Gene Expression Noise in Embryonic Spatial Patterning

Reliable formation of the head-to-tail axis in the fruit fly

David M. Holloway
Mathematics Department
British Columbia Institute of Technology
Burnaby, B.C., Canada
David_Holloway@bcit.ca

Alexander V. Spirov
Computer Science and CEWIT
State University of New York at Stony Brook
Stony Brook, N.Y., USA
alexander.spirov@gmail.com

***Abstract*—Fruit flies serve as a model for understanding the genetic regulation involved in specifying the complex body plans of higher animals. The head-to-tail (anterior-posterior) axis of the fly (*Drosophila*) is established in the first hours of development. Maternally supplied factors form concentration gradients which direct embryonic (zygotic) genes where to be activated to express proteins. These protein patterns specify the positions and cell types of the body's tissues. Recent research has shown, comparing between embryos, that the zygotic gene products are much more precisely positioned than the maternal gradients, indicating an embryonic error reduction mechanism. Within embryos, there is the additional aspect that DNA and mRNA operate at very low copy number, and the associated high relative noise has the potential to strongly affect protein expression patterns. In recent work, we have focused on the noise aspects of positional specification within individual embryos. We simulate activation of *hunchback* (*hb*), a primary target of the maternal Bicoid (Bcd) protein gradient, which forms an expression pattern dividing the embryo into anterior and posterior halves. We use a master equation approach to simulate the stochastic dynamics of *hb* regulation, using the known details of the *hb* promoter, the region of DNA responsible for transcribing *hb* mRNA. This includes the binding/unbinding of Bcd molecules at the promoter, *hb* transcription, subsequent translation to Hb protein, binding/unbinding of Hb at the promoter (self-regulation), and diffusion of the Bcd and Hb proteins. Model parameters were set by deterministically matching large scale pattern features for a series of experimental expression patterns: wild-type (WT) embryos; *hb* mutants lacking self-regulation; and constructs in which portions of the *hb* promoter were used to express a reporter gene (lacZ). The model was then solved stochastically to predict the noise output in these different experiments. In subsequent noise measurements we experimentally corroborated a number of the predictions. These include that mRNA is noisier than protein, and that Hb self-regulation reduces noise. Results indicate that WT (self-regulatory) Hb output noise is predominantly dependent on the transcription and translation dynamics of its own expression, and is uncorrelated with Bcd fluctuations. This contradicts prior work, which had assumed a complete dependence of Hb fluctuations on Bcd fluctuations. In the constructs and mutant, which lack self-regulation, we find that increasing the number and strength of Bcd binding sites (there are 6 in the core *hb* promoter) provides a rudimentary level of noise reduction. The model is robust to the various Bcd binding site numbers seen across different fly species. New directions in the project include incorporating a known inhibitor of *hb*, Krüppel, into the model to study its effect on the noise dynamics. Our study has identified particular ways in which *hb* output noise is controlled. Since these involve common modes of gene regulation (e.g. multiple regulatory sites, self-regulation), these results contribute to the general understanding of the reproducibility and determinacy of spatial patterning in early development.**

Keywords-Drosophila; gene regulation; spatial pattern; transcription noise; embryo development

I. INTRODUCTION

The development of animal body plans depends on the coordinated expression of genes in well-defined regions of the embryo to produce properly differentiated tissues. Development must be robust to variation in embryo size, geometry, temperature, dosage of maternal factors, and errors in the timing of events. Our work focuses on quantifying between-embryo variability and within-embryo noise in order to understand the mechanisms by which development achieves such robustness. Using the well-characterized genetics and molecular biology of the fruit fly, *Drosophila melanogaster*, we are investigating the robustness of very early spatial patterns (of mRNAs and proteins) which determine where specific tissues form later in development.

