1 Supplementary Text

1.1 The PRR7-PRR9-Y model

As with previously published biological models (Leloup & Goldbeter, 2003; Locke *et al.*, 2005; Ueda *et al*, 2001), Hill functions are used to describe transcription rates and Michaelis-Menten kinetics are invoked for RNA and protein degradation rates. In contrast to mass action kinetics, this simplifies the reaction kinetics and reduces the number of kinetic parameters. For both purposes of model accuracy as well as numerical efficacy, we modeled cytosolic and nuclear protein pools instead of introducing an explicit delay.

The following ordinary differential equations (ODEs) describe our mathematical model of the Arabidopsis circadian oscillator, interlocked with a hypothetical Y component and including PRR7 and PRR9. $c_i^{(j)}(t)$ describes the cellular concentration of the i^{th} gene. $i = L$ labels LHY, $i = T$ labels TOC1, $i = X$ labels X, $i = Y$ labels Y, $i = P7$ labels PRR7 and $i = P9$ labels PRR9. mRNA, protein in the cytoplasma or nucleus are distinguished by $j = m, c, n$ respectively.

$$\frac{dc_L^{(m)}}{dt} = ld \cdot q_1 \cdot c_P^{(n)} + \frac{n_1 \cdot c_X^{(n)^a}}{g_1^a + c_X^{(n)^a}} \cdot \frac{g_7^h}{g_7^h + c_{P7}^{(n)^h}} \cdot \frac{g_8^i}{g_8^i + c_{P9}^{(n)^i}} - \frac{m_1 \cdot c_L^{(m)}}{k_1 + c_L^{(m)}}$$

$$\frac{dc_L^{(c)}}{dt} = p_1 \cdot c_L^{(m)} - r_1 \cdot c_L^{(c)} + r_2 \cdot c_L^{(n)} - \frac{m_2 \cdot c_L^{(c)}}{k_2 + c_L^{(c)}}$$

$$\frac{dc_L^{(n)}}{dt} = r_1 \cdot c_L^{(c)} - r_2 \cdot c_L^{(n)} - \frac{m_3 \cdot c_L^{(n)}}{k_3 + c_L^{(n)}}$$

$$\frac{dc_T^{(m)}}{dt} = \frac{n_2 \cdot c_Y^{(n)^b}}{g_2^b + c_Y^{(n)^b}} \cdot \frac{g_3^c}{g_3^c + c_L^{(n)^c}} - \frac{m_4 \cdot c_T^{(m)}}{k_4 + c_T^{(m)}}$$

$$\frac{dc_T^{(c)}}{dt} = p_2 \cdot c_T^{(m)} - r_3 \cdot c_T^{(c)} + r_4 \cdot c_T^{(n)} - ((1 - ld) \cdot m_5 + m_6) \frac{c_T^{(c)}}{k_5 + c_T^{(c)}}$$

$$\frac{dc_T^{(n)}}{dt} = r_3 \cdot c_T^{(c)} - r_4 \cdot c_T^{(n)} - ((1 - ld) \cdot m_7 + m_8) \frac{c_T^{(n)}}{k_6 + c_T^{(n)}}$$

$$\frac{dc_X^{(m)}}{dt} = \frac{n_3 \cdot c_T^{(n)^d}}{g_4^d + c_T^{(n)^d}} - \frac{m_9 \cdot c_X^{(m)}}{k_7 + c_X^{(m)}}$$

$$\frac{dc_X^{(c)}}{dt} = p_3 \cdot c_X^{(m)} - r_5 \cdot c_X^{(c)} + r_6 \cdot c_X^{(n)} - \frac{m_{10} \cdot c_X^{(c)}}{k_8 + c_X^{(c)}}$$

$$\frac{dc_X^{(n)}}{dt} = r_5 \cdot c_X^{(c)} - r_6 \cdot c_X^{(n)} - \frac{m_{11} \cdot c_X^{(n)}}{k_9 + c_X^{(n)}}$$

