PNAS RESEARCH ARTICLE | PHYSICS

Deriving a genetic regulatory network from an optimization principle

Thomas R. Sokolowski[a,b], Thomas Gregor[c,d], William Bialek[c,e,1], and Gašper Tkačik[a,1]

Affiliations are included on p. 10.

Contributed by William Bialek; received February 13, 2024; accepted November 13, 2024; reviewed by Dmitri B. Chklovskii, Angela H. DePace, Michael M. Desai, and Jane Kondev

Many biological systems operate near the physical limits to their performance, suggesting that aspects of their behavior and underlying mechanisms could be derived from optimization principles. However, such principles have often been applied only in simplified models. Here, we explore a detailed mechanistic model of the gap gene network in the *Drosophila* embryo, optimizing its 50+ parameters to maximize the information that gene expression levels provide about nuclear positions. This optimization is conducted under realistic constraints, such as limits on the number of available molecules. Remarkably, the optimal networks we derive closely match the architecture and spatial gene expression profiles observed in the real organism. Our framework quantifies the tradeoffs involved in maximizing functional performance and allows for the exploration of alternative network configurations, addressing the question of which features are necessary and which are contingent. Our results suggest that multiple solutions to the optimization problem might exist across closely related organisms, offering insights into the evolution of gene regulatory networks.

gene regulatory networks | optimization | evolution | *Drosophila*

Optimization provides a mathematical framework for addressing fundamental problems across the physical and statistical sciences. In evolutionary theory, stochastic optimization underpins our understanding of how natural selection improves organismal fitness by favoring adaptive traits that push against physical constraints (1, 2). Many biological systems, such as photon counting in vision (3), diffraction-limited imaging in insect eyes (4), and molecule counting in bacterial chemotaxis (5) exhibit performance that approaches these physical limits.

Observations of near-optimal performance raise the possibility of using optimization as a principle to derive nontrivial predictions about biological systems' functional behavior and underlying mechanisms (6, 7). Such approaches have been applied successfully to simplified models, including coding efficiency in sensory processing (8–11), growth rates in metabolic networks (12), matter flux in transport networks (13, 14), information transmission in regulatory networks (15), and the design of molecular machines and assemblies (16). However, these successes have often been limited to models that omit many biological details or focus on optimizing single, isolated components of complex systems.

A more ambitious goal is to use optimization to derive accurate, first-principles predictions for complex, multicomponent biological systems, where interactions are described in molecular detail. Here, we test the hypothesis that optimization, when correctly formulated and constrained, can quantitatively predict the behavior of a complex molecular system: a well-characterized genetic regulatory network involved in *Drosophila* embryonic development. Specifically, we will search for network parameters that maximize positional information (17)—the information, in bits, that local gene expression levels provide about the position of a cell along the anterior–posterior (AP) axis of the embryo. Information is limited by noise, and noise ultimately arises from the randomness of individual molecular events, so we constrain our optimization by fixing the mean numbers of molecules (mRNAs and proteins) to those observed experimentally. In words, we are asking for the network that squeezes as much information as possible out of a limited number of molecules. When made mathematically precise, this constrained optimization makes quantitative predictions about the spatial patterns of gene expression, their noise levels, their dynamics, and the regulatory interactions that control these patterns. Perhaps surprisingly, we find solutions to our optimization problem that closely resemble the real gap gene network in all these ways.

Significance

Information crucial for life is represented by surprisingly low concentrations of key molecules, yet cells use these small signals to make reliable decisions. Could the mechanisms of life have been tuned to extract as much information as possible from a limited number of molecules? We apply this physical principle to a genetic network that controls pattern formation in early development of the fruit fly embryo, searching for the settings of 50+ internal parameters that give each cell the maximum possible information about its position in the embryo. The resulting optimal networks recapitulate many features of the real network, quantitatively. This approach makes tradeoffs explicit, rationalizes the network architecture, and provides perspectives on the interplay of chance and necessity.

Author contributions: T.R.S.,T.G., W.B., and G.T. designed research; T.R.S. and G.T. performed research; T.R.S., W.B., and G.T. contributed new reagents/analytic tools; and T.R.S., T.G., W.B., and G.T. wrote the paper.

Reviewers: D.B.C., Simons Foundation; A.H.D., Harvard; M.M.D., Harvard University Department of Organismic and Evolutionary Biology; and J.K., Brandeis University.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: wbialek@princeton.edu or gtkacik@ist.ac.at.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2402925121/-/DCSupplemental.

Published January 3, 2025.

We emphasize at the outset that there are many technical problems to address in actually solving the optimization problem. At each setting of the parameters, we must be able to estimate the positional information. We must explore the parameter space in some systematic way. We might want to use the observed patterns of gene expression to help guide our search but we need to be sure that the solutions we find do not depend on this guidance. We expect multiple solutions and would like at least a preliminary characterization of this diversity. We give a guide to these issues in the main text, with details in *SI Appendix*.

If the optimization hypothesis holds, it could illuminate many open questions about the evolution of genetic regulatory networks (18). For example, we could explore whether the presence of specific genes, interactions, or regulatory elements is due to evolutionary necessity or historical contingency. While this question is central to evolutionary biology (19), it cannot be answered by genetic experiments alone, as these cannot rule out alternative evolutionary pathways. Instead, we require a theoretical framework that simulates alternative evolutionary outcomes. Our optimization scheme allows us to add or remove individual components of the network and reoptimize other parameters, thereby exploring what is possible, for example, in the absence of regulatory feedback loops. We will see that features that might have seemed accidental or redundant are likely necessary.

Additionally, we are interested in understanding the interplay between optimization and developmental constraints. Could different evolutionary outcomes have emerged given the known physical constraints, particularly those related to the availability of molecular resources? We can address these questions in our optimization framework, which explicitly balances functional performance against resource costs and constraints.

To be concrete we focus on the early stages of development in the *Drosophila* embryo (20). Approximately two hours postfertilization, the gap genes *hunchback*, *Krüppel*, *giant*, and *knirps* exhibit intricate spatial and temporal expression patterns along the embryo's AP axis (21). These genes form a network with inputs from the maternal morphogens (Bicoid, Nanos, and Torso-like) (20, 22), and the output of the network in turn drives the expression of downstream genes essential for body segmentation (23).

At a critical developmental stage, the combination of four gap gene expression levels encodes 4.3 ± 0.1 bits of information about position along the AP axis (24, 25), sufficient for specifying positions with a precision of about 1% of embryo length (EL) (26); this matches the precision of downstream events, such as the positions of stripes in pair-rule gene expression and the formation of the cephalic furrow. Evidence suggests that this flow of information—from encoding to readout—may operate near optimal efficiency, subject to constraints on molecular counts and mean resource usage (24, 26, 27). These observations led us to focus on the gap gene network as a system in which to test the optimization hypothesis.

While previous work has uncovered many features of the gap gene network through quantitative experiments, genetic manipulations, and mathematical modeling (28–37), fundamental questions remain. Is the network's behavior shaped primarily by its evolutionary history (38), or by the developmental and physical constraints, such as limited molecular resources (39–41)? Are all components, including the three maternal morphogens and four gap genes, necessary? Most crucially, can we derive the network's behavior from an underlying theoretical principle, rather than fitting its parameters to empirical data?

Optimization in a Realistic Context

To answer the questions outlined above, we formulated a detailed and realistic spatial-stochastic model of gap gene regulation in the *Drosophila* embryo. This model captures the regulation of gap genes by maternal morphogens, cross-regulation among gap genes, nuclear divisions, transcription, translation, degradation processes, and the diffusion of gene products (Fig. 1, Box 1, and *SI Appendix*, section 1). To apply our optimization principle we search within this large class of models for those that generate gene expression patterns encoding the maximum positional information while respecting geometrical constraints, the known temporal schedule of nuclear divisions, maternal input properties, and the physical limits on the number of molecules in the gap gene network that can be synthesized.

We take the spatial profiles and absolute levels of the three maternal morphogens, as well as maximal gap gene transcription, translation, and degradation rates to be fixed at their measured or estimated values (Box 1 and *SI Appendix*, Table S2). This still leaves more than 50 parameters which govern how gap genes integrate transcriptional regulatory signals from other gap genes

