# SIMULATION OF GENETIC INTERACTION FOR DROSOPHILA LEG FORMATION

# KOJI KYODA<sup>†\*</sup> and HIROAKI KITANO<sup>‡\*</sup>

<sup>†</sup>Department of Computer Science, Keio University 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223, JAPAN

<sup>‡</sup>Sony Computer Science Laboratory 3-14-13 Higashi-Gotanda, Shinagawa, Tokyo 141, JAPAN

\* Systems Biology Group, Kitano Symbiotic Systems Project, ERATO, JST 6-31-15 6A, Jingu-Mae, Shibuya-ku, Tokyo 150, JAPAN

The formation of Drosophila wings and legs are major research topics in Drosophila development, and several hypotheses, such as the polar-coordinate model and the boundary model, has been proposed to explain mechanisms behind these phenomena. A series of recent studies have revealed complex interaction among genes involved in establishing three principal axes (A-P, D-V, and P-D) of leg formation. In this paper, we present a simulation system for leg formation, simulating the genes interactions involved. We use this simulator to investigate a mathematical framework of leg formation which is otherwise well-founded from a molecular perspective. Particularly, we focus on the formation of the expression patterns of dpp, wg, dll, dac, al, en, hh and ci genes, which are involved in the development of the third instar Drosophila leg disc. The most interesting part of this research is showing how the coaxial gene expression patterns behind the P-D axis can be formed, and how positional information, as postulated in the polar-coordinate model, can be conveyed to each cell. Our results suggest that P-D axis can be formed by a set of genes with different activation thresholds ; the process involves different chemical gradients of dpp and wg products, forming a bi-polar contour. Interestingly, this combination of chemical gradients can specify unique positions of cells for the hemisphere, leaving the A-P axis determiner to decide only whether the cells are anterior or posterior. All in all, our so-called Bi-Polar Model describes axial formation of the leg disc well.

# 1 Introduction

Historically, *Drosophila* has been an extremely popular animal for molecular and developmental biology. There is a substantial accumulation of genetic and cytological knowledge and detailed analysis for the organism. Research on the development of *Drosophila* is particularly important because *Drosophila* exhibits complex morphological changes, some of which may share underlying mechanisms with animals with more complex body structures. The formation of leg, wing, eye, and other structures from imaginal discs is particularly interesting as these discs undergo dramatic changes in structure. Leg formation in *Drosophila* is especially important not only for *Drosophila* research but also vertebrate development research, as they may share some common mechanisms in forming limb axes.

Past models of *Drosophila* leg formation have been proposed by many researchers. In 1976, French *et al.* proposed the *polar coordinate model* to describe limb regeneration in insects and vertebrates [French, *et al.*, 76]. Unfortunately, the polar coordinate model is only empirical, and does not suggest underlying molecular mechanisms. Meinhardt proposed the *boundary model* [Meinhardt, 83] which postulates that the positional information of a cell in the leg disc is determined by both cartesian and polar coordinates. Meinhardt argued that the boundary of three territorial sections serves a unique role in leg generation, and perhaps a diffusive factor is discharged from the region where the territorial sections intersect. However, recent studies have cast doubt on the assumption of such diffusive factors. Recently, Kondo applied the polar coordinate model to *Drosophila* leg disc [Kondo, 92] with *reaction-diffusion theory* [Turing, 52]. However, such models are not very biologically-grounded, especially with regard to actual genetics. So far, no model fully accounts for leg formation in genetically well-founded manner.

The goal of this paper is to propose a biologically faithful model in the light of recent of molecular biology studies, involving gene regulation and axis determination in the *Drosophila* leg disc. The work reported in this paper is a part of the Virtual *Drosophila* Project [Kitano, et al., 97] at Sony Computer Science Laboratory Inc., which has created detailed models of early embryogenesis [Hamahashi and Kitano, 98] and eye formation [Morohashi and Kitano, 98], [Ueda and Kitano, 98] of *Drosophila*.

## 2 Imaginal Leg Disc

### 2.1 The Formation of Leg Disc

During an embryogenesis, the epithelium folds, producing the the leg disc. It is believed that the Anterior-Posterior axis (A-P axis) and the Dorsal-Ventral axis (D-V axis) are formed at this early stage. In later stages of the third instar leg disc, coaxial patterns of gene expression produce gene products with positional information along the Proximal-Distal axis (P-D axis). Later, in the third instar larva, imaginal disc elongation starts from the center of the disc, which grows into the adult *Drosophila* leg. The fate map of the *Drosophila* leg disc is shown in Figure 1 [Bryant, 93].



Figure 1: (a). The fate map of leg imaginal disc. The center of the disc becomes the distal tip and the periphery becomes the proximal boundary [Bryant, 93]. (b). The mechanism of dll expression regulation. Wg and Dpp activities required for Dll induction [Lecuit and Cohen, 97].

