Supporting Information

Khatri et al. 10.1073/pnas.0812260106

SI Text

Morphogen Concentration Profile. The simplest mechanism by which a morphogen profile is generated in development is from a morphogen diffusing from a source (such as maternally implanted mRNA transcripts in the case of bicoid) that are degraded in a concentration-dependent manner. If the morphogen has a diffusion constant D and degradation rate k, both being uniform in space, and its source is at $x = 0$ this mechanism gives rise to an exponential gradient in concentration $\mathsf{M}(x) = \mathsf{M}_0 e^{-\alpha x}$, where $\alpha = \sqrt{k/D}$ and M_0 is related to the rate of production of the morphogen, J. However, this expression is only valid if $\alpha L \ll 1$, such that the boundary of the embryo at $x = L$ is not important. For a more general expression valid for any α, the reaction-diffusion equation $\partial_t \mathsf{M} = D\partial_x^2 \mathsf{M} - k\mathsf{M}$ can be solved with boundary conditions $\partial_x \mathsf{M}|_{x=0} = -J/D$ and $\partial_x \mathsf{M}|_{x=L} = 0$ to give the steady-state morphogen gradient as,

$$\mathsf{M}(x) = \mathsf{M}_0 \frac{\cosh\, \alpha(x - L)}{\sinh\, \alpha L}. \qquad \textbf{[s1]}$$

We use Eq. **s1** to calculate how the concentration profile of the morphogen changes across the embryo as the morphogen steepness α is mutated.

Probability of Transcription Factor Binding. To calculate this probability, we use the canonical ensemble of statistical mechanics, for which the partition function $\mathcal{Z}$ is most simply expressed in terms of a spin-like variable, which represents the occupation of each binding site, $\sigma_j = \{0, \mathsf{R}, \mathsf{M}\}$,

$$\mathcal{Z} = \sum_{\sigma_\mathsf{P}} \sum_{\sigma_1} e^{-\beta(E_{\sigma_\mathsf{P}\mathsf{P}} + E_{\sigma_1 1} - \mu_{\sigma_\mathsf{P}} - \mu_{\sigma_1} + \delta E_{\sigma_\mathsf{P}\sigma_1})}$$

with $E_{0j} = 0$, $\delta E_{ii'} = 0$, for either $i = 0$ or $i' = 0$ and $\mu_0 = 0$, where $\mu_{\sigma_j} = \frac{1}{\beta}\ln[\sigma_j]$ represents the chemical potentials of the protein species at the jth binding site, β is the inverse of thermal temperature and 0 represents a free binding site. Formally this construction is known as a 3-state Pott's model. So given a "genome" $\boldsymbol{G} = [\boldsymbol{b}_\mathsf{R}, \boldsymbol{b}_\mathsf{M}, \boldsymbol{r}_\mathsf{P}, \boldsymbol{r}_1, \boldsymbol{g}_\mathsf{R}, \boldsymbol{g}_\mathsf{M}]$ from which the binding and glue energies can be calculated (E_{ij} and $\delta E_{ii'}$ given by Eq. **1** and Eq. **2**, respectively, in the main text), p_{RP} is given schematically by

$$p_{\mathsf{RP}} = p(\sigma_1 - \mathsf{R}\ulcorner)$$
$$= \frac{1}{\mathcal{Z}}((0 - \mathsf{R}\ulcorner) + (\mathsf{M} - \mathsf{R}\ulcorner) + (\mathsf{R} - \mathsf{R}\ulcorner)) \qquad \textbf{[s2]}$$

where, for example,

$$(\mathsf{M} - \mathsf{R}\ulcorner) \equiv e^{-\beta(E_{\mathsf{RP}} + E_{\mathsf{M}1} - \mu_\mathsf{R} - \mu_\mathsf{M} + \delta E_{\mathsf{RM}})}$$

is the Boltzmann factor for cooperative binding of the morphogen and RNAP to the regulatory region of DNA.

