Supporting Information for *The Presence of Nuclear Cactus in the Early Drosophila Embryo May Extend the Dynamic Range of the Dorsal Gradient*

Michael D. O’Connell¹, Gregory T. Reeves¹

¹North Carolina State University Department of Chemical and Biomolecular Engineering, Raleigh NC.

S1 Text. Detailed description of model formulation

Original model formulation

\[ \frac{d(V_{n}C_{dl,n}^{h})}{dT} = A_{n}(k_{in}C_{dl,c}^{h} - k_{out}C_{dl,n}^{h}) \] (1)

\[ \frac{d(V_{C}C_{dl,c}^{h})}{dT} = A_{m}\Gamma_{dl}(C_{dl,c}^{h-1} - 2C_{dl,c}^{h} + C_{dl,c}^{h+1}) + V_{C}(k_{D}C_{dc,c}^{h} - k_{b}C_{dl,c}^{h}C_{Cact,c}^{h}) - A_{n}(k_{in}C_{dl,c}^{h} - k_{out}C_{dl,n}^{h}) \] (2)

\[ \frac{d(V_{C}C_{dc,c}^{h})}{dT} = A_{m}\Gamma_{dc}(C_{dc,c}^{h-1} - 2C_{dc,c}^{h} + C_{dc,c}^{h+1}) - V_{C}(k_{D}C_{dc,c}^{h} - k_{b}C_{dl,c}^{h}C_{Cact,c}^{h}) \] (3)

\[ \frac{d(V_{C}C_{Cact,c}^{h})}{dT} = A_{m}\Gamma_{Cact}(C_{Cact,c}^{h-1} - 2C_{Cact,c}^{h} + C_{Cact,c}^{h+1}) + V_{C}(k_{D}C_{dc,c}^{h} - k_{b}C_{dl,c}^{h}C_{Cact,c}^{h} - k_{Deg}C_{Cact,c}^{h}) + P_{Cact} \] (4)

where \( k_{D} = A\exp\left(-\frac{1}{2}\left(\frac{\phi}{2}\right)^{2}\right) + B. \)

The original model formulation features four equations representing nuclear dl, cytoplasmic dl, cytoplasmic dl/Cact complex and cytoplasmic Cact (Eqns 1-4, respectively). In these equations, \( C_{s}^{h} \) is the concentration of species \( s \) within nuclear/cytoplasmic compartment \( h \), where \( s = \{dl,n; dl,c; dc,c; Cact,c\}. \)

The three cytoplasmic species are exchanged between neighboring compartments at a rate \( \Gamma_{s} \) in proportion to the surface area of the two touching faces, \( A_{m} \) (transport in the anterior-posterior direction is ignored, as the dl gradient is assumed to be approximately unchanging along this direction). Free dl is exchanged between the nucleus and cytoplasm at rates \( k_{out} \& k_{in} \) in proportion to the surface area of each nucleus, \( A_{n} \). Cytoplasmic dl and Cact bind to form dl/Cact complex at a rate \( k_{b} \), and the unbinding rate of dl/Cact
complex is B. Toll-mediated degradation of dl/Cact complex is governed by a Gaussian function with parameters $A$ and $\phi$. Finally, Cact is produced uniformly at a rate $P_{Cact}$ and degraded at a rate $k_{Deg}$.

In this model formulation, the nuclei are assumed well-mixed, and are initially empty at the start of each interphase. During interphase, the system is governed by all four equations. At the end of interphase, the nuclei dissolve and nuclear dl is mixed with cytoplasmic dl within each compartment. During mitosis, the nuclei are undefined and the system is governed by the three cytoplasmic equations, minus the terms for nuclear import/export. At the end of mitosis, the number of nuclei/compartments increases instantly, and the total concentration of each protein is interpolated into the new compartments. The concentration of nuclear dl starts at zero once again, and the next interphase begins. (See also [1].)

**Updated Model Formulation**

To allow for the presence of dl protein within nuclei at the start of interphase, we updated the model by removing the assumption that nuclei begin interphase empty. This updated model formulation is governed by six differential equations, now including equations for nuclear dl/Cact complex and nuclear Cact. At the beginning of each interphase, nuclear envelopes encapsulate a volume of cytoplasm as they reform. The concentration of each species is thus the same between the nucleus and cytoplasm before nuclear import/export dynamics take over.

