

# Complementary approaches to understanding the plant circadian clock

Ozgur E. Akman<sup>1,\*</sup> Maria Luisa Guerriero<sup>1</sup> Laurence Loewe<sup>1</sup> Carl Troein<sup>1,2</sup>

<sup>1</sup> Centre for Systems Biology at Edinburgh, University of Edinburgh, UK

<sup>2</sup> School of Biological Sciences, University of Edinburgh, UK

\* Current address: School of Engineering, Computing & Mathematics, University of Exeter, UK

Circadian clocks are oscillatory genetic networks that help organisms adapt to the 24-hour day/night cycle. The clock of the green alga *Ostreococcus tauri* is the simplest plant clock discovered so far. Its many advantages as an experimental system facilitate the testing of computational predictions.

We present a model of the *Ostreococcus* clock in the stochastic process algebra Bio-PEPA and exploit its mapping to different analysis techniques, such as ordinary differential equations, stochastic simulation algorithms and model-checking. The small number of molecules reported for this system tests the limits of the continuous approximation underlying differential equations. We investigate the difference between continuous-deterministic and discrete-stochastic approaches. Stochastic simulation and model-checking allow us to formulate new hypotheses on the system behaviour, such as the presence of self-sustained oscillations in single cells under constant light conditions.

We investigate how to model the timing of dawn and dusk in the context of model-checking, which we use to compute how the probability distributions of key biochemical species change over time. These show that the relative variation in expression level is smallest at the time of peak expression, making peak time an optimal experimental phase marker. Building on these analyses, we use approaches from evolutionary systems biology to investigate how changes in the rate of mRNA degradation impacts the phase of a key protein likely to affect fitness. We explore how robust this circadian clock is towards such potential mutational changes in its underlying biochemistry. Our work shows that multiple approaches lead to a more complete understanding of the clock.

## 1 Introduction

The daily cycles in sunlight, temperature and other environmental parameters are highly important to most organisms. To follow and anticipate these cycles, living cells generate biochemical rhythms with a period of approximately 24 hours (*circadian*). The majority of the known circadian clocks, including those in eukaryotes, are based on one or more interlocking transcriptional feedback loops between a set of key genes. Crucial to the function of the clock is its ability to *entrain* to environmental signals (i.e. to adjust its internal rhythm by synchronising with external cycles), so that the phase of gene expression is maintained under changes to the length of the day (photoperiod). Such entrainment acts through various photoreceptor pathways, where light affects kinetic parameters of the core clock. In addition to circadian entrainment, a defining feature of circadian clocks is that they exhibit continued oscillations in constant light conditions [11].

The circadian clocks of many organisms are organised around complex feedback loop architectures, making the determination of design principles a challenging computational problem. Although research has revealed much about the clock of the foremost model plant organism, *Arabidopsis thaliana*, there are still unidentified components and inconsistencies between computational models and experimental observations [16]. For this reason, to increase our mechanistic understanding of circadian clocks, it is

desirable to investigate simpler systems that possess functional similarity with more complex networks. The circadian clock of the green alga *Ostreococcus tauri* [8] is the simplest plant clock discovered so far, and is thus an ideal model system for understanding plant circadian function with the help of experiments, simulations and theory. A quantitative model describing the biochemical reactions of the *Ostreococcus* clock can serve as a focal point for this research, yielding a low-dimensional test system for various mathematical analysis techniques.

Bio-PEPA [6] is a stochastic process algebra specifically defined to model and analyse biochemical systems. Exploiting the defined formal mappings of Bio-PEPA models into a number of equivalent representations, it is possible to analyse Bio-PEPA models using different mathematical and computational techniques, including ordinary differential equations (ODEs), stochastic simulation algorithms (SSAs) and model-checking.

In previous work we used both ODEs and SSAs to model the clock of the fungus *Neurospora crassa*, demonstrating that combining different analysis methods is important for fully quantifying the relationship between feedback architecture and circadian behaviour [1]. Here we build on this approach, applying a broader range of computational techniques to the *Ostreococcus* clock. We develop and analyse a Bio-PEPA model of the clock, focusing on various stochastic methods which are the most appropriate in this case due to the low copy numbers characteristic of the system. In particular, we exploit the automatic generation of PRISM models from Bio-PEPA to carry out a novel application of the PRISM model-checker [15] to a circadian model, computing time-dependent probability distributions for the clock components. We use the model to quantify the variability and robustness of the clock's functional behaviour with respect to the following factors: (i) internal stochastic noise, the inevitable consequence of a system comprising a small number of molecules; (ii) environmental changes, such as photoperiod variations and transitions between constant light/darkness; and (iii) mutational changes that affect the biochemical reaction rates of our model, representing perturbations to the system that occur on an evolutionary timescale.

The rest of our paper is structured as follows. After an overview of Bio-PEPA in Section 2, the *Ostreococcus* clock is introduced in Section 3, followed by the description of the corresponding Bio-PEPA model. In Section 4 we analyse the model using various approaches. We first use stochastic simulation to investigate how different light conditions affect the oscillations of the clock. We then explore approaches for modelling light entrainment in a continuous-time Markov chain (CTMC) before using model-checking to compute the time-dependent probability distributions of protein levels. This enables us to identify the phase markers that are most robust to stochastic fluctuations. Finally we use ideas from a recently developed framework for evolutionary systems biology [18] to test how mutational changes in mRNA degradation rate affect the phase of oscillations in comparison to the inherent stochastic noise that is present at the individual cell level. The full Bio-PEPA model is given in Appendix A.

## 2 An overview of Bio-PEPA

Bio-PEPA [6] is a stochastic process algebra, recently developed for the modelling and analysis of biological systems. We give here a brief overview of the main features of the language. For a detailed presentation of its syntax and semantics, see [6].

The main components of a Bio-PEPA system are the *species components*, describing the behaviour of each species, and the *model component*, specifying all interactions and initial amounts of species. The syntax of Bio-PEPA components is given by:

$$S ::= (\alpha, \kappa) \text{ op } S \mid S + S \mid C \quad \text{with op} = \downarrow \mid \uparrow \mid \oplus \mid \ominus \mid \odot \quad P ::= P \stackrel{t}{\bowtie} P \mid S(x)$$

where  $S$  is the *species component* and  $P$  is the *model component*. In the prefix term  $(\alpha, \kappa) \text{ op } S$ ,  $\kappa$  is the *stoichiometry coefficient* of species  $S$  in reaction  $\alpha$ , and the *prefix combinator* “op” represents the role of  $S$  in the reaction. Specifically,  $\downarrow$  indicates a *reactant*,  $\uparrow$  a *product*,  $\oplus$  an *activator*,  $\ominus$  an *inhibitor* and  $\odot$  a generic *modifier*. The notation  $\alpha \text{ op}$  is a shorthand for  $(\alpha, \kappa) \text{ op } S$  when  $\kappa = 1$ . The operator “+” expresses a choice between possible actions, and the constant  $C$  is defined by an equation  $C \stackrel{\text{def}}{=} S$ . The process  $P \bowtie_{\mathcal{I}} Q$  denotes synchronisation between components  $P$  and  $Q$ ; the set  $\mathcal{I}$  determines the activities on which the operands are forced to synchronise, with  $\bowtie_{*}$  denoting a synchronisation on all common action types. In the model component  $S(x)$ , the parameter  $x \in \mathbb{N}$  represents the initial number of molecules  $S$  present. In addition to species and model components, a Bio-PEPA system consists of kinetic rates, parameters and, if needed, locations, events and other auxiliary information for the species.

The formal representation offered by Bio-PEPA allows for different kinds of analysis through the defined mapping into continuous-deterministic and discrete-stochastic analysis methods (see [6] for details). More on Bio-PEPA can be found at [5], including two software tools, the Bio-PEPA Eclipse Plug-in and the Bio-PEPA Workbench [10]. Both tools process Bio-PEPA models automatically and either compute time-series results directly using various SSA or ODE solvers, or generate representations that can be used by other tools.

