1 Supplementary Text

1.1 Model One: LHY/CCA1-TOC1-X Network

The first network modelled (Fig 1) includes two extra components compared to the LHY/CCA1-TOC1 network modelled in (Locke et al., 2005); An extra gene, called $X$, was added to the pathway, and a constant light activation term was added to $TOC1$. As in previous clock models (Locke et al., 2005; Leloup et al., 1999; Leloup and Goldbeter, 2003; Ueda et al., 2001; Kurosawa and Iwasa, 2002; Kurosawa et al., 2002) Michealis-Menten kinetics were used to describe enzyme mediated degradation of proteins, and Hill functions were used to describe the transcriptional activation term of the mRNA for $LHY$ and $TOC1$. We use the cytosolic and nuclear pools of our model proteins to represent all the processes between the accumulation of an mRNA and the regulation of the next gene in the network by an active form of the cognate protein. There is some evidence that Michaelis-Menten kinetics may not be an accurate approximation of processes in higher organisms, and other clock models (Forger and Peskin, 2003) include additional processes, such as mRNA export from the nucleus, but there is currently no data to specify their dynamics in plants. The converse approach is to combine all intermediate steps as a time delay between the synthesis of RNA and active protein. This aids intuitive understanding by reducing the number of model components, so simplified versions of our models will be described elsewhere (JCL, MST and AJM, unpublished results). Given that the time delays can hamper subsequent mathematical analysis, we present the more detailed models here.

As $LHY$ and $CCA1$ are indistinguishable for our purposes, we retain only one gene, $LHY$, in our model. Quantitative differences in LHY and CCA1 regulation have sometimes been reported (Mizoguchi et al., 2002), though their qualitative behaviour is very similar. Differences in the response to LHY
and CCA1 overexpression might occur but current data (Fowler et al., 1999) include potentially confounding effects of overexpression level, genetic background, and developmental stage. Combining LHY and CCA1 genes removes 16 parameters and 3 equations from the model, which can be included when they are informed by further data.

We took the following as our mathematical model for the central circadian network: a LHY-TOC1-X feedback loop which involves the cellular concentrations $c_i^{(j)}(t)$ of the products of the $i^{th}$ gene ($i = L$ labels LHY, $i = T$ labels TOC1, $i = X$ labels X) where $j = m, c, n$ denotes that it is the corresponding mRNA, or protein in the cytoplasm or nucleus respectively.

