Supporting Text

A More Realistic Model of Circadian Rhythms in Drosophila

Currently about a dozen proteins have been identified as key players in generating circadian rhythms in *Drosophila melanogaster*. Experts believe that two interlocked feedback loops create a robust biochemical oscillator with a period close to 24 h. One loop involves negative feedback, whereby PERIOD (PER) and TIMELESS (TIM) proteins (1-7) enter the nucleus, interact with transcription factors, dCLOCK (dCLK) and CYCLE (CYC) (8), and inhibit their own transcription (9). PER and TIM proteins are synthesized in the cytoplasm, where they are phosphorylated by DOUBLETIME (DBT) and SHAGGY (SGG), respectively (10-13). Phosphorylated PER monomers are rapidly degraded, but PER in complex with TIM is more stable. PER also forms homodimers in low abundance, and homodimers may also be stable (14). PER and TIM enter the nucleus with independent transport rate (15), but TIM facilitates the nuclear transport of PER (16). PER enters the nucleus in multimeric complexes with TIM and DBT. Once in the nucleus, TIM appears to dissociate from the complex, and then DBT participates in intranuclear turnover of PER. In the nucleus, PER interacts with dCLK and CYC via PAS domains on all three proteins, and inhibits the function of dCLK and CYC to promote transcription of *per* and *tim* genes. Light interacts with this mechanism by down regulating a photoreceptor protein, CRYPTOCHROME (CRY), whose role is to bind to and stabilize TIM (17-19). Light-induced degradation of TIM creates either phase delays or advances, depending on the time of the light pulse.

A second feedback loop also involves dCLK and CYC, which activate transcription of PAR domain protein 1 (PDP1) and VRILLE (VRI) as well as PER and TIM. dCLK and CYC are positively regulated by PDP1 (20) and negatively regulated by VRI (20, 21). The interlocked feedback loops, through PDP1, VRI and PER, introduce complicated dynamical behaviors in this biochemical control system.

Based on these facts, we propose a new model of the CR mechanism that retains the autocatalytic accumulation of PER which is a defining feature of the simple model (Box 1, main text). The new, more realistic model, with 12 variables and 31 parameters, is related to recent studies by Leloup and Goldbeter (22, 23), and Forger and Peskin (24). Several assumptions of the model are noteworthy. First of all, we assume that it is a tetrameric form

of PER/TIM complex that enters the nucleus. This assumption preserves the property of autocatalytic accumulation of PER. (Alternative assumptions, which preserve the autocatalytic accumulation of PER, are possible, e.g. PER homodimers interact with TIM to produce PER/TIM heterodimers plus PER monomer.) Second, we assume that dCLK-regulated transcription is inhibited by nuclear PER only, because it has been shown that PER without a cytoplasmic localization domain (CLD) performs its function in the absence of TIM. At present, we do not include the dynamics of VRI and PDP1, but dCLK's self-inhibition via VRI is included. Also, the effect of light can be simulated by the degradation of TIM, thereby excluding another variable, CRY. A schematic diagram of the more complete model is represented in Fig. 5, and a set of rate constants is proposed in Table 2.

Bifurcation analysis of the larger model reveals many qualitative similarities to the simple model. For example, Fig. 1A of the main text and Fig. 6 show that both models exhibit a region of bistability with a SNIC bifurcation to large-amplitude, long-period, stable limit cycle oscillations. This behavior seems to be a common property of regulatory systems with both positive and negative feedbacks. As for the simple model in the text, the expanded model presented here, with a resetting mechanism for k_{in}, exhibits cycles (dash-dot line in Fig. 6) with a period, $T = \mu^{-1} \ln \sigma = 24$ h, that is largely independent of the kinetic constants of the reaction network. A detailed analysis of the expanded model is left for future work.

Kinetic Equations of Expanded Model

a) mRNAs of *per*, *tim*, and *dClk*

$$\frac{dP_m}{dt} = v_{mp} \cdot \frac{F^m}{K_1^m + F^m} - k_{dmp} \cdot P_m$$

$$\frac{dT_m}{dt} = v_{mt} \cdot \frac{F^m}{K_1^m + F^m} - k_{dmt} \cdot T_m$$

$$\frac{dC_m}{dt} = v_{mc} \cdot \frac{1}{1 + k \cdot F^n} - k_{dmc} \cdot C_m$$

b) PER monomer, homodimer, and nuclear PER

$$\frac{dP_1}{dt} = v_p \cdot P_m - k_{p3} \cdot P_1 - \frac{k_{p1} \cdot P_1}{J_p + P_{tot}} + 2k_{p3} \cdot P_2 + 2k_{p3} \cdot P_2T_1 + \frac{2k_{p2} \cdot P_2}{J_p + P_{tot}} + \frac{2k_{p2} \cdot P_2T_1}{J_p + P_{tot}}$$