$$\frac{dc_Y^{(m)}}{dt} = (ld \cdot q_2 \cdot c_P^{(n)} + \frac{(ld \cdot n_4 + n_5) \cdot g_5^e}{g_5^e + c_T^{(n)^e}}) \cdot \frac{g_6^f}{g_6^f + c_L^{(n)^i}} - \frac{m_{12} \cdot c_Y^{(m)}}{k_{10} + c_Y^{(m)}}$$

$$\frac{dc_Y^{(c)}}{dt} = p_4 \cdot c_Y^{(m)} - r_7 \cdot c_Y^{(c)} + r_8 \cdot c_Y^{(n)} - \frac{m_{13} \cdot c_Y^{(c)}}{k_{11} + c_Y^{(c)}}$$

$$\frac{dc_Y^{(n)}}{dt} = r_7 \cdot c_Y^{(c)} - r_8 \cdot c_Y^{(n)} - \frac{m_{14} \cdot c_Y^{(n)}}{k_{12} + c_Y^{(n)}}$$

$$\frac{dc_P^{(n)}}{dt} = (1 - ld) \cdot p_5 - \frac{m_{15} \cdot c_P^{(n)}}{k_{13} + c_P^{(n)}} - q_3 \cdot ld \cdot c_P^{(n)}$$

$$\frac{dc_{P7}^{(m)}}{dt} = \frac{n_6 \cdot c_L^{(n)^j}}{g_9^j + c_L^{(n)^j}} - \frac{m_{16} \cdot c_{P7}^{(m)}}{k_{14} + c_{P7}^{(m)}}$$

$$\frac{dc_{P7}^{(c)}}{dt} = p_6 \cdot c_{P7}^{(m)} - r_9 \cdot c_{P7}^{(c)} + r_{10} \cdot c_{P7}^{(n)} - \frac{m_{17} \cdot c_{P7}^{(c)}}{k_{15} + c_{P7}^{(c)}}$$

$$\frac{dc_{P7}^{(n)}}{dt} = r_9 \cdot c_{P7}^{(c)} - r_{10} \cdot c_{P7}^{(n)} - \frac{m_{18} \cdot c_{P7}^{(n)}}{k_{16} + c_{P7}^{(n)}}$$

$$\frac{dc_{P9}^{(m)}}{dt} = \frac{n_7 \cdot c_L^{(n)^k}}{g_{10}^k + c_L^{(n)^k}} - \frac{m_{19} \cdot c_{P9}^{(m)}}{k_{17} + c_{P9}^{(m)}}$$

$$\frac{dc_{P9}^{(c)}}{dt} = p_7 \cdot c_{P9}^{(m)} - r_{11} \cdot c_{P9}^{(c)} + r_{12} \cdot c_{P9}^{(n)} - \frac{m_{20} \cdot c_{P9}^{(c)}}{k_{18} + c_{P9}^{(c)}}$$

$$\frac{dc_{P9}^{(n)}}{dt} = r_{11} \cdot c_{P9}^{(c)} - r_{12} \cdot c_{P9}^{(n)} - \frac{m_{21} \cdot c_{P9}^{(n)}}{k_{19} + c_{P9}^{(n)}}$$

The constants in the reaction rates denote transcription (n_j, g_j), degradation (m_j, k_j), translation (p_j), transport between nucleus and cytoplasma (r_j) and acute light input (q_j). Hill coefficients (a, b,... j, k), describing the degree of cooperativity, are allowed to vary between $(1,5]$. We use non-cooperative binding terms for activators and/or repressors acting on the same gene, for example PRR7, PRR9 and X acting on LHY. Thus PRR7 and PRR9 repress LHY independently and do not enhance one another.

