SmaI,SacI |
| pJRC05 | I | NotI | $P_{trc}$ | SphI,BamHI | *yfp* | SacI |
| pJRC06 | II | Xba | $P_{trc}$ | -- | *cfp* | SalI |
| pJRC08 | II | Xba | $P_{trc}$ | -- | *yfp* | SalI |
| pJRC09 | I | NotI | $P_{trc}$ | SphI,BamHI | *yfp* | SacI |
| pJRC10 | I | NotI | $P_{trc}$ | SphI,BamHI | *cfpLVA* | SacI |
| pJRC11 | II | Xba | $P_{trc}$ | -- | *cfpLVA* | SalI |
| pJRC12 | II | Xba | $P_{trc}$ | -- | *yfpLVA* | SalI |
| pJRC13 | I | NotI | $P_{trc}$ | SphI,BamHI | *yfpLVA* | SacI |
| pJRC14 | I | NotI | $P_{kaiA}$ | BamHI | *cfp* | SacI |
| pJRC15 | I | NotI | $P_{kaiBC}$ | -- | *cfp* | SacI |
| pJRC16 | I | NotI | *kaiC*USR | BamHI | *cfp* | SacI |
| pJRC17 | I | NotI | $P_{psbAI(s)}$ | BamHI | *cfp* | SacI |
| pJRC18 | I | NotI | $P_{kaiA}$ | BamHI | *yfp* | SacI |
| pJRC19 | I | NotI | $P_{kaiBC}$ | BamHI | *yfp* | SacI |
| pJRC20 | I | NotI | *kaiC*USR | BamHI | *yfp* | SacI |
| pJRC21 | I | NotI | $P_{psbAI(s)}$ | BamHI | *yfp* | SacI |
| pJRC22 | I | NotI | $P_{kaiA}$ | BamHI | *cfpLVA* | SacI |
| pJRC23 | I | NotI | $P_{kaiBC}$ | -- | *cfpLVA* | SacI |
| pJRC24 | I | NotI | *kaiC*USR | BamHI | *cfpLVA* | SacI |
| pJRC25 | I | NotI | $P_{psbAI(s)}$ | BamHI | *cfpLVA* | SacI |
| pJRC26 | I | NotI | $P_{kaiA}$ | BamHI | *yfpLVA* | SacI |
| pJRC27 | I | NotI | $P_{kaiBC}$ | -- | *yfpLVA* | SacI |
| pJRC28 | I | NotI | *kaiC*USR | BamHI | *yfpLVA* | SacI |
| pJRC29 | I | NotI | $P_{psbAI(s)}$ | BamHI | *yfpLVA* | SacI |
| pJRC30 | II | XbaI | $P_{kaiA}$ | -- | *yfp* | SalI |
| pJRC31 | II | XbaI | $P_{kaiBC}$ | -- | *yfp* | SalI |
| pJRC32 | II | XbaI | *kaiC*USR | -- | *yfp* | SalI |
| pJRC33 | II | XbaI | $P_{psbAI(s)}$ | -- | *yfp* | SalI |
| pJRC34 | II | XbaI | $P_{kaiA}$ | -- | *yfpLVA* | SalI |
| pJRC35 | II | XbaI | $P_{kaiBC}$ | -- | *yfpLVA* | SalI |