### 3 The *Ostreococcus* clock

*Ostreococcus tauri* is an exceptionally small green alga with a highly reduced genome [9]. Experiments and homology searches indicate that its circadian clock is very simple compared to higher plants, such as *Arabidopsis thaliana*. Only a handful of the clock genes identified in other plants have been found in *Ostreococcus*, and only two of these appear to be central to the clock. The first of these, which we refer to as *TOC1*, is homologous to *Arabidopsis TOC1 (TIMING OF CAB EXPRESSION 1)* and other *PRRs (PSEUDO RESPONSE REGULATORS)*. The other gene, here called *LHY*, is homologous to *Arabidopsis LHY (LATE ELONGATED HYPOCOTYL)* and *CCA1 (CIRCADIAN CLOCK ASSOCIATED 1)* [8]. An ODE model of the *Ostreococcus* clock as a negative feedback loop between these two genes was introduced in [20], where it was applied to drug treatments and other perturbations. The full model includes details of the luciferase assay used to measure mRNA and protein levels, but here we use only the central parts of the model, which describe the dynamics of the native mRNAs and proteins. The model is illustrated in Figure 1.

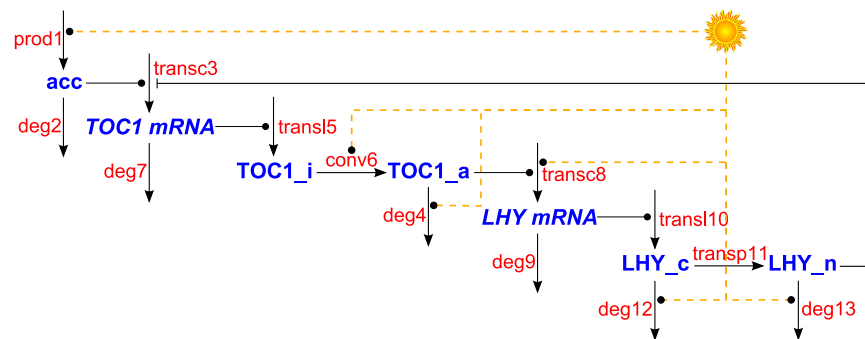


Figure 1: The genetic regulatory network underlying our model of the *Ostreococcus* clock. The network comprises a single negative feedback loop involving the *LHY* and *TOC1* genes, augmented by 5 light inputs which synchronise the endogenous oscillations in gene expression to the day/night cycle [20].

*TOC1* transcription requires light, which is buffered by a “light accumulator” (*acc*) and is inhibited by the presence of nucleic LHY protein (*LHY\_n*) for most of the day. *TOC1* activates *LHY* transcription through an unknown mechanism, proposed in [20] to work as follows: *TOC1* mRNA is translated into inactive *TOC1* protein (*TOC1\_i*), which is activated slowly during the day but quickly after dusk. The active form (*TOC1\_a*) drives *LHY* transcription throughout the night but is quickly degraded after dawn. *LHY* mRNA is translated into cytosolic LHY (*LHY\_c*), which is quickly translocated to the nucleus, thereby closing the feedback loop. Light also accelerates the rate of LHY degradation.

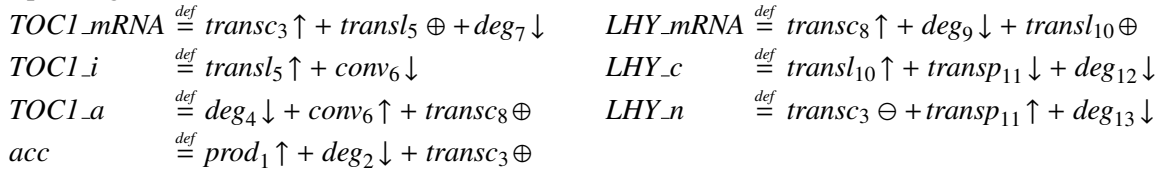
The model parameters were estimated by fitting simulated time-courses to equivalent data obtained from experiments over a wide range of light conditions. Some experiments alternated 12 hours of light and dark (denoted LD 12:12), others used longer or shorter days (such as LD 16:8 or 8:16), and many included transitions between different conditions, often into constant light (LL) [20].

### 3.1 A Bio-PEPA model of the clock

A model of the clock as described above was implemented in Bio-PEPA. Here we describe its main features; for the full model, including kinetic laws and parameters, see Appendix A.

One of the key issues involved in obtaining a realistic stochastic model is the correct scaling of the initial concentrations and kinetic parameters of the continuous ODE model in [20] so as to obtain the respective molecule counts and rate constants for the Bio-PEPA discrete-state model. Since the absolute values of the initial concentrations are not known, the initial values in the original ODE model are given in arbitrary relative units. However, the peak number of *TOC1* and LHY protein molecules was estimated experimentally over a number of free-running cycles in LL conditions using a TopCount luminometer. From this, approximate initial values for our discrete-state model were computed, yielding a rough estimate for the scaling factor of  $\Omega = 50$ . After such rescaling the Bio-PEPA model can be analysed by ODEs and SSAs, which both give results in molecule counts.

The proteins and mRNAs shown in Figure 1 are modelled as the following Bio-PEPA species components that describe the possible reactions they can participate in and how their amounts are affected by the occurrence of each reaction. Reactions are associated with functional rates representing the corresponding kinetic law.



For instance, the transcription of *TOC1\_mRNA* is modelled by reaction *transc*<sub>3</sub>, which involves three different species (*TOC1\_mRNA*, *LHY\_n*, and *acc*), and is positively regulated by the light-accumulator *acc* and negatively regulated by *LHY\_n*. The kinetic law for this reaction is given by a Hill function, commonly used for describing transcription in clock models [1, 2, 13, 19]:

$$\Omega \cdot \frac{tmp\_toc1\_transcription}{1 + tmp\_toc1\_transcription + \left( \frac{R\_toc1\_lhy}{\Omega} \cdot LHY\_n \right)^{H\_toc1\_lhy}} .$$

Here, species names represent molecule counts,  $tmp\_toc1\_transcription = L\_toc1 + acc \cdot R\_toc1\_acc / \Omega$  and  $L\_toc1$ ,  $R\_toc1\_acc$ ,  $R\_toc1\_lhy$  and  $H\_toc1\_lhy$  are parameters. For comparisons with experiments we also defined the observables  $Total\_LHY = LHY\_c + LHY\_n$  and  $Total\_TOC1 = TOC1\_i + TOC1\_a$ .

Bio-PEPA functional rates allow the definition of general kinetic laws. We use this facility to represent the entrainment of the system to light/dark cycles through the time-dependent function below:



