

Modelling Developmental Regulatory Networks

Tommy Krul¹, Jaap A. Kaandorp¹, and Joke G. Blom²

¹Section Computational Science, Faculty of Science, University of Amsterdam,
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

{tkrul, jaapk}@science.uva.nl

²CWI (Center for Mathematics and Informatics), P.O. Box 94079, 1090 GB
Amsterdam, The Netherlands

joke.blom@cwi.nl

Abstract. This paper introduces a model for simulating regulatory networks that is capable of reproducing spatial and temporal expression patterns in developmental processes. The model is a generalization of the standard connectionist model used for modelling genetic interactions, where the terms for the regulation of gene products and the diffusion term have been separated. This model can be coupled with biomechanical models of cell aggregates and used to study the formation of spatial and temporal expression patterns of gene products during development in cellular systems.

1 Introduction

One of the most important goals in biology is to understand the chemical and cellular processes that govern the development of a fertilized egg cell to a fully-grown organism. Currently the genetic regulation and the specification of the body plan in development in a number of model organisms, as for example *Drosophila*, is known in great detail [4, 14]). The development of the body plan is typically controlled by large networks of regulatory genes. And although there are many examples available of how individual genes affect the developmental process, there are no cases available where a line of causality can be mapped from the genomic sequence to a developmental process [5]. An important issue is therefore to understand the structure and dynamics of regulatory networks. And in addition to experimental observations, the only other option to study how genetic information is translated into a spatial-temporal expression pattern that eventually gives rise to the complete morphology of the fully-grown organism, is to use simulation models.

A number of different models have been proposed to capture the notion of gene networks. And although most of these models were originally created for the reverse engineering of genetic networks [1, 2, 13, 15, 16], it is usually trivial to adapt these models for simulation purposes. To date, the most frequently used is the Boolean network model, introduced by Kauffman [11] in 1969 and used, among others, by Jackson et al. [9] in 1986 to show that pattern formation in both space and time is possible when using active feedback loops and limited

communication among cells. But although these methods generally possess great expressiveness, they use a high level of abstraction, which limits their biological explanatory power. Also, examining gene expression data reveals that genes usually spend a great deal of their time at intermediate concentration levels. This means that discretizing gene expression levels tends to result in a significant loss of information. Furthermore, there are some concepts in control theory that seem indispensable for genetic regulatory systems that either can not be implemented using discrete variables or that have radically different dynamic behavior. This includes amplification, subtraction and addition of signals, maintaining equilibrium using negative feedback and smoothly varying the period of a periodic phenomenon like the cell cycle. More recently, continuous models, also referred to as connectionist [12] or additive models [6], have been proposed. These models closely resemble neural-like connectionist architectures or biochemical networks of interacting chemicals [10]. Reinitz and Sharp [16] have successfully used a system of differential equations to reproduce the formation of stripes of expression of the pair-rule gene *eve* in the *Drosophila* embryo. Even more recently, Salazar-Ciudad et al. [17, 18] used a very similar model to: (1) compare diffusion and direct-contact induction processes as mechanisms of pattern formation, (2) identify the possible range of behavior of real gene networks and (3) suggest causal mechanisms to generate known patterns. With these models it was also demonstrated that different types of body segmentation patterns during the development of insects, a process which is basically controlled by an identical set of genes, can be explained by 'rewiring' the regulatory network. But a significant limitation of the models by Reinitz and Sharp and Salazar-Ciudad et al. is that cells are modelled as static discrete lattice sites on a grid. In for example the work by Salazar-Ciudad et al. the *Drosophila* embryo is represented by a linear cellular automaton, while in the actual developing organism the spatial and temporal patterns of gene products during development will be formed in an aggregate of cells with a usually highly complex geometry.

In this paper we describe a model for simulating regulatory networks that is capable of reproducing spatial and temporal expression patterns as found during embryonic development. The model is based on the standard connectionist model proposed by Mjølness et al. [12] and is to some extent similar in approach to the models used by Salazar-Ciudad and Reinitz and Sharp. Yet, our model does not represent cells as static discrete lattice sites on a grid. Cells are modelled as independent objects, treating the intra- and extracellular environment as explicitly separated chemical spaces. Basically this model can be combined with a biomechanical model of a cell aggregate. In our current version cells are represented as point shaped sources and sinks in the space of the extracellular matrix. As a test case we use the formation of stripes of expression of the pair-rule gene *eve* in the *Drosophila* embryo.