#### 2.2 Gene Regulation Network

In the posterior compartment (the back half) of the disc, cells continuously express the engrailed(en) gene, which encodes a homeodomain protein. En induces the expression of hedgehog(hh) in the same compartment (in this paper, DNA genes such as engrailed are indicated with their short forms italicized (en), and the proteins they produce are capitalized but without italics (En)). Hh diffuses to the anterior compartment (the front half) because hh encodes a secreted protein [Dominguez, et al., 96]. In an anterior compartment, cells express cubitus interruptus(ci), which encodes a zinc-finger protein. En, which is only in the posterior compartment, represses ci in the posterior cells [Eaton and Kornberg, 90]. Consequently, cells express ci only in the anterior compartment. Ci protein is required to repress hh expression in anterior cells [Dominguez, et al., 96]. The A-P axis is determined by these genes, but the D-V axis forming mechanism is not understood clearly. In the wing disc, dpp is expressed in the A-P boundary and wg is expressed in the D-V boundary. They form the A-P axis in the wing disc.

In third instar disc, Hh induces the expression of wingless(wg) in the ventral anterior wedge. At the same time, Hh induces the expression of decapentaplegic(dpp) in the dorsal A-P boundary [Basler and Struhl, 94]. Both wg and dpp encode secreted proteins, which create gradients of concentration shown in Figure 1. Wg signaling inhibits dpp expression, whereas Dpp signaling inhibits wg expression in the leg disc. This mutual repression maintains separate territorial areas of wg and dpp expression.

The P-D axis determines which part of the imaginal disc turns into which

Gene	Symbol	Classification	Cellular location
distal-less	dll	homeodomain	nuclear
dachshund	dac	novel	nuclear
aristaless	al	$\operatorname{homeodomain}$	$\operatorname{nuclear}$
cubitus interruptus	ci	zinc finger	$\operatorname{cytoplasmic}$
engrailed	en	homeodomain	nuclear
decapentaplegic	dp p	TGF-beta-like	secreted
hedgehog	hh	TGF beta family	secreted
wingless	wg	WNT family	secreted

Table 1: Eight regulatory genes involved in Drosophila leg formation.

part of the leg. This is determined by genes expressed as coaxial rings, including *al*, *dll*, and *dac*. Lecuit and Cohen [Lucuit and Cohen, 97] experimentally show that Wg and Dpp directly activate *dll*, as shown in Figure 1. The domain, which is defined by the overlap of Wg and Dpp signals in the center of the disc, forms circles of different diameters and provides positional information along the P-D axis. While Wg and Dpp induce *dac* expression at a low level of combined activity, they repress its expression at high levels of combined activity. Similarly, the expression pattern of *aristaless (al)* appears to be under the control of both *wg* and *dpp* signaling [Campbell, *et al.*, 93].

### 3 Modeling

An actual *Drosophila* leg disc is composed of more than 10,000 cells and is shaped like an ellipse. Our model consists of 1,200 cells where approximately 8 cells in real real imaginal disc are represented by one virtual cell in the simulation. In the third instar leg disc, a process called *folding* occurs in the epithelium. Foldings occur at several different diameters. The cells which are positioned opposite ends of a fold generally do not have no interactions each other, so it is possible simulate only part of the folding, without considering the full spatial structure of the leg disc folding. Thus, we simulated an imaginal disc as an circular assembly of cells on the same plane.

We have implemented eight major genes and their regulations involved in the formation of the *Drosophila* leg disc. The regulatory relationship between genes and their products are shown in Table 2. The gene regulatory network in Table 2 is based on a lot of literature. However, some regulatory relationships, marked in parentheses (), are only hypothetical relationships introduced in order to make expression patterns consistent with actual data. We discuss this in more detail later. In our model, we defined four processes: transcription, translation, protein diffusion and protein decay.

	en	ci	hh	dpp	wg	al	dll	dac
Engrailed (En)		—	+	_	_			
Cubitus interruptus (Ci)			_	(-)				
Hedgehog (Hh)				+	+			
Decapentaplegic (Dpp)					-	+	+	+
Wingless (Wg)				-		+	+	+

Table 2: Transcriptional Regulation in the Simulation. "+" means that a protein acts as activator of transcription of the target gene. Consequently, "-" means that it acts as repressor of transcription of the target gene. "()" means the hypothetical regulation.

#### 3.1 Transcription

We have modeled gene transcription as a stochastic model. Each gene has a promoter region, a threshold which determines whether the gene starts to be transcribed or not, and variable indicating the amount of mRNA transcribed from the gene. Genes produce specific proteins as their products. Each protein has a binding affinity to each promoter region. With a strong binding affinity, the protein can bind to a promoter region with high probability. Conversely, it is hard to bind to a promoter region when the binding affinity is weak.