$$+ \frac{2k_{p2} \cdot P_2T_2}{J_p + P_{tot}} - 2k_{app} \cdot P_1^2 + 2k_{dpp} \cdot P_2 + 2k_{p3} \cdot P_2T_2 + k_{out} \cdot P_N$$

$$\frac{dP_2}{dt} = k_{app} \cdot P_1^2 - k_{dpp} \cdot P_2 - 2k_{p3} \cdot P_2 - \frac{2k_{p2} \cdot P_2}{J_p + P_{tot}} - 2k_{apt} \cdot P_2 \cdot T_1 + k_{dpt} \cdot P_2T_1 + k_{t3} \cdot P_2T_1$$

$$\frac{dP_N}{dt} = 2k_{in} \cdot P_2T_2 - k_{out} \cdot P_N - k_{aitf} \cdot F \cdot P_N + k_{ditf} \cdot F_{inac} + k_{dc} \cdot F_{inac} - k_{p3} \cdot P_N - \frac{k_{p1} \cdot P_N}{J_p + P_{tot}}$$

c) TIM monomer

$$\frac{dT_1}{dt} = v_t \cdot T_m - k_{t3} \cdot T_1 + 2k_{p3} \cdot P_2T_1 + \frac{2k_{p2} \cdot P_2T_1}{J_p + P_{tot}} + \frac{4k_{p2} \cdot P_2T_2}{J_p + P_{tot}} + 2k_{apt} \cdot P_2 \cdot T_1 + k_{dpt} \cdot P_2T_1$$

$$-k_{apt} \cdot P_2T_1 \cdot T_1 + 2k_{dpt} \cdot P_2T_2 + 4k_{p3} \cdot P_2T_2 + 2k_{in} \cdot P_2T_2$$

d) dCLK monomer

$$\frac{dC}{dt} = v_c \cdot C_m - k_{dC} \cdot C - k_{acc} \cdot C \cdot Y + k_{dcc} \cdot F$$

e) Complex: PER/TIM trimer, PER/TIM tetramer, transcription factor (dCLK and CYC), and inactive transcription factor (dCLK, CYC, and nuclear PER)

$$\frac{dP_2T_1}{dt} = 2k_{apt} \cdot P_2 \cdot T_1 - k_{dpt} \cdot P_2T_1 - k_{t3} \cdot P_2T_1 - 2k_{p3} \cdot P_2T_1 - \frac{2k_{p2} \cdot P_2T_1}{J_p + P_{tot}} - k_{apt} \cdot P_2T_1 \cdot T_1 + 2k_{dpt} \cdot P_2T_2 + 2k_{t3} \cdot P_2T_2$$

$$\frac{dP_2T_2}{dt} = k_{apt} \cdot P_2T_1 \cdot T_1 - 2k_{dpt} \cdot P_2T_2 - 2k_{t3} \cdot P_2T_2 - 2k_{p3} \cdot P_2T_2 - \frac{2k_{p2} \cdot P_2T_2}{J_p + P_{tot}} - k_{in} \cdot P_2T_2$$

$$\frac{dF}{dt} = k_{acc} \cdot C \cdot Y - k_{dcc} \cdot F - k_{aitf} \cdot F \cdot P_N + k_{ditf} \cdot F_{inac} - k_{dc}F + k_{p3} \cdot F_{inac} + \frac{k_{p2} \cdot F_{inac}}{J_p + P_{tot}}$$

$$\frac{dF_{inac}}{dt} = k_{aitf} \cdot F \cdot P_N - k_{ditf} \cdot F_{inac} - k_{p3} \cdot F_{inac} - k_{dc} \cdot F_{inac} - \frac{k_{p2} \cdot F_{inac}}{J_p + P_{tot}}$$

f) Total Concentrations

$$P_{tot} = P_1 + 2P_2 + 2P_2T_1 + 2P_2T_2 + P_N + F_{inac}$$

$$T_{tot} = T_1 + P_2T_1 + 2P_2T_2$$

$$C_{tot} = C + F + F_{inac}$$

$$Y_{tot} = Y + F + F_{inac}$$

Robustness Analysis

For a fair comparison of the robustness of CR period of the three models RS, LG and TH described in the main text, we first need to equip LG and TH with compensation relations analogous to $\sigma = e^{-24\cdot\mu}$ (a constraint on the RS model). To this end, we first rank for each model the sensitivity coefficients, $\frac{\partial \log T}{\partial \log p_i}$, in order to identify the parameters that most influence the period (see Fig. 7). It is worthwhile to note from Fig. 7 that, while RS concentrates virtually all of the sensitivity into just two parameters (by design), both LG and TH have a more even distribution. The fact that LG has a more even distribution than TH is significant in the following results (as it will fair worse in both tests conducted). From these rankings, we choose as compensatory parameters: v_m and k_s for LG and J_p and k_m for TH.