| Name | NS (I,II) | cut | Promoter | cut | Gene | cut |
|---|---|---|---|---|---|---|
| pJRC36 | II | XbaI | *kaiC*USR | -- | *yfpLVA* | SalI |
| pJRC37 | II | XbaI | $P_{psbAI(s)}$ | -- | *yfpLVA* | SalI |
| pJRC38 | I | NotI | $P_{kaiA}$ | BamHI | *yfp* | SacI |
| pJRC39 | I | NotI | $P_{kaiBC}$ | BamHI | *yfp* | SacI |
| pJRC40 | I | NotI | *kaiC*USR | BamHI | *yfp* | SacI |
| pJRC41 | I | NotI | $P_{psbAI(s)}$ | BamHI | *yfp* | SacI |
| pJRC42 | I | NotI | $P_{purF}$ | -- | *yfp* | SacI |
| pJRC43 | I | NotI | $P_{purF}$ | -- | *yfpLVA* | SacI |
| pJRC44 | I | NotI | $P_{psbAI(f)}$ | BamHI | *cfp* | SacI |
| pJRC45 | I | NotI | $P_{psbAI(f)}$ | BamHI | *yfp* | SacI |
| pJRC46 | I | NotI | $P_{psbAI(f)}$ | BamHI | *cfpLVA* | SacI |
| pJRC47 | I | NotI | $P_{psbAI(f)}$ | BamHI | *yfpLVA* | SacI |
| pJRC49 | II | -- | $P_{psbAI(f)}$ | -- | *cfp* | SalI |
| pJRC50 | II | -- | $P_{psbAI(f)}$ | -- | *yfp* | SalI |
| pJRC51 | II | -- | $P_{psbAI(f)}$ | -- | *cfpLVA* | SalI |
| pJRC52 | II | -- | $P_{psbAI(f)}$ | -- | *yfpLVA* | SalI |
| pJRC53 | II | XbaI | $P_{kaiA}$ | -- | *yfp* | SalI |
| pJRC54 | II | XbaI | $P_{kaiBC}$ | -- | *yfp* | SalI |
| pJRC55 | II | XbaI | $P_{psbAI(s)}$ | -- | *yfp* | SalI |
| PIL001 | II | -- | $P_{lacIQ}$ | -- | *lacIQ* | SalI |

NOTES: NS = neutral site
*kaiC*USR means ~500 bases upstream of the *kaiC* start codon
"cut" means a unique restriction site in the position indicated (e.g., before promoter)
$P_{psbAI(s)}$=100 bases upstream of psbAI start codon
$P_{psbAI(f)}$=500 bases upstream of psbAI start codon
BamHI not unique in NSII vector
$P_{kaiBC}$ and $P_{purF}$ contain BamHI site; inserted into vectors using BglII downstream (removes restriction site)
$P_{psbAI(f)}$ and $P_{lacIQ}$::lacIQ contain XbaI site; inserted into vectors using NheI upstream (removes restriction site)
Plasmid pIL001 constructed by Inna Lipchin