The updated model consists of the following six differential equations:

$$\frac{d}{dT} \left( V_n C_{dl,n}^h \right) = A_n \left( k_{in,dl} C_{dl,c}^h - k_{out,dl} C_{dl,n}^h \right) - V_n \left( k_{b} C_{dl,n}^h C_{Cact,n}^h - k_{D0} C_{dc,n}^h \right)$$  \hspace{1cm} (5)  

$$\frac{d}{dT} \left( V_c C_{dl,c}^h \right) = A_m \Gamma_{dl} \left( C_{dl,c}^{h-1} - 2 C_{dl,c}^h + C_{dl,c}^{h+1} \right) + V_c \left( k_D C_{dc,c}^h - k_b C_{dl,c}^h C_{Cact,c}^h \right) - A_n \left( k_{in,dl} C_{dl,c}^h - k_{out,dl} C_{dl,n}^h \right)$$  \hspace{1cm} (6)  

$$\frac{d}{dT} \left( V_n C_{dc,n}^h \right) = A_n \left( k_{in,dc} C_{dc,e}^h - k_{out,dc} C_{dc,n}^h \right) + V_n \left( k_{b} C_{dc,n}^h C_{Cact,n}^h - k_{D0} C_{dc,n}^h \right)$$  \hspace{1cm} (7)  

$$\frac{d}{dT} \left( V_c C_{dc,c}^h \right) = A_m \Gamma_{dc} \left( C_{dc,c}^{h-1} - 2 C_{dc,c}^h + C_{dc,c}^{h+1} \right) - V_c \left( k_D C_{dc,c}^h - k_b C_{dc,e}^h C_{Cact,c}^h \right) - A_n \left( k_{in,dc} C_{dc,e}^h - k_{out,dc} C_{dc,n}^h \right)$$  \hspace{1cm} (8)  

$$\frac{d}{dT} \left( V_n C_{Cact,n}^h \right) = A_n \left( k_{in,Cact} C_{Cact,c}^h - k_{out,Cact} C_{Cact,n}^h \right) - V_n \left( k_b C_{Cact,n}^h C_{Cact,n}^h - k_{D0} C_{Cact,n}^h \right)$$  \hspace{1cm} (9)  

$$\frac{d}{dT} \left( V_c C_{Cact,c}^h \right) = A_m \Gamma_{Cact} \left( C_{Cact,c}^{h-1} - 2 C_{Cact,c}^h + C_{Cact,c}^{h+1} \right) + V_c \left( k_D C_{Cact,c}^h - k_b C_{Cact,c}^h C_{Cact,c}^h \right) - k_{Deg} C_{Cact,c}^h$$  \hspace{1cm} (10)  

where $k_D = k_{D0} \exp \left( -\frac{1}{2} \left( \frac{\phi}{\sigma} \right)^2 \right) + k_{D0}$.  

}\n
Non-dimensionalizing, we have:

\[
\begin{align*}
\frac{d (\bar{V}_n U_n^h)_{\text{nuc}}}{dt} &= \bar{A}_n (\zeta_{\text{dl}} U_{\text{cyt}}^h - \xi_{\text{dl}} U_{\text{nuc}}^h) - \bar{V}_n (\gamma U_{\text{nuc}}^h Z_{\text{nuc}}^h - \beta_0 W_{\text{nuc}}^h) \\
\frac{d (\bar{V}_c U_c^h)_{\text{cyt}}}{dt} &= \bar{A}_m \lambda_{\text{dl}} (U_{\text{cyt}}^{h-1} - 2U_{\text{cyt}}^h + C_{\text{cyt}}^{h+1}) + \bar{V}_C \left( \beta W_{\text{cyt}}^h - \gamma U_{\text{cyt}}^h Z_{\text{cyt}}^h \right) - \bar{A}_n (\zeta_{\text{dl}} U_{\text{cyt}}^h - \xi_{\text{dl}} U_{\text{nuc}}^h) \\
\frac{d (\bar{V}_n W_n^h)_{\text{nuc}}}{dt} &= \bar{A}_n (\zeta_{\text{dl}} U_{\text{cyt}}^h - \xi_{\text{dl}} U_{\text{nuc}}^h) + \bar{V}_n (\gamma U_{\text{nuc}}^h Z_{\text{nuc}}^h - \beta_0 W_{\text{nuc}}^h) \\
\frac{d (\bar{V}_c W_c^h)_{\text{cyt}}}{dt} &= \bar{A}_m \lambda_{\text{dl}} (W_{\text{cyt}}^{h-1} - 2W_{\text{cyt}}^h + W_{\text{cyt}}^{h+1}) - \bar{V}_C \left( \beta W_{\text{cyt}}^h - \gamma U_{\text{cyt}}^h Z_{\text{cyt}}^h \right) - \bar{A}_n (\zeta_{\text{dl}} U_{\text{cyt}}^h - \xi_{\text{dl}} U_{\text{nuc}}^h) \\
\frac{d (\bar{V}_n Z_n^h)_{\text{nuc}}}{dt} &= \bar{A}_n (\zeta_{\text{cact}} Z_{\text{cyt}}^h - \xi_{\text{cact}} Z_{\text{nuc}}^h) - \bar{V}_n \psi (\gamma U_{\text{nuc}}^h Z_{\text{nuc}}^h - \beta_0 W_{\text{nuc}}^h) \\
\frac{d (\bar{V}_c Z_c^h)_{\text{cyt}}}{dt} &= \bar{A}_m \lambda_{\text{cact}} (Z_{\text{cyt}}^{h-1} - 2Z_{\text{cyt}}^h + Z_{\text{cyt}}^{h+1}) + \bar{V}_C \left( \psi \beta W_{\text{cyt}}^h - \psi \gamma U_{\text{cyt}}^h Z_{\text{cyt}}^h \right) - \alpha Z_{\text{cyt}}^h - \bar{A}_n (\zeta_{\text{cact}} Z_{\text{cyt}}^h - \xi_{\text{cact}} Z_{\text{nuc}}^h) + 1
\end{align*}
\]

where \( \bar{\beta} = \beta \exp \left( -\frac{1}{2} \left( \frac{\Sigma}{\sigma} \right)^2 \right) \) + \( \beta_0 \) and

\[
\begin{align*}
\zeta_s &= \frac{A_s^h K_{\text{in,s}}}{V_n^h T} \quad \xi_s = \frac{A_s^h K_{\text{out,s}}}{V_n^h T} \\
\lambda_s &= \frac{G_s A_s^h}{V_n^h T} \\
\gamma &= k_b c_{\text{cact}} T \\
\alpha &= k_{\text{Deg}} T \\
\beta &= k_D^{\text{max}} T \\
\beta_0 &= k_D^0 T
\end{align*}
\]

\[
\begin{align*}
U_{\text{nuc}} &= \frac{C_{\text{dl},n}}{C_{\text{dl}}} U_{\text{cyt}} \\
W_{\text{nuc}} &= \frac{C_{\text{dc},n}}{C_{\text{dc}}} W_{\text{cyt}} \\
Z_{\text{nuc}} &= \frac{C_{\text{Cact},n}}{C_{\text{Cact}}} Z_{\text{cyt}} \\
\bar{A}_n &= \frac{A_n}{A_s} \quad \bar{A}_m = \frac{A_m}{A_s} \\
\bar{V}_n &= \frac{V_n}{V_n^0} \\
\bar{V}_C &= \frac{V_C}{V_n^0} \\
\bar{T} &= 1 \text{ min} \\
\bar{\psi} &= \frac{\sigma}{\Sigma} \\
\psi &= \frac{C_{\text{cact},c}}{C_{\text{Cact}}} \\
\end{align*}
\]

for \( s = \{ \text{dl, n}; \text{dl, c}; \text{dc, n}; \text{dc, c}; \text{Cact, n}; \text{Cact, c} \} \).