2 Methods

2.1 The Case Study: Eve Stripe Formation in *Drosophila*

As a proof of concept we tried to replicate the basic pattern forming processes that are known to occur during early embryonic bodyplan formation and selected the well-studied regulatory network for the generation of Eve stripe two of *Drosophila* for this purpose [3, 4, 14, 20]. The bodyplan of *Drosophila* is, like all arthropods, segmented. It contains 14 segments, of which three make up the head, including its antennae and mouth parts. The next three segments make up the thorax and each of these segments will develop a set of legs. The last eight segments will develop into the abdominal. The initial gradients that guide the developmental process are a result of the presence of mRNA molecules in the anterior and posterior poles of the embryo. These are deposited in the egg by the mother before it was fertilized. After fertilization the anterior mRNA molecules activate the transcription of the Bicoid protein, while the posterior mRNA activates the translation of the Nanos protein. In the initial phase of the developmental process of *Drosophila* all cells are in a syncytium. This means that cells lack a cell membrane and it is therefore that during this stage all gene products can diffuse freely throughout the embryo. Diffusion of the Bicoid proteins forms a gradient going from a high concentration on the anterior pole to a low concentration on the posterior pole. The Nanos protein forms a gradient in exactly the opposite direction, going from a high concentration on the posterior pole to a low concentration on the anterior pole. In turn a Hunchback protein gradient is formed through the activation of Hunchback by Bicoid in the anterior region and the repression of Hunchback in the posterior pole by Nanos. This leads to the formation of a high concentration of Hunchback in the anterior region with a sharp cut-off towards the posterior. Also Caudal mRNA molecules can be found throughout the embryo and as Bicoid represses its transcription a final posterior-anterior gradient is formed.

Once the main anterior-posterior gradients are formed, other proteins get activated that produce more finely tuned patterns. The Eve protein is expressed in a pattern of seven stripes. It has been shown [3, 4] that the expression of each of these stripes is regulated independently. For example, the expression of Eve stripe two is regulated by the activating and repressing influences of four proteins: Bicoid, Hunchback, Giant and Krüppel. The binding of Bicoid and Hunchback stimulates the transcription of Eve, while the expression of Giant and Krüppel have a negative influence on its translation. Right in between a gap of high levels of Giant and Krüppel the expression of the second Eve stripe finally becomes a narrow band of cells. This is also illustrated in figure 1.

2.2 The Model

Biological cells are modelled as independent objects that can absorb and exude chemical substances from and into the extracellular space. The model makes no assumption on the shape of cells and treats them as point shaped sources

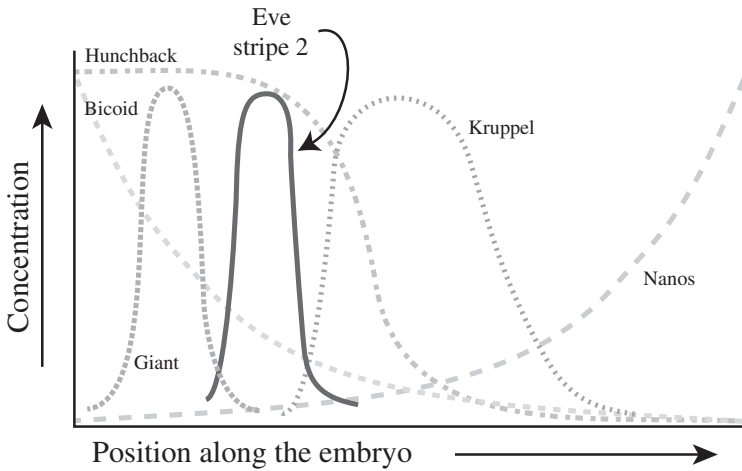


Fig. 1. The location of Eve stripe two is specified by four proteins. Bicoid and Hunchback activate Eve, while Giant and Krüppel have a repressing influence. Trapped between high concentrations of Giant and Krüppel Eve stripe two narrows down to a band of only a few cells thick. Modified after [3]