Light in the first basic model is included as suggested in (Locke *et al*, 2005). This approach is supported by the fact that it is experimentally possible to separate between sustained and prompt light induction mechanisms. CCA1/LHY, for example, are acutely induced by light after a dark period but light per se is not necessary for their expression. PRR9, in contrast, is induced by light after a dark period but light is also necessary for PRR9 expression. The different light input mechanisms are thus also separated in the mathematical model. Direct constant light during the day is mediated by the parameter *ld*: $ld = 1$ during the day and $ld = 0$ at night. Intermediate levels of *ld* can be associated with different light intensities, from darkness to a high luminescence level. The acute light effect had been modeled by introducing a hypothetical protein P. P is translated in the dark, active only in the presence of light and displays light dependent degradation. This creates a pulse of activity at the dark to light transition that is quickly degraded to zero. In the PRR7-PRR9-Y model, both LHY and Y are activated by an acute light pulse mediated by P.

ZTL is not explicitly modeled, but the effect of ZTL mediated degradation of TOC1 protein in darkness is implicitly included through darkness-dependent degradation of TOC1 (Locke *et al*, 2005; Mas *et al*, 2003).

1.2 The PRR7-PRR9light-Y model

In the next iteration, light input is introduced on PRR9 activation. With the exception of PRR9 RNA, all the other equations remain as described for the PRR7-PRR9-Y model. PRR9 RNA concentration is described by the following ODE:

$$\frac{dc_{P9}^{(m)}}{dt} = (ld \cdot q_4 \cdot c_P^{(n)} + ld \cdot n_7 + n_8)\frac{c_L^{(n)^k}}{g_{10}^k + c_L^{(n)^k}} - \frac{m_{19} \cdot c_{P9}^{(m)}}{´k_{17} + c_{P9}^{(m)}}$$

As PRR9 expression is dependent on LHY (Farre *et al*, 2005), all activation terms are coupled with LHY repression. With constant and acute light induction and light independent activation, the model is allowed three degrees of freedom for the transcription of PRR9 and the optimization emphasizes them according to the desired characteristics.

1.3 The PRR7-PRR9light-Y' model

For the last model iteration, acute light induction is removed from Y. Y RNA concentration is described by the following equation:

$$\frac{dc_Y^{(m)}}{dt} = (ld \cdot n_4 + n_5) \cdot \frac{g_5^e}{g_5^e + c_T^{(n)^e}} \cdot \frac{g_6^f}{g_6^f + c_L^{(n)^{fi}}} - \frac{m_{12} \cdot c_Y^{(m)}}{´k_{10} + c_Y^{(m)}}$$

The remaining ODEs remain unchanged.

1.4 Optimization process

Parameter estimation is a critical aspect of mathematical modeling as there is typically only sparse and noisy experimental data available. We therefore apply an Evolutionary Strategy (ES) to identify the 85/87 parameters characterizing the 19 ODEs underlying the model dynamics. This is a well-established optimization routine that is based on the idea of biological evolution and adaptation (Weicker, 2002).

The initial sets of oscillatory parameter ranges were identified using a random search in parameter space. Solutions were visualized in scatter plots, all possible parameters combinations plotted against each other. Rhythms were demanded for free-running conditions and light-dark cycles. However, the system displays oscillatory behavior for singular points

evenly distributed in parameter space. Ranges were therefore set to intuitively reasonable values:

$(\mathrm{n}_j, m_j, p_j, q_j) \in [0,10], (k_j) \in [0,50], (g_j, r_j) \in [0,30]$

To guarantee an oscillatory initialization, solutions from a random search of 10,000 parameters sets were used to start the optimization algorithm.

Like the Evolutionary Algorithms (EA), the ES uses the following operators, applied in order: mating selection, recombination, mutation fitness function evaluation and environmental selection. Each iteration loop is called a generation and the iteration is continued until a termination condition is met, in our case a predefined number of generations. The ES is distinguished from the EA by using strategy parameters to vary the step size or mutation strength during the routine, the so-called self-adaptation. We use a population based ES with 12 individuals and a selection strength of 5. The optimal parameter set found in the ES, is finally optimized by a Hill procedure, a local optimizer that tracks the minimum in the environment of the starting point (Mitchell, 1996).