The gene expression model equations take the general form:

\[
\frac{d}{dt} [\text{mRNA}^h_i] = \frac{1}{T_i} \left( f_i (U_{\text{nuc}}^h_i, [\text{sna}^h]) - [\text{mRNA}^h_i] \right) \tag{17}
\]

where \( f_i \) is defined for each species \( i \) as the product of the appropriate on/off terms. To avoid discontinuities, we use a Hill function with \( n_H = 100 \) to approximate the Heaviside step function: \( C^{n_H} / (\theta^{n_H} + C^{n_H}) \) for production and \( \theta^{n_H} / (\theta^{n_H} + C^{n_H}) \) for repression:

\[
\begin{align*}
\text{f}_{\text{sna}} &= \frac{(U_{\text{nuc}}^h)^{n_H}}{(\theta_{\text{dl,sna}})^{n_H} + (U_{\text{nuc}}^h)^{n_H}}, \\
\text{f}_{\text{vnd}} &= \frac{\left( U_{\text{nuc}}^h \right)^{n_H}}{(\theta_{\text{dl,vnd}})^{n_H} + (U_{\text{nuc}}^h)^{n_H}}, \\
\text{f}_{\text{sog}} &= \frac{\left( U_{\text{nuc}}^h \right)^{n_H}}{(\theta_{\text{dl,sog}})^{n_H} + (U_{\text{nuc}}^h)^{n_H}}, \\
\text{f}_{\text{zen}} &= \frac{(\theta_{\text{dl,zen}})^n}{(\theta_{\text{dl,zen}})^n + (U_{\text{nuc}}^h)^n}, \\
\text{f}_{\text{zen}} &= \frac{(\theta_{\text{dl,zen}})^n}{(\theta_{\text{dl,zen}})^n + (U_{\text{nuc}}^h)^n}.
\end{align*}
\]

To simulate noise we use \( U_{\text{eff}}^h = U_{\text{nuc}}^h(t) + \eta N(0,1) \sqrt{U_{\text{nuc}}^h(t)} \), where \( N(0,1) \) is a random number selected from the standard normal distribution. Values that fall below zero are set to zero.
Nuclear/cytoplasmic dimensions

For NCs 10-13, the volume and surface area of each nucleus are calculated assuming the nucleus is a sphere with constant radius $r$ based on measurements in [2]. For nuclear cycle 14, the volume and surface area of each nucleus are calculated assuming the nucleus is a prolate spheroid with major and minor axes $b$ and $a$, respectively, where $b$ was approximated such that $b = 2a$. The surface area and volume parameters are normalized to the area and volume measurements of a nucleus at the end of NC14 ($A^{14}_{n} \approx 160 \mu m^2$ and $V^{14}_{n} \approx 190 \mu m^3$).

The volume of the cytoplasm is simply the total volume of each compartment, $V = H \left( \frac{L}{n_i} \right)^2$, where $H = 25 \mu m$ represents the constant height of the simulated array of compartments, $L = 270 \mu m$ is the length of the simulated array (i.e., the length of half the embryo’s circumference), and $n_i$ is the number of nuclei in nuclear cycle $i$.

Estimation of physical parameters determining noise levels

Derivation of $\eta$

To simulate the stochastic behavior of gene expression downstream of dl, we added artificial noise to the simulated dl nuclear concentration, $\delta U_{\text{nuc}}$, in proportion to the square root of the nuclear dl concentration, $U_{\text{nuc}}$, with $\eta$ the proportionality constant. Here we provide a detailed derivation of $\eta$.

Berg and Purcell [3] argued that a microorganism could, at best, determine the concentration, $c$, of molecule X within a spherical volume of radius $a$ with a fractional error of

$$\frac{\delta c}{c} \sim (1.61 DaTC)^{-1/2} \tag{18}$$

where $T$ is the measurement timescale, and $D$ is the diffusivity of molecule X (see also [4]). Thus, we assume that the relative fluctuations in concentration of dl at its binding site, $C_{\text{dl},n}$, are

$$\frac{\delta C_{\text{dl},n}}{C_{\text{dl},n}} \approx \frac{1}{\sqrt{1.61 DaC_{\text{dl},n}T}}, \tag{19}$$

where $D$ is the diffusivity of dl within the nucleus, $a$ is the length of the DNA binding site, and $T$ is an averaging time for the cell to measure this concentration [3, 4]. If we apply non-dimensionalization to our dl nuclear concentration by substituting $C_{\text{dl},n} = U_{\text{nuc}}C_{\text{dl}}$, we arrive at