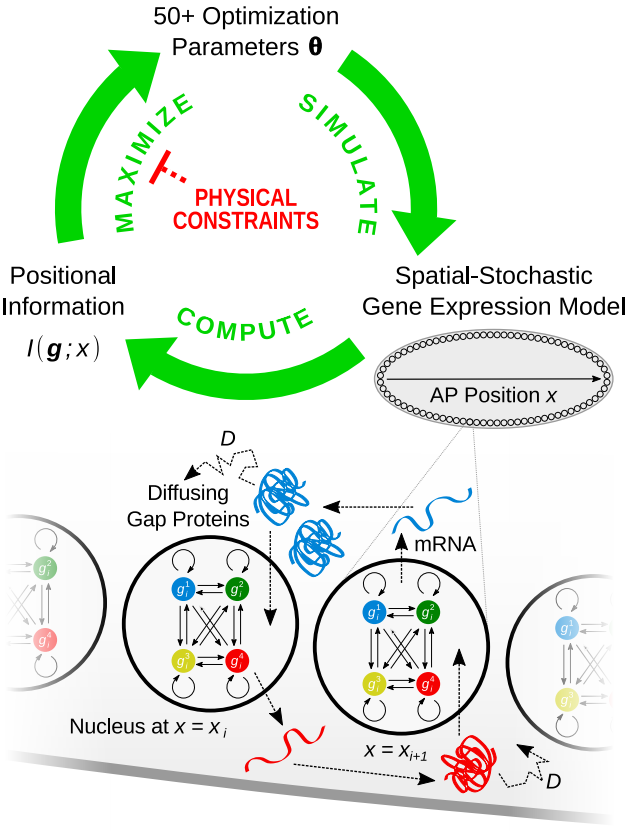


Fig. 1. Deriving a genetic regulatory network from an optimization principle. We simulate patterning during early fly development in a biophysically realistic, spatial–stochastic gap gene expression model (*Bottom*; see Box 1) that accounts for the stochastic gene expression dynamics in individual nuclei along the anterior–posterior (AP) axis of the embryo. Regulatory interactions among four gap genes (arrows between colored circles in each nucleus), their response to three maternal morphogen gradients, and spatial coupling between neighboring nuclei are parameterized by a set of over 50 parameters $\boldsymbol{\theta}$. For each parameter set, we numerically simulate the resulting noisy gap gene expression patterns, compute the system's positional information $I(\mathbf{g};x)$, and adjust $\boldsymbol{\theta}$ using stochastic optimization to iteratively maximize the encoded I subject to physical constraints (*Top*). For details, see Box 1 and *SI Appendix*, sections 1–5.

Box 1.

Model Description and Assumptions

We model the expression of gap genes up to the readout time, 40 min into nuclear cycle 14 ($T = 166$ min postfertilization) under physical constraints (*SI Appendix*, sections 1 and 5). First, we assume that the dynamics of gap gene expression can be described by effective rates for mRNA and protein synthesis and degradation. mRNA dynamics is assumed to set the slowest time scale (lifetime $\tau_M = 20$ min), so that the corresponding protein concentrations (lifetime $\tau_P = 10$ min) track the mRNA levels. The maximal mRNA production rate at full activation, $\rho_{\max}$, reproduces the maximal mRNA counts per nuclear volume reported for gap gene *hb* in early nuclear cycle 14 ($M_{\max} \approx 5{\cdot}10^2$) (39). Proteins are produced from mRNA in bursts (burst size $\beta = 12$ per mRNA), leading to maximal average protein number per nucleus $N_{\max} \approx 6 \cdot 10^3$. These parameters are assumed to be the same for all gap genes (40). Second, while we have considered a description of nuclei distributed over the two-dimensional surface of the embryo, a one-dimensional description provides a tractable approximation. In this model of pattern formation along the anterior–posterior axis $N = 70$ nuclei are spaced uniformly at positions $x_i = i \cdot \Delta$ (with $\Delta = 8.5$ μm) along the length $L = N\Delta$ of the embryo (42). During the simulated time period, the embryo is a syncytium, allowing expression levels in neighboring cells to be coupled via an effective diffusion constant D (baseline value $D = 0.5\ \mu m^2/s$, varied in Fig. 3*E*) that includes both cytoplasmic diffusion and transport across the nuclear membrane. Third, the spatial profiles of maternal inputs to the gap gene network (A = anterior, P = posterior, and T = terminal system; see Box image) are established early and are assumed to be constant throughout the relevant time period. Fourth, the rate of mRNA synthesis for each gene expressed at x_i is modulated between zero and $\rho_{\max}$ by local gap gene expression levels $\mathbf{g}_i = (g_i^1, g_i^2, g_i^3, g_i^4)$ and the local maternal input concentrations, $\mathbf{c}_i = (c_i^A, c_i^P, c_i^T)$, as described by regulation functions f_α that parametrically differ between gap genes α but are the same for all positions i. Regulatory functions are inspired by the Monod–Wyman–Changeux (MWC) model, where the expression of gene α switches between active (probability f_α) and inactive ($1 - f_\alpha$) states; these probabilities depend on regulatory inputs as follows (43, 44):

$$f_\alpha(\mathbf{g}_i, \mathbf{c}_i, \theta) = \frac{\rho_\alpha}{1 + \exp\left(-F_\alpha(\mathbf{g}_i, \mathbf{c}_i, \theta)\right)}$$

$$F_\alpha(\mathbf{g}_i, \mathbf{c}_i, \theta) = \underbrace{\sum_{\zeta \in \{1..4\}} H_G^{\alpha\zeta} \log\left(1 + \frac{g_i^\zeta}{K_G^{\alpha\zeta}}\right)}_{\text{Self- and Cross-Reg.}} + \underbrace{\sum_{\kappa \in \{A,P,T\}} H_M^{\alpha\kappa} \log\left(1 + \frac{c_i^\kappa}{K_M^{\alpha\kappa}}\right)}_{\text{Feed Forward (FF) Reg.}} + \underbrace{F_0}_{\text{Base Expr.}} .$$

Here F_0 is the bias toward active state in absence of any regulatory inputs; $H_G^{\alpha\zeta}$ and $H_M^{\alpha\kappa}$ are the strengths of the regulatory inputs ($H > 0$ is activating and $H < 0$ repressive; $|H| < H_{\max}$ cf. Fig 4*C*), by gap proteins (for self- and cross-regulation) and maternal morphogen proteins (for feed forward regulation), respectively, while $K_G^{\alpha\zeta}$ and $K_M^{\alpha\kappa}$ are the associated concentration thresholds. Taken together, we obtain a system of coupled stochastic differential equations (*SI Appendix*, section 1),

$$\partial_t g_i^\alpha = \underbrace{f_\alpha(\mathbf{g}_i, \mathbf{c}_i, \theta)}_{\text{Regulated Production}} - \underbrace{\frac{1}{\tau_\alpha} g_i^\alpha}_{\text{Degradation}} + \underbrace{\frac{D}{\Delta^2} \sum_n (g_n^\alpha - g_i^\alpha)}_{\text{Diffusion}} + \underbrace{\sum_k \Gamma_i^k(\theta)}_{\text{Noise}},$$

where n runs over all neighbors of nucleus i and Γ_i^k represent stochastic forces whose magnitude we derive in *SI Appendix*, section 1 to account for the following noise processes: i) "input noise" caused by the random arrival of TFs to gap gene CREs (45); ii) "output noise" due to stochastic mRNA and protein synthesis and degradation; iii) "diffusion noise" due to stochastic spatial exchange of gap gene proteins between neighboring nuclei. Importantly, the variances of these noise terms typically scale (inversely) with the number of gap gene products. We phenomenologically add iv) "extrinsic noise" due to maternal morphogen variability as well as other sources of embryo-to-embryo variation, where the contribution to the gap gene expression variance is assumed to be proportional to the squared mean expression (46), with the coefficient of variation ($CV^2 = 0.02$) estimated directly from measurements (*SI Appendix*, Fig. S1).

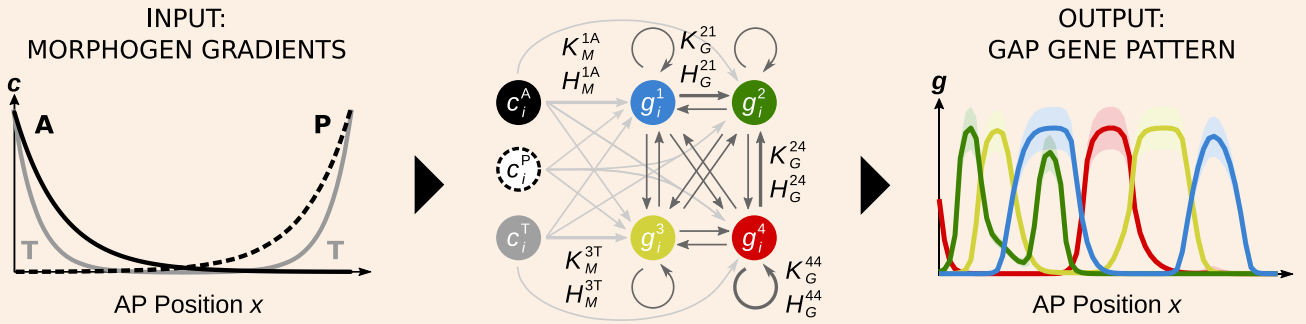
The model is solved in two steps (*SI Appendix*, section 2). In step one, we numerically integrate the deterministic part of the equation system defined above to obtain the mean expression levels at each position along the embryo axis at time T. Nuclear divisions are incorporated by doubling the maximal expression rate at the experimentally determined division times (i.e., $\rho_{\max}$ is only reached after the last division before T). In step two, the means are used to compute the full covariance matrix of the noise fluctuations in the gap protein levels, describing noise magnitude (on-diagonal matrix elements) and correlations (off-diagonal matrix elements) across space and gap gene species.

(see Box 1, continued)

Box 1. continued.

These two quantities are used to compute the decoding map (26) and the corresponding positional information $I(\mathbf{g};x)$ (25, 27) (*SI Appendix*, Fig. S2 and section 3).