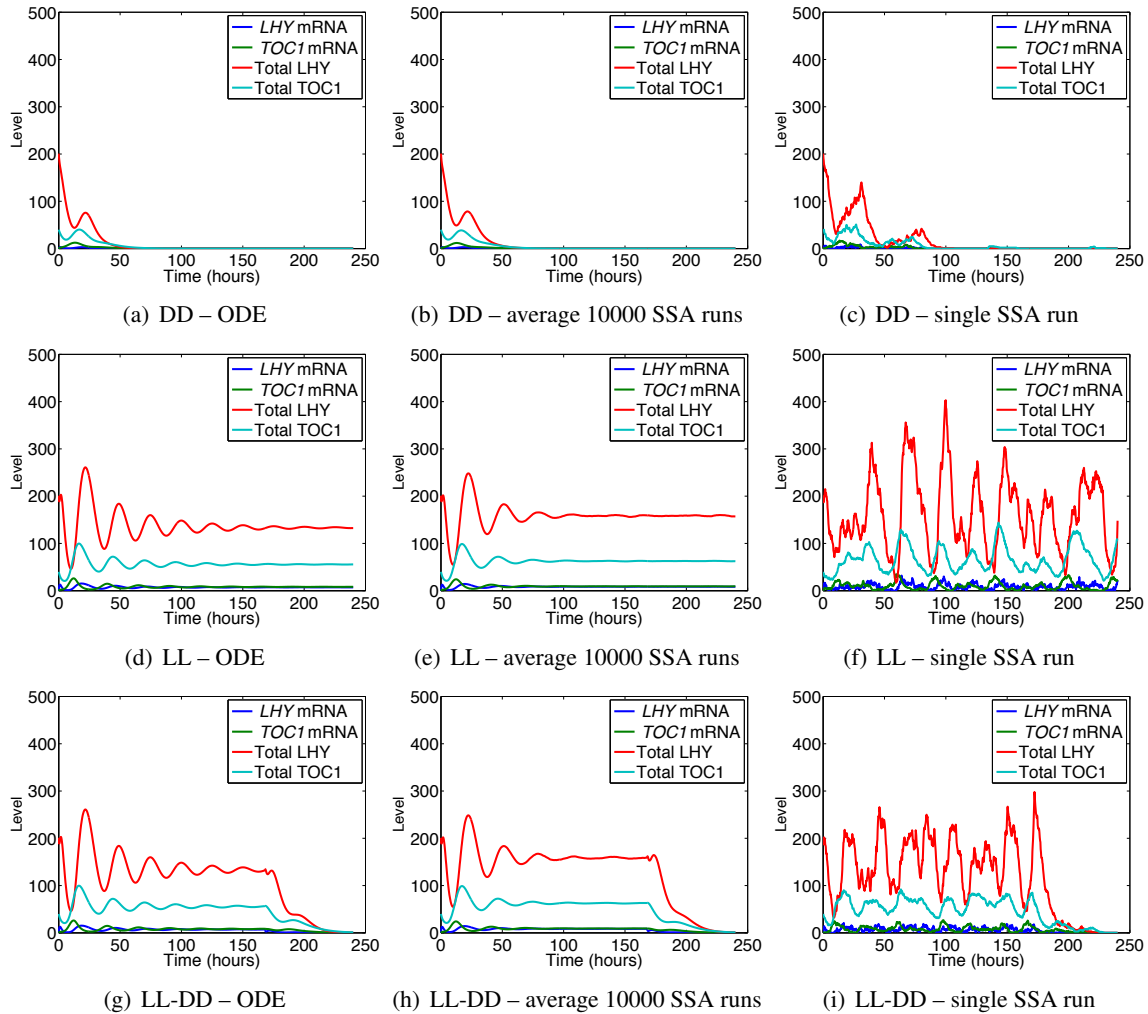


Figure 2: Comparison of the deterministic and stochastic models in different light conditions.

individual SSA runs we note that occasionally a non-zero number of molecules can be briefly observed, even after several circadian cycles (Figure 2(c)).

**LL system.** In constant light (LL), the deterministic system also exhibits damped oscillations, with all species tending to non-zero constant values after about 7-8 days (Figure 2(d)). Similar behaviour is observed by averaging over 10000 SSA runs (Figure 2(e)), though the oscillations damp more rapidly and the steady-state value is slightly different (e.g. the LHY copy number is about 130 in ODEs compared to about 160 in the SSA). However, a 10-fold increase of the scaling factor to  $\Omega = 500$  yields a perfect quantitative agreement between the SSA and ODEs (see Figure 8 in Appendix C). Although the precise source of the discrepancy is not known at present, we hypothesise that it is caused by a breakdown of the continuous approximation underpinning ODEs, due to the low copy numbers obtained at small  $\Omega$  values.

The other, most notable, difference between the deterministic and stochastic models is the behaviour of individual SSA runs (Figure 2(f)), for which persistent irregular oscillations (in both phase and amplitude) are observed. Because of phase diffusion effects, however, these oscillations cannot be detected in

the mean behaviour: hence, in this case, neither the simple average over multiple SSA runs nor the ODE solution gives us a correct indication of the real behaviour of the system, emphasising the importance of observing single realisations.

In view of these findings, we hypothesise that single cell experimental data may exhibit sustained oscillations, whereas the behaviour observed in a large population of cells would be closer to the rapid damping reproduced by both the ODE and SSA average. This is due to the fact that free-running oscillations are unlikely to be synchronised over different cells. Furthermore, visual inspection of the solutions obtained by averaging over different numbers of SSA runs suggests that it should be possible to discern the stochastic effects with a population of around 100 cells. The experimental data currently available for this system are at the level of a population comprising at least 10000 cells. However, we anticipate that the development of new experimental techniques for measuring gene expression in single cells or small populations will enable this hypothesis to be tested experimentally in the not-too-distant future.

**LL-DD system.** As an additional experiment, we consider a system which is kept in constant light (LL) for 160 hours, and then transferred into constant dark (DD). The time-series results are reported in Figures 2(g)–2(i). It can be seen that single realisations of the SSA — approximating the behaviour of individual cells — exhibit immediate cessation of self-sustained oscillations following the LL-DD transfer. This behaviour can be understood by considering the fixed points of the deterministic model.

The LL fixed point is located far from the origin in phase space and is of the stable focus type. Trajectories of the ODE — approximating the behaviour of a large population of cells — spiral around the fixed point as they converge to it, producing slowly damping oscillations. In the corresponding stochastic model, fluctuations kick individual realisations of the system between these spiralling trajectories, thereby preventing the system from remaining close to the fixed point for long periods (see Figure 3). This leads to the irregular self-sustained oscillations observed. By contrast, the DD fixed point of the ODE system is located at the origin. As species concentrations must be positive, the fixed point cannot be a stable focus, and is instead a stable node. Trajectories of the ODE converge directly onto it, generating oscillations that quickly damp to zero. Individual realisations of the stochastic model are thus repeatedly perturbed between rapidly convergent trajectories. Consequently, they quickly approach the DD steady-state following the LL-DD transition, remaining in its vicinity thereafter (see Figure 3).

The model thus predicts that the DD behaviour of the LL-DD system at the single-cell level mirrors that at the population level. If further experiments were to reveal that self-sustained oscillations are observed during the DD phase, this would indicate that the model requires modification to convert the DD fixed point into a stable focus bounded away from the origin.

**LD system.** The light conditions considered so far are experimental settings useful for observing the system's endogenous dynamics. It is also informative, however, to observe the behaviour of the clock under natural conditions (alternating 24-hour cycles of light and dark). We present results obtained for three different photoperiods: 6 hours light/18 hours dark (LD 6:18), 12 hours light/12 hours dark (LD 12:12), and 18 hours light/6 hours dark (LD 18:6).

As described in Section 3, exposure to periodic external stimuli such as light/dark cycles has the effect of resetting the free-running oscillations observed in constant light, so as to establish stable phase relationships with the forcing stimulus. Compared with the free-running LL system, the entrainment to LD cycles regularises the dynamics of the system, markedly reducing the variability of oscillations, particularly in terms of phase. As a consequence, persistent regular oscillations with a stable phase relationship to the light/dark cycle are observed in both ODEs (Figures 4(a), 4(d), 4(g)) and when averaging over multiple SSA runs (Figures 4(b), 4(e), 4(h)). This phase regularisation can also be seen in indi-

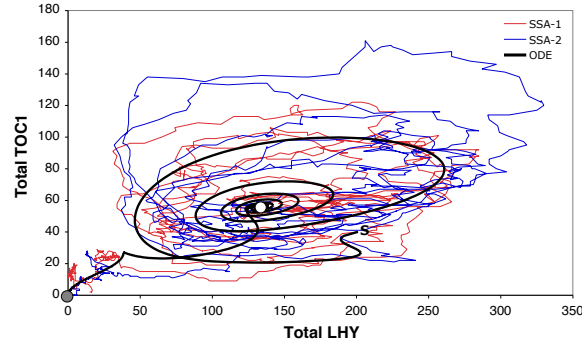


Figure 3: Phase-space for the LL-DD transfer experiment. White and grey dots indicate the fixed points (steady-states) of the deterministic system in LL and DD conditions, respectively. The eigenvalues of the model obtained by linearising the deterministic equations around a fixed point determine the behaviour of the system in its neighbourhood. These are listed for both fixed points in Table 2 of Appendix C.

vidual SSA runs of the entrained system (compare Figures 4(c), 4(f), 4(i)) with the simulations of the free-running clock in Figure 2(f)). These observations are consistent with previous stochastic analyses of clock models [13, 1]. We also note that, as for the LL system, the deterministic and mean stochastic behaviour, whilst very similar, are not in perfect agreement.

## 4.2 Model-checking: time-dependent probability distributions

Model-checking [7] is a formal verification method that allows modellers to state properties of a given model and to automatically check whether they are met. Probabilistic model-checking (see, for instance, [4, 15]) adds probabilistic measures in the evaluation of queries. Recently, model-checking has grown in popularity within the field of systems biology due to its ability to directly answer complex questions on a model's behaviour. Traditional model-checking verification differs from simulation-based analysis in that the verification of a property is obtained from a computation over the entire state-space of the continuous-time Markov chain (CTMC) underlying the model. The major drawback of this approach is the state-space explosion problem: the model's dimension is often too large for computational viability.