In *Drosophila*, the mother deposits the mRNA of the *bicoid* gene at the anterior (head) end of the embryo. Translation to protein (Bicoid, Bcd) and posterior transport forms a spatial concentration gradient (Fig. 2A). Since Bcd regulates the transcription of genes in the embryo in a concentration-dependent manner, this conveys anterior-posterior (AP) positional information to the embryo. The *hunchback* (*hb*) gene is one of the primary targets of Bcd, responding to high Bcd concentrations to form a broad anterior expression pattern (*hb* mRNA, Fig. 2B; Hb protein, Fig. 2C), which then serves as a cue for anterior-specific differentiation. Observations that the Hb mid-embryo domain boundary forms with lower positional variability than that of Bcd [1, 2], and that other embryonic patterns decrease positional variability during the course of development [3], indicate that there are embryonic mechanisms which filter variability in factors differing between embryos, particularly variation in geometry, size, and maternal dosage.

Supported by the US NIH (GM072022) and NSERC Canada.

Within embryos, the reliable formation of gene expression patterns additionally depends on controlling the noise intrinsic to transcribing genes (DNA) to mRNA and translating mRNA to protein, as well as controlling noise in the transport processes involved in pattern formation. Noise is especially relevant for the low concentrations involved (several copies to several thousand per cell), the nonlinear amplification of signals, and the inhomogeneity of embryonic tissue. These effects can strongly contribute to between-cell differences in gene expression levels. A great deal has been learned about transcriptional/translational noise in single-celled organisms in the past decade (e.g. [4]); our work builds on this, to understand how noise is controlled in the coordinated development of complex body plans.

Anterior *hb* expression is controlled by the proximal *hb* promoter, a region of about 1500 base pairs upstream of the *hb* coding region (transcription start site) on the DNA. This region of DNA can be used to make synthetic constructs in which non-*hb* reporter genes form *hb*-like anterior pattern [5]. This region contains 6 binding sites for the Bcd protein, and 2 self-regulatory sites for the Hb protein. We have formulated a dynamic model of regulation at the *hb* promoter (Fig. 1) with: binding and unbinding of the protein regulators; transcription of *hb* mRNA; translation of Hb protein; decay of *hb* mRNA, Hb protein and Bcd protein; and diffusion of the Hb and Bcd proteins (*bcd* translation at the anterior pole is also modelled, to form the AP Bcd gradient). Solving this model deterministically, in conjunction with experimental measurements, we showed that Hb self-regulation, through the 2 sites in the *hb* promoter, creates a dynamic bistability which underlies the sharpness of the mid-embryo domain border [6]. We have now solved the model stochastically, to characterize the noise generated in the process of *hb* pattern formation. We have experimentally corroborated a number of the modelling results, and present a number of other predictions (see also [7]). Since the type of regulatory interactions seen in *hb* expression are common to other developmental genes, these results should provide insight into more general mechanisms by which development is made robust to gene expression noise.

II. METHODS

We formulated the interactions summarized in Fig. 1 into a system of reaction-diffusion differential equations for each bound-state of the DNA and the mRNA and protein concentrations. This model was solved stochastically with the MesoRD software package [8]. Geometry was specified as a one-dimensional series of 100 subvolumes (each a 5μm cube), corresponding to the nuclei and their surrounding cytoplasm along the length of the AP axis (this stage of *Drosophila* development is precellular). MesoRD solves the reaction-diffusion master equation (RDME), in which each reaction and diffusion event has a probability (set by the macroscopic rate constants) of occurring in a unit of time. The software implements the next subvolume queuing method [9] to significantly improve memory and processing requirements (e.g. compared to [10]), making computation possible for the number of species and subvolumes in the *hb* model. Model parameters were determined by fitting macroscopic features of published data [5, 6]: *hb* boundary position and angle, expression levels, and timescales. Parameters are available in [7].

Stochastic predictions were tested against our own data. Fly embryos were heat fixed and immunostained with fluorescent antibodies for Hb protein. Fluorescent in situ hybridization (FISH) was used for *hb* mRNA determination, following the method of [11]. Images were collected by confocal microscopy. Raw images were cropped and rotated for standardization. Each nucleus plus its cytoplasmic neighbourhood ('energid') was identified by Voronoi tessellation [12]. Averaged pixel intensities within each energid were used for comparison to simulation output. Data was used from a 10% DV (dorsoventral) strip, centred on the AP midline, in order to minimize geometric distortion from the embryo periphery. Background fluorescence was removed by a Genetic Algorithms approach as described in [7].