Maximal copy number of gene products per nucleus and extrinsic noise imply that positional information must be upper bounded. To reach its absolute maximum value of $\log_2(N) \approx 6$ bits (error-free specification of N nuclei), these constraints would have to be lifted, implying a slower developmental process (due to lower protein and mRNA degradation rates) and/or a higher metabolic cost (due to higher transcription and translation rates). Within set rate constraints, various networks differ in the amount of actual gene expression, which we quantify by "resource utilization" (RU), the average expression across all gap genes and positions. RU $= 1$ means that gap gene expression is fully induced, proceeding at the maximal rate, in every nucleus; for the *Drosophila* WT pattern, RU ≈ 0.2 (*SI Appendix*, section 5).



Box Fig. 1: Spatial-stochastic model for gap gene expression. The gap gene regulatory network (center; colored circles = gap genes; grayscale circles = maternal morphogens) in each nucleus transforms maternal inputs with known spatial profiles (*Left*) into a gap gene expression pattern at readout time T (*Right*; solid lines = computed mean expression; shade = computed SD). Each interaction in the network stands either for the feed forward (FF) regulation of a gap gene by a morphogen input (light gray arrows), or for the regulation of a gap gene by other gap genes or by itself (cross- and self-regulation, dark gray arrows), and is described by two parameters (concentration threshold K and regulatory strength H; several parameter pairs are shown, corresponding to nearby thicker arrows). All parameters denoted by regulatory arrows are jointly optimized to maximize positional information $I(\mathbf{g};x)$.

and their morphogen inputs; we refer to all these parameters together as $\boldsymbol{\theta}$ (*SI Appendix*, Table S3). As an example, for each gene regulated by another, there is a parameter that measures the concentration at which the regulator exerts a half-maximal activating or repressive effect on its target and another parameter that measures the strength of this regulatory interaction. Different points in this 50+ dimensional space describe a wide spectrum of regulatory networks and their diverse expression patterns, most of which are very unlike the real one, even though they are generated by networks that are physically possible given the known component parts.

For any set of parameters, we simulate the time evolution of our model, evaluating the mean spatial pattern of expression for all four gap genes and their (co)variability at every nuclear location along the AP axis. These calculations, carried out in the Langevin formalism, are complex yet numerically tractable (*SI Appendix*, section 2); they properly account for maternal morphogen gradient variability and intrinsic biochemical stochasticity.

Positional information $I(\mathbf{g};x)$ is the mutual information between the set of gap gene expression levels $\mathbf{g} \equiv \{g_1, g_2, g_3, g_4\}$ and the AP coordinate x (7, 24, 25, 27). This quantity can be computed from the means and covariances of gap gene expression, which are the results of our model simulation at fixed $\boldsymbol{\theta}$ (Box 1 and *SI Appendix*, section 3 and Fig. S2). If the gap gene system indeed has been strongly selected to maximize positional information at some readout time T, then the real network should be near the optimal setting of parameters, $\boldsymbol{\theta}^* = \text{argmax}_{\boldsymbol{\theta}}\, I(\mathbf{g}(T);x)$. This problem is well posed because there are physical limits on the resources available to represent and transmit information: The maximal rates of molecular synthesis combine with degradation rates to limit the maximum number of molecules for each chemical species, setting the scale of the noise which in turn limits information transmission. In addition to constraining the maximal synthesis rates, it also makes sense to constrain the mean number of molecules actually used by the embryo, and again we can estimate this number from experiment.

We have previously solved simplified versions of this constrained optimization problem on small subnetworks, which inform the regulatory topology choices we make here (42, 43, 48–51). For example, maximizing the information that the steady-state gene expression level can provide about a single regulatory input leads to a nontrivial induction curve; this happens because maximizing information involves a compromise between minimizing noise and making full use of the available dynamic range. In addition, we have shown that where multiple genes respond to a common activating input (as with gap genes and their maternal inputs), they should mutually cross-repress to reduce redundancy. Last, we also understood the beneficial role of gap gene product diffusion and subcritical self-activation in simple genetic circuits, providing intuitions and an interpretative framework for the results we report below. Taken together, these previous results suggest that we will find nontrivial optimal solutions also in the

more complex case studied here and motivate the rich MWC parameterization we chose to use (Box 1).

Despite this extensive theoretical groundwork, understanding the entire network at the level where comparisons with data are possible required a new computational strategy (Box 1). We thus employed a large scale numerical approach, combining simulation and optimization (Fig. 1). We emphasize that the class of mechanistic regulatory network models that we explore is rich enough to predict experimentally measurable noisy spatial gene expression patterns but also bounded by quantitative biophysical constraints. Taken together, this opens up a route toward deriving the first ab initio prediction for a gene regulatory network in a biophysically realistic context.

Comparing Optimal Networks with the Real Network

We used a custom simulated annealing code to optimize the gap gene system for positional information (Fig. 2*A*). Briefly, we treat the positional information as the (negative) energy in a Monte Carlo simulation, with parameters as the underlying state. Changes are proposed one parameter at a time and accepted or rejected by the Metropolis criterion; as we iterate we slowly lower the effective temperature of the simulation. Once we have "cooled" sufficiently, the optimum is refined using a final round of gradient descent. Details are in *SI Appendix*, section 4.

We first focused on our main hypothesis: Is there a gap gene network that maximizes positional information and generates patterns of gap gene expression similar to those of the wild-type (WT) fruit fly? This question is complicated for multiple reasons. On the one hand, our optimization problem may not have any nontrivial solutions, or it may have solutions that we cannot find using our stochastic optimization, due to the curse of dimensionality. Alternatively, our optimization problem may have a multitude of locally optimal solutions, each corresponding to a maximum of positional information in the parameter space, but with very different gene expression patterns. In this latter case, it is technically challenging to find a particular local optimum or ascertain whether such an optimum exists close to the observed pattern. In what follows, we first describe our search for a *Drosophila*-like optimal solution, and second, explore the ensemble of *all* optimal solutions we could find.

To look for a *Drosophila*-like optimal solution, we first biased our search toward the known patterns of gene expression using a method that interpolates smoothly between optimization and pure data fitting (52); we then removed the bias to be sure that we have found a true optimum. This approach works by