Statistical model-checking (see, for instance, [21]) is an alternative query-based approach: it estimates the probability distributions and computes approximate results of queries (together with an estimate of the error) by generating random realisations of the CTMC and averaging the results obtained by evaluating the queries on each of them. The advantage of statistical model-checking over exact verification approaches is that it does not need to build the explicit state-space of the model, which is often intractable, and it does not rely on the transient solution of the CTMC. In essence, statistical model-checking is a verification technique which allows modellers to perform additional analyses of a stochastic system by automatically evaluating queries over multiple simulation traces. The obvious drawback of statistical model-checking is that it only considers a finite number of behaviours of the system (i.e. paths in the CTMC) and, hence, the accuracy of the results is strongly related to that number. However, exact verification of biological systems is generally infeasible, and statistical model-checking often represents a good practical solution. Another issue of probabilistic model-checking is that the transient solution of the CTMC can incur the same averaging effect discussed previously relating to ODEs and mean SSA behaviour: computing the expected value of the model variables might not be sufficient because this would be exactly the same as the deterministic behaviour. Results of reward-based properties, for instance, are



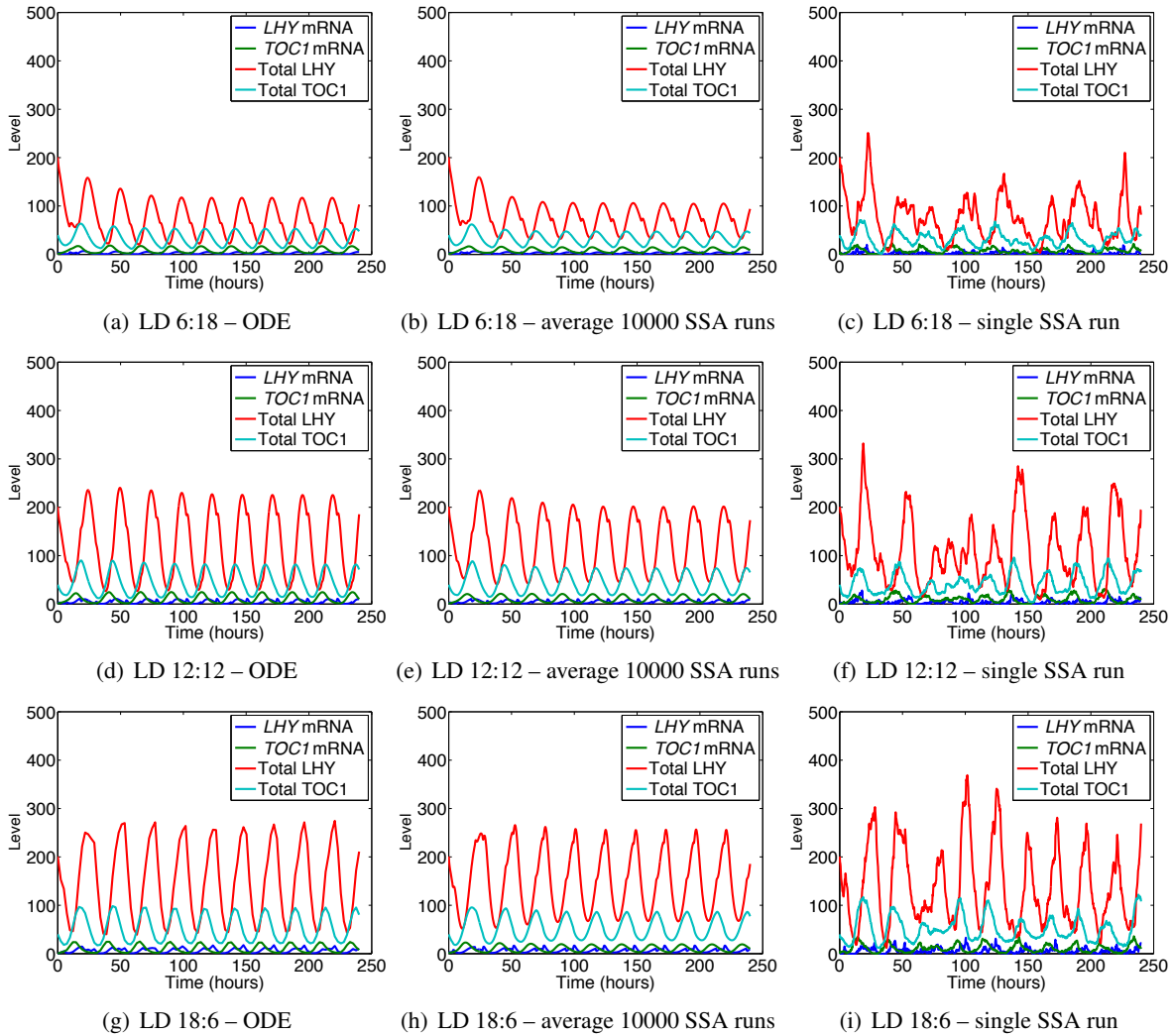


Figure 4: Comparison of the deterministic and stochastic models for different photoperiods.

computed in terms of expected values and this, especially in the case of oscillations which are out of phase, does not give satisfactory results.

PRISM [15] is a probabilistic model-checker, which can be used to verify properties of a CTMC model. It also includes a discrete-event simulator for statistical model-checking. PRISM has been used to analyse systems from a wide range of application domains, and recently also biochemical systems [14]. Models are described using the state-based PRISM language, and it is possible to specify quantitative properties of the system using a property specification language which includes the temporal logic CSL (Continuous Stochastic Logic) [3, 4].

Using the Bio-PEPA Workbench [5], we generated a PRISM model of the clock (together with a set of reward structures and some standard CSL properties which are automatically generated). In the PRISM model, one module is defined for each species, and module local variables are used to record the current quantity of each species. The transitions correspond to the activities of the Bio-PEPA model and the updates take the stoichiometry into account. Transition rates are specified in an auxiliary module

which defines the functional rates corresponding to all the reactions. In order to have a finite CTMC, lower and upper bounds are defined for each variable. In the following, we focus on statistical model-checking: in this case, the choice of the values for the bounds has no effect on the performance of the analysis (since the CTMC is not built) and so we can use arbitrarily high values that are guaranteed not to be reached.

**Modelling the light in PRISM.** In order to model the entrained clock (LD system), time-dependent events must be represented. However, because of the intrinsic nature of model-checking algorithms, which involve the numerical solution of the CTMC underlying stochastic models, deterministic events and time-dependent functions cannot be explicitly specified in PRISM.

We investigated several approaches to address this problem. A first possibility is to split the model-checking algorithm into different (two or more, depending on the time window we are interested in) analysis steps over different time intervals, each with constant light conditions: two different CTMCs would be considered (one for the day-time system and one for the night-time one) with the algorithm switching back and forth between them. The main issue with this approach is how to merge the results obtained over the different time periods. For some particular queries, such as those relating to reachability, this can be done by splitting them into a number of subqueries such that the result (i.e. probability distributions) of one query can be used as the initial state for the next one. For an arbitrary CSL query, however, this cannot be done, and this strongly limits the kind of queries that could be verified.

The alternative approach we consider here is to represent the light by approximating time using a monotonically increasing stochastic variable. The main issue of this approach is that we introduce an additional stochastic effect which is absent in the system we have described so far (i.e. where light is modelled as a deterministic on/off switch). In practice this does not matter, provided that the stochastic variability introduced is kept smaller than the variability of any experiments the model may be compared to. The introduction of an additional variable to model time also causes an increase in the state-space, but this is not an issue here since we focus on statistical model-checking only. An extract from the PRISM model showing how we model time and the day/night switch is provided in Appendix B.

**Time-dependent probability distributions of protein levels.** Using model-checking we can compute the time-dependent probability distributions for each of the model species. For instance, by verifying the CSL property

$$\mathcal{P}_{=?}[\mathbf{F}^{[T,T]} (LHY\_c + LHY\_n = i)]$$

for time instant  $T \in [0, 96]$  and protein level  $i \in [0, 500]$ , we can observe how the probability distribution for LHY protein changes over time during the first 96 hours of simulation.

Each of the plots in Figure 5 refers to a different light condition (DD, LL, and LD 12:12), showing how the probability of being at a particular level changes over time in each case. The plots also report the mean value  $\mu$  and the standard deviation  $\sigma$  of LHY expression. In all cases, the initial amount is  $LHY = 200$ , and then the probability mass gradually moves away from this initial value. In DD, as expected from the results of Section 4.1, the bulk of the probability mass rapidly moves close to zero. In LL we can clearly observe the effect of phase diffusion, resulting in a probability distribution spread almost equally across a wide range of values. By contrast, in LD we are able to observe clear oscillations in the probability distribution; we also note that the amplitude of peak expression is more variable than that of trough expression, with a much broader spread of the probability mass around the mean.