Data trends were found by 2D Singular Spectrum Analysis (2D-SSA) [13], a non-parametric technique with an adaptive filter, on the 2D (AP and DV) fluorescence intensity surface; the leading components of the decomposition give the pattern's trend. Noise (values in Table 1) was then quantified from the difference of each energid's intensity to the trend value at each position (i.e. local residuals). Noise measures were calculated from the anterior expressing regions (15-45 % egg length (%EL), m positions) as $\sqrt{\frac{\sum[(data-trend)/trend]^2}{m-1}}$, where *data* is the average pixel intensity for an energid and *trend* is the SSA-extracted trend at that position (i.e. the noise measure is a standard deviation for the relative residuals). Noise was calculated similarly for simulation output, from the difference of the stochastic output and the deterministic solution at each position.

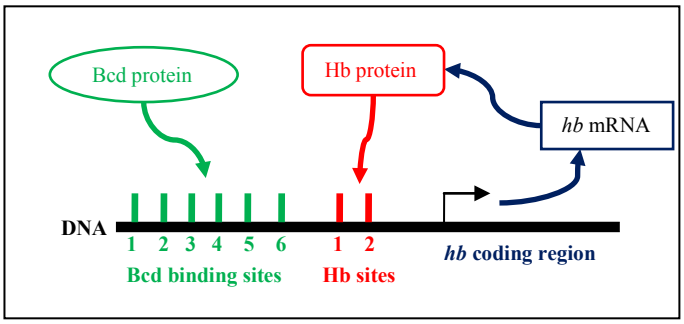

Figure 1. Schematic of model interactions at the *hb* promoter. All events – Bcd and Hb proteins binding/unbinding at specific sites on the DNA; transcription of *hb* mRNA; translation to Hb protein; and diffusion of Bcd and Hb protein between nuclei – are modelled stochastically via a master equation approach. These interactions are modelled for 100 nuclei along the length of the anteroposterior (AP; head-to-tail) axis of the embryo.