and sinks on the extracellular matrix. The extracellular matrix is represented by a grid that is overlaid on the cells and the space between them. Furthermore, protein levels are maintained within each cell. Paracrine protein levels are also maintained in the extracellular matrix, where they are allowed to diffuse. Note that we assume that these gene substances do not go into direct chemical interaction with each other. This means that we prohibit the formation of Turing-like reaction-diffusion patterns and that in our model chemical patterns can only arise through (modelled) genetic regulation. The model allows proteins to activate and inhibit each other and also makes a distinction between proteins that stay within a cell and only function for intracellular regulation (non-paracrine factors) and proteins that can diffuse into the extracellular matrix and are members of signaling pathways and exercise short- and long-range effects (paracrine factors).

For our model we adapt the equations as used by Salazar-Ciudad et al. [17, 18] to explicitly separate the intra- and extracellular environment. This separation is accomplished by removing the diffusion term from the intracellular equation and introducing extra equations that model the extracellular diffusion of paracrine gene substances. This separation does not necessarily preserve the exact dynamics of the equations as proposed by Salazar-Ciudad et al. but does preserve its pattern forming ability. The extracellular matrix is represented by a simple grid on which the diffusion equation is solved. This allows gene substances that are exuded from cells to diffuse through the extracellular matrix. The change in concentration for all intracellular gene products over time, whether paracrine or non-paracrine, is described by:

$$\frac{\partial g_{ij}(t)}{\partial t} = \frac{\phi(h_{ij})}{k_j + \phi(h_{ij})} - \lambda_j g_{ij}(t)$$

$$\text{Where } h_{ij} = \sum_{k=1}^{N_g} W_{jk} g_{ik} + h_j \text{ and } i = 1, \dots, N_c \text{ and } j = 1, \dots, N_g \quad (1)$$

Here g_{ij} is the intracellular concentration of gene substance j within cell i . Here k_j , λ_j and h_j are respectively the binding rate constant of gene j , the intrinsic rate of degradation of gene j and the activation threshold level of gene j . All enhancers and protein binding sites are described by the matrix W . Each element W_{ab} of W characterizes the regulatory effect of one gene on another by a single real number for each possible pair a and b . A positive value for W_{ab} meaning that gene b activates gene a while a negative value represses gene a . In the case that W_{ab} is zero, gene b has no effect on gene a . N_c is the total number of cells in the system and N_g is the total number of genes. ϕ is introduced to ensure that inhibiting interactions do not lead to negative concentration values. In our case ϕ is chosen to be:

$$\phi(x) = \begin{cases} x & \forall x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

In the model we make a distinction between paracrine (diffusing) and non-paracrine (non-diffusing or transcription) factors. Non-paracrine factors stay within the boundaries of a cell, while paracrine products are allowed to be transported through the cell membrane and diffuse through the extracellular matrix. The dynamics of the extracellular proteins is governed by a set of partial differential equations:

$$\frac{\partial c_j(\mathbf{x}, t)}{\partial t} = \mathcal{D}_j \nabla^2 c_j(\mathbf{x}, t) - \lambda_j c_j(\mathbf{x}, t) \quad (3)$$

Here c_j is the concentration of the paracrine gene product j , \mathcal{D}_j is the diffusion velocity and λ_j is the rate of degradation of c_j . Equations (1) and (3) are coupled by the constraint

$$g_{ij}(t) = c_j(\mathbf{x}_i, t) \quad (4)$$

which defines the exchange of paracrine gene products between the intra- and extracellular environment. Here $c_j(\mathbf{x}_i, t)$ expresses the concentration of the extracellular paracrine gene substance j at the location \mathbf{x}_i of cell i at time t . The implicit assumption here is that the timescale of the rate of exchange of gene substances between the extracellular matrix and the intracellular environment of a cell is much smaller than the timescale at which the extracellular diffusion of substances occurs.