\[
\begin{align*}
\frac{dc_i^{(m)}}{dt} &= q_1c_T^{(n)}\Theta(t) + \frac{n_1c_X^{(n)a}}{g_1^a + c_X^{(n)a}} - \frac{m_1c_L^{(m)}}{k_1 + c_L^{(m)}} \\
\frac{dc_L^{(c)}}{dt} &= p_1c_L^{(m)} - r_1c_L^{(c)} + r_2c_L^{(n)} - \frac{m_2c_L^{(c)}}{k_2 + c_L^{(c)}} \\
\frac{dc_L^{(n)}}{dt} &= r_1c_L^{(c)} - r_2c_L^{(n)} - \frac{m_3c_L^{(n)}}{k_3 + c_L^{(n)}} \\
\frac{dc_T^{(m)}}{dt} &= \frac{(n_2 + \Theta(t)n_3)g_2^b}{g_2^b + c_L^{(n)b}} - \frac{m_4c_T^{(m)}}{k_4 + c_T^{(m)}} \\
\frac{dc_T^{(c)}}{dt} &= p_2c_T^{(m)} - r_3c_T^{(c)} + r_4c_T^{(n)} - ((1 - \Theta(t))m_5 + m_6)c_T^{(c)} \\
\frac{dc_T^{(n)}}{dt} &= r_3c_T^{(c)} - r_4c_T^{(n)} - ((1 - \Theta(t))m_7 + m_8)c_T^{(n)} \\
\frac{dc_X^{(m)}}{dt} &= \frac{n_4c_T^{(n)}}{g_3^c + c_T^{(n)c}} - \frac{m_9c_X^{(m)}}{k_7 + c_X^{(m)}} \\
\frac{dc_X^{(c)}}{dt} &= p_3c_X^{(m)} - r_5c_X^{(c)} + r_6c_X^{(n)} - \frac{m_{10}c_X^{(c)}}{k_8 + c_X^{(c)}} \\
\frac{dc_X^{(n)}}{dt} &= r_5c_X^{(c)} - r_6c_X^{(n)} - \frac{m_{11}c_X^{(n)}}{k_9 + c_X^{(n)}} \\
\frac{dc_P^{(n)}}{dt} &= (1 - \Theta)p_4 - \frac{m_{12}c_P^{(n)}}{k_{10} + c_P^{(n)}} - q_2c_P^{(n)}
\end{align*}
\]
Here the various rate constants $n_j$, $g_j$ etc parameterise transcription ($n_j$, $g_j$), degradation ($m_j$, $k_j$), translation ($p_j$), and the nuclear ↔ cytoplasmic protein transport ($r_j$). There is evidence that LHY and CCA1 proteins bind as a dimer to the promoter of $TOC1$ (Daniel et al., 2004), and that there is only one active binding site on the $TOC1$ promoter (Alabadi et al., 2001), so the Hill coefficient for $TOC1$ inhibition by LHY protein, $b$, was set to 2. As there is no experimental evidence to support different values for the Hill coefficients $a$ and $c$ these were also set to 2. The acute light effect appears through the term $q_1 c_P(n) \Theta(t)$. Light is known to give an acute, transient activation response for expression of $LHY$ and $CCA1$ (Kim et al., 2003; Kaczorowski and Quail, 2003; Doyle et al., 2003). This was modelled as in (Locke et al., 2005), using a simple mechanism involving an interaction of a light sensitive protein $P$, with concentration $c_P(n)$ with the $LHY$ gene promoter. $\Theta = 1$ when light is present, 0 otherwise. The values of the four parameters that appear in the equation for $c_P(n)$ are chosen so as to give an acute light activation profile which is close to that observed in experiment. The essential features of Eq 10 are that $P$ is produced only when light is absent and is degraded strongly when light is present.

A constant light activation term was added to $X$ mRNA production, $\Theta(t) n_3$, and the effect of ZTL is modelled by the degradation terms for $TOC1$ in the cytoplasm and the nucleus, which are dark activated as suggested in (Mas et al., 2003).

1.2 Model Two: The interlocked feedback loop model

For the interlocked feedback loop model an extra loop is added to the network structure of the LHY/CCA1-TOC1-X circuit. Here a hypothetical gene $Y$ activates $TOC1$, and TOC1 then feeds back to repress $Y$. The light input into this loop is moved from $TOC1$ to $Y$ as there is no evidence of light activation of
TOC1 (Makino et al., 2002). Y was allowed to be both acutely light activated, in the manner of LHY, so to explain the extremely light sensitive response seen experimentally in the PTC of the cca1;lhy mutant (Figure 2), and to have a constant light activation term, to allow the clock to sense photoperiod.

We used a non-cooperative binding term for the activation and repression of TOC1 by Y and LHY respectively, and for the repression of Y by TOC1 and LHY. This means that LHY represses TOC1 transcription irrespective of the levels of Y.

The Hill coefficients in the equations were allowed to vary between 1 and 4 in the optimisation procedure. In order to obtain a compromise between flexibility and overall number of free parameters the Hill coefficients of activation and repression of TOC1 were set to the same value, i.e. b = c.