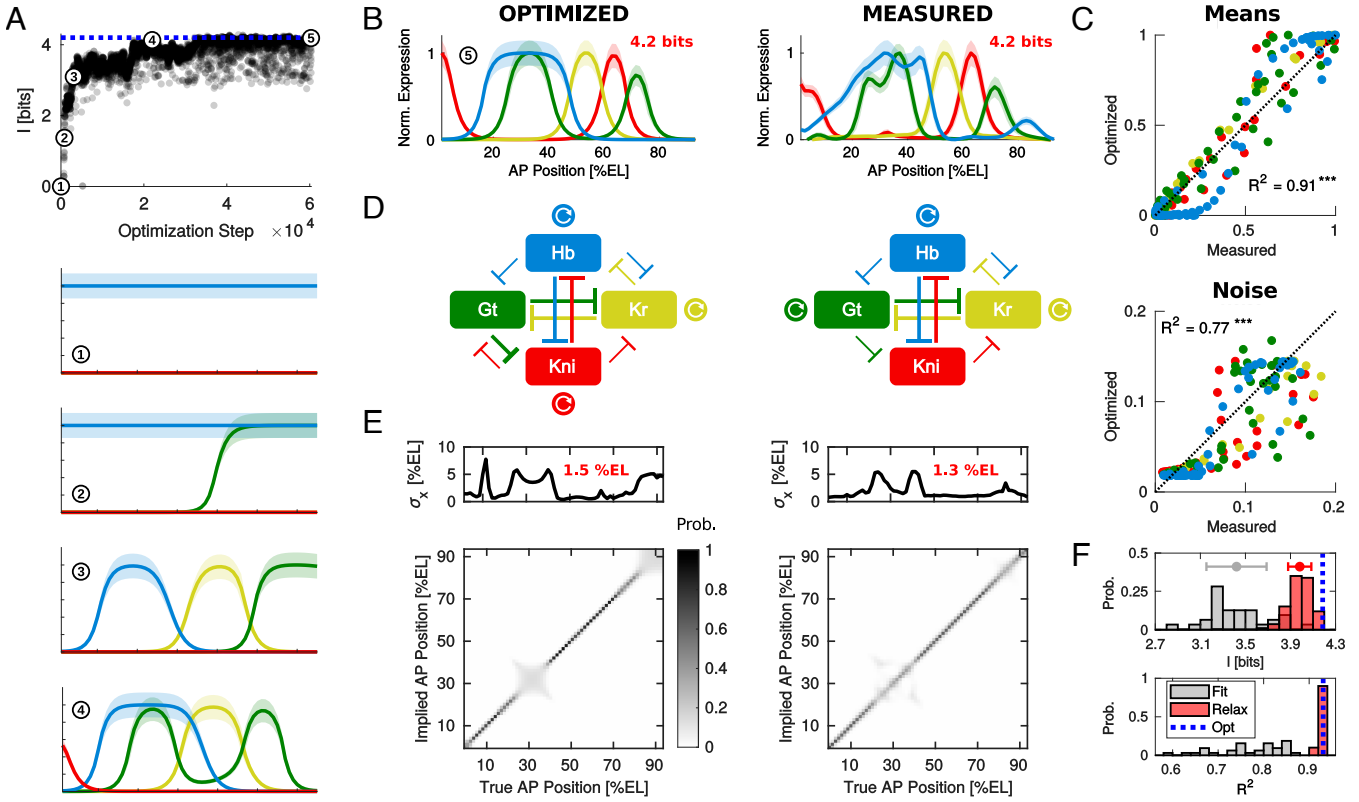


Fig. 2. Networks that maximize information transmission recapitulate the measured gap gene expression patterns and the regulatory network interactions. (*A*) Positional information increases during a single optimization run, starting with the homogeneous profile at 0 bits (encircled 1), proceeding through more complex spatial patterns (encircled 2 to 4), to the final solution (encircled 5, pattern in panel *B*) that reaches ~4.2 bits (dashed blue line). (*B*) Predicted optimal (*Left*) vs. measured gap gene expression pattern (*Right*; 47), 40 min into NC14 (blue = *hunchback*/Hb; green = *giant*/Gt; yellow = *Krüppel*/Kr; red = *knirps*/Kni; shade = SD in gene expression). Positional information estimate from data is consistent with that reported in ref. 24. (*C*) Measured vs. predicted mean expression (*Top*) and variability (*Bottom*) are highly correlated (color code as in *B*; Pearson $p < 10^{-3}$). (*D*) Predicted gap gene regulatory network (*Left*; blunt arrows = repression; circular arrows = self-activation) vs. literature-based reconstruction (*Right*; 21). Interaction strength is depicted by the width of the arrows. (*E*) Predicted (*Left*) vs. measured (*Right*) decoding map (*Bottom*) shows a nearly unambiguous code (diagonal band) with ~1.5% median positional error and few outlier positions (*Top Inset*) (26). (*F*) Fitting the model to mean WT gap gene expression profiles yields lower positional information (upper histogram; gray bars = distribution over replicate fits starting with random parameters, N = 32; red bars = distribution over replicate fits "relaxed" from the optimized solution, N = 250; horizontal error bars show mean ± SD), compared to the optimized solution (blue dashed line). The optimal solution better recapitulates the mean WT pattern compared to model fitting with random initial parameters; "relaxation" from the optimal solution does not significantly improve the fit (lower histogram showing goodness-of-fit as R^2; colors as above).

augmenting the function we are trying to optimize—positional information—with a weak "force" that pulls the gene expression patterns toward those observed in WT embryos. A key innovation here is to keep this effective force small by a proper choice of its Lagrange multiplier, thereby ensuring that stochastic annealing is dominated by optimization of positional information and not by fitting (*SI Appendix*, section 4). We emphasize that once a solution is found in this manner, we verify explicitly that it is indeed a true maximum, by numerically evaluating the derivatives of positional information with respect to all its parameters (*SI Appendix*, Fig. S3). Fig. 2*A* shows the optimization trajectory (increasing positional information and corresponding expression patterns) that found an optimal solution very close to the experimentally measured one, compared next.

Fig. 2*B* compares the gap gene expression profiles generated by the optimized network to data. The match in mean expression profiles is very good (Fig. 2*C*) although not perfect, a point we briefly return to later. The predicted gap gene variability similarly recapitulates the measured behavior.

The mechanistic nature of our model allows us to understand how the optimal pattern emerges. For example, the precision of the system output, manifested in the low variability (∼10%) of gap gene expression levels at a fixed position, is achieved through a combination of temporal averaging and spatial averaging via diffusion, which substantially reduces noise components transmitted from upstream regulators and morphogens (42, 53, 54). The spatial patterns of expression in the optimal solution are shaped significantly by mutual repression and self-activation, closely mimicking what has been inferred about the structure of the network from genetic interventions (Fig. 2*D*). Optimization correctly predicts strong mutual repression between *hunchback* and *knirps*, between *giant* and *Krüppel*, as well as most weak repressive interactions and self-activation of *hunchback* (54); see *SI Appendix*, Fig. S4 for detailed analysis of regulatory interactions. Together, these factors combine to encode positional information nearly unambiguously, with a median positional error of ∼1.5% (Fig. 2*E*); even the elevated positional uncertainty around the cephalic furrow and in the far posterior is consistent between the optimal prediction and the real embryo (26).

We emphasize that physical constraints are an essential building block for the predictions we report here; they are also the only route by which quantitative empirical data inform our calculations. Constraints enter not only in the form of a fixed embryo and nuclear geometry and readout time (Box 1) but specifically as maximal molecular production and degradation rates for gap gene mRNA and proteins. These rates limit the maximal number of independent gap gene products per nucleus and correspond, on a molecular level, to fully induced transcription and translation unfolding at the maximum possible pace. Within these maximal constraints, which we hold fixed to experimentally estimated values (Box 1 and *SI Appendix*, sections 1 and 5), expression still varies across gap genes and spatial locations. Averaging over these variations defines the *average* resource utilization (RU, Box 1) at each setting of the parameters. Our optimal solutions are therefore further constrained to match the empirically determined RU of the *Drosophila* WT pattern; we explore alternative RU values further below.

Ab initio optimization performed here makes only minimal use of prior knowledge to derive a wide range of predictions: while we fix a small number of resource constraints to measured values as noted above, optimization predicts the values of many more free parameters that determine the gap gene interaction network and its resulting expression patterns. The biasing procedure described above weakly couples to measured expression data to make our search more efficient, but it is not essential for testing our primary hypothesis. Taken together, the optimization approach stands in stark contrast to traditional model fitting, which uses gap gene expression data in full to set the constraints as well as infer all of the regulatory parameters (52). This distinction has three important consequences.

First, in traditional model fitting the objective functions are purely statistical (e.g., maximum likelihood in some model, mean-square-error, ...), lacking any biological interpretation. In contrast, our strategy of optimizing positional information is grounded by a meaningful and independently measurable phenotype of the patterning system. Indeed, our optimal solution (Fig. 2 *A* and *B*) reaches 4.2 bits, to be compared with 4.2 to 4.3 bits estimated directly from data (24, 25). The actual information transmitted through this network thus agrees with the optimum within ∼0.1 bit, or ∼2.5% of the total.

Second, if we perform a conventional fitting of parameters, minimizing the mean-square-error of the predicted mean gap gene expression, the best fits underestimate the positional information (Fig. 2 *F*, *Top*). This is because such fitting fails to take into account the structure of the noise in the system, and its functional consequences. Furthermore, the goodness-of-fit landscape is very rugged, so that many fits starting from random initial parameters get trapped in suboptimal solutions; optimizing the positional information regularizes this ruggedness and leads to solutions whose goodness-of-fit cannot be significantly improved further (Fig. 2 *F*, *Bottom*). In other words, the network that we predict through optimization cannot be brought significantly closer to the real network even by fitting to data, either starting with random parameters (Fig. 2*F*, "Fit") or relaxing from the information-maximizing solution ("Relax") . As an aside, this implies that any quantitative deviations we observe in Fig. 2*B* between the optimal prediction and the WT pattern arise because the class of models we consider still is a bit too restrictive in its expressive power.

Third, optimization with no biasing toward the WT expression pattern can identify novel, locally optimal solutions that maximize positional information and are thus functionally near degenerate, yet are qualitatively different from the gene expression patterns observed in the real embryo. These solutions are found by independent optimization runs from random initial parameter combinations and can terminate at very different points in the high-dimensional parameter space. If we focus on fitting to the observed patterns, we would have no way of finding or assessing these alternative networks. We examine these solutions next, as a segue into related evolutionary questions.

Multiple optimization runs produce diverse solutions that locally maximize positional information while not exceeding the resource utilization of the wild-type pattern. Together, these solutions constitute the "optimal ensemble". A natural comparison is provided by the "random ensemble", where parameters $\boldsymbol{\theta}$ are drawn independently and uniformly from broad but realistic intervals. The vast majority of the random ensemble forms no patterns ($I \approx 0$ bits); we thus focus our comparisons in Fig. 3 on a functionally relevant nontrivial subset of the random ensemble. Optimization for positional information automatically leads to significantly lower positional error (Fig. 3*A*), a higher number of boundaries where gap gene expression switches from low to high or vice-versa ("slopes"), more uniform utilization of resources across gap genes (i.e., the four gap genes each drive close to one-quarter of total expression), as well a slight but significant