In the following, we focus on the case of alternating light/dark cycles (LD 12:12). In Figure 6(a) we report the probability distribution for LHY protein, together with its mean  $\mu$  and standard deviation  $\sigma$  in

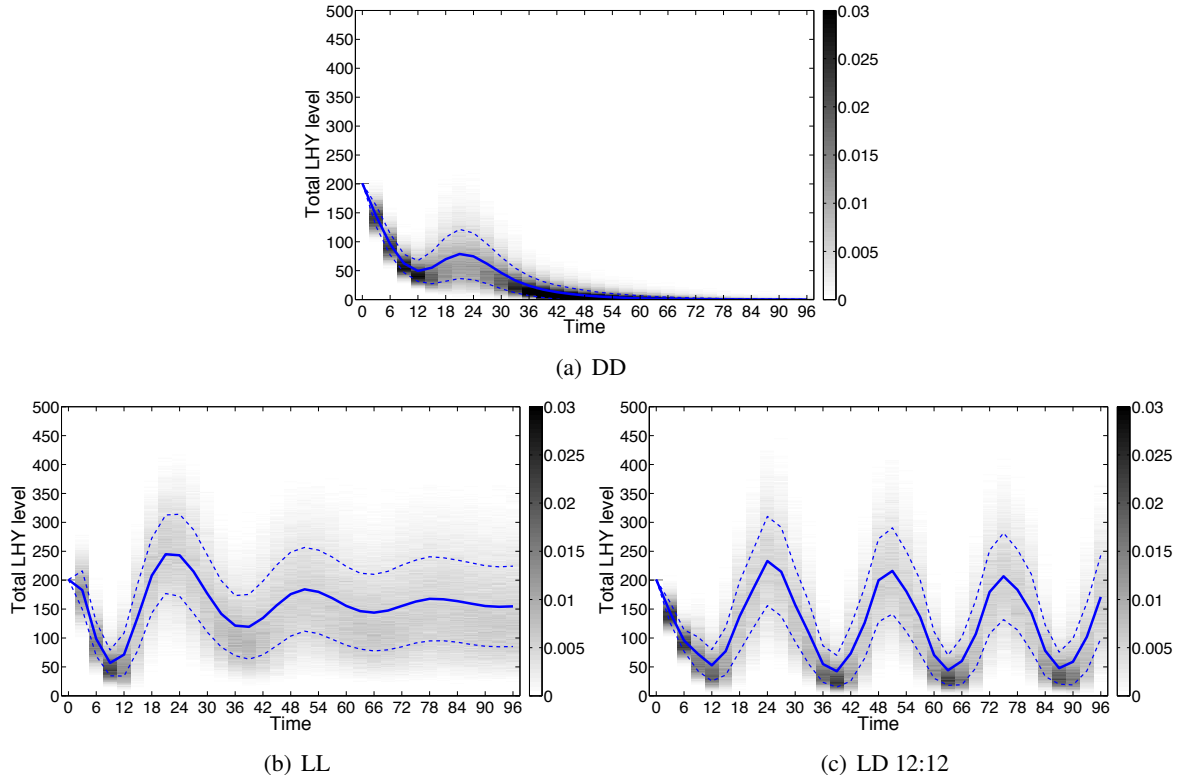


Figure 5: Probability distribution of total LHY level over the first 4 days (0–96 hours, 10000 runs). The heatmap represents the probability distribution: a darker colour corresponds to a higher probability. Blue lines show average and standard deviation ( $\mu \pm \sigma$ ). Similar changes in distribution with time were obtained for TOC1 protein.

a single 24-hour cycle (from 120 to 144 hours). Figure 6(b) plots the oscillation in the corresponding coefficient of variation  $c_v = \frac{\sigma}{\mu}$ . This provides a normalised measure of the sensitivity of the LHY protein oscillation to stochastic fluctuations as a function of circadian time. Small values of  $c_v$ , therefore, correspond to robust phase markers (a high signal-to-noise ratio) and large values poor phase markers (a low-signal-to-noise ratio). Commonly used phase measures in circadian research are the times of peak and trough expression together with the time at which the oscillation falls to its midpoint level. It can be seen in Figure 6(b) that the coefficient of variation is minimal around the peak, suggesting that the latter is the best of the standard phase markers for analysing experimental LHY data.

As discussed in Section 4.1, for the DD system, both the deterministic solution and SSA average quickly attain a constant value. Species amounts can, however, be greater than zero for short time intervals in individual SSA runs (see Figure 2(c)). The following CSL property computes the probability that the total LHY level remains in the range  $[0, e]$  between 96 and 500 hours.

$$\mathcal{P}_{=?}[\mathbf{G}^{[96,500]} (LHY\_c + LHY\_n \leq 0 + e)]$$

$e$	0	2	4	6	8	10	12	14	16	18	20
$\mathbf{P}$	0.8768	0.9276	0.9545	0.9737	0.9833	0.9903	0.9931	0.9964	0.9987	0.9993	0.9998

Table 1: Probability of total LHY to stay below the threshold  $e$  in DD (96–500 hours, 10000 runs).

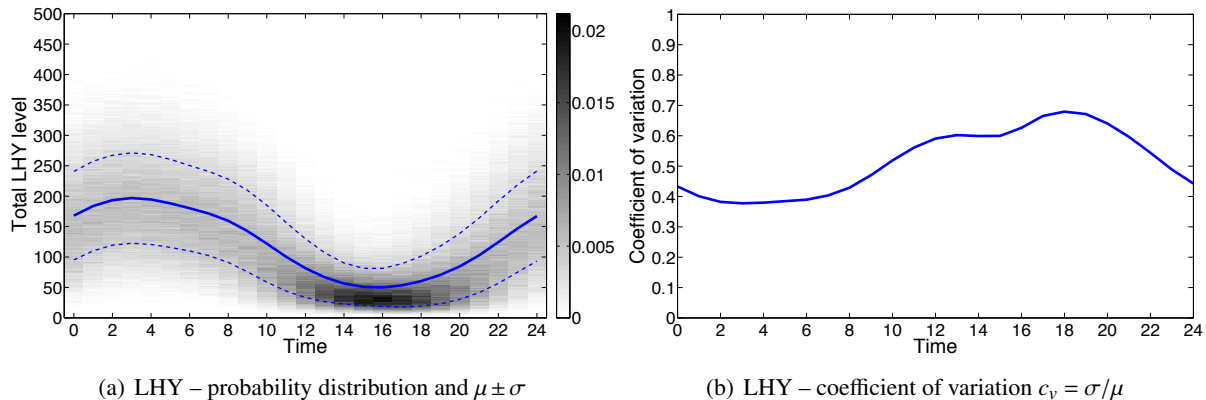


Figure 6: Probability distribution of LHY level over one day (120–144 hours) in LD 12:12 (10000 runs). The distribution of TOC1 exhibits a qualitatively similar variation.

The results reported in Table 1 quantify the observed trend, showing that there is a small probability that LHY exceeds  $e$  for  $0 \leq e \leq 20$ , and that this probability decreases with increasing  $e$ .

### 4.3 Distribution of Mutational Effects: robustness analysis

Thus far we have investigated aspects of the inherent random noise caused by the small size of the system and the discrete nature of its components. We have also explored the consequences of variations in the light environment. We now consider the impact of mutational noise caused by DNA changes that alter reaction rates by affecting the structure of corresponding proteins. This analysis differs from the above in that it requires simulation of vast numbers of different parameter combinations. We do this in the context of a recently developed framework for evolutionary systems biology that describes how the effects of DNA changes propagate through various levels of organisation and abstraction until they impact fitness at the highest level of biological functionality [18]. We limit our analysis to what has been described as the “third” level of the adaptive landscape [18], which maps a combination of biochemical reaction rates to a computable system-level property likely to affect fitness via a quantitative mechanistic model.