\[
\begin{align*}
\frac{dc_{L}^{(m)}}{dt} &= \Theta(t) q_1 c_P^{(n)} + \frac{n_1 c_X^{(n)a}}{g_1 + c_X^{(n)a}} - \frac{m_1 c_L^{(m)}}{k_1 + c_L} \\
\frac{dc_{L}^{(c)}}{dt} &= p_1 c_L^{(m)} - r_1 c_L^{(c)} + r_2 c_L^{(n)} - \frac{m_2 c_L^{(c)}}{k_2 + c_L} \\
\frac{dc_{L}^{(n)}}{dt} &= r_1 c_L^{(c)} - r_2 c_L^{(n)} - \frac{m_3 c_L^{(n)}}{k_3 + c_L} \\
\frac{dc_{T}^{(m)}}{dt} &= \left( \frac{n_2 c_Y^{(n)b}}{g_2 + c_Y^{(n)b}} \right) \left( \frac{g_5}{g_5 + c_X^{(n)a}} \right) - \frac{m_4 c_T^{(m)}}{k_4 + c_T} \\
\frac{dc_{T}^{(c)}}{dt} &= p_2 c_T^{(m)} - r_3 c_T^{(c)} + r_4 c_T^{(n)} - ((1 - \Theta(t))m_5 + m_6) \frac{c_T^{(c)}}{k_5 + c_T} \\
\frac{dc_{T}^{(n)}}{dt} &= r_3 c_T^{(c)} - r_4 c_T^{(n)} - ((1 - \Theta(t))m_7 + m_8) \frac{c_T^{(n)}}{k_6 + c_T} \\
\frac{dc_{X}^{(m)}}{dt} &= \frac{n_3 c_T^{(n)d}}{g_4 + c_T^{(n)d}} - \frac{m_9 c_X^{(m)}}{k_7 + c_X} \\
\frac{dc_{X}^{(c)}}{dt} &= p_3 c_X^{(m)} - r_5 c_X^{(c)} + r_6 c_X^{(n)} - \frac{m_{10} c_X^{(c)}}{k_8 + c_X}
\end{align*}
\]
\[
\frac{dc^{(n)}_X}{dt} = r_5 c^{(c)}_X - r_6 c^{(n)}_X - \frac{m_{11} c^{(n)}_X}{k_9 + c^{(n)}_X} \\
\frac{dc^{(m)}_Y}{dt} = \left( \Theta(t) q_2 c^{(n)}_P + \frac{(\Theta(t) n_4 + n_5) g_5^f}{g_5^f + c^{(n)}_T} \right) \left( \frac{g_6^f}{g_6^f + c^{(n)}_L} \right) - \frac{m_{12} c^{(m)}_Y}{k_{10} + c^{(m)}_Y} \\
\frac{dc^{(c)}_Y}{dt} = p_4 c^{(m)}_Y - r_7 c^{(c)}_Y + r_8 c^{(n)}_Y - \frac{m_{13} c^{(c)}_Y}{k_{11} + c^{(c)}_Y} \\
\frac{dc^{(n)}_Y}{dt} = r_7 c^{(c)}_Y - r_8 c^{(n)}_Y - \frac{m_{14} c^{(n)}_Y}{k_{12} + c^{(n)}_Y} \\
\frac{dc^{(n)}_P}{dt} = (1 - \Theta(t)) p_5 - \frac{m_{15} c^{(n)}_P}{k_{13} + c^{(n)}_P} - q_3 \Theta(t) c^{(n)}_P
\]

\(1.3\) \textit{Optimisation process}

We follow the optimisation technique as described (Locke et al., 2005). We summarise the technique here, but for a full description please refer to (Locke et al., 2005). There is significant noise in the experimental data for the mRNA levels of the key genes in the clock of Arabidopsis, and very little data for protein abundance, making a direct fit to the data difficult. This motivated us to construct an empirical cost function designed to give a value for the goodness of fit of our solution to qualitative features that are consistent in the data.