# Appendix 2: Strains

| Name | NSI | NSII | Resist | Background |
|---|---|---|---|---|
| JRCS02 | $P_{trc}$::*cfp* | | Sp | |
| JRCS03 | $P_{trc}$::*gfpmut3.1* | | Sp | |
| JRCS04 | $P_{trc}$::*gfpuv* | | Sp | |
| JRCS05 | $P_{trc}$::*yfp* | | Sp | |
| JRCS06 | | $P_{trc}$::*cfp* | Kan | |
| JRCS07 | $P_{trc}$::*yfp* | $P_{trc}$::*cfp* | Sp, Kan | |
| JRCS08 | | $P_{trc}$::*yfp* | Kan | |
| JRCS10 | $P_{trc}$::*cfpLVA* | | Sp | |
| JRCS11 | | $P_{trc}$::*cfpLVA* | Kan | |
| JRCS13 | $P_{trc}$::*yfpLVA* | | Sp | |
| JRCS14 | $P_{trc}$::*yfpLVA* | $P_{trc}$::*cfpLVA* | Sp, Kan | |
| JRCS16 | $P_{kaiA}$::*cfp* | | Sp | |
| JRCS17 | $P_{kaiBC}$::*cfp* | | Sp | |
| JRCS18 | *kaiC*USR::*cfp* | | Sp | |
| JRCS19 | $P_{psbAI(s)}$::*cfp* | | Sp | |
| JRCS22 | $P_{kaiA}$::*cfpLVA* | $P_{kaiA}$::*yfpLVA* | Sp, Kan | |
| JRCS23 | $P_{kaiBC}$::*cfpLVA* | $P_{kaiA}$::*yfpLVA* | Sp, Kan | |
| JRCS24 | $P_{psbAI(s)}$::*cfpLVA* | $P_{kaiA}$::*yfpLVA* | Sp, Kan | |
| JRCS25 | $P_{kaiA}$::*cfpLVA* | $P_{kaiBC}$::*yfpLVA* | Sp, Kan | |
| JRCS26 | $P_{kaiBC}$::*cfpLVA* | $P_{kaiBC}$::*yfpLVA* | Sp, Kan | |
| JRCS27 | $P_{psbAI(s)}$::*cfpLVA* | $P_{kaiBC}$::*yfpLVA* | Sp, Kan | |
| JRCS28 | $P_{kaiA}$::*cfpLVA* | $P_{psbAI(s)}$::*yfpLVA* | Sp, Kan | |
| JRCS29 | $P_{kaiBC}$::*cfpLVA* | $P_{psbAI(s)}$::*yfpLVA* | Sp, Kan | |
| JRCS30 | $P_{psbAI(s)}$::*cfpLVA* | $P_{psbAI(s)}$::*yfpLVA* | Sp, Kan | |
| JRCS31 | $P_{kaiA}$::*yfpLVA* | | Sp | |
| JRCS32 | $P_{kaiBC}$::*yfpLVA* | | Sp | |
| JRCS33 | $P_{psbAI(s)}$::*yfpLVA* | | Sp | |
| JRCS34 | | $P_{kaiA}$::*yfpLVA* | Kan | |
| JRCS35 | | $P_{kaiBC}$::*yfpLVA* | Kan | |
| JRCS37 | | $P_{psbAI(s)}$::*yfpLVA* | Kan | |
| JRCS38 | $P_{kaiA}$::*yfp* | | Sp | |
| JRCS39 | $P_{kaiBC}$::*yfp* | | Sp | |
| JRCS40 | *kaiC*USR::*yfp* | | Sp | |
| JRCS41 | $P_{psbAI(s)}$::*yfp* | | Sp | |

| Name | NSI | NSII | Resist | Background |
|---|---|---|---|---|
| JRCS42 | $P_{purF}$::*yfp* | | Sp | |
| JRCS43 | $P_{purF}$::*yfpLVA* | | Sp | |
| JRCS44 | $P_{psbAI(f)}$::*cfp* | | Sp | |
| JRCS45 | $P_{psbAI(f)}$::*yfp* | | Sp | |
| JRCS46 | $P_{psbAI(f)}$::*cfpLVA* | | Sp | |
| JRCS47 | $P_{psbAI(f)}$::*yfpLVA* | | Sp | |
| JRCS48 | $P_{trc}$::*yfpLVA* | $P_{lacIQ}$::*lacIQ* | Sp, Kan | |
| JRCS49 | | $P_{psbAI(f)}$::*cfp* | Kan | |
| JRCS50 | | $P_{psbAI(f)}$::*yfp* | Kan | |
| JRCS51 | | $P_{psbAI(f)}$::*cfpLVA* | Kan | |
| JRCS52 | | $P_{psbAI(f)}$::*yfpLVA* | Kan | |
| JRCS53 | $P_{kaiBC}$::*yfp* | | Sp | |
| JRCS54 | $P_{psbAI(s)}$::*yfp* | | Sp | |
| JRCS55 | $P_{purF}$::*yfp* | | Sp | |
| JRCS56 | $P_{kaiA}$::*cfp* | $P_{kaiA}$::*yfp* | Sp, Kan | |
| JRCS57 | $P_{kaiBC}$::*cfp* | $P_{kaiBC}$::*yfp* | Sp, Kan | |
| JRCS58 | $P_{psbAI(s)}$::*cfp* | $P_{psbAI(s)}$::*yfp* | Sp, Kan | |
| JRCS59 | | $P_{kaiA}$::*yfp* | Kan | |
| JRCS60 | | $P_{kaiBC}$::*yfp* | Kan | |
| JRCS61 | | $P_{psbAI(s)}$::*yfp* | Kan | |
| JRCS62 | $P_{kaiBC}$::*yfpLVA* | | Sp | Δ*kaiA* (AMC702) |
| JRCS63 | $P_{kaiBC}$::*yfpLVA* | | Sp | Δ*kaiB* (AMC703) |
| JRCS64 | $P_{kaiBC}$::*yfpLVA* | | Sp | Δ*kaiC* (AMC704) |
| JRCS65 | $P_{kaiBC}$::*yfpLVA* | | Sp | Δ*kaiBC* (AMC705) |
| JRCS66 | | $P_{kaiBC}$::*yfpLVA* | Kan | Δ*kaiA* (AMC702) |
| JRCS67 | | $P_{kaiBC}$::*yfpLVA* | Kan | Δ*kaiB* (AMC703) |
| JRCS68 | | $P_{kaiBC}$::*yfpLVA* | Kan | Δ*kaiC* (AMC704) |
| JRCS69 | | $P_{kaiBC}$::*yfpLVA* | Kan | Δ*kaiBC* (AMC705) |
| JRCS70 | $P_{kaiBC}$::*yfpLVA* | $P_{kaiBC}$::*yfpLVA* | Sp, Kan | |
| IL001 | $P_{trc}$::*yfp* | $P_{lacIQ}$::*lacIQ* | Sp, Kan | |