The cost function that is minimized by the optimization procedure evaluates the simulation in comparison to the desired main characteristics of the clock. The equations are solved for 600 hours, calculating the cost terms based on the second 300 simulation hours. This assures that the system has reached its limit cycle and is not evaluated in a transitory state. Our cost function is almost exclusively based on wild-type characteristics, which allows us to use most phenotype characteristics for model evaluation. The following cost function was used for all three models:

$$f = \delta_\tau^{ld} + \delta_\phi^{ld} + \delta_\Delta^{ld} + \delta_\tau^{dd} + \delta_\tau^{ll} + \delta_{DM}^{ll} + \delta_\sigma^{ld} + \delta_{amp}$$

ld , *ll* and *dd* denote, if the error term was determined in light-dark cycles, constant light or constant darkness conditions. The error terms are weighted differently according to the importance of the described characteristic using a different weighting factor (k).

The first term

$$\delta_\tau^{ld} = \sum_{i=L,T,P7,P9} (24.0 - \tau_i)^2 / k_i, \quad \text{with } k_i = 0.25 \ \forall i$$

sums up deviations of LHY (L) , TOC1 (T), PRR7 (P7) and PRR9 (P9) RNA oscillations from a 24 hr period in light-dark cycles. The weighting is increased by factor 4, as the 24 hr

period is one of the two most important characteristics of the system in addition to the circadian peak, which means that an error of half an hour adds with value 1 to the cost function.

The second term

$$\delta_{\tau}^{ld} = \sum_{i=L,T,P7,P9} (\phi_i^{data} - \phi_i)^2 / k_i, \quad \text{with } k_L = 0.25,\ k_T = 0.25,\ k_{P7} = 0.5,\ k_{P9} = 0.5$$

fits the circadian peaks of LHY, TOC1, PRR7 and PRR9 RNA to the peak expression times observed in experimental data, measured from dawn in light-dark cycles. The target phases are $\phi_L = 1\text{h},\ \phi_T = 11\text{h},\ \phi_{P7} = 7\text{h},\ \phi_{P9} = 4\text{h}$ with stronger weighted LHY and TOC1 peaks.

The third term

$$\delta_{\Delta}^{ld} = \sum_{i=L,T} (\Delta_i^{data} - \Delta_i)^2 / k_i, \quad \text{with } k_i = 4\ \forall i$$

represents a measure of the RNA transcription and degradation rates, using the broadness of the RNA peak. Δ_i denotes the time difference between points of the RNA level at 2/3 rd's of its maximum value before and after the peak. The approximate experimental values are $\Delta_L^{data} = 2$ and $\Delta_L^{data} = 6$ showing that LHY has a sharp peak in light-dark cycles and TOC1 a broad one. However, this is just an approximate value from experimental data and we therefore reduce the weighting, relating a 2 hr error to a fitness value of 1.

The next two terms were used to fit the free running period under constant conditions. TOC1 RNA oscillations were used to determine the period and calculate the error terms.
The fourth term

$$\delta_{\tau}^{dd} = (24.0 - \tau_T)^2 / k, \quad \text{with } \begin{cases} k = 2 & \text{for } \tau_T > 24 \\ k = 0.5 & \text{for } \tau_T < 24 \end{cases}$$

was used to set the free running period under constant darkness in the PRR7-PRR9-Y and PRR7-PRR9light-Y models. These two first models were unable to achieve a longer period in constant dark than in constant light conditions, contrary to the biological observation. Convergence of the ES was achieved by selecting periods longer than 24 h in darkness and longer than 26h in constant light (see below). The PRR7-PRR9Light-Y' model corrected this discrepancy and the fourth error term was set as follows:

$$\delta_{\tau}^{dd} = (26.0 - \tau_T)^2 / k, \quad \text{with } \begin{cases} k = 0.5 \text{ for } \tau_T > 26 \\ k = 2 \quad \text{for } \tau_T < 26 \end{cases}$$

For the PRR7-PRR9-Y and PRR7-PRR9light-Y models the fifth term of the cost function was described as:

$$\delta_{\tau}^{ll} = (26.0 - \tau_T)^2 / k, \quad \text{with } \begin{cases} k = 0.5 \text{ for } \tau_T > 26 \\ k = 2 \quad \text{for } \tau_T < 26 \end{cases}$$