We constructed our cost function \(\Delta\) as a sum of terms that each quantify the agreement between our model and a qualitative experimental feature. Small values of the cost function correspond to a model (or set of parameter values) that give a good qualitative agreement with the corresponding experimental features. The weighting of each term in the cost function was chosen so that an acceptable error within the range of experimental variability would add on the order of 1 unit to the cost function. In order to evaluate the terms in the cost function we solved the equations numerically over 600 hours, 300 hours in 12 hour light 12 hour dark cycles (LD), followed by 300 hours in darkness.
(DD) (the first 200 hours of the LD cycles of each solution are discarded as transitory). In order to find a set of optimal solutions for each network studied, the cost function was calculated for a cross section of parameter space chosen using a Sobol quasi-random number generator (Press et al., 1996). The best fifty solutions were then put through a further optimisation step using a simulated annealing routine (Brooks and Morgan, 1995).

For the LHY/CCA1-TOC1-X network the cost function is essentially the same as that used in (Locke et al., 2005). We reproduce here the description of the cost function from Appendix A of that paper. The cost function is given by:

\[ \Delta = \delta_{\tau_d} + \delta_{\tau_d} + \delta_{\phi} + \delta_{cL} + \delta_{\text{size}}. \]  

(24)

Firstly,

\[ \delta_{\tau_d} = \sum_{i=L,T} \langle \frac{(24 - \tau_i^{(m)})^2}{0.15} \rangle_{ld} \]  

(25)

Is the summed error in the period, \( \tau \), for LHY (L) and TOC1 (T) mRNA levels in light:dark cycles (LD), where \( \langle \rangle_{ld} \) gives the average over the cycles between 200 < \( t < 300 \), and a “marginally acceptable” period difference of \( \approx 25 \text{mins} \) contributes O(1) to the cost function.

Secondly,

\[ \delta_{\tau_d} = \sum_{i=L,T} \langle \frac{(25 - \tau_i^{(m)})^2}{f} \rangle_{d} \]  

(26)

where the average of \( \langle \rangle_{d} \) is now over 300 < \( t < 600 \) (DD). The biological evidence strongly indicates that the free running period of the clock is greater than 24 (Millar et al., 1995), probably about 25, but we have less confidence in assigning a precise value hence we adopt values of \( f = 0.05 \) if \( \tau_i^{(m)} \leq 25 \) and \( f = 2 \) if \( \tau_i^{(m)} > 25 \).
Thirdly,

\[
\delta_{\phi} = \sum_{i=L,T} \left[ (\Delta \Phi_i)_{ld}^2 + \left( \frac{\sigma[c_i^{(m)}(t_{ip})]_{ld}}{0.05(c_i^{(m)}(t_{ip}))_{ld}} \right)^2 + \left( \frac{\sigma[\Delta \Phi_i]}{5/60} \right)^2 \right] + \delta_{ent}
\]  

The first term compares the mean difference in phase over the LD cycles, where \( \Delta \Phi_i = \Phi_i - \Phi_i \), \( \Phi_i \) is the phase (from dawn) of the RNA peak in the model and \( \Phi_L = 1h, \Phi_T = 11h \) are the target phases of the peaks in \( c_L^{(m)} \) and \( c_T^{(m)} \) respectively. We assume a cost that is O(1) for solutions that differ by an hour. The next two terms ascribe a cost of O(1) for limit cycle solutions in LD cycles whose peak heights are within 5 percent, and whose variation in peak times is 5 minutes. \( \sigma[]_{ld} \) is the standard deviation for the cycles in LD. The term \( \delta_{ent} \) checks that the solution is truly entrained to the light/dark cycle, i.e. is not oscillating with the correct phase simply because of the initial conditions chosen, as follows: The solution is rerun for 75 hours, taking the solution at 202 hours and shifting it back 3 hours, i.e initialising the \( t = 202 \) solution as the \( t = 199 \) solution. The new phase of the second peak is compared to the original phase of the second peak. If the phase difference is still near 3 hours, then the solution is too weakly entrained, and the solution is pathological. The LD cycles have failed to phase shift the response. Hence \( \delta_{ent} \) takes the form of \( \log(0.5)/\log(\delta \phi/3) \), where \( \delta \phi \) is the phase difference in hours between the shifted and original solution, and \( \delta \phi/3 \) is therefore the fraction of the imposed 3 hour phase shift remaining after 2 periods. The term \( \log(0.5) \) gives the acceptable remaining phase difference of 1.5 hours for the second cycle, which results in an O(1) contribution to the cost function.