NOTES: "Background" means the strain into which the constructs were inserted
(If this field is blank, the original strain was PCC7942)
NSI and NSII refer to the construct in neutral site I or II respectively
Resist: Sp=40 μg/mL spectinomycin; Kan=50 μg/mL kanamycin
Sp,Kan=2 μg/mL spectinomycin and 2.5 μg/mL kanamycin
Strain IL001 prepared by Inna Lipchin

# Bibliography


1. Copinschi, G., Spiegel, K., Leproult, R., and Van Cauter, E., Pathophysiology of human circadian rhythms, *Novartis Found Symp, 227*, 143 (2000).

2. deMairan, J., Observation botanique, in *Histiore de l'Academie Royale des Science de Paris*, pp. 35 (1729).

3. Pittendrigh, C. S., Circadian systems: entrainment, in *Handbook of behavioral neurobiology: biological rhythms*, Vol. 4: Biological rhythms, Aschoff, J., Ed., Plenum Press, New York, pp. 95 (1981).

4. Pittendrigh, C. S., Circadian systems: general perspective, in *Handbook of behavioral neurobiology*, Vol. 4: Biological rhythms, Aschoff, J., Ed., Plenum Press, New York, pp. 57 (1981).

5. Duffy, J. F., and Dijk, D. J., Getting through to circadian oscillators: why use constant routines?, *J Biol Rhythms, 17*, 4 (2002).

6. Richter, C. P., Inherent twenty-four hour and lunar clocks of a primate--the squirrel monkey, *Comm. Behav. Biol., 1*, 305 (1968).

7. Moore-Ede, M. C., Sulzman, F. M., and Fuller, C. A., *The clocks that time us: physiology of the circadian timing system*, Harvard University Press, Cambridge, MA (1982).

8. Young, M. W., Jackson, F. R., Shin, H. S., and Bargiello, T. A., A biological clock in *Drosophila, Cold Spring Harb Symp Quant Biol, 50*, 865 (1985).

9. Rutila, J. E., Suri, V., Le, M., So, W. V., Rosbash, M., and Hall, J. C., CYCLE is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of *Drosophila period* and *timeless*, *Cell, 93*, 805 (1998).

10. Jury, S. H., Chabot, C. C., and Watson, W. H., Daily and circadian rhythms of locomotor activity in the American lobster, *Homarus americanus*, *J. Exp. Mar. Biol. Ecol., 318*, 61 (2005).

11. Bianchi, D. E., An endogenous circadian rhythm in *Neurospora crassa*, *J Gen Microbiol, 35*, 437 (1964).

12. Partensky, F., Hess, W. R., and Vaulot, D., *Prochlorococcus*, a marine photosynthetic prokaryote of global significance, *Microbiol Mol Biol Rev, 63*, 106 (1999).