For the spatial discretization of the equation (3) on the extracellular grid we use a straightforward finite difference approximation. As an integrator we

use in all cases an explicit forward Euler scheme. Because of the fact that cells can be positioned freely throughout the extracellular matrix there is no longer a one-to-one correspondence between grid points and biological cells. We therefore need to interpolate gene substances from the cell locations to the grid and vice versa. First the concentration c_{ij}^{new} of gene substance j at position \mathbf{x}_i of cell i is determined. This is done by calculating the average of the intracellular concentration $g_{ij}(t)$ and the extracellular concentration $c_j(\mathbf{x}_i, t)$, which in turn is determined using linear interpolation. Then the mass m_{ij} of the absorbed or exuded substance is determined by multiplying the difference between the intracellular concentration $g_{ij}(t)$ and the newly-found average concentration c_{ij}^{new} with the volume v_i of the corresponding cell. Finally, all that is left to do is setting the new concentration values. The new intracellular concentration of gene j of cell i becomes c_{ij}^{new} , while the new concentration values in the grid can be found by linearly distributing the concentration $m_{ij}/volume_{gridcell}$ over the neighboring lattice points.

The algorithm starts by integrating the time dependent diffusion equation (3) over one time step. Thereafter the new grid values are used to update the paracrine protein concentration levels at the positions of the cells using the interpolation method described earlier (4). Then the intracellular variables are advanced one timestep using equation (1). Finally all paracrine gene substances are updated once again.

Model parameters consisted mainly of regulatory relationships between genes and activation threshold levels and were tuned to correspond roughly to known interaction properties of the regulatory network of *Drosophila*. For reasons of simplicity only a subset of the proteins known to be involved in generation of Eve stripe two were taken into account (see also figure 2). All systems were simulated until an equilibrium was reached.

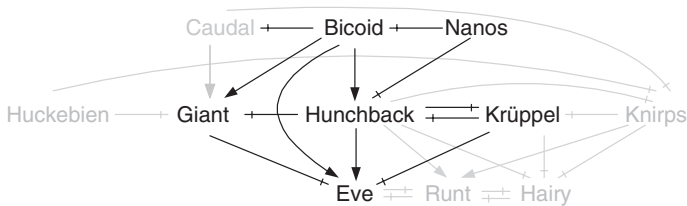


Fig. 2. A selection of important regulatory interactions between genes involved in the generation of Eve stripe two. An arrow indicates a positive regulatory relationship, while a line crossed at its end indicates a repressive relationship. Black interactions were taken into account, while interactions colored gray were not simulated. Modified after [3]

3 Results

As can be seen in figure 3 for a one- and in figure 4 for a two-dimensional domain we are able to reproduce expression patterns that resemble the specification of Eve stripe two. Differences when compared to biological data in the expression of the Giant and Krüppel protein are in part due to the fact that not all proteins involved in the specification of Eve stripe two have been taken into account. The Krüppel protein should only be expressed in the exact center of the embryo. This deviation probably comes from the absence of the repressive influence of the Knirps (kni) protein in the posterior region. Another difference with biological data comes from the Giant protein. This may be caused by the absence of the repressive influence of the Huckebien (hkb) protein on both of the poles. This last deviation is nevertheless quite significant. Giant does not form sharp borders, resulting in a band of cell expressing Eve stripe two that is too wide when compared to biological data. In two dimensional simulations this error becomes even more dominant. Yet, it is important to note here that the regulatory relationships were tuned by hand. And it should therefore come as no surprise that it is likely that a much better match can be found when using a more sophisticated parameter tuning method.