Next,

\[
\delta_{size} = \sum_{i=L,T} \left( \frac{1}{(\Delta c_i^{(m)})_{ld}} \right)^2 + \left( \frac{\tau_o}{\tau_e} \right)^2
\]  

The first term costs for solutions in LD cycle with oscillation sizes, \( (\Delta c_i^{(m)}) = \)
\( c_i^{(m) \text{max}} - c_i^{(m) \text{min}} \), less than 1nm, and the second term checks that the oscillations do not decay too quickly when entering DD as follows: \( \tau_o \) is a decay constant over the 300 hours in DD, \( \tau_o = -300/\log((\Delta c_{T \text{ld}}^{(m)} - \Delta c_{T \text{d}}^{(m)})/\Delta c_{T \text{ld}}^{(m)}) \), and \( \tau_e \) gives the acceptable decay constant, that \( TOC1 \) oscillations size has dropped by 1/4 over 300 hours, \(-300/\log(0.75)\).

Finally,

\[
\delta_{cL} = \sum_{i=2,-2} \left\{ \left( \frac{2/3c_{L, (m)}(t_p)}{c_{L, (m)}(t_p) - c_{L, (m)}(t_p + i)} \right)^2 \right\}_{\text{ld}} + \ldots \\
\left\{ \left( \frac{0.05(c_{L, (m)}(t_p - 2) - c_{L, (m)}(t_m))}{c_{L, (m)}(t_m) - c_{L, (m)}(t_m + i)} \right)^2 \right\}_{\text{ld}} + 10 \left( \frac{\langle c_{L, (m)}(t_{pd}) \rangle_{\text{ld}}}{\langle c_{L, (m)}(t_{pl}) \rangle_{\text{ld}}} \right)^4
\]

The first term checks that the \( LHY \) mRNA expression profile has a sharp peak in LD cycles, with an \( O(1) \) contribution if \( LHY \)’s expression level has dropped by 2/3 of its oscillation size within 2 hours before and after its peak of expression. The second term checks that \( LHY \) mRNA expression has a broad minimum, with an \( O(1) \) contribution if 2 hours before and after the minimum point \( LHY \)’s expression has only increased to 5 percent of the level 2 hours before \( LHY \)’s peak. The last term checks that the peak of \( LHY \) mRNA expression drops from LD into DD, as it loses its light activation. An additional cost term was added to shape fit the \( TOC1 \) mRNA profile, as suggested in (Locke et al., 2005) in order to stop spurious solutions where \( TOC1 \) mRNA expression is saturated, but this term was found to be unnecessary for optimising a network where \( TOC1 \) is light activated, and was not used for further optimisations.

Throughout the implementation the cost function was “capped” at \( \Delta_{\text{max}} = 10^4 \), such that \( \Delta \rightarrow \text{Min}(10^4, \Delta) \).

As discussed in the computational methods, we added new terms to the cost function in order to optimise the interlocked feedback loop model to both WT
and *cca1;lhy* mutant data. The equations were re-solved with the translation rate of *LHY* reduced to a thousandth of its WT value in order to simulate the double mutant. The cost function now becomes

\[ \Delta = \delta_{\tau_d} + \delta_{\tau_r} + \delta_{\phi} + \delta_{c_L} + \delta_{\text{size}} + \delta_{\phi_d} + \delta_{c_{L_d}} + \delta_{c_{L_r}} + \delta_{\phi_{Y}} + \delta_{\phi_{cy}} + \delta_{\text{size}} \]  

(30)

where the label \((dm)\) denotes the cost function for the *cca1;lhy* double mutant. One new WT cost function term \(\delta_{\phi_d}\) added represents a minor change to constrain an appropriate phase difference between the peak levels of *LHY* and *TOC1* mRNA, \(\Delta \Phi_d = \phi_T - \phi_L\) (modulo half the period), with a characteristic prefactor of 10h. This term makes no discernable difference to the cost function when applied to the optimised one loop models. See term below:

\[ \delta_{\phi_d} = \left( \frac{10}{\Delta \Phi_d} \right)^2 \]  