13. Pandey, K. D., Shukla, S. P., Shukla, P. N., Giri, D. D., Singh, J. S., Singh, P., and Kashyap, A. K., Cyanobacteria in Antarctica: ecology, physiology and cold adaptation, *Cell Mol Biol (Noisy-le-grand), 50*, 575 (2004).

14. Shestakov, S. V., and Khyen, N. T., Evidence for genetic transformation in blue-green alga *Anacystis nidulans*, *Mol Gen Genet, 107*, 372 (1970).

15. Mitsui, A., Kumazawa, S., Takahashi, A., Ikemoto, H., Cao, S., and Arai, T., Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically, *323*, 720 (1986).

16. Chen, T.-H., Chen, T.-L., Hung, L.-M., and Huang, T.-C., Circadian rhythm in amino acid uptake by *Synechococcus* RF-1, *Plant physiol, 97*, 55 (1991).

17. Sweeney, B. M., and Borgese, M. B., Note. A circadian rhythm in cell division in a prokaryote, the cyanobacterium *Synechococcus* WH7803, *J Phycol, 25*, 183 (1989).

18. Kondo, T., Strayer, C. A., Kulkarni, R. D., Taylor, W., Ishiura, M., Golden, S. S., and Johnson, C. H., Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria, *Proc Natl Acad Sci U S A, 90*, 5672 (1993).

19. Mori, T., Saveliev, S. V., Xu, Y., Stafford, W. F., Cox, M. M., Inman, R. B., and Johnson, C. H., Circadian clock protein KaiC forms ATP-dependent hexameric rings and binds DNA, *Proc Natl Acad Sci U S A, 99*, 17203 (2002).

20. Hayashi, F., Suzuki, H., Iwase, R., Uzumaki, T., Miyake, A., Shen, J. R., Imada, K., Furukawa, Y., Yonekura, K., Namba, K., and Ishiura, M., ATP-induced hexameric ring structure of the cyanobacterial circadian clock protein KaiC, *Genes Cells, 8*, 287 (2003).

21. Kondo, T., Tsinoremas, N. F., Golden, S. S., Johnson, C. H., Kutsuna, S., and Ishiura, M., Circadian clock mutants of cyanobacteria, *Science, 266*, 1233 (1994).

22. Nishiwaki, T., Iwasaki, H., Ishiura, M., and Kondo, T., Nucleotide binding and autophosphorylation of the clock protein KaiC as a circadian timing process of cyanobacteria, *Proc Natl Acad Sci U S A, 97*, 495 (2000).

23. Ishiura, M., Kutsuna, S., Aoki, S., Iwasaki, H., Andersson, C. R., Tanabe, A., Golden, S. S., Johnson, C. H., and Kondo, T., Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobacteria, *Science, 281*, 1519 (1998).

24. Lorne, J., Scheffer, J., Lee, A., Painter, M., and Miao, V. P., Genes controlling circadian rhythm are widely distributed in cyanobacteria, *FEMS Microbiol Lett, 189*, 129 (2000).

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

26. Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., Arellano, A., Coleman, M., Hauser, L., Hess, W. R., Johnson, Z. I., Land, M., Lindell, D., Post, A. F., Regala, W., Shah, M., Shaw, S. L., Steglich, C., Sullivan, M. B., Ting, C. S., Tolonen, A., Webb, E. A., Zinser, E. R., and Chisholm, S. W., Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation, *Nature, 424*, 1042 (2003).

27. Williams, S. B., Vakonakis, I., Golden, S. S., and LiWang, A. C., Structure and function from the circadian clock protein KaiA of *Synechococcus elongatus*: a potential clock input mechanism, *Proc Natl Acad Sci U S A, 99*, 15357 (2002).

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

29. Katayama, M., Kondo, T., Xiong, J., and Golden, S. S., *ldpA* encodes an iron-sulfur protein involved in light-dependent modulation of the circadian period in the cyanobacterium *Synechococcus elongatus* PCC 7942, *J Bacteriol, 185*, 1415 (2003).

30. Aschoff, J., Circadian rhythms in man, *Science, 148*, 1427 (1965).

31. Aschoff, J., and Wever, R., Circadian period and phase-angle difference in chaffinches (Fringilla coelebs L.), *Comp Biochem Physiol, 18*, 397 (1966).