We generated a StochKit [17] model using the Bio-PEPA Workbench [5] in order to build on the code base of an analysis framework that has previously been used to investigate the evolutionary systems biology of a different, highly simplified circadian clock system that lacks entrainment [19]. We added a class for computing phase in particularly noisy circadian clocks, using two thresholds to block stochastic noise from generating artificially short ‘cycles’ with very low amplitude. Thresholds were set at 20% and 35% of the distance between the highest peak and lowest trough observed over 10 days, since minima are less variable than maxima (see Figure 4(f)). We used the peak of total TOC1 as a phase marker since our model checking results showed peaks to have higher signal/noise ratios (Figure 6(b)).

We consider the parameter combination used in the rest of the paper to be the wild-type and introduce mutational noise by multiplying degradation rates of all mRNAs in the system (deg7 and deg9 in Figure 1) by a uniformly distributed factor ([0.5,1.5], where 1.0 represents the wild-type). To differentiate between internal stochastic noise and mutational noise we ran 40000 simulations in two sets, each analysing 10000 time-courses for the wild-type and 10000 for mutants. In the first set (Figure 7(a–c)), the system size  $\Omega = 50$  corresponds to that of a single cell. The resulting noise is substantial as indicated by the large width of the distribution of observed phase values (Figure 7(a)). In the second set of

simulations (Figure 7(d–f)), we increased  $\Omega$  a million fold. This results in an excellent approximation of corresponding time-courses produced by ODEs; these simulations are thus denoted as ‘mean’ here. The resulting internal stochastic noise is minimal (see the extremely narrow distribution in Figure 7(d)).

Our ‘mean’ results show how phase is affected by mutational changes in the mRNA degradation rate (Figure 7(d,e)). If the same mutational changes are introduced in the noisy single cell system, they are much more difficult to detect (see Figure 7(a,b)). As a different way of visualising results, Figure 7(c,f) plots the high-level consequences of mutations (phase) against their low-level effect (the factor affecting mRNA degradation). The resulting graphs for the ‘mean’ system show clearly how a change in mRNA degradation rate is expected to affect the mean change in phase (Figure 7(f)). Figure 7(c) shows the corresponding plot for single cell simulations. It demonstrates that a wave of TOC1 peaks starts around the time of dusk (18h in these simulations) and continues for a few hours. The time at which the wave starts is minimally affected by the mutations we investigated, in stark contrast to the variance, which is much lower for higher mRNA degradation rates. In other words, a lower rate of mRNA degradation leads

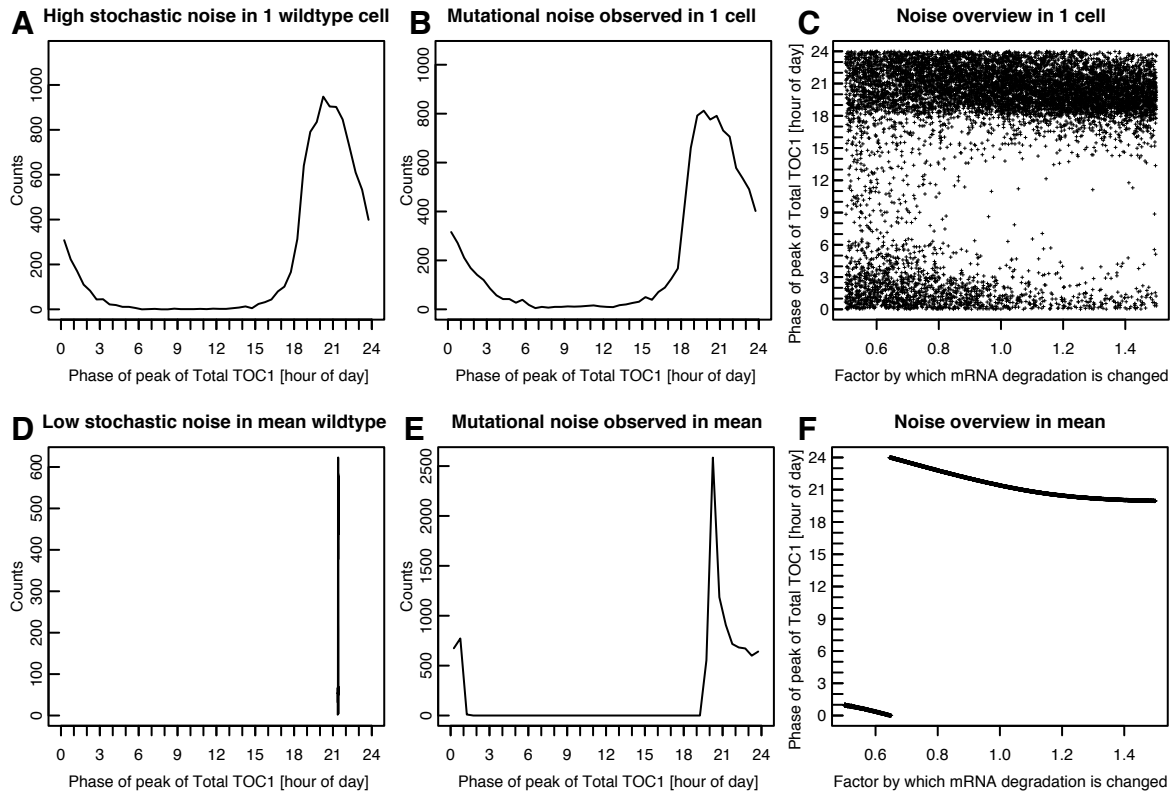


Figure 7: Robustness analysis showing how inherent stochastic noise and mutational noise affect the peak phase of total TOC1 in 4 sets of 10000 runs. Here, noise is measured by the width of the corresponding distributions in phase values, where wider widths indicate greater noise. Plots (a–c) are dominated by the high internal stochastic noise observed in single-cell simulations ( $\Omega = 50$ ). Plots (d–f) observe population averages as computed with  $\Omega = 50 \times 10^6$  and show virtually no internal stochastic noise. The first column (a,d) plots the behaviour of the wild-type, the second (b,e) adds mutational noise by changing the mRNA degradation rate by a factor drawn from a uniform distribution [0.5,1.5], where 1.0 is the wild-type. The third column (c,f) shows how phase depends on mutational effects on mRNA degradation.

to greater internal stochastic noise. This increase in variance explains why the average phase is strongly shifted towards later hours for smaller mRNA degradation rates in the ‘mean’ system (Figure 7(f)), since the mean is strongly affected by large values in skewed distributions as found in (Figure 7(c)). Taken together, these results demonstrate that a higher mRNA degradation rate will on average move the phase forward and make it more reliable (decrease the phase variance), whereas a decrease will move it backwards and make it less reliable.

These subtle patterns in the mutational robustness of the clock could not have been uncovered without running large numbers of simulations with varying parameter combinations. Such work requires a different infrastructure for data analysis from projects that analyse only a few parameter sets.

## 5 Conclusions

We have studied the circadian network of *Ostreococcus tauri* by developing a process algebra model of the clock, based on an existing deterministic representation that was parameterised according to quantitative experimental data. We have investigated several key aspects of the clock, such as the conditions necessary for persistent oscillations in its constituent genes and proteins, as well as the effect of different environmental and mutational changes on the phases of these oscillations. We used the Bio-PEPA stochastic process algebra as a modelling language and applied a range of the analysis methods supported. Because of the low copy numbers of the molecular species involved in the clock network, we focused on stochastic analysis methods which enable the system’s intrinsic variability to be observed.

In particular, we used stochastic simulation to explore how the clock responds to changes in the light environment, and compared the results obtained against the behaviour of the corresponding deterministic system. We predict that the qualitative behaviour of the free-running (LL) clock will be dependent on the size of the cellular population; while damped oscillations will be observed in large populations (simulated by the SSA average and the deterministic model), self-sustained oscillations may be detectable in single cells (simulated by individual runs of the SSA). Model-checking was further used to investigate how the variability of the clock’s behaviour changes over a circadian cycle. By computing the time-dependent probability distributions of the clock proteins, we identified the time of peak expression as the most robust phase marker, suggesting its use as an experimental measure.

Finally, we added mutational noise to our system by randomly changing the overall rate of mRNA degradation and observing how this affects the phase of the oscillations of a key clock protein, likely to have an impact on fitness. We found that the large amount of stochastic noise at the single cell level makes it hard to observe functional changes that may be induced by mutations, without averaging over many observations.