(31)

\(\delta_{\text{size}}\) was also altered slightly in order to ensure both *LHY* mRNA and *TOC1* mRNA oscillations do not decay too quickly when entering DD. This is necessary as in the interlocked feedback loop model *TOC1* mRNA levels can oscillate through *TOC1*’s feedback loop with \(Y\) whilst *LHY* mRNA levels are arrhythmic. \(\delta_{\text{size}}\) becomes

\[ \delta_{\text{size}} = \sum_{i=L,T} \left[ \left( \frac{1}{\langle \Delta e_{i}^{(m)} \rangle_{ld}} \right)^2 + \left( \frac{\tau_o}{\tau_e} \right)^2 \right]. \]  

(32)

The first term remains the same, and the second term is now summed over *LHY* \((L)\) and *TOC1* \((T)\). All the other WT cost function terms remain the same as for the one loop model optimisation.

Using the same methodology as for the WT terms, we define below the new
double mutant terms of the cost function. The first new term,

$$\delta_{\tau_{ld}}^{dm} = \sum_{i=Y,T} \langle (24 - \tau_i^{(m)})^2/0.15 \rangle_{ld}$$  \hspace{1cm} (33)$$

is the summed error in the period, \(\tau\), for \(Y\) (Y) and \(TOC1\) (T) mRNA levels in LD cycles.

We penalise solutions with a period of \(TOC1\) greater than 18 hours in the dark. \(\delta_{\tau_{ld}}^{(m)} = 0\) if the period is less than 18 hours, otherwise:

$$\delta_{\tau_{ld}}^{dm} = \langle (18 - \tau_T^{(m)})^2/0.1 \rangle_d$$  \hspace{1cm} (34)$$

Next,

$$\delta_{\phi}^{dm} = \sum_{i=Y,T} \left[ \langle \Delta \Phi_i^2 \rangle_{ld} + (\sigma[\Delta \Phi_i])^2 \right]$$  \hspace{1cm} (35)$$

The first term compares the mean difference in phase over the LD cycles, where \(\Delta \Phi_i = \tilde{\phi}_i - \phi_i\), \(\phi_i\) is the phase (from dawn) of the RNA peak in the model and \(\tilde{\phi}_Y = 1h, \tilde{\phi}_T = 5h\) are the target phases of the peaks in \(c_Y^{(m)}\) and \(c_T^{(m)}\) respectively. The second term describes a cost of \(O(1)\) for solutions whose variations in peak phases are \(1h\). Next,

$$\delta_{\text{size}}^{dm} = \sum_{i=Y,T} \left( \frac{1}{\langle \Delta c_i^{(m)} \rangle_{ld}} \right)^2$$  \hspace{1cm} (36)$$

This term costs for solutions in LD cycle with oscillation sizes, \(\langle \Delta c_i^{(m)} \rangle = c_i^{(m)\text{max}} - c_i^{(m)\text{min}}\), less than \(1\text{nm}\). Finally,

$$\delta_{c_Y}^{dm} = \sum_{i=2,-2} \langle \left( \frac{2/3c_Y^{(m)}(t_p)}{c_Y^{(m)}(t_p) - c_Y^{(m)}(t_p + i)} \right)^2 \rangle_{ld}$$  \hspace{1cm} (37)$$

The first term checks that the \(Y\) mRNA expression profile has a sharp peak
in LD cycles, with an O(1) contribution if Y’s expression level has dropped by 2/3 of its oscillation size within 2 hours before and after its peak of expression. As for the single loop optimisations, throughout the implementation the cost function was “capped” at $\Delta_{\text{max}} = 10^4$, such that $\Delta \rightarrow \text{Min}(10^4, \Delta)$. The sum of the double mutant cost function terms was also capped at $10^3$.

References