We then 'continue' the limit cycle period in the two chosen parameters, generating a locus of (p_i, p_j) pairs that maintain $T = \text{constant} \cong 24$ h. From this parametric curve, we derive a compensation relationship for p_i as a function of p_j. For LG, we find that $v_m = 0.43\cdot(k_s - 0.3)^{-0.16}$ gives constant period close to 24 h, and for TH, $J_p = 3.2\times10^{-5}\cdot k_m^{-3.2}$. These functions fix the period in two parameters to within half-a-percent (data not shown), so we may confidently begin our analysis with the three models on equal footing.

Now, Test A aims to assess how the period of a model co-varies (measured by the coefficient of variation of the distribution of resulting periods, CV) in response to varying strengths, σ_p, of vector perturbations applied to the basal parameter set, simulating variations among individuals within a population. An abbreviated version of the algorithm is

```
// initialize
b = < bj >, the basal parameter vector of length n

// loop through a number of perturbation strengths, σ
for each σ = 0.01, 0.02, …, 0.4

      // collect 10000 randomly perturbed individuals and their periods
      for i = 1 to 10000

            1) draw n random numbers from N(1, σ) to get r = < rj >
            2) set the individual's parameters, p = < rj * bj >
            3) apply the compensation relationship
            4) compute period and save
      end

      // calculate statistic; (CV, σ) is what we plot
      compute CV from the mean and variance of the collected periods
end
```

Because certain perturbations move us outside of the domain of oscillation, especially for higher perturbation strengths, we threw out many samples in order to get 10^4 valid period values. The number of samples thrown out indicates how likely the oscillation is to be preserved under given perturbation strength, providing another statistic of robustness (unrelated to temperature compensation).

Test B is meant to simulate a sample of single genetic mutations and measure the ability of the affected individuals to temperature compensate. Recall that $k_i = k_i^0 e^{-\Delta G_i^\dagger / R\theta}$, where $\Delta G_i^\dagger = \Delta H_i^\dagger - \theta \cdot \Delta S_i^\dagger$ is the free energy of activation of reaction *i*. The enthalpy of activation, usually called the 'activation energy,' $E_{a,i} = \Delta H_i^\dagger$, is always > 0, and the entropy of activation, $\Delta S_i^\dagger$, is always < 0. We write the rate constant expression in dimensionless form: $k_i = k_i^o e^{s_i^\dagger} e^{-\varepsilon_i^\dagger \cdot 298/\theta}$ where $s_i^\dagger = \Delta S_i^\dagger / R$ and $\varepsilon_i^\dagger = \Delta H_i^\dagger / R \cdot 298$. Since values for $s_i^\dagger$ and $\varepsilon_i^\dagger$ aren't experimentally known, we suppose (not unreasonably) that $s_i^\dagger = -\varepsilon_i^\dagger$ and choose $\varepsilon_i^\dagger$ values in the range [3, 20] in order to obtain good temperature compensation for each model (see Table 3). $\Delta H_i^\dagger$ and $\Delta S_i^\dagger$ are dependent on the biophysical properties of the enzyme that catalyzes reaction *i*; hence, their values are genetically determined and subject to random variations by single gene mutations.

Binding constants are treated similarly. According to the Gibbs equation, $K_i = e^{s_i^o} e^{-\varepsilon_i^o \cdot 298/\theta}$, where s_i^o and ε_i^o are dimensionless standard entropies and enthalpies of the binding reaction. These values are also subject to modification by mutations that change the binding properties of specific proteins.

With this in mind, we generate 'mutant' individuals by perturbing s_i and ε_i for a particular choice of *i*. (Because the models have already been temperature compensated by our choices of ε_i's, we do not constrain the parameters any further by the compensation relations of Test A.) For each mutant individual, we calculate the maximum excursion of period, $\Delta T = T_{\max} - T_{\min}$, over a 10-degree temperature range ($293 < \theta < 303$) as a measure of the mutant's ability to temperature compensate[*]. The algorithm is:

[*] Although the variation of the period over the stated temperature range may not be strictly increasing or decreasing and may in fact be highly non-linear, this measure was considered adequate given a sampling of perturbations. A more accurate statistic would depend on the shape of the curve in addition to the maximum and minimum values.

```
// initialize
a = < ε_i >, the basal vector of ε_i values, of length n
b = < s_i >, the basal vector of s_i values, of length n

// loop through a number of perturbation strengths, σ
for each σ = 0.01, 0.02, …, 0.4

      // for each parameter (that can be mutated)
      for i = 1 to n

            // collect 10 random samples of individuals with p_i-mutations
            // and their periods in the temperature range 293-303 K
            for j = 1 to 10

                  1) draw r_1 and r_2 from N(1,σ)
                  2) set the individual's ε = a, s = b
                  3) perturb ε_i = a_i * r_1, s_i = b_i * r_2
                  4) compute periods, T_k, for θ = 293, 294, …, 303 K
                  5) compute temp. comp. stat, ΔT_i,j = max{T_k} - min{T_k},
            end


            // calculate intermediate statistic
            compute ΔT_i = arithmetic mean of ΔT_i,j from p_i-mutation stat
      end

      // calculate final statistic; (ΔT, σ) is what we plot
      compute ΔT as the harmonic mean of ΔT_i
end
```

References

1. Darlington, T. K., Wager-Smith, K., Ceriani, M. F., Staknis, D. S., Gekakis, N., Steeves, T. D. L., Weitz, C. J., Takahashi, J. S. & Kay, S. A. (1998) *Science* **280,** 1599-1603.
2. Gekakis, N., Saez, L., Delahaye-Brown, A., Myers, M. P., Sehgal, A., Young, M. W. & Weitz, C. J. (1995) *Science* **270,** 811-814.
3. Myers, M. P., Wager-Smith, K., Wesley, C. S., Sehgal, A. & Young, M. W. (1995) *Science* **270,** 805-808.
4. Sehgal, A., Price, J. L., Man, B. & Young, M. W. (1994) *Science* **263,** 1603-1606.
5. Sehgal, A., Rothenfluh-Hilfiker, A., Hunter-Ensor, M., Chen, Y., Myers, M. P. & Young, M. W. (1995) *Science* **270,** 808-810.
6. Vosshall, L. B., Price, J. L., Sehgal, A., Saez, L. & Young, M. W. (1994) *Science* **263,** 1606-1609.
7. Zeng, H., Hardin, P. E. & Rosbash, M. (1994) *EMBO* **13,** 3590-3598.
8. Rothenfluh, A., Young, M. W. & Saez, L. (2000) *Neuron* **26,** 505-514.
9. Lee, C., Bae, K. & Edery, I. (1999) *Molec. Cell. Biol.* **19,** 5316-5325.
10. Kloss, B., Price, J. L., Saez, L., Blau, J., Rothenfluh, A., Wesley, C. S. & Young, M. W. (1998) *Cell* **94,** 97-107.
11. Kloss, B., Rothenfluh, A., Young, M. W. & Saez, L. (2001) *Neuron* **30,** 699-706.
12. Price, J. L., Blau, J., Rothenfluh, A. & Young, M. W. (1998) *Cell* **94,** 83-95.
13. Martinek, S., Inonog, S., Manoukian, A. S. & Young, M. W. (2001) *Cell* **105,** 769-779.
14. Huang, Z. J., Curtin, K. D. & Rosbash, M. (1995) *Science* **267,** 1169-1172.
15. Meyer, P., Saez, L. & Young, M. W. (2006) *Science* **311,** 226-229.

16. Cyran, S. A., Yiannoulos, G., Buchsbaum, A. M., Saez, L., Young, M. W. & Blau, J. (2005) *J. Neurosci.* **25,** 5430.
17. Ceriani, M. F., Darlington, T. K., Staknis, D., Mas, P., Petti, A. A., Weitz, C. J. & Kay, S. A. (1999) *Science* **285,** 553-556.
18. Emery, P., So, W. V., Kaneko, M., Hall, J. C. & Rosbash, M. (1998) *Cell* **95,** 669-679.
19. Myers, M. P., Wager-Smith, K., Rothenfluh-Hilfiker, A. & Young, M. W. (1996) *Science* **271,** 1736.
20. Cyran, S. A., Buchsbaum, A. M., Reddy, K. L., Lin, M., Glossop, N. R. J., Hardin, P. E., Young, M. W., Storti, R. W. & Blau, J. (2003) *Cell* **112,** 329-341.
21. Blau, J. & Young, M. W. (1999) *Cell* **99,** 661-671.
22. Leloup, J. & Goldbeter, A. (1998) *J. Biol. Rhythms* **13,** 70-87.
23. Leloup, J. & Goldbeter, A. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 7051-7056.
24. Forger, D. & Peskin, C. S. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 14806-14811.