We assigned a higher weight to this error term if the period was longer than 26 hr but allowed the period to be shorter than 26h by choosing a weighting factor higher than one, if the period was smaller than 26 hr. In agreement with the experimental data the PRR7-PRR9Light-Y' model was able to generate oscillations with a longer period in constant light than in constant darkness, and the error term describing the period in constant light was set as follows:

$$\delta_{\tau}^{ll} = (25.0 - \tau_T)^2 / k, \quad \text{with } k = 1$$

The sixth term

$$\delta_{DM}^{ll} = (30.0 - \tau_L)^2 / k, \quad \text{with } \begin{cases} k = 25 \quad \text{for } \tau_L > 30 \\ k = 1 \quad \text{for } \tau_L < 30 \end{cases}$$

denotes the error for the free running period of LHY RNA oscillations in the long period phenotype of a *prr7prr9* double mutant. The experimentally determined period in constant light varies considerably between experiments but is consistently longer than 30 hr (Farre *et al*, 2005; Nakamichi *et al*, 2005; Salome & McClung, 2005). A 5 hr longer period therefore still contributes as 1, whereas a 1h smaller period is normally weighted.

The seventh term

$$\delta_{\sigma}^{ld} = \sum_{i=L,T,P7,P9} (\sigma_i^{\phi} + \sigma_i^{\max})$$

introduces standard deviations for the circadian peak and the maximal value of LHY, TOC1, PRR7 and PRR9 RNA levels in light-dark cycles. This term assures that the height and timing is continuous and does not vary significantly over the considered time period.

The last term

$$\delta_{amp} = \sum_{i=L,T,P7,P9} \frac{1}{(c_i^{\max,ld} - c_i^{\min.ld})^2} + \frac{0.5}{(c_T^{\max,ll} - c_T^{\min.ll})^2}$$

selects against small amplitude oscillations that might not be detectable or distinguished from arrhythmicity in experimentation. The difference between maximal and minimal RNA levels for LHY, TOC1, PRR7 and PRR9 in light-dark cycles and constant light are added. Differences under free-running conditions are weighted less than under entrainment.

References

Farre EM, Harmer SL, Harmon FG, Yanovsky MJ, Kay SA (2005) Overlapping and distinct roles of PRR7 and PRR9 in the Arabidopsis circadian clock. *Curr Biol*: **15** 47-54

Leloup JC, Goldbeter A (2003) Toward a detailed computational model for the mammalian circadian clock. *Proc Natl Acad Sci U S A*: **100** 7051-7056

Locke J, Southern M, Kozma-Bognr L, Hibberd V, Brown P, Turner M, Millar A (2005) Extension of a genetic network model by iterative experimentation and mathematical analysis

Molecular Systems Biology: **1** msb4100018-E4100011-msb4100018-E4100019

Mas P, Kim WY, Somers DE, Kay SA (2003) Targeted degradation of TOC1 by ZTL modulates circadian function in Arabidopsis thaliana. *Nature*: **426** 567-570

Mitchell M (1996) *An introduction to genetic algorithms*. MIT Press, Cambridge, MA, USA.

Nakamichi N, Kita M, Ito S, Yamashino T, Mizuno T (2005) PSEUDO-RESPONSE REGULATORS, PRR9, PRR7 and PRR5, together play essential roles close to the circadian clock of Arabidopsis thaliana. *Plant Cell Physiol*: **46** 686-698

Salome PA, McClung CR (2005) PSEUDO-RESPONSE REGULATOR 7 and 9 are partially redundant genes essential for the temperature responsiveness of the Arabidopsis circadian clock. *Plant Cell*: **17** 791-803

Ueda HR, Hagiwara M, Kitano H (2001) Robust oscillations within the interlocked feedback model of Drosophila circadian rhythm. *J Theor Biol*: **210** 401-406

Weicker K (2002) *Evolutionäre Algorithmen*. Teubner, Stuttgart.