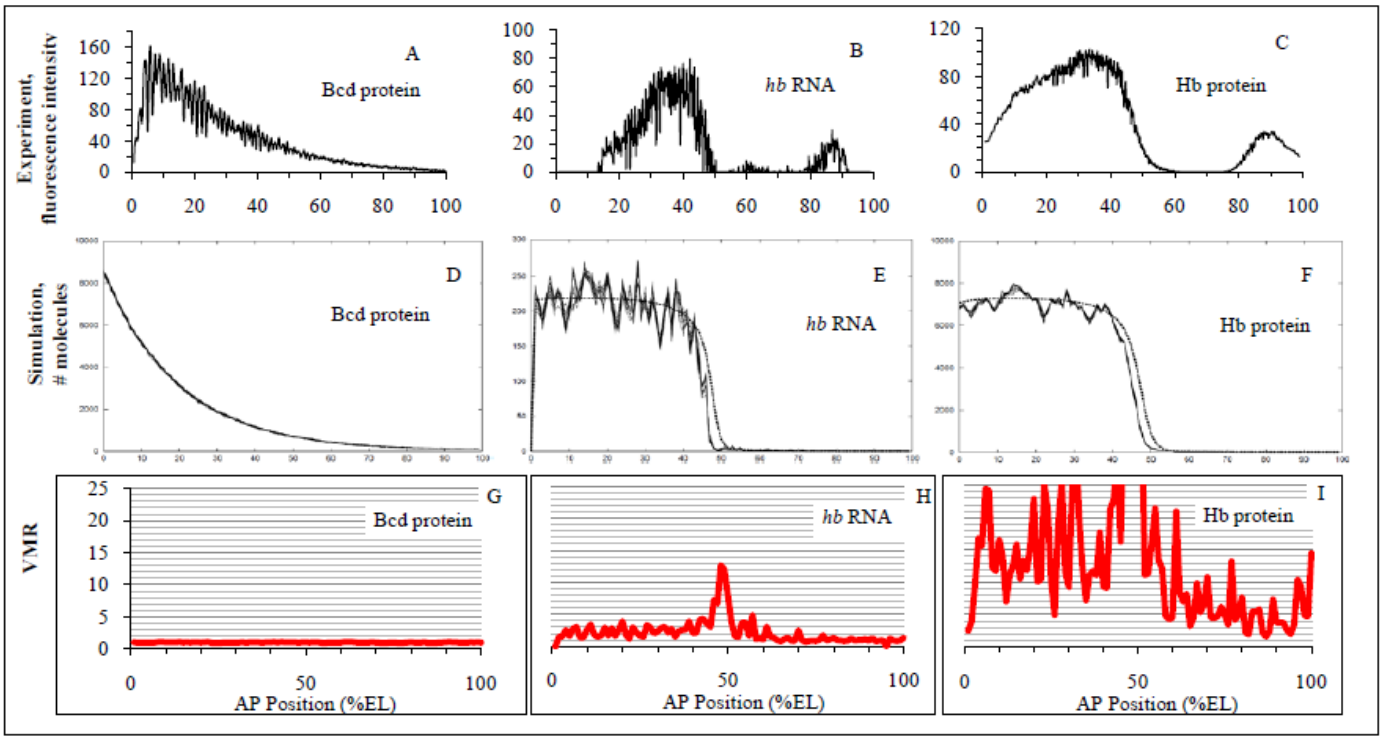

Figure 2. Noise characteristics of *hb* gene expression. A-C) Experimental data for Bcd, *hb* RNA and Hb protein, respectively; concentration (from fluorescence intensity) vs. AP position (in percent egg length, %EL). Relative noise increases with transcription, from Bcd to *hb* RNA, and decreases with translation, from *hb* RNA to Hb protein. (Data in A from [15]; B,C data from [7].) D-F) Stochastic simulation of the anterior activation (Fig. 1 interactions) captures these trends. Noise simulations are overlaid at 5 second intervals (at t=29-30 minutes) to show temporal stability; dashed line - deterministic solution. G-I) Variance-to-mean ratio (VMR) for the simulations shows the different character of the fluctuations: Bcd is Poisson distributed (VMR=1); anterior *hb* RNA begins to deviate from Poisson (VMR~2-3); anterior Hb protein shows strong deviations from Poisson (mean VMR~16). This trend is characteristic of 'bursting' in gene expression, due to the transcription and translation dynamics [4].

III. RESULTS

A. *hb mRNA is noisier than protein*

Fig. 2D-F shows a simulation result for the Fig. 1 model: Bcd protein, *hb* mRNA and Hb protein, respectively (6 replicates run, Fig. 2 shows typical noise levels). The Bcd concentration was set to experimentally determined levels [14]. Fluctuations in Bcd number are expected to be relatively low in this concentration range. The transcription of *hb* mRNA, depending on a stochastic production term, which in turn depends on the (highly stochastic) number of regulators bound to the promoter, is predicted to have higher relative noise. A key finding is that protein noise is significantly lower than mRNA noise: this is corroborated by our experimental data (Figs. 2B vs. 2C). The *hb* mRNA and protein noise levels were found to be largely independent of Bcd noise for the given concentration range: 6 replicates with Bcd noise had very similar Hb noise to 19 replicates without Bcd noise. Table 1 gives noise levels for experiments and for simulations (here without Bcd input noise, the noise is entirely generated by *hb* transcription and translation; data from [7]). For normal (wild-type, WT) flies, and simulations, the drop in noise from mRNA to protein is statistically significant. We predict this is due to the concentration difference resulting from translational efficiency, since each mRNA molecule produces on average 35 protein molecules (model parameters).