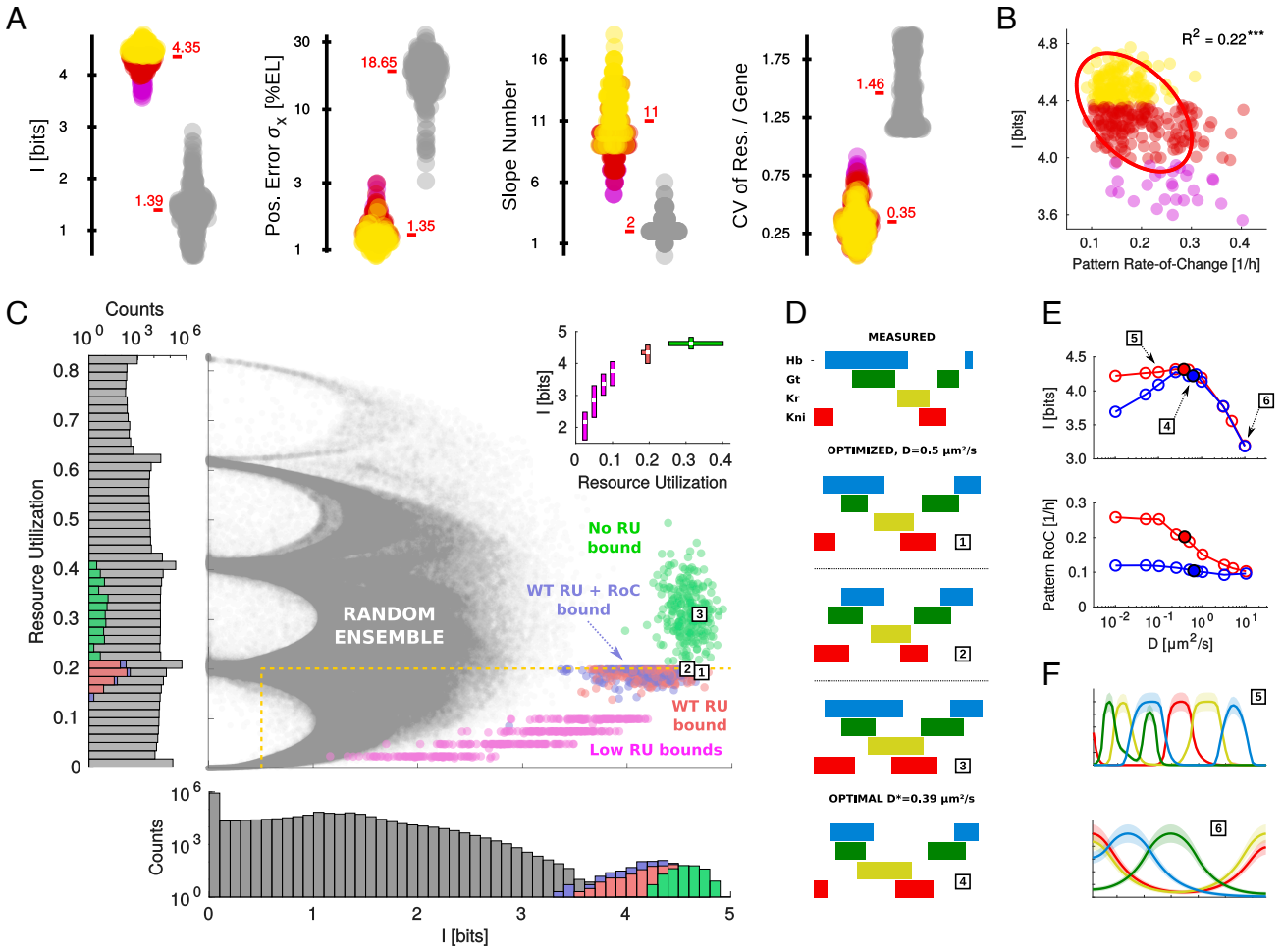


Fig. 3. Optimal and random gap gene network ensembles. (*A*) Patterning phenotypes for optimal ensemble (color, solutions from "WT RU" ; N = 324) vs. random ensemble (gray, including only solutions >0.5 bit that are at or below WT resource utilization; random choice of N = 319 solutions meeting this criterion) reveal that high positional information (leftmost; violet, red, yellow indicate lowest, middle, highest third of the information interval) correlates with low positional error, high number of gap gene "slopes," and a more uniform utilization of resources across the four gap genes (red numbers = ensemble medians). (*B*) Within the optimal ensemble, higher information correlates with higher dynamical stability, i.e., lower pattern rate-of-change [each dot = one optimal solution; red ellipse = 1 SD contour in the I vs. RoC plane; color code and optimal ensemble as in (*A*); N = 324]. (*C*) Random (gray, N = 1,580,129) and various optimal ensembles (red = resource utilization bounded by *Drosophila* WT denoted by a dashed horizontal yellow line, N = 350; magenta = progressively smaller resource utilization; blue = WT resource utilization plus a bound on pattern rate-of-change, N = 607; green = no resource utilization or rate-of-change bounds, N = 239) depicted in the information vs. resource utilization plane (each dot = unique parameter combination). Histograms in the margins show the raw counts of evaluated parameter combinations. *Inset*: Information vs. resource utilization (median, 0.1 to 0.9-quantile intervals over ensembles in the main panel shown as central white squares and ribbons, respectively). Three example solutions are depicted by boxed numbers and shown in panel (*D*). (*D*) Example optimal solutions (denoted by boxed 1 to 3) from panel (*C*), optimized at fixed gap product diffusion (D = 0.5 $\mu m^2/s$). An example solution (boxed 4) where D was also optimized, from the ensemble shown in panel (*E*). All these solutions qualitatively match WT gap gene expression domains (*Top*) and the regulatory network architecture. (*E*) Mean positional information (*Top*) and pattern rate-of-change (*Bottom*) as a function of the gap gene diffusion constant D (empty circles = mean across optimal ensembles), capped at WT resource utilization (red) or with an additional rate-of-change constraint (blue). Solid red and blue circles = mean values for the case where D itself is also a free optimization parameter; yellow shade = broad range of D consistent with literature reports. Example solution with optimal D is denoted by boxed 4 and is shown in panel (*D*); two example solutions at suboptimal D are denoted by boxed 5 and 6 and are shown in panel (*F*). (*F*) Two example solutions optimized at lower-than-optimal (*Top*) and higher-than-optimal (*Bottom*) diffusion constant values. Unless specified otherwise, the averages plotted in this figure were computed over $N \gtrsim 100$ independent optimization runs starting from random initial parameters.

overallocation of resources in the anterior, as can be seen in data as well (see *SI Appendix*, Fig. S5 for further patterning phenotypes). Within the optimal ensemble, solutions with higher information tend to be more dynamically stable at readout time (Fig. 3*B*), which we quantify by pattern rate-of-change (RoC), the mean (absolute) temporal derivative of gap gene expression (55). Low RoC is relevant since pair-rule genes appear to read out gap gene expression via fixed decoding rules (26, 29), implying that temporally varying solutions could cause larger spatial drifts in pair-rule stripes.

Networks in the random ensemble that transmit large amounts of information are exceedingly rare: The probability of drawing a network with positional information of 4 bits or more by chance is $\ll 10^{-6}$ (Fig. 3*C*). In contrast, optimization strongly and robustly enriches solutions above 4 bits (Fig. 3*C*). In our optimization, we have constrained the maximum numbers of molecules, and the real embryo uses ~20% (RU = 0.2) of this maximum, on average. This resource utilization appears necessary for high-information solutions, whereas permitting more utilization within the same maximal rate limits does not noticeably increase positional information. In fact, among $>10^3$ optimization runs we never found a solution exceeding 5 bits, indicating that such information values likely cannot be accessed within realistic constraints.

The random and the optimal ensemble are closely related to the evolutionary concepts of the *neutral* and the *selected* phenotype distributions (56). The random ensemble delineates what is physically possible in the absence of selection for function: the entire space of expression patterns and derived phenotypes that can be achieved with a fixed cast of molecular players by sampling a broad range of regulatory parameters. This is analogous to the definition of a neutral distribution over phenotypes in evolutionary biology, which likewise samples the range of phenotypes reachable by randomly sampling the corresponding genotypes, either computationally with some model genotype–phenotype map (57) or experimentally (58, 59). In contrast to the random ensemble, the optimal ensemble delineates solutions that maximize function within fixed physical constraints. How closely natural selection *could* approach this optimality (as quantified by the selected phenotype distribution), or indeed *has* approached it (via the actual WT pattern), depends on selection strength and its history, genetic load, linkage disequilibrium, and other limitations that are of negligible concern to in silico optimization. Successful prediction of the pattern in Fig. 2*B* confirms our primary hypothesis and implies that selection was sufficiently strong to overcome such limitations and push the gap gene system beyond neutral evolution and evolutionary tinkering (60) toward optimality (6, 52, 61). Even in strictly ab initio runs with zero bias toward the WT pattern we repeatedly find solutions that closely reproduce the overall size and placement of expression domains in *Drosophila* (Fig. 3*D* and *SI Appendix*, Fig. S6), the encoded positional information, as well as the regulatory interactions. Tantalizing early experimental work suggests that dipteran species related to *Drosophila* may feature a broadly consistent gap gene domain arrangement whose expression domains are, however, shifted (62, 63) or swapped (64), as we find in our optimal ensembles (*SI Appendix*, Fig. S6).

Taken together, our results imply a nuanced interplay of evolutionary necessity and contingency. In the 50+ dimensional parameter space of possible networks, there is a highly nonrandom, locally optimal solution that produces expression patterns very similar to what we see in real fly embryos, but there are many other local optima that transmit about the same amount of positional information; all of these solutions are rare in the random ensemble. Careful analysis of optimal ensembles suggests that three or more interacting gap genes are necessary for a wide spectrum of optimal solutions, while with less genes the optimal patterns are much more stereotyped (*SI Appendix*, Figs. S8–S10). It is an open question whether network architectures that enable such a wide spectrum could themselves be evolutionarily favored due to higher evolvability (65). It is also an open question whether alternative optima quantitatively recapitulate gap gene patterns seen in other dipterans or whether the degeneracy is removed by selection for additional phenotypes beyond positional information.

Alternatives to the Real Network

Our theory provides a framework within which we can explore tradeoffs beyond the structure of the gap gene network. In our standard optimal ensemble ("WT RU" in Fig. 3*C*), we have taken the effective diffusion constant of gap gene products to be a fixed physical property of the embryo, $D = 0.5\ \mu m^2/s$, in line with existing measurements (41). In fact, D lumps together multiple microscopic processes, including actual diffusion of gap gene products through the cytoplasm, import and export across the nuclear membrane, and binding/unbinding to various cytoplasmic components. It is therefore likely that the effective D could change and be tuned by evolution, so long as it remains far below the "speed limit" set by the bare cytoplasmic diffusion, which appears to be the case (66). Therefore, we can view D as one more parameter to be optimized.