$$\frac{\delta U_{\text{nuc}}}{U_{\text{nuc}}} \approx \frac{1}{\sqrt{1.61 DaU_{\text{nuc}}C_{\text{dl}}T}}, \tag{20}$$

or, equivalently:

$$\delta U_{\text{nuc}} = \eta \sqrt{U_{\text{nuc}}}, \tag{21}$$

Thus, the noise levels are $\delta U_{\text{nuc}} = \eta \sqrt{U_{\text{nuc}}}$, and the proportionality constant is

$$\eta = (1.61 DaC_{\text{dl}}T)^{-1/2} \tag{22}$$

This is the formula employed in our gene expression simulations, in which, at each time step, a random level of Gaussian noise with a mean of zero and a standard deviation of $\delta U_{\text{nuc}}$ is added to the concentration of dl in each nucleus.

We estimate the following values for the physical parameters included in the calculation of $\eta$. First, the length of the dl binding site is 10 base pairs [5], meaning $a \approx 3$ nm. Second, it seems reasonable to
assume $C_{dl} \approx 1 - 10 \text{nM} \approx 0.6 - 6 \text{mole/\mu m}^3$ based on concentration measurements of Bicoid (FlyBase: FBgn0000166) [4]. Third, $D$ is based on fluorescence correlation spectroscopy studies of Bicoid, in which its diffusivity was found to be roughly $1 - 10 \mu m^2/s$ [6]. Finally, we also argue a timescale of approximately $T \approx 1 - 10 \text{ min}$ is an appropriate averaging time for transcription during the early Drosophila embryo (see [7]). Therefore, using the mean value for each term in (22), we find that

$$\eta \approx [1.61 \ (5.5 \mu m^2 s^{-1}) \ (3 \text{ nm}) \ (2.7 \mu m^{-3}) \ (5.5 \text{ min})]^{-1/2} \approx 0.2.$$  

(23)

While this calculation does not prove that $\eta = 0.2$, it shows that our simulations are in accord with independent estimates of noise levels.

**Estimation of $T$**

For our estimate of $T$, we begin with the assumption that the averaging time for the concentration of dl to be read-out by a gene locus (ie, to result in a transcription decision) is related to the time between transcriptional bursts [8, 9]. Recent work on transcriptional bursting events in the early Drosophila embryo suggests the timescale for bursting events is roughly 1-10 min [7, 10, 11].

**References**

[1] Kanodia JS, Rikhy R, Kim Y, Lund VK, DeLotto R, Lippincott-Schwartz J, et al. Dynamics of the Dorsal morphogen gradient. PNAS. 2009 Dec;106(51):21707–12.

[2] Gregor T, Wieschaus EF, McGregor AP, Bialek W, Tank DW. Stability and nuclear dynamics of the bicoid morphogen gradient. Cell. 2007 Jul;130(1):141–52.

[3] Berg HC, Purcell EM. Physics of chemoreception. J Biophys. 1977 Nov;20(2):193–219.

[4] Gregor T, Tank DW, Wieschaus EF, Bialek W. Probing the limits to positional information. Cell. 2007 Jul;130(1):153–64.

[5] Ip YT, Kraut R, Levine M, Rushlow CA. The dorsal morphogen is a sequence-specific DNA-binding protein that interacts with a long-range repression element in Drosophila. Cell. 1991 Jan;64(2):439–446.

[6] Abu-Arish A, Porcher A, Czerwonka A, Dostatni N, Fradin C. High mobility of bicoid captured by fluorescence correlation spectroscopy: implication for the rapid establishment of its gradient. J Biophys. 2010 Aug;99(4):L33–L35.

[7] Bothma JP, Garcia HG, Esposito E, Schlissel G, Gregor T, Levine M. Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. PNAS. 2014 Jul;

[8] Molina N, Suter DM, Cannavo R, Zoller B, Gotic I, Naef F. Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. PNAS. 2013 Dec;110(51):20563–20568.