Calculation of Fitness Landscape for Fixed Genome. To calculate the fitness landscape for spatial patterning of the anterior of an embryo, based on the fitness functional given by Eq. **4** in the main text, we approximate the concentration profile of the TF by

$$\mathsf{T}(x, \alpha) \sim \frac{1 + a\mathsf{M}(x)}{1 + b\mathsf{M}(x)}. \qquad \textbf{[s3]}$$

This arises from directly evaluating Eq. **s2**, but ignores the term in the denominator related to the cooperative binding of morphogen binding with itself to the regulatory region of T, which is responsible for repression of transcription for large $\mathsf{M}(x)$. In relation to the results in the main paper (Fig. 2, main text) this amounts to ignoring the small negative curvature in $\mathsf{T}(x)$ for $x \approx 0$. The parameters a and b are functions of the energies $\{E_{ij}\}$ and $\{\delta E_{ii'}\}$ or equivalently the genome $\boldsymbol{G}$. Calculating the functional in Eq. **4** from the main text and using $\mathsf{M}(x) \approx \mathsf{M}_0 e^{-\alpha x}$, gives

$$F(\alpha) \approx \frac{1}{\alpha L}\frac{b - a}{a} \ln\left[\frac{(1 + be^{-\alpha L/2})^2}{(1 + b)(1 + be^{-\alpha L})}\right].$$

In Fig. S1*A*, we plot Eq. **s4** with random values of the parameters a and b, restricted to a range equivalent to constraining $0\, k_\mathsf{B}T \le E_{ij} \le 20 k_\mathsf{B}T$ and $-5k_\mathsf{B}T \le \delta E_{ii'} \le 0k_\mathsf{B}T$, consistent with the ranges used in the simulations in the main paper; each separate curve is a different random pair of a and b. It is clear that each of the curves look qualitatively like the curves seen in Fig. 5 in the main text, suggesting that each of these curves is the fitness landscape $F(\alpha)$ for different genotypes or fixed values of $\{E_{ij}\}$ and $\{\delta E_{ii'}\}$, with α allowed to vary. In Fig. S1*B*, we also show a scatter plot for a simulation of length 10^7 mutations or Monte Carlo steps at a population size of $N = 10$; comparison to Fig. S1*A* gives qualitative agreement, further supporting that each of these curves is a fitness landscape for fixed genotype.

Threshold Mechanism of Cooperative Binding of M–R. For our simple model of spatial gene regulation, the approximate expression for the concentration profile of the TF given by Eq.**s3** gives rise to a simple sigmoidal response with respect to $\ln(\mathsf{M})$ as shown schematically in Fig. S2*A* for 2 different sets of values of a and b, for $a/b \gg 1$, such that T is switched on ($\mathsf{T} \to a/b$) when M is large and off ($\mathsf{T} \to 1$) when it is small. It is simple to show the threshold concentration is $\mathsf{M}^* \propto 1/(b - 2a) \propto e^{\beta E_{\mathsf{M}1}}$, where $E_{\mathsf{M}1}$ is the binding energy of the morphogen binding to the first binding site. Now an exponentially decaying morphogen gradient across the field of cells of the embryo leads to a linear mapping of the sigmoidal profile shown in Fig. S2*B*, where it is clear that the "rate" of mapping ($\mathrm{d}\mathsf{T}/\mathrm{d}x$) is proportional the morphogen gradient steepness α. So as the morphogen gradient becomes steeper, the switch from high to low T is sharper, as shown in Fig. 2*C*, leading to an increase in fitness. However, as shown in Fig. S2*B*, for a given genome $\boldsymbol{G}$ and in particular the binding energy $E_{\mathsf{M}1}$, we see that only one optimum value of α will give a spatial switch positioned exactly at the midpoint of the embryo; from the functional Eq. **4** in the main text, for any α smaller or larger, the fitness must decrease, and thus we see where the general form of Eq. **4** arises. In addition, we note that the gene regulatory module has a maximum sensitivity set roughly by $E_{\mathsf{M}1} = 0$, so that there is, in addition, an ultimate limit to fitness imposed by the largest optimum value of α to which this corresponds. This is seen by a decrease in number of viable solutions or phenotypes at large α in Fig. 1 and Fig. 5 in the main text or equivalently by the decrease in free fitness for large α in Fig. 4 in the main text.