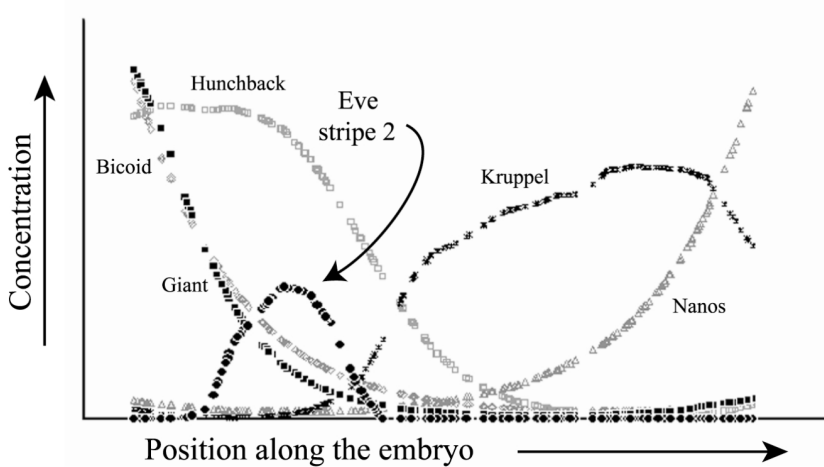


Fig. 3. One-dimensional concentration gradients showing an approximation of the specification of Eve stripe two. The Bicoid, Nanos and Hunchback gradients are shown in gray, while the Giant, Krüppel and Eve gradients are black. When comparing this figure to figure 1 some significant differences are observed. These difference can in part be explained by the absence of certain proteins that do have regulatory relationships with some of simulated proteins but were not taken into account during the simulation

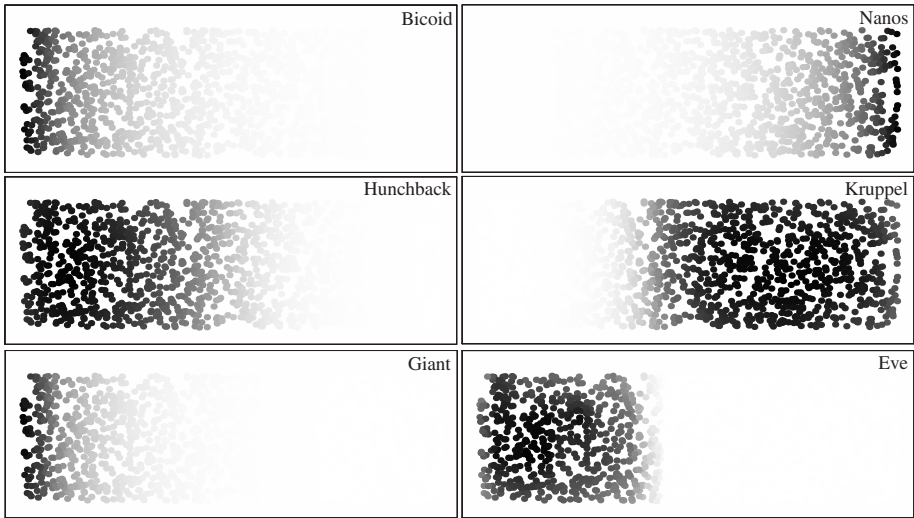


Fig. 4. Two-dimensional concentration gradients for six proteins involved in the specification of Eve stripe two. These graphs were produced using the same parameters as used for figure 3

4 Discussion

The examples in Figure 3 and 4 show that our model of regulatory networks is capable of forming relatively complex and one- and two-dimensional spatial and temporal morphogene expression patterns that correspond to morphogene gradients generated during early embryonic development of *Drosophila*. This model was coupled with a model of cells (or in the case of *Drosophila* nuclei) that can be positioned freely throughout the extracellular matrix, meaning there is no longer a one-to-one correlation between grid points and biological cells, as in many previous models [15–18]. This combination results in a more flexible model that can be used to simulate spatial and temporal expression patterns in a variety of cellular systems. The flexibility of the model mainly comes from the fact that we separated the terms that model the genetic regulation from the term that equates the diffusion of gene substances. This in contrast to many of the previous projects where connectionist models were used to simulate gene regulation [6, 12, 16–18]. Yet this simple modification allows us to perform much more realistic simulations. For example, it elegantly solves the strange correlation found in many previous connectionist models for gene interactions were a grid cell corresponds directly to a biological cell, meaning that increasing the accuracy of the generated pattern by using a finer grid, automatically results in an increase of the total number of simulated cells. In our model the biological cells and grid points are separated so that using a finer grid will result in an increased accuracy of the pattern. There is a catch however. We model cells as volume-less objects in the extracellular matrix, meaning that gene substances are