[9] Hao N, O’Shea EK. Signal-dependent dynamics of transcription factor translocation controls gene expression. Nature Struct & Mol Bio. 2012;19(1):31–39.

[10] He F, Ren J, Wang W, Ma J. A multiscale investigation of bicoid-dependent transcriptional events in Drosophila embryos. PLoS One. 2011;6(4):e19122.

[11] Little SC, Tikhonov M, Gregor T. Precise developmental gene expression arises from globally stochastic transcriptional activity. Cell. 2013 Aug;154(4):789–800.
| Parameter | Description |
|-----------|-------------|
| $A_n$     | Surface area of a nucleus |
| $A_m$     | Area available for intercompartmental exchange |
| $V_n$     | Volume of a nucleus |
| $V_C$     | Volume of a compartment |
| $A_{14}^n$| Surface area of a nucleus in NC14 |
| $V_{14}^n$| Volume of a nucleus in NC14 |
| $k_{in,j}$| Nuclear import of species $j$ |
| $k_{out,j}$| Nuclear export of species $j$ |
| $\Gamma_i$| Intercompartmental exchange of species $i$ |
| $k_b$     | Binding rate for dl and Cact |
| $k_{D0}$  | Unbinding rate for dl/Cact complex |
| $k_{Dmax}$| Maximum Toll-mediated degradation of dl/Cact complex |
| $P_{Cact}$| Production rate of Cact |
| $k_{Deg}$ | Degradation rate of Cact |
| $C_i^0$   | Initial concentration of species $i$ |
| $x$       | DV axis coordinate |
| $\phi$    | Width of Toll signal |
| $L$       | Length of the simulated region |
| $T$       | Time scale |

Table S1: S1 Table. Summary of model parameters

| Nuclear Cycle | 10 | 10m | 11 | 11m | 12 | 12m | 13 | 13m | 14 |
|---------------|----|-----|----|-----|----|-----|----|-----|----|
| $t(end)$ (min)| 3.61 | 7.70 | 11.3 | 15.4 | 21.6 | 25.7 | 37.8 | 43.0 | 97.7 |
| Duration (min) | 3.61 | 4.09 | 3.6 | 3.9 | 6.0 | 4.1 | 12.1 | 5.2 | 54.7 |
| $n$ | 13 | 13 | 19 | 19 | 26 | 26 | 36 | 36 | 51 |
| $A_n$ | 1.4 | - | 1.5 | - | 1.1 | - | 1.2 | - | 1.0 |
| $A_m$ | 6.5 | 6.5 | 4.4 | 4.4 | 3.2 | 3.2 | 2.3 | 2.3 | 1.7 |
| $V_n$ | 1.7 | - | 1.8 | - | 1.1 | - | 1.3 | - | 1.0 |
| $V_C$ | 56 | 58 | 25 | 27 | 13 | 14 | 6.3 | 7.6 | 2.8 |

Table S2: S2 Table. Dimensional quantities used in model formulation

Figure S1: S1 Fig. Effect of noise on gene expression
(Left to right) Increasing the noise parameter, $\eta$ from 0 to 1 shows that the slopes of the gene expression boundaries approach infinity at $\eta = 0$, and become very noisy above $\eta = 0.2$. (Note: each run is an average of 10 runs for each parameter adjustment to reduce randomness in the plot due to noise. This comports with the experimental data, which are the average of 10+ embryos. The same is true for Fig. S2-S4.)
Using free dl as the input to the gene expression model, a sensitivity analysis shows little sensitivity to changes in the dl threshold parameters ($\theta_{dl:mRNA}$), lifetime parameters ($\tau_i$), and noise parameter ($\eta$) for our genes of interest. (Hill coefficient $n_H = 100$.)
Using both free dl and dl/Cact complex as the input to the gene expression model, a sensitivity analysis shows high sensitivity to changes in the dl threshold parameters ($\theta_{dl:mRNA}$) for both Type III genes ($sog$ and $zen$; green and yellow, respectively), and little sensitivity to changes in lifetime and noise parameters. (Hill coefficient $n_H = 100$).
Using a soft threshold ($n_H = 4$) does not change the conclusions of our sensitivity analysis.

Figure S4: S4 Fig. Sensitivity analysis, free dl case, soft threshold