Ergodicity of Simulations at Low Population Sizes. A system that is ergodic is one in which time averages of quantities are equal to ensemble or phase–space averages. To test for ergodicity of the Monte Carlo simulations of evolution, we compared the free fitness (Eq. **6** in the main text) landscape calculated using an average over a time series (Fig. S2 *A* and *B*, solid lines), from

simulations of length 10^7 attempted mutations (Monte Carlo steps) to the free fitness calculated using an ensemble average (Fig. 2 *A* and *B*, red circles) over the end state of 100 shorter simulations of length 2×10^4 attempted mutations. Fig. S2 *A* and *B* shows this comparison for the population sizes of $N = 20$ and $N = 110$, respectively, for an initial value morphogen gradient $\alpha_0 = 10$. The time series of the morphogen gradient α for the 100 shorter simulations are shown in Fig. 2 *C* and *D* for $N = 20$ and $N = 110$, respectively. It is clear that there is very good overlap of the free fitness calculated via a time average and ensemble average (the greater variation in the ensemble average is due to the lower sample size used compared to the time average), and we conclude that these simulations are ergodic for these range of population sizes. In addition, we also repeated the longer simulations (10^7 mutations) for initial morphogen gradients of $\alpha_0 = 1$ and $\alpha_0 = 7$, as shown in Fig. S2 *A* and *B*, where it is seen that the simulations are also independent of initial conditions for these range of parameters, further indicating the ergodicity of the simulations.

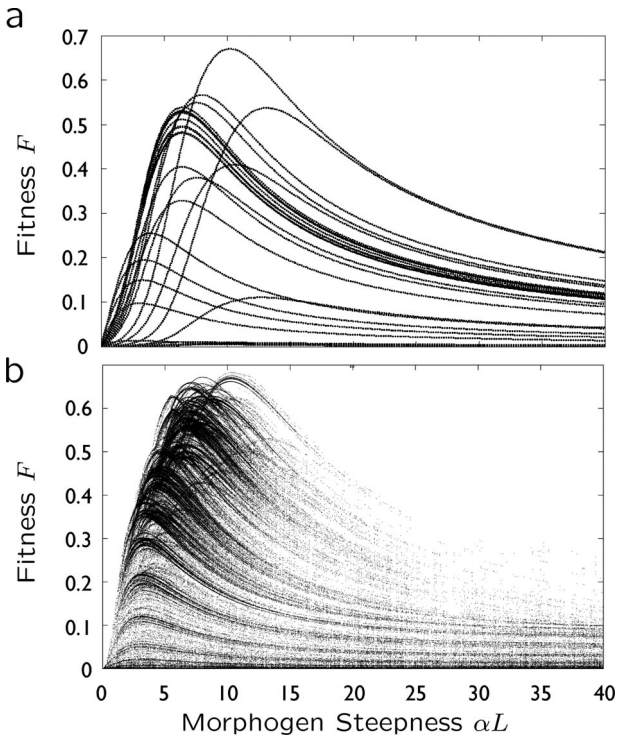

Fig. S1. Fitness landscape for fixed genotype. (*A*) Eq. **s3** plotted many times, each with random values of the parameters with constraints on the parameters *a* and *b* equivalent to the constraints used for the binding and glue energies in the main text. (*B*) Scatter plot of fitness vs. morphogen gradient (F vs. α) from a simulation for patterning the anterior of an embryo for a population size of $N = 10$. The congruence of the two plots and Fig. 5 in the main text, indicate that the landscape is structured into multiple solutions $F_i(\alpha)$ for fixed genotype.