B. *Hb self-regulation decreases noise*

The hb^{14F} mutant forms a truncated Hb protein which cannot bind DNA. These embryos have reduced expression (15% of WT) and lack a sharp domain boundary [6]. Blocking binding of the 2 Hb sites in the model produces noisier *hb* mRNA and protein than WT. This is corroborated experimentally: higher noise is observed in hb^{14F} embryos (Table 1). *hb* mRNA concentrations in the model are 30 copies/nucleus for hb^{14F} and 200 copies/nucleus for WT: the low mutant levels are firmly in a stochastic regime; the WT levels are more deterministic. Mutant and WT protein levels, in the thousands of copies per nucleus, could be expected to be fairly low noise. But mutant protein is created from noisy mRNA and propagates this noise at relatively high concentration.

C. *Transcription and translation create unique noise characteristics*

Fig. 2G-I shows the variance-to-mean ratio (VMR) for the simulation in Fig. 2D-F. Bcd protein displays the characteristic VMR=1 of a Poisson distribution. This can be expected from the simple point-release, diffusion, decay kinetics forming the Bcd gradient. Solving the RDME for the *hb* mRNA and Hb protein kinetics shows increasing departure from Poisson VMR in the anterior activated region: transcription gives 2-3 fold increase in VMR (Fig. 2H); translation produces a further

5-6 fold increase (Fig. 2I). The increase at the transcriptional level is due to self-feedback: with no binding at the Hb sites, *hb* mRNA shows Poisson VMR (but translation still produces a 5-6 fold increase in Hb protein VMR). These results demonstrate the unique noise arising during transcription and translation: *hb* output noise cannot be predicted from Bcd input fluctuations.

D. The number and strength Bcd binding sites can affect noise

Driever et al. [5] created a line of synthetic constructs in which portions of the *hb* promoter were used to drive expression of the lacZ reporter. A number of these constructs were driven by only Bcd binding sites (no Hb sites). Modelling these shows the degree to which binding site number and binding strength affects noise. Table 1 shows this, for between 1 and 9 binding sites, and for strong (A) and weak (X) sites. Increasing the number of sites, and increasing the strength of binding, decreases expression noise. While Bcd-only constructs are still much noisier than WT expression with Hb self-feedback, multiple Bcd sites and binding strength may provide a basal buffering against binding/unbinding noise at single sites. This may be evolutionarily conserved: other fly species have between 4 and 10 Bcd binding sites in the *hb* promoter.

IV. CONCLUSION

This project has investigated the ways in which regulatory interactions, which depend on the binding site structure of the gene promoter, affect noise in spatial expression patterns in developing embryos. Many developmental genes are regulated in similar ways to *hb*: we predict that the principles found here - for the contribution of self-feedback, multiple binding sites, transcription and translation to expression noise - are important for reliability throughout development.

TABLE I. HUNCHBACK EXPRESSION, PERCENT NOISE

WT RNA (expt.)	***WT protein (expt.)***	***WT RNA (simul.)***	***WT protein (simul.)***	***hb^{14F} RNA (simul.)***	***hb^{14F} protein (simul.)***	***hb^{14F} protein (expt.)***	***lacZ[e] 1A[f]***	***lacZ 3A***	***lacZ 3X[g]***	***lacZ 4A***	***lacZ 4X***	***lacZ 2x(3X)[h]***	***lacZ 3x(3X)[i]***
47 (22)[a]	5.1 (.9)[a]	11 (1)[b]	5.3 (.9)[b]	26 (4)[c]	11 (2)[c]	8.9 (3)[d]	82	64	82	47	102	48	30

a. Mean and (standard deviation) for 3 WT embryos. RNA is noisier than protein in each embryo (and $p<0.05$ for t-test).
b. Mean and (standard deviation) for 19 WT simulations. RNA is noisier than protein in every simulation (and $p<0.01$ for t-test).
c. Mean and (standard deviation) for 17 simulations of a mutant lacking self-regulation. Both RNA and protein are higher noise than in WT (t-test, $p<0.01$).
d. Mean (and standard deviation) for 3 hb^{14F} mutant embryos. Noise is higher than in WT (t-test, $p<0.05$).
e. lacZ is an expression reporter for constructs using portions of the *hb* promoter. All lacZ results given are from simulations.
f. Construct with 1 strong (A) Bcd binding site.
g. Construct with 3 weak (X) Bcd binding sites.
h. Construct with two copies of the 3X motif.
i. Construct with three copies of the 3X motif.