32. Aschoff, J., and von Saint Paul, U., Circadian rhythms of brain temperature in the chicken, measured at different levels of constant illumination, *Jpn J Physiol, 23*, 69 (1973).

33. Aschoff, J., Figala, J., and Poppel, E., Circadian rhythms of locomotor activity in the golden hamster (*Mesocricetus auratus*) measured with two different techniques, *J Comp Physiol Psychol, 85*, 20 (1973).

34. Aschoff, J., Naps as integral parts of the wake time within the human sleep-wake cycle, *J Biol Rhythms, 9*, 145 (1994).

35. Katayama, M., Tsinoremas, N. F., Kondo, T., and Golden, S. S., *cpmA*, a gene involved in an output pathway of the cyanobacterial circadian system, *J Bacteriol, 181*, 3516 (1999).

36. Iwasaki, H., Williams, S. B., Kitayama, Y., Ishiura, M., Golden, S. S., and Kondo, T., A KaiC-interacting sensory histidine kinase, SasA, necessary to sustain robust circadian oscillation in cyanobacteria, *Cell, 101*, 223 (2000).

37. Kutsuna, S., Kondo, T., Aoki, S., and Ishiura, M., A period-extender gene, *pex*, that extends the period of the circadian clock in the cyanobacterium *Synechococcus* sp. strain PCC 7942, *J Bacteriol, 180*, 2167 (1998).

38. Liu, Y., Tsinoremas, N. F., Johnson, C. H., Lebedeva, N. V., Golden, S. S., Ishiura, M., and Kondo, T., Circadian orchestration of gene expression in cyanobacteria, *Genes Dev, 9*, 1469 (1995).

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

40. Barkai, N., and Leibler, S., Circadian clocks limited by noise, *Nature, 403*, 267 (2000).

41. Vilar, J. M., Kueh, H. Y., Barkai, N., and Leibler, S., Mechanisms of noise-resistance in genetic oscillators, *Proc Natl Acad Sci U S A, 99*, 5988 (2002).

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

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

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

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

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

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

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

49. Yamaguchi, S., Isejima, H., Matsuo, T., Okura, R., Yagita, K., Kobayashi, M., and Okamura, H., Synchronization of cellular clocks in the suprachiasmatic nucleus, *Science, 302*, 1408 (2003).

50. Kiss, I. Z., Zhai, Y., and Hudson, J. L., Emerging coherence in a population of chemical oscillators, *Science, 296*, 1676 (2002).

51. Kondo, T., Mori, T., Lebedeva, N. V., Aoki, S., Ishiura, M., and Golden, S. S., Circadian rhythms in rapidly dividing cyanobacteria, *Science, 275*, 224 (1997).

52. Mori, T., and Johnson, C. H., Independence of circadian timing from cell division in cyanobacteria, *J Bacteriol, 183*, 2439 (2001).

53. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B., Gene regulation at the single-cell level, *Science, 307*, 1962 (2005).

54. Young, M. W., and Kay, S. A., Time zones: a comparative genetics of circadian clocks, *Nat Rev Genet, 2*, 702 (2001).

55. Nakajima, M., Imai, K., Ito, H., Nishiwaki, T., Murayama, Y., Iwasaki, H., Oyama, T., and Kondo, T., Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro, *Science, 308*, 414 (2005).

56. Fogg, G. E., Stewart, W. D. P., Fay, P., and Walsby, A. E., *The blue-green algae*, Academic Press, London (1973).

57. Allen, M. B., and Arnon, D. I., Studies in nitrogen-fixing blue-green algae. I. Growth and nitrogen fixation by *Anabaena cylindrica* Lemm., *Plant physiol, 30*, 366 (1955).

58. Bustos, S. A., and Golden, S. S., Expression of the *psbDII* gene in *Synechococcus* sp. strain PCC 7942 requires sequences downstream of the transcription start site, *J Bacteriol, 173*, 7525 (1991).