A number of the novel hypotheses we have formulated in this modelling study may provide new biological insights into the behaviour of the *Ostreococcus* clock and will hopefully inspire subsequent experimental research. In addition, further theoretical work can build on the novel model-checking results reported here, to explore additional ways in which systems biology models can be automatically analysed using approaches based on concurrency theory. Our results demonstrate that the integration of different computational techniques is critical for fully quantifying the architectural [2] and mutational [18] robustness of the circadian clock.

## Acknowledgements

The authors thank Gerben van Ooijen for TopCount data and Jane Hillston and Andrew Millar for their helpful comments. The Centre for Systems Biology at Edinburgh is a Centre for Integrative Systems Bi-

ology (CISB) funded by BBSRC and EPSRC, ref. BB/D019621/1. CT is supported by The International Human Frontier Science Program Organization.

## References

- [1] O.E. Akman, F. Ciocchetta, A. Degasperi & M.L. Guerriero (2009): *Modelling Biological Clocks with Bio-PEPA: Stochasticity and Robustness for the Neurospora Crassa Circadian Network*. In: *Proc. of CMSB'09, LNCS 5688*, Springer, pp. 52–67.
- [2] O.E. Akman, J.C.W. Locke, S. Tang, I. Carré, A.J. Millar & D.A. Rand (2008): *Isoform switching facilitates period control in the Neurospora crassa circadian clock*. *Mol. Sys. Biol.* 4, p. 64.
- [3] A. Aziz, K. Kanwal, V. Singhal & V. Brayton (1996): *Verifying continuous time Markov chains*. In: *Proc. of CAV'96, LNCS 1102*, Springer, pp. 269–276.
- [4] C. Baier, B. Haverkort, H. Hermann & J.-P. Katoen (2003): *Model-Checking Algorithms for Continuous-Time Markov Chains*. *IEEE Trans. on Software Eng.* 29(6), pp. 524–541.
- [5] Bio-PEPA Home Page. <http://www.biopepa.org/>.
- [6] F. Ciocchetta & J. Hillston (2009): *Bio-PEPA: a Framework for the Modelling and Analysis of Biological Systems*. *Theoretical Computer Science* 410(33-34), pp. 3065–3084.
- [7] E.M. Clarke, O. Grumberg & D. Peled (1999): *Model Checking*. The MIT Press.
- [8] F. Corellou, C. Schwartz, J.-P. Motta, E.B. Djouani-Tahri, F. Sanchez & F.-Y. Bouget (2009): *Clocks in the Green Lineage: Comparative Functional Analysis of the Circadian Architecture in the Picoeukaryote Ostreococcus*. *Plant Cell* 21, pp. 3436–3449.
- [9] E. Derelle, C. Ferraz, S. Rombauts, P. Rouzé, A.Z. Worden, Robbins S. et al. (2006): *Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features*. *Proc. Natl. Acad. Sci. USA* 103(31), p. 11647.
- [10] A. Duguid, S. Gilmore, M.L. Guerriero, J. Hillston & L. Loewe (2009): *Design and development of software tools for Bio-PEPA*. In: *Proc. of WSC'09*, pp. 956–967.
- [11] J.C. Dunlap, J.J. Loros & P.J. DeCoursey (2004): *Chronobiology: Biological Timekeeping*. Sinauer.
- [12] D.T. Gillespie (1977): *Exact stochastic simulation of coupled chemical reactions*. *J. Phys. Chem.* 81(25), pp. 2340–2361.
- [13] D. Gonze, J. Halloy & A. Goldbeter (2002): *Robustness of circadian rhythms with respect to molecular noise*. *Proc. Natl. Acad. Sci. USA* 99(2), pp. 673–678.
- [14] J. Heath, M. Kwiatkowska, G. Norman, D. Parker & O. Tymchyshyn (2008): *Probabilistic Model Checking of Complex Biological Pathways*. *Theoretical Computer Science* 319, pp. 239–257.
- [15] A. Hinton, M. Kwiatkowska, G. Norman & D. Parker (2006): *PRISM: A tool for automatic verification of probabilistic systems*. In: *Proc. of TACAS'06, LNCS 3920*, pp. 441–444.
- [16] M.A. Jones (2009): *Entrainment of the Arabidopsis Circadian Clock*. *J. Plant Biology* 52(3), pp. 202–209.
- [17] H. Li, Y. Cao, L.R. Petzold & D.T. Gillespie (2008): *Algorithms and Software for Stochastic Simulation of Biochemical Reacting Systems*. *Biotechnology Progress* 24(1), pp. 56–61.
- [18] L. Loewe (2009): *A framework for evolutionary systems biology*. *BMC Systems Biology* 3(27).
- [19] L. Loewe & J. Hillston (2008): *The distribution of mutational effects on fitness in a simple circadian clock*. In: *Proc. of CMSB'08, LNBI 5307*, Springer, pp. 156–175.
- [20] J.S. O'Neill, G. van Ooijen, C. Troein, L. Dixon, F.-Y. Bouget, F. Corellou & A. Millar (2009): *Circadian rhythms persisting in the absence of transcription in a eukaryote*. Submitted.
- [21] H.L.S. Younes & R.G. Simmons (2002): *Probabilistic verification of discrete event systems using acceptance sampling*. In: *Proc. of CAV'02, LNCS 2404*, Springer, pp. 223–235.

## Appendix

### A The full Bio-PEPA model

#### Kinetic parameters:

<i>acc_rate</i>	=	0.085759993119922787	<i>R_toc1_lhy</i>	=	0.80473130211377397
<i>H_toc1_lhy</i>	=	2.4786793492076216	<i>L_toc1</i>	=	0.0001028030683282734
<i>R_toc1_acc</i>	=	0.40030354494924164	<i>D_mrna_toc1</i>	=	0.33395900070057227
<i>T_toc1</i>	=	0.65069237578254624	<i>Di_toc1_ia_l</i>	=	0.11696163098006726
<i>Di_toc1_ia_d</i>	=	0.34434576584349563	<i>D_toc1_a_l</i>	=	0.53999998111757508
<i>D_toc1_a_d</i>	=	0.3587344573844497	<i>H_lhy_toc1</i>	=	2.4123768479176113
<i>R_lhy_toc1_a_l</i>	=	3.3859126401378155	<i>R_lhy_toc1_a_d</i>	=	1.1074418532202324
<i>D_mrna_lhy</i>	=	1.9405472466939	<i>T_lhy</i>	=	6.5204407183218498
<i>Di_lhy_cn</i>	=	7.0630744698933485	<i>D_lhy_l</i>	=	0.34866585983482207
<i>D_lhy_d</i>	=	0.21098655584281875			

#### Time of the day at which dawn and dusk occur:

<i>t<sub>dawn</sub></i>	=	6
<i>t<sub>dusk</sub></i>	=	18 // for the LD 12:12 system

#### Time-dependent function representing light in LD system:

$$\text{light\_time} = H \left( \left( \left( \text{time} - 24 \cdot \left\lfloor \frac{\text{time}}{24} \right\rfloor \right) - t_{\text{dawn}} \right) \cdot \left( t_{\text{dusk}} - \left( \text{time} - 24 \cdot \left\lfloor \frac{\text{time}}{24} \right\rfloor \right) \right) \right)$$

#### Scaling factor:

$$\Omega = 50$$

#### Initial values:

<i>acc_init</i>	=	[0.99897249736755245 · Ω]
<i>TOC1_mRNA_init</i>	=	[1.9315264449894309e <sup>-06</sup> · Ω]
<i>TOC1_i_init</i>	=	[0.34581773957827311 · Ω]
<i>TOC1_a_init</i>	=	[0.47960829226604956 · Ω]
<i>LHY_mRNA_init</i>	=	[9.999999999999999e <sup>-07</sup> · Ω]
<i>LHY_c_init</i>	=	[4.0361051173018776 · Ω]
<i>LHY_n_init</i>	=	[6.7029410613103796e <sup>-06</sup> · Ω]

#### Additional functions:

<i>tmp_toc1_transcription</i>	=	$L_{toc1} + acc \cdot \frac{R_{toc1\_acc}}{\Omega}$
<i>toc1_a_decay</i>	=	$\text{light\_time} \cdot D_{toc1\_a\_l} + (1 - \text{light\_time}) \cdot D_{toc1\_a\_d}$
<i>toc1_i_a_conversion</i>	=	$\text{light\_time} \cdot Di_{toc1\_ia\_l} + (1 - \text{light\_time}) \cdot Di_{toc1\_ia\_d}$
<i>lhy_decay</i>	=	$\text{light\_time} \cdot D_{lhy\_l} + (1 - \text{light\_time}) \cdot D_{lhy\_d}$
<i>lhy_toc1_reg</i>	=	$TOC1\_a \cdot \left( \text{light\_time} \cdot \frac{R_{lhy\_toc1\_a\_l}}{\Omega} + (1 - \text{light\_time}) \cdot \frac{R_{lhy\_toc1\_a\_d}}{\Omega} \right)$