allowed to diffuse throughout the space unhindered by cells. Yet this also means that decreasing the size of the grid increases the effect of the error introduced by this simplification. We also showed that the model can be extended to more dimensions. Simulations have been carried out on both one and two-dimensional domains using exactly the same parameters for the regulatory network. Currently we are extending these simulations to three dimensions, a major problem here are the computational requirements. Because of the separation between the cells and the extracellular space it is possible to extend the model in such a way that it also models the effect of physical interactions between cells and their environment. One of our future plans is to combine our model of regulatory network for development with a bio-mechanical model of moving and interacting cells. In the simulation results shown in this paper, the model parameters describing regulatory relationships between genes and activation threshold levels, were tuned to correspond roughly to known interaction properties of the regulatory network of *Drosophila*. Currently we are investigating a method based on genetic algorithms to do a guided search through parameter spaces (see also [15, 17, 18]).

References

1. Akutsu, T., Miyano, S., Kuhara, S.: Inferring qualitative relations in genetic networks and metabolic pathways, *Bioinformatics* **16** (2000) 727–734
2. Brazma, A., Schlitt, T.: Reverse engineering of gene regulatory networks: a finite linear model ()
3. Carroll, S.B., Grenier, J.K., Weatherbee, S.D.: From DNA to diversity: molecular genetics and the evolution of animal design, (2001)
4. Davidson, E.H.: Genomic regulatory systems: development and evolution, Academic Press, London (2001)
5. Davidson, E.H. et al.: A genomic regulatory network for development *Science* **295** (2002) 1669–1678
6. D’haeseleer, P., X. Wen, Fuhrman, S., Somogyi, R.: Linear Modelling of mRNA expression levels during CNS development and injury, *Pacific Symposium on Bio-computing* (1999)
7. Glazier, J.A., Graner, F.: Simulation of the Differential Adhesion Driven Rearrangement of Biological Cells , *Physical Review E* **47** (1993) 2128–2154
8. Graner, F., Glazier, J.A.: Simulation of Biological Cell Sorting Using a Two-Dimensional Extended Potts Model, *Physical Review Letters* **69** (1992) 2013–2016
9. Jackson, E.R., Johnson, D., Nash, W.G.: Gene networks in development, *Journal of Theoretical Biology* **119** (1986) 379–396
10. Kaneko, K., Yomo, T. Isologous deversification: a theory of cell differentiation. *Bull. Math. Biol.* **59** (1997) 139–196
11. Kauffman, S.A., *J. Theoretical Biology* **44** (1974) 167–190
12. Mjølness, E., Sharp, D.H., Reinitz, J.A.: connectionist model of development. *J. Theoretical Biology* **152** (1991) 429–453
13. Murphy, K. and Mian, S., Modelling gene expression data using dynamic Bayesian networks, Technical report, Computer Science Division, University of California, Berkeley, CA. (1999)

14. St Johnston, D, Nüsslein-Volhard, C.: The origin of pattern and polarity in the *Drosophila* embryo, *Cell* **68** (1992) 201–209
15. Reinitz, J., Kosman, D., Vanario-Alonso, C.E., Sharp D.H.: Stripe forming architecture of the gap gene system. *Developmental Genetics* **23** (1998) 11–27
16. Reinitz, J., Sharp, D.H.: Mechanism of formation of eve stripes. *Mechanisms of Development* **49** (1995) 133–158
17. Salazar-Ciudad, I, Newman, S.A., Sole, R.V.: Phenotypic and dynamical transitions in model genetic networks I. emergence of patterns and genotype-phenotype relationships, *Evolution and Development* **3** (2001) 84–94
18. Salazar-Ciudad, I, Sole, R.V., Newman, S.A.: Phenotypic and dynamical transitions in model genetic networks II. application to the evolution of segmentation mechanisms, *Evolution and Development* **3** (2001) 95–103
19. Turing, A.M., The chemical basis of morphogenesis, *Transactions R. Soc. Lond B.* **237** (1952) 37–72. Reprinted in *Bull. Math. Biol.* **52** (1991) 153–197
20. Wolpert, L., Beddinton R., Brooks J., Jessel, T.: *Principles of development*, Current Biology Ltd, Oxford University Press (1998)