Remarkably, we find that there is a broad optimum for D at the experimentally estimated value (Fig. 3*E*). Larger diffusion constants lead to a precipitous drop in information even when all other parameters $\boldsymbol{\theta}$ are reoptimized because high diffusion smooths gap gene profiles to the extent that adjacent nuclei can no longer be distinguished reliably (Fig. 3 *F*, *Bottom*). On the other hand, slower diffusion does not serve as effectively to average over local super-Poisson noise sources; the optimization algorithm compensates by finding parameters that generate more and steeper transitions between high and low expression levels (Fig. 3 *F*, *Top*), but even these unrealistic patterns do not transmit quite as much information. Thus, a single parameter displaced away from its optimum causes significant decreases in positional information; to lessen the impact the optimization algorithm adjusts other network parameters, driving the predicted patterns of gene expression away from what we see in the real embryo.

We next address the question of evolutionary necessity and sufficiency. To this end, we make structural changes to the network and then reoptimize all of its parameters to explore "alternative evolutionary histories" that could have unfolded with changed molecular components or mechanisms. As an example, Fig. 4*A* characterizes solutions obtained using 1, 2, $\cdots$, 5 gap genes, subject to the same *total* resource utilization as the WT, plotting the positional information vs. the rate at which expression patterns are changing at readout time. Networks that transmit 4 bits or more—as in the real embryo—are completely inaccessible using only one or two gap genes, even though these networks are allowed to utilize the same total number of molecules as in the optimal four-gene networks above. Such networks also do not produce a wide spectrum of optimal solutions (*SI Appendix*, Figs. S8 and S9).

With three gap genes, the optimized networks grow significantly more diverse and can transmit a total information comparable to what is seen experimentally, but detailed analysis reveals that the three-gene networks all have local defects where the positional error spikes above 5 to 10% of the embryo length (*SI Appendix*, Fig. S7), in contrast to the much more uniform distribution of precision along the length of the real embryo (24); we can quantify this by looking at the variations in the positional error along the AP axis (Fig. 4 *A*, *Inset*). This failure of the three-gene networks arises because they cannot realize a sufficient number of slopes or switches between high and low expression levels. Four gap genes thus are necessary to ensure that high positional information translates into defect-free patterning not just on average, but uniformly across the entire AP axis of every embryo (24). The marginal benefit of a putative fifth gap gene appears small and may not be sufficient to establish the required additional regulatory mechanisms or to maintain them at mutation–selection balance (67).

We can explore, in the same spirit, the role of the multiple maternal morphogens. In the fly embryo, the anterior (A, Bicoid), posterior (P, Nanos), and terminal (T, Torso-like) systems jointly regulate gap gene expression (26). In our model, we can remove one or two of these inputs and reoptimize all the parameters of the gap gene network, and find that there are moderate yet statistically significant losses in both positional information and stability (Fig. 4*B*). The impact of primary morphogen deletions is limited because the optimization algorithm adjusts the gap gene cross-regulation parameters to restore informative spatial patterns. This ability, however, disappears entirely if gap gene cross-regulation

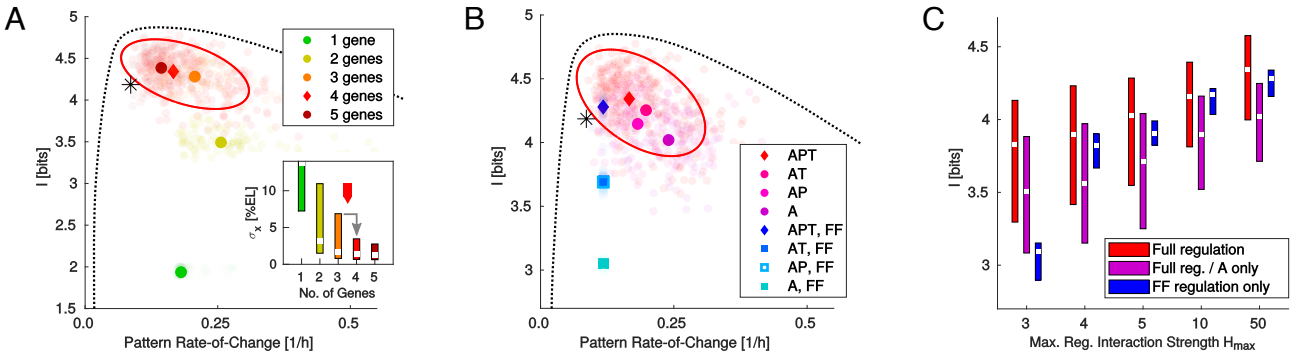


Fig. 4. Necessity and sufficiency of gap gene regulatory network mechanisms. (*A*) Optimal ensembles (transparent symbols = individual optimal solutions; solid symbols = ensemble medians) for networks with 1, 2, ⋯, 5 gap genes (legend colors) optimized at the WT resource utilization (for reference, red diamond + red ellipse at 1 SD = WT-like optimal ensemble from Fig. 3*B*). Solutions delineate the accessibility frontier (dotted black line for visual guidance) in positional information (I) vs. pattern rate-of-change (RoC) plane. (*Inset*) While the median positional error (white squares) plateaus for optimal networks with 3 gap genes or more, the variability in positional error (ribbons denote 0.1- and 0.9-quantiles across AP positions in individual embryos) significantly shrinks only with 4 gap genes or more (red arrow). (*B*) Optimal ensembles for networks responding to different subsets of the three morphogens (A = anterior; P = posterior; T = terminal; red/purple circle symbols = ensemble median; red dots, diamond, ellipse = WT-like ensemble as in *A*). Optimal ensembles for purely feed-forward (FF) networks, i.e., no gap gene self- or cross-regulation, denoted in bluish hues (legend). (*C*) Positional information in optimal ensembles with gap gene self- and cross-regulation (red = with APT morphogens, purple = with A morphogen only) or without self- and cross-regulation (blue = "FF only" networks with APT morphogens), for different maximal regulatory strength, H_{max}. White squares denote the median over optimal ensembles in (*A* and *B*); ribbons denote the corresponding 0.1- and 0.9- quantile ranges. Lower H_{max} values potentiate the advantage of optimal networks with self-regulation, cross-regulation, and three maternal morphogens, compared to architectures deficient in these mechanisms. Unless specified otherwise, the averages plotted in this figure were computed over $N \gtrsim 100$ independent optimization runs starting from random initial parameters.

is not permitted and the gap gene network is feed-forward (FF) only (light gray arrows in Box figure, Fig. 4*B*); in the absence of cross-regulation, removal of each primary morphogen system is associated with a large decrease in positional information.

Fig. 4*B* suggests that stable, high information patterns could be generated by utilizing all three maternal morphogens even without the ability of gap genes to regulate one another. But in the absence of cross-regulation, the time scale for variations in the pattern is determined solely by the intrinsic lifetime of the most stable species (mRNA). In contrast, gap gene interactions allow for the emergence of longer time scales which both slow the variations and can reduce noise by temporal averaging (49); possible evidence for these effects has been discussed previously (55). Evolutionarily, adding gap gene cross-regulation creates variability in the rate-of-change phenotype that could additionally be selected for. Indeed, the WT-like solution of Fig. 2*B* falls close to the accessibility frontier of Fig. 4*B* (star symbol), suggesting such a preference.

Last, we varied the maximal allowed strength of regulatory interactions, H_{max} (Box 1), in our model. This parameter determines how strongly each individual input, either a morphogen or a gap gene acting via self- or cross-regulation, can impact the expression of a target gap gene. While in simple thermodynamic models of gene regulation H_{max} sets the maximal number of transcription factor molecules that bind cooperatively to their target sites as they regulate a single gene, such molecular interpretation is too narrow for nonequilibrium regulatory schemes that are likely relevant for metazoans (44, 68, 69). Even though higher values of H_{max} do lead to steeper induction curves and more sensitive regulatory response around the half-induction point, mimicking implications of a higher "effective cooperativity," H_{max} should be construed more broadly than as requiring the binding of many TF molecules simultaneously.

Optimizations presented so far used $H_{max} = 50$, sufficiently high not to impose any functional constraint. As H_{max} is lowered and the constraint kicks in, the optimal feed-forward solution of Fig. 4*B* (dark blue) suffers large drops in encoded positional information (Fig. 4*C*). Optimal feed-forward architectures are thus heavily reliant on regulatory interaction strengths that appear unrealistic; at $H_{max} = 50$ a change in single regulatory input concentration of 10 percent can change the target gene induction by more than five-fold. Further, one might have been tempted to interpret Fig. 4*B* by saying that cross-regulation and multiple input morphogens provide alternative or even redundant paths to high information transmission, but we see that this degeneracy is lifted when we limit the regulation strength to more realistic levels. From an evolutionary perspective, gap gene cross-regulation, therefore, is favorable for two reasons: First, it generates temporally stable phenotypes at the accessibility frontier (as in Fig. 4*B*); and second, it permits high information solutions also in networks where the strength of individual regulatory interactions is limited (as in Fig. 4*C*).