**Functional Rates:**

$prod_1$	: $acc\_rate \cdot \Omega \cdot light\_time$	// light accumulator increase: mass action
$deg_2$	: $acc\_rate \cdot acc$	// light accumulator decrease: mass action
$transc_3$	: $\Omega \cdot \frac{tmp\_toc1\_transcription}{1 + tmp\_toc1\_transcription + \left(\frac{R\_toc1\_lhy}{\Omega} \cdot LHY\_n\right)^{H\_toc1\_lhy}}$	// TOC1 transcription: Hill kinetics
$deg_4$	: $toc1\_a\_decay \cdot TOC1\_a$	// TOC1 degradation: mass action
$transl_5$	: $T\_toc1 \cdot TOC1\_mRNA$	// TOC1 translation: mass action
$conv_6$	: $TOC1\_i\_a\_conversion \cdot TOC1\_i$	// TOC1 conversion: mass action
$deg_7$	: $D\_mrna\_toc1 \cdot TOC1\_mRNA$	// TOC1 mRNA degradation: mass action
$transc_8$	: $\Omega \cdot \frac{lhy\_toc1\_reg^{H\_lhy\_toc1}}{1 + lhy\_toc1\_reg^{H\_lhy\_toc1}}$	// LHY transcription: Hill kinetics
$deg_9$	: $D\_mrna\_lhy \cdot LHY\_mRNA$	// LHY mRNA degradation: mass action
$transl_{10}$	: $T\_lhy \cdot LHY\_mRNA$	// LHY translation: mass action
$transp_{11}$	: $Di\_lhy\_cn \cdot LHY\_c$	// LHY nuclear transport: mass action
$deg_{12}$	: $lhy\_decay \cdot LHY\_c$	// LHY degradation, cytosol: mass action
$deg_{13}$	: $lhy\_decay \cdot LHY\_n$	// LHY degradation, nucleus: mass action

**Species components:**

$LHY\_c$	$\stackrel{def}{=} transl_{10} \uparrow + transp_{11} \downarrow + deg_{12} \downarrow$
$LHY\_mRNA$	$\stackrel{def}{=} transc_8 \uparrow + deg_9 \downarrow + transl_{10} \oplus$
$TOC1\_a$	$\stackrel{def}{=} deg_4 \downarrow + conv_6 \uparrow + transc_8 \oplus$
$TOC1\_mRNA$	$\stackrel{def}{=} transc_3 \uparrow + transl_5 \oplus + deg_7 \downarrow$
$acc$	$\stackrel{def}{=} prod_1 \uparrow + deg_2 \downarrow + transc_3 \oplus$
$TOC1\_i$	$\stackrel{def}{=} transl_5 \uparrow + conv_6 \downarrow$
$LHY\_n$	$\stackrel{def}{=} transc_3 \ominus + transp_{11} \uparrow + deg_{13} \downarrow$

**Species observables:**

$Total\_LHY$	$= LHY\_c + LHY\_n$
$Total\_TOC1$	$= TOC1\_i + TOC1\_a$

**Model component:**

$LHY\_c(LHY\_c\_init)$	$\boxtimes$	$LHY\_mRNA(LHY\_mRNA\_init)$	$\boxtimes$	$TOC1\_a(TOC1\_a\_init)$	$\boxtimes$	
$TOC1\_mRNA(TOC1\_mRNA\_init)$	$\boxtimes$	$acc(acc\_init)$	$\boxtimes$	$TOC1\_i(TOC1\_i\_init)$	$\boxtimes$	$LHY\_n(LHY\_n\_init)$

## B Modelling light/dark cycles in PRISM

```

const int max_days_simulated = 21;

module time

min : [0..59] init 0;
hour : [0..23] init 0;
day : [0..max_days_simulated] init 0;
light_time : [0..1] init 0;

[change_min] (min < 59) -> 60: (min'=min + 1);

[change_hour_dawn] (min = 59 & hour = (tdawn-1) ) -> 60: (min'=0) & (hour' = hour + 1) & (light_time' = 1);
[change_hour_dusk] (min = 59 & hour = (tdusk-1) ) -> 60: (min'=0) & (hour' = hour + 1) & (light_time' = 0);
[change_hour] (min = 59 & hour < 23 & hour != (tdawn-1) & (hour != tdusk-1) ) ->
60: (min'=0) & (hour' = hour + 1);

[time_change_day] (min = 59 & hour = 23 & day < max_days_simulated) ->
60: (min'=0) & (hour'=0) & (day' = day + 1);
[time_end_day] (min = 59 & hour = 23 & day = max_days_simulated) -> 60: (min'=0) & (hour'=0) & (day'=0);

endmodule

```

## C Additional simulation and analysis results

Figure 8 shows the comparison between the ODE results and the mean SSA behaviour with scaling factors  $\Omega = 50$  and  $\Omega = 500$ . We observe a slight difference between the SSA results with the different scaling factors. The reference value is  $\Omega = 50$  (estimated from experimental protein counts), and consequently the predicted biological behaviour is that shown in Figure 8(b). Note, instead, that the ODE results (Figure 8(a)) agree perfectly with the SSA results for the larger value  $\Omega = 500$  (Figure 8(c)).

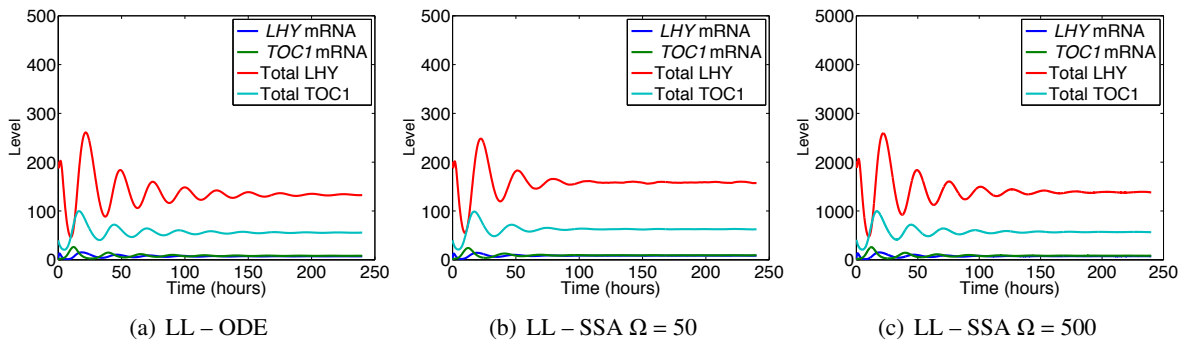


Figure 8: Comparison of deterministic and stochastic models for constant light (LL) with different scaling factors  $\Omega$ .

Table 2 lists the eigenvalues of linearisation at the fixed points of the deterministic model. These determine the dynamics in the vicinity of the steady-states [1]. All eigenvalues of the DD fixed point are negative and real, identifying it as a stable node [1]. The LL fixed point retains 3 of the DD eigenvalues; the remaining ones comprise two complex conjugate pairs with negative real parts. The steady-state is therefore a stable focus [1]. The positions of the fixed points were estimated using the Nelder-Mead simplex algorithm [2], as implemented in the MATLAB routine `fminsearch`. Derivatives were computed analytically using the MATLAB Symbolic Math Toolbox.

DD	LL
-0.0858	-0.0858
-0.2110	-0.0200 - 0.2509i
-0.3340	-0.0200 + 0.2509i
-0.3443	-0.6447 - 0.3216i
-0.3587	-0.6447 + 0.3216i
-1.9405	-1.9506
-7.2741	-7.4117

Table 2: Eigenvalues of the linearised ODE system about the DD and LL fixed points.

## References

- [1] J. Guckenheimer & P. Holmes (1983): *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*. Springer.
- [2] J.C.Lagarias, J. A. Reeds, M. H. Wright & P. E. Wright (1998): *Convergence properties of the Nelder-Mead simplex method in low dimensions*. *SIAM J. Optimiz.* 9(1), pp. 112–147.