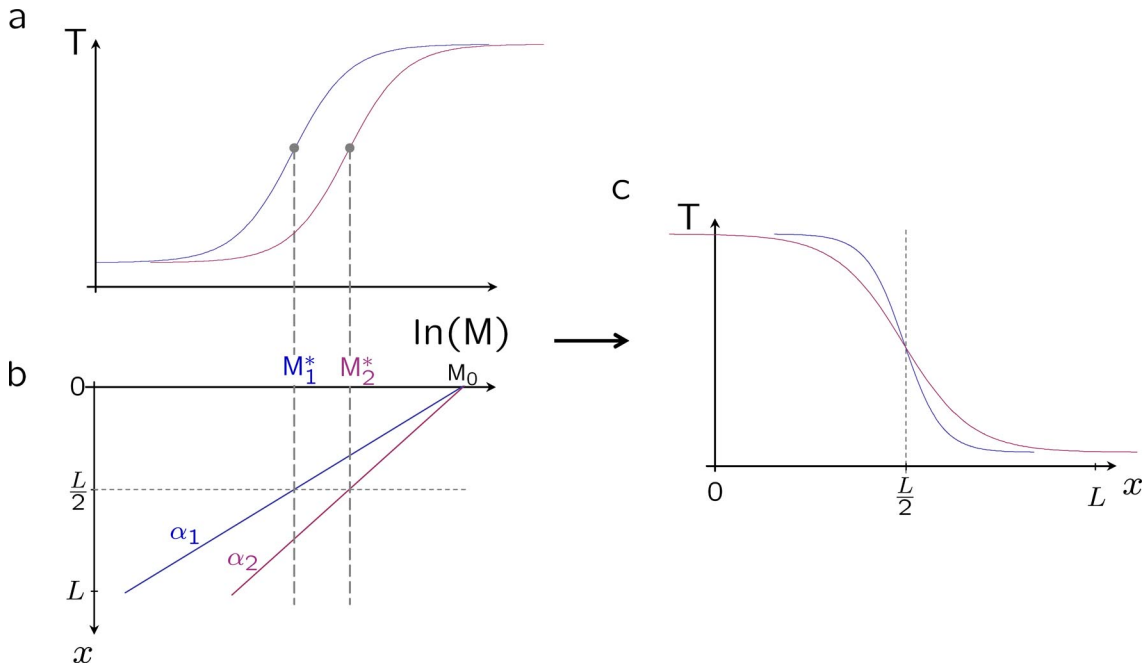

Fig. S2. Threshold patterning mechanism. (*A*) A schematic linear log plot of the sigmoidal relationship of T to M, for 2 different sets of values of the parameters *a* and *b*. The threshold value of the morphogen concentration M*, represent when T is approximately half its maximum value. (*B*) A schematic linear-log plot of *x* vs. ln(M), illustrating that different genotypes (represented by the parameters *a* and *b*) which gives rise to different sigmoidal responses in *A*), require a different morphogen gradient steepness α, in order that the profile of T is mapped to *x* with the threshold at the midpoint of the embryo. (*C*) A schematic plot (on a linear–linear scale) of T(*x*) showing that the larger morphogen steepness α_2 gives rise to a sharper profile, than α_1. We see that sharper patterning means higher values of α, which in turn requires a regulatory apparatus that is more sensitive to morphogen concentrations (i.e., lower M*).

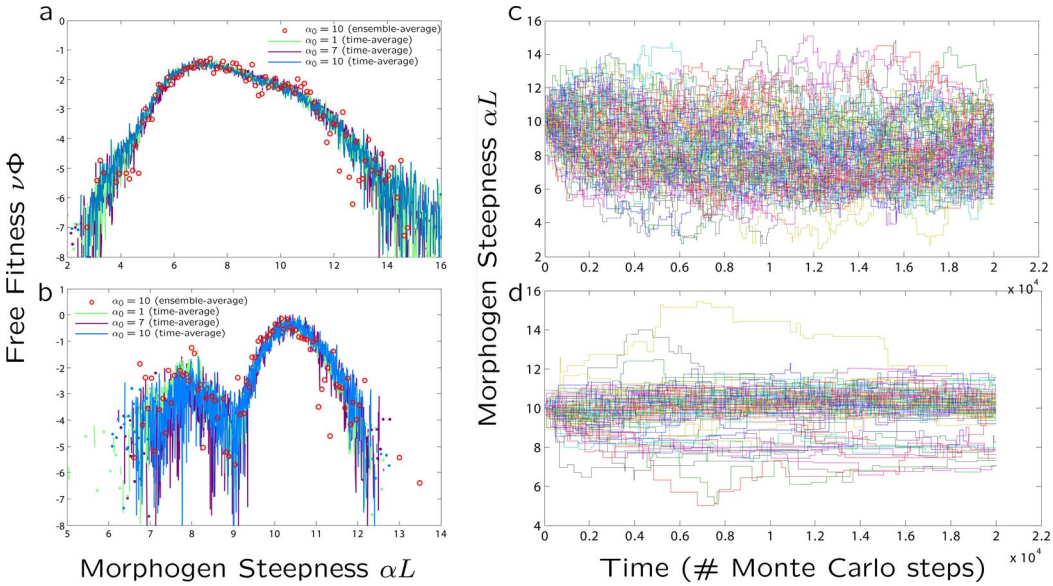

Fig. S3. Ergodicity of simulations. (*A* and *B*) Free fitness landscapes for populations sizes of $N = 20$ (*A*) and $n = 110$ (*B*), calculated using a time average of Monte Carlo simulations of length 10^7 mutations (solid lines) and using an ensemble-average from the end point of 100 simulations of length 2×10^4 (shown as time series of α in *C* and *D* for $N = 20$ and $N = 110$, respectively). (*C* and *D*) Time series of the morphogen gradient α from 100 independent simulations of length 2×10^4 mutations, for populations sizes of $N = 20$ (C) and $N = 110$ (*D*).