59. Welsh, D. K., and Kay, S. A., Bioluminescence imaging in living organisms, *Curr Opin Biotechnol, 16*, 73 (2005).

60. Hollis, R. P., Lagido, C., Pettitt, J., Porter, A. J., Killham, K., Paton, G. I., and Glover, L. A., Toxicity of the bacterial luciferase substrate, n-decyl aldehyde, to *Saccharomyces cerevisiae* and *Caenorhabditis elegans*, *FEBS Lett, 506*, 140 (2001).

61. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C., Green fluorescent protein as a marker for gene expression, *Science, 263*, 802 (1994).

62. Atkins, D., and Izant, J. G., Expression and analysis of the green fluorescent protein gene in the fission yeast *Schizosaccharomyces pombe*, *Curr Genet, 28*, 585 (1995).

63. Yeh, E., Gustafson, K., and Boulianne, G. L., Green fluorescent protein as a vital marker and reporter of gene expression in *Drosophila*, *Proc Natl Acad Sci U S A, 92*, 7036 (1995).

64. Kunert, A., Hagemann, M., and Erdmann, N., Construction of promoter probe vectors for *Synechocystis* sp. PCC 6803 using the light-emitting reporter systems Gfp and LuxAB, *J Microbiol Methods, 41*, 185 (2000).

65. Xu, X., and Wolk, C. P., Role for *hetC* in the transition to a nondividing state during heterocyst differentiation in *Anabaena* sp, *J Bacteriol, 183*, 393 (2001).

66. Yoon, H. S., and Golden, J. W., Heterocyst pattern formation controlled by a diffusible peptide, *Science, 282*, 935 (1998).

67. Vellanoweth, R. L., and Rabinowitz, J. C., The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli in vivo*, *Mol Microbiol, 6*, 1105 (1992).

68. Becskei, A., Boselli, M. G., and van Oudenaarden, A., Amplitude control of cell-cycle waves by nuclear import, *Nat Cell Biol, 6*, 451 (2004).

69. Andersen, J. B., Sternberg, C., Poulsen, L. K., Bjorn, S. P., Givskov, M., and Molin, S., New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria, *Appl Environ Microbiol, 64*, 2240 (1998).

70. Berg, O. G., A model for the statistical fluctuations of protein numbers in a microbial population, *J Theor Biol, 71*, 587 (1978).

71. Muller-Hill, B., Crapo, L., and Gilbert, W., Mutants that make more *lac* repressor, *Proc Natl Acad Sci U S A, 59*, 1259 (1968).

72. McClure, W. R., Mechanism and control of transcription initiation in prokaryotes, *Annu Rev Biochem, 54*, 171 (1985).

73. Arbuzov, V. A., and Ivanova, L. E., [Rate of ribosome movement along messenger RNA in *E. coli* under normal and inhibited translation], *Biokhimiia, 41*, 768 (1976).

74. Thattai, M., and van Oudenaarden, A., Intrinsic noise in gene regulatory networks, *Proc Natl Acad Sci U S A, 98*, 8614 (2001).

75. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., and van Oudenaarden, A., Regulation of noise in the expression of a single gene, *Nat Genet, 31*, 69 (2002).

76. Paulsson, J., Summing up the noise in gene networks, *Nature, 427*, 415 (2004).

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

78. Pedraza, J. M., and van Oudenaarden, A., Noise propagation in gene networks, *Science, 307*, 1965 (2005).

79. Pedraza, J. M., Noise in gene regulatory networks, in *Complex systems science in biomedicine*, Deisboeck, T., Ed., Kluwar Academic/Plenum Publishers, New York (2005).

80. Rowe, H. E., *Signals and noise in communication systems*, D. van Nostrand Company Inc., Princeton, N.J. (1965).

81. Selinger, D. W., Saxena, R. M., Cheung, K. J., Church, G. M., and Rosenow, C., Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation, *Genome Res, 13*, 216 (2003).

82. Gillespie, D., Exact stochastic simulation of coupled chemical reaction, *Journal of Physical Chemistry, 81*, 2340 (1977).