ACKNOWLEDGMENT

We thank our collaborators Francisco Lopes, Luciano da Fontoura Costa, Bruno Travençolo, Nina Golyandina and Konstantin Usevich.

REFERENCES

[1] B. Houchmandzadeh, E. Wieschaus, and S. Leibler, "Establishment of developmental precision and proportions in the early *Drosophila* embryo," Nature, vol. 415, pp. 798-802, 2002.

[2] A.V. Spirov and D.M. Holloway, "Making the body plan: precision in the genetic hierarchy of *Drosophila* embryo segmentation", In Silico Biol., vol. 3, pp. 89-100, 2003.

[3] D.M. Holloway, L.G. Harrison, D. Kosman, C.E. Vanario-Alonso, A.V. Spirov, "Analysis of pattern precision shows that *Drosophila* segmentation develops substantial independence from gradients of maternal gene products," Dev. Dynam., vol. 235, pp. 2949-2960, 2006.

[4] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, et al., "Noise in protein expression scales with natural protein abundance," Nat. Genet., vol. 38, pp. 636 – 643, 2006.

[5] W. Driever, G. Thoma, C. Nüsslein-Volhard, "Determination of spatial domains of zygotic gene-expression in the *Drosophila* embryo by the affinity of binding-sites for the *Bicoid* morphogen," Nature, vol. 340, pp. 363-367, 1989.

[6] F.J.P. Lopes, F.M.C. Vieira, D.M. Holloway, P.M. Bisch, A.V. Spirov, "Spatial bistability generates *hunchback* expression sharpness in the *Drosophila* embryo," PLoS Comput. Biol., vol. 4, e1000184, 2008.

[7] D.M. Holloway, F.J.P. Lopes, L. da Fontoura Costa, B.A.N. Travençolo, N. Golyandina, K. Usevich, and A.V. Spirov, "Gene expression noise in spatial patterning: *hunchback* promoter structure affects noise amplitude and distribution in *Drosophila* segmentation," PLoS Comput. Biol., vol. 7, e1001069, 2011.

[8] J. Hattne, D. Fange, and J. Elf, "Stochastic reaction-diffusion simulation with MesoRD," Bioinformatics, vol. 21, pp. 2923 – 2924, 2005.

[9] J. Elf, A. Dončić, and M. Ehrenberg, "Mesoscopic reaction-diffusion in intracellular signaling," Proc. SPIE, vol. 5110, pp. 114-124, 2003.

[10] D. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," J. Comput. Phys., vol. 22, pp. 403 – 434, 1976.

[11] H. Janssens, D. Kosman, C.E. Vanario-Alonso, J. Jaeger, M. Samsonova, et al., "A high- throughput method for quantifying gene expression data from early *Drosophila* embryos," Dev. Genes Evol., vol. 215, pp. 374-381, 2005.

[12] L.dF. Costa and R.M.Jr. Cesar, Shape Classification and Analysis: Theory and Practice, 2nd Ed. Boca Raton, Fla: CRC Press, 2009.

[13] N. Golyandina and K. Usevich, "2D-extension of Singular Spectrum Analysis: algorithm and elements of theory," in Matrix Methods: Theory, Algorithms, Applications, V. Olshevsky and E. Tyrtyshnikov, Eds. Singapore: World Scientific Publishing, 2010, pp.450-474.

[14] T. Gregor, D.W. Tank, E.F. Wieschaus, and W. Bialek, "Probing the limits to positional information," Cell, vol. 130, pp. 153-164, 2007.

[15] http://urchin.spbcas.ru/flyex/; A. Pisarev, E. Poustelnikova, M. Samsonova, and J. Reinitz, "FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution," Nucl. Acids Res., vol. 37, pp. D560 - D566, 2009.