Discussion

The idea that living systems can approach fundamental physical limits to their performance, and hence optimality, goes back at least to explorations of the diffraction limit in insect eyes and the ability of the human visual system to count single photons (6). The specific idea that biological systems optimize information transmission emerged shortly after Shannon's formulation of information theory, in the context of neural coding (7, 70). Despite this long history, most efforts to derive the behaviors and mechanisms of biological systems from an optimization principle have been carried out in very simplified contexts, using functional models with a small number of parameters. Here we have instantiated these ideas in a realistic context, using models that permit interpretation in terms of molecular mechanisms and interactions, as well as direct connections to rich experimental observations.

We focused on the *Drosophila* gap gene system, one of the paradigms for developmental biology and for physical precision measurements in living systems (71). Our work extends previous mathematical models of this system (28–36, 72–75), as well as attempts to predict it ab initio (76–78). In particular, we systematically incorporate the unavoidable physical sources of noise, highlighting how patterning precision can emerge from noisy signals by a synergistic combination of multiple mechanisms. Crucially, we do not *fit* the parameters of the model

to data, but rather *derive* them via optimization. In contrast to previous prediction attempts, our constraints and comparisons to data are not stylized, but fully quantitative and commensurate with the precision of the corresponding experiments.

We have found networks that maximize positional information with a limited number of molecules, and there is at least one local optimum quantitatively matching a large range of observations in the wild-type *Drosophila* system: the spatial patterns of expression and variability, the resulting decoding map, the molecular architecture of the network, and even subtle biases in spatial resource utilization. This result validates the optimization hypothesis.

While the observed gap gene network is a solution to the problem of optimizing positional information, this does not mean that the evolutionary trajectory leading to modern flies is anything like our optimization algorithm. When viewed in the broader context of arthropod developmental patterning systems, the *Drosophila* network can be seen as a synchronous limiting version of a more dynamic, clock-and-wavefront-like patterning strategy (35, 36, 76, 79, 80). Importantly, these two viewpoints are not incompatible: Multiple evolutionary routes could lead toward the same local optimum. Our approach optimizes just for the information content of the final pattern, without making statements about the intermediate patterns that a single embryo passes through. Moreover, some of our results, such as the rate-of-change stability (Fig. 3*B*), dynamical slow-down (*SI Appendix*, Fig. S5), and timescale lengthening of the best-scoring patterns could tentatively point to the emergence of dynamical systems that maximize instantaneous positional information by momentarily "freezing," within a defined readout time window, an otherwise quite dynamic gene expression profile.

Our optimization framework provides a platform for exploring the necessity and sufficiency of various network components that ensure maximal information transmission. Using this framework to deliver on our introductory questions, we have established that four gap genes appear necessary for defect-free patterning and that the apparent redundancy between the three maternal morphogens and gap gene cross-regulation is lifted under reasonable constraints on the strength of regulatory interactions.

Numerical optimization is not evolutionary adaptation: our results depend on the postulated fitness proxy, i.e., positional information, as well as on the scope of the search space within which we look for optimal solutions. We believe that there are strong justifications for our choices. First, we previously reported empirical signatures of optimality in positional encoding (15, 24, 26). Second, we consider a space of models for the gap gene network at least as rich as has been tried in the past, with parameters that could plausibly be tuned by evolutionary changes in gap gene enhancers, which is believed to be the dominant, fast mode of adaptation for gene regulatory networks.

Despite these caveats, our results provide an intriguing perspective on evolutionary questions. Discussions about the interplay of evolutionary optimization and developmental constraints, necessity versus contingency, and limits to selection have a venerable history (38, 81). Rather than discussing these questions in qualitative terms, here we explored the role of physical constraints and tradeoffs quantitatively, in the context of an expressive mechanistic model, using the powerful concepts of the random and the optimal ensembles. In the words of Jacob (60), the random ensemble delineates the space of the "possible." Within this space, our optimization principle is equivalent to the hypothesis that positional information is a proxy for fitness and selection pressure is strong, identifying a much more restricted optimal ensemble. It is surprising that this principle alone is sufficient to ensure that the optimal ensemble contains a solution very close to Jacob's "actual," the *Drosophila* gap gene network that we observe and measure.

Data, Materials, and Software Availability. There are no data underlying this work.

ACKNOWLEDGMENTS. We thank Nicholas H. Barton for his comments on the manuscript, Benjamin Zoller for helpful discussions, and Aleksandra Walczak and Curtis Callan for early collaborations that shaped this work. Special thanks to Eric F. Wieschaus for many persistently inspiring conversations. This work was supported in part by the Human Frontiers Science Program; the Austrian Science Fund (FWF P28844); by the European Research Council grant DynaTrans (101118866); by U.S. NSF, through the Center for the Physics of Biological Function (PHY–1734030); by NIH Grants R01GM097275, U01DA047730, and U01DK127429; by the John Simon Guggenheim Memorial Foundation; and by the LOEWE priority program "Center for Multiscale Modeling in Life Sciences" (CMMS), sponsored by the Hessian Ministry for Science and Research, Arts and Culture (HMWK).

Author affiliations: [a]Institute of Science and Technology Austria, Klosterneuburg AT-3400, Austria; [b]Frankfurt Institute for Advanced Studies, Frankfurt am Main DE-60438, Germany; [c]Joseph Henry Laboratory of Physics and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544; [d]Department of Stem Cell and Developmental Biology, UMR3738, Institut Pasteur, Paris FR-75015, France; and [e]Center for Studies in Physics and Biology, Rockefeller University, New York, NY 10065

1. G. A. Parker, J. M. Smith, Optimality theory in evolutionary biology. *Nature* **348**, 27–33 (1990).
2. B. Walsh, M. Lynch, *Evolution and Selection of Quantitative Traits* (Oxford University Press, 2018).
3. F. M. Rieke, D. A. Baylor, Single photon detection by rod cells of the retina. *Rev. Mod. Phys.* **70**, 1027–1036 (1998).
4. H. B. Barlow, The size of ommatidia in apposition eyes. *J. Exp. Biol.* **29**, 667–674 (1952).
5. H. C. Berg, E. M. Purcell, Physics of chemoreception. *Biophys. J.* **20**, 193–219 (1977).
6. W. Bialek, *Biophysics: Searching for Principles* (Princeton University Press, Princeton, NJ, 2012).
7. G. Tkačik, W. Bialek, Information processing in living systems. *Annu. Rev. Condens. Matter Phys.* **7**, 89–117 (2016).
8. Y. Karklin, E. P. Simoncelli, Efficient coding of natural images with a population of noisy linear-nonlinear neurons. *Adv. Neural. Inf. Process. Syst.* **24**, 999 (2011).
9. B. A. Olshausen, D. J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).
10. E. C. Smith, M. S. Lewicki, Efficient auditory coding. *Nature* **439**, 978–982 (2006).
11. W. Młynarski, J. H. McDermott, Ecological origins of perceptual grouping principles in the auditory system. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25355–25364 (2019).
12. R. U. Ibarra, J. S. Edwards, B. O. Palsson, Escherichia coli k-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **420**, 186–189 (2002).
13. A. Tero *et al*., Rules for biologically inspired adaptive network design. *Science* **327**, 439–442 (2010).
14. E. Katifori, G. J. Szöllősi, M. O. Magnasco, Damage and fluctuations induce loops in optimal transport networks. *Phys. Rev. Lett.* **104**, 048704 (2010).
15. G. Tkačik, C. G. Callan, W. Bialek, Information flow and optimization in transcriptional regulation. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12265–12270 (2008).
16. Y. Savir, E. Noor, R. Milo, T. Tlusty, Cross-species analysis traces adaptation of rubisco toward optimality in a low-dimensional landscape. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3475–3480 (2010).
17. L. Wolpert, Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25**, 1–47 (1969).
18. L. N. de la Rosa, G. B. Müller, *Evolutionary Developmental Biology* (Springer, 2021).
19. Z. D. Blount, R. E. Lenski, J. B. Losos, Contingency and determinism in evolution: Replaying life's tape. *Science* **362**, 655 (2018).
20. C. Nüsslein-Volhard, E. Wieschaus, Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795–801 (1980).
21. J. Jaeger, The gap gene network. *Cell. Mol. Life Sci.* **68**, 243–274 (2011).
22. J. Briscoe, S. Small, Morphogen rules: Design principles of gradient-mediated embryo patterning. *Development* **142**, 3996–4009 (2015).
23. P. A. Lawrence, *The Making of a Fly: The Genetics of Animal Design* (Blackwell Scientific Publications Ltd, 1992).
24. J. O. Dubuis, G. Tkačik, E. F. Wieschaus, T. Gregor, W. Bialek, Positional information, in bits. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16301–16308 (2013).
25. G. Tkačik, J. O. Dubuis, M. D. Petkova, T. Gregor, Positional information, positional error, and readout precision in morphogenesis: A mathematical framework. *Genetics* **199**, 39–59 (2015).
26. M. D. Petkova, G. Tkačik, W. Bialek, E. F. Wieschaus, T. Gregor, Optimal decoding of cellular identities in a genetic network. *Cell* **176**, 844–855 (2019).
27. G. Tkačik, T. Gregor, The many bits of positional information. *Development* **148**, dev176065 (2021).
28. L. Sánchez, D. Thieffry, A logical analysis of the drosophila gap-gene system. *J. Theor. Biol.* **211**, 115–141 (2001).

29. J. Jaeger *et al*., Dynamic control of positional information in the early *Drosophila* embryo. *Nature* **430**, 368–371 (2004).
30. J. Jaeger *et al*., Dynamical analysis of regulatory interactions in the gap gene system of *Drosophila melanogaster*. *Genetics* **167**, 1721–1737 (2004).
31. T. J. Perkins, J. Jaeger, J. Reinitz, L. Glass, Reverse engineering the gap gene network of *Drosophila melanogaster*. *PLoS Comput. Biol.* **2**, e51 (2006).
32. Manu, *et al*., Canalization of gene expression and domain shifts in the *Drosophila* blastoderm by dynamical attractors. *PLoS Comput. Biol.* **5**, e1000303 (2009).
33. Manu, *et al*., Canalization of gene expression in the *Drosophila* blastoderm by gap gene cross regulation. *PLoS Biol.* **7**, e1000049 (2009).
34. M. Ashyraliyev *et al*., Gene circuit analysis of the terminal gap gene huckebein. *PLoS Comput. Biol.* **5**, e1000548 (2009).
35. B. Verd, A. Crombach, J. Jaeger, Dynamic maternal gradients control timing and shift-rates for *Drosophila* gap gene expression. *PLoS Comput. Biol.* **13**, e1005285 (2017).
36. B. Verd *et al*., A damped oscillator imposes temporal order on posterior gap gene expression in *Drosophila*. *PLoS Biol.* **16**, e2003174 (2018).
37. R. Seyboldt *et al*., Latent space of a small genetic network: Geometry of dynamics and information. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2113651119 (2022).
38. J. M. Smith *et al*., Developmental constraints and evolution: A perspective from the Mountain Lake conference on development and evolution. *Q. Rev. Biol.* **60**, 265–287 (1985).
39. S. C. Little, M. Tikhonov, T. Gregor, Precise developmental gene expression arises from globally stochastic transcriptional activity. *Cell* **154**, 789–800 (2013).
40. B. Zoller, S. C. Little, T. Gregor, Diverse spatial expression patterns emerge from unified kinetics of transcriptional bursting. *Cell* **175**, 835–847 (2018).
41. T. Gregor, D. W. Tank, E. F. Wieschaus, W. Bialek, Probing the limits to positional information. *Cell* **130**, 153–164 (2007).
42. T. R. Sokolowski, G. Tkačik, Optimizing information flow in small genetic networks. IV. Spatial coupling. *Phys. Rev. E* **91**, 062710 (2015).
43. A. M. Walczak, G. Tkačik, W. Bialek, Optimizing information flow in small genetic networks. II. Feed-forward interactions. *Phys. Rev. E* **81**, 041905 (2010).
44. R. Grah, B. Zoller, G. Tkačik, Nonequilibrium models of optimal enhancer function. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 31614–31622 (2020).
45. G. Tkačik, T. Gregor, W. Bialek, The role of input noise in transcriptional regulation. *PLoS One* **3**, e2774 (2008).
46. P. S. Swain, M. B. Elowitz, E. D. Siggia, Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12795–12800 (2002).
47. J. O. Dubuis, R. Samanta, T. Gregor, Accurate measurements of dynamics and reproducibility in small genetic networks. *Mol. Syst. Biol.* **9**, 639 (2013).
48. G. Tkačik, A. M. Walczak, W. Bialek, Optimizing information flow in small genetic networks. *Phys. Rev. E* **80**, 031920 (2009).
49. G. Tkačik, A. M. Walczak, W. Bialek, Optimizing information flow in small genetic networks. III. A self-interacting gene. *Phys. Rev. E* **85**, 041903 (2012).
50. P. Hillenbrand, U. Gerland, G. Tkačik, Beyond the French flag model: Exploiting spatial and gene regulatory interactions for positional information. *PLoS One* **11**, e0163628 (2016).
51. T. R. Sokolowski, A. M. Walczak, W. Bialek, G. Tkačik, Extending the dynamic range of transcription factor action by translational regulation. *Phys. Rev. E* **93**, 022404 (2016).
52. W. Młynarski, M. Hledík, T. R. Sokolowski, G. Tkačik, Statistical analysis and optimality of neural systems. *Neuron* **109**, 1227–1241 (2021).
53. T. Erdmann, M. Howard, P. R. ten Wolde, Role of spatial averaging in the precision of gene expression patterns. *Phys. Rev. Lett.* **103**, 258101 (2009).
54. T. R. Sokolowski, T. Erdmann, P. R. ten Wolde, Mutual repression enhances the steepness and precision of gene expression boundaries. *PLoS Comput. Biol.* **8**, e1002654 (2012).
55. D. Krotov, J. O. Dubuis, T. Gregor, W. Bialek, Morphogenesis at criticality. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3683–3688 (2014).
56. M. Hledík, N. Barton, G. Tkačik, Accumulation and maintenance of information in evolution. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2123152119 (2022).
57. R. Grah, C. C. Guet, G. Tkacik, M. Lagator, Linking molecular mechanisms to their evolutionary consequences: A primer. *Genetics*, https://doi.org/10.1093/genetics/iyae191 (2024).
58. C. G. de Boer *et al*., Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nat. Biotechnol.* **38**, 56–65 (2020).
59. M. Lagator *et al*., Predicting bacterial promoter function and evolution from random sequences. *eLife* **11**, e64543 (2022).
60. F. Jacob, *The Possible and the Actual* (University of Washington Press, 1982).
61. G. Sella, A. E. Hirsh, The application of statistical physics to evolutionary biology. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9541–9546 (2005).
62. A. Crombach, M. A. García-Solache, J. Jaeger, Evolution of early development in dipterans: Reverse-engineering the gap gene network in the moth midge *Clogmia albipunctata* (psychodidae). *Biosystems* **123**, 74–85 (2014).
63. K. R. Wotton *et al*., Quantitative system drift compensates for altered maternal inputs to the gap gene network of the scuttle fly *Megaselia abdita*. *eLife* **4**, e04785 (2015).
64. Y. Goltsev, W. Hsiong, G. Lanzaro, M. Levine, Different combinations of gap repressors for common stripes in anopheles and *Drosophila* embryos. *Dev. Biol.* **275**, 435–446 (2004).
65. J. L. Payne, A. Wagner, The causes of evolvability and their evolution. *Nat. Rev. Genet.* **20**, 24–38 (2019).
66. T. Gregor, W. Bialek, R. R. de Ruyter van Steveninck, D. W. Tank, E. F. Wieschaus, Diffusion and scaling during early embryonic pattern formation. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18403–18407 (2005).
67. U. Gerland, T. Hwa, On the selection and evolution of regulatory DNA motifs. *J. Mol. Evol.* **55**, 386–400 (2002).
68. J. Estrada, F. Wong, A. DePace, J. Gunawardena, Information integration and energy expenditure in gene regulation. *Cell* **166**, 234–244 (2016).
69. B. Zoller, T. Gregor, G. Tkačik, Eukaryotic gene regulation at equilibrium, or non? *Curr. Opin. Syst. Biol.* **31**, 100435 (2022).
70. H. B. Barlow, "Sensory mechanisms, the reduction of redundancy, and intelligence" in *Proceedings of the Symposium on the Mechanization of Thought Processes, Volume II*, D. V. Blake, A. M. Uttley, Eds. (HM Stationery Office, London, UK, 1959), pp. 537–574.
71. T. Gregor, H. G. Garcia, S. C. Little, The embryo as a laboratory: Quantifying transcription in *Drosophila*. *Trends Genet.* **30**, 364–375 (2014).
72. M. D. Schroeder *et al*., Transcriptional control in the segmentation gene network of *Drosophila*. *PLoS Biol.* **2**, e271 (2004).
73. E. Segal, T. Raveh-Sadka, M. Schroeder, U. Unnerstall, U. Gaul, Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* **451**, 535–540 (2008).
74. J. Bieler, C. Pozzorini, F. Naef, Whole-embryo modeling of early segmentation in *Drosophila* identifies robust and fragile expression domains. *Biophys. J.* **101**, 287–296 (2011).
75. T. Duque *et al*., Simulations of enhancer evolution provide mechanistic insights into gene regulation. *Mol. Biol. Evol.* **31**, 184–200 (2014).
76. P. François, V. Hakim, E. D. Siggia, Deriving structure from evolution: Metazoan segmentation. *Mol. Syst. Biol.* **3**, 154 (2007).
77. P. François, E. D. Siggia, Predicting embryonic patterning using mutual entropy fitness and in silico evolution. *Development* **137**, 2385–2395 (2010).
78. J. B. Rothschild, P. Tsimiklis, E. D. Siggia, P. François, Predicting ancestral segmentation phenotypes from *Drosophila* to anopheles using in silico evolution. *PLoS Genet.* **12**, e1006052 (2016).
79. B. Verd, N. A. Monk, J. Jaeger, Modularity, criticality, and evolvability of a developmental gene regulatory network. *eLife* **8**, e42832 (2019).
80. E. Clark, A. D. Peel, M. Akam, Arthropod segmentation. *Development* **146**, dev17 0480 (2019).
81. A. S. Wilkins, *The Evolution of Developmental Pathways* (Sinauer Associates, 2002).