Through binding to a promoter region, a protein either activates or represses the gene's ability to transcribe mRNA, and hence ultimately its ability to produce more protein product. When activators and repressors compete some gene X, the function of transcription of gene X is determined by the inequality below:

$$\frac{\sum_{i} \alpha_{A} U_{A}}{\sum_{i} \alpha_{A} U_{A} + \sum_{j} \alpha_{R} U_{R} + C} > \text{Threshold}(gene_{X})$$
(1)  
$$amount_{X} = \begin{cases} M_{X} & \text{if } inequality (1) is true} \\ 0 & \text{else} \end{cases}$$

where  $\alpha_A$  and  $\alpha_R$  are the activation and inhibition rates of gene products *i* and *j*, respectively.  $U_A$  and  $U_R$  are the concentration of activator and repressor proteins. In case neither activator nor the repressor binds to the promoter region of target gene, we introduce a constant value C to adjust the probability of transcription.  $M_X$  is a constant value which determines the amount of mRNA transcribed in each calculation time step.

### 3.2 Translation

In the *Drosophila* leg disc, translational control is not understood well enough to take it into consideration. In our model, the amount of transcribed mRNA is the same as the amount of resultant translated protein.

# 3.3 Diffusion

The secreted protein diffuses into a given concentration gradient. This longrange signaling molecule acts as a *morphogen*, specifying cell fate through membrane receptors and cellular transduction mechanisms. The diffusion equation is defined by:

$$\frac{\partial U_A}{\partial t} = D_A \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) U_A \tag{2}$$

where  $U_A$  is the concentration of secreted protein A at cell-position (x, y).  $D_A$  is a constant value which is individually determined for each protein.

#### 3.4 Model Equation

In the simulation, we implemented transcription/translation and diffusion as described above. In addition, we also considered the degradation protein which is caused by methylation. We unify these processes with the following equation.

$$\frac{\partial U_i}{\partial t} = D_i \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) U_i + f(\mathbf{U}) - g \cdot U_i \tag{3}$$

$U_i$	:	concentration of protein i	$D_i$	:	diffusion constant
t	:	$\operatorname{time}$	f	:	protein production function
x	:	position on x axis	$\mathbf{U}$	:	concentration vector
y	:	position on y axis	g	:	degradation rate $(g = 0.1)$

Currently we hand-optimize the binding affinity, protein diffusion constant value, etc. In the future we hope to use an optimization algorithm to determine ideal parameter settings for the simulator.

## 4 System Architecture

We have implemented the simulation system in C++ and OpenGL. The system is composed of core model and a visualization system. A diagram is shown in Figure 2.

The system core is composed of the Leg, Cell, Gene and Protein classes. Each cell in this system contains a list of proteins. First, the system checks whether a protein is secretable. If the protein secretable, the amount of diffusing protein is calculated in comparison with the protein concentration in the adjacent cells by the Diffusion Engine. Second, the Reaction Engine calculates the amount of protein which is produced from genes through transcription and translation in the presence of activator and repressor protein concentrations.



Figure 2: System architecture of the leg simulator.

The visualization system has two kinds of windows. The gene expression patterns and protein concentration patterns are visualized in the first window. The leg disc of the system can be rotated and observed from any viewpoint. The protein concentration in a cell is represented by the brightness of the cell sphere. The second window is a control panel which allows us to modify simulation parameters.

#### 4.1 Initial Conditions

Our simulator begins with settings appropriate for when the epithelium is being folded and the leg disc is created. At this point, ci has already been expressed in the whole anterior compartment, while en is expressed in all of the posterior compartment, and wg is localized near the ventral anterior-posterior boundary [Bate and Arias, 93].

## 5 Results

The simulator successfully reproduces the expression patterns for the genes involved and the localization patterns of their products. Figure 3 shows simulated and actual expression patterns for the genes ci, hh, en, wg, dpp, dll, dac, al. These results agree well with experimental biological data. In the case of the secreted protein, the protein diffuses in all directions and the localization pattern extends out from the expression pattern. For nuclear and cytoplasmic



Figure 3: The expression patterns for genes involved and localization patterns of their products. (a). The simulated patterns of gene expressions. (b). The localization patterns of gene products. (c). The actual patterns of gene expressions. (i),(ii). [Eaton and Kornberg, 90] hh expression is same as en, (iii). [Brook and Cohen, 96], (iv). [Zecca, et al., 96], (v). [Campbell, et al., 93], (vi). Dll domain (red), Dac domain (green) and the overlap of the Dll and Dac domain (yellow), [Lecuit and Cohen, 97].

protein, which are localized in a single cell, the localization patterns are the same as the expression patterns.

## 6 Discussion

#### 6.1 Ci Inhibits Dpp Expression

We implemented the regulation network shown in Table 2. In our model, we hypothetically assigned Ci the role of an inhibitor for dpp. This caused the simulated dpp expression pattern to be consistent with the actual dpp expression pattern of the leg disc. Unless Ci represses dpp expression, the expression pattern of dpp forms a wedge expression rather than a sharp expression. The expression patterns of dpp are shown in Figure 4.



Figure 4: The *dpp* expression patterns in the leg disc. (i). actual pattern [Brook and Cohen, 96], (ii). simulated pattern with the repression of Ci, (iii). without the repression of Ci.

In the wing disc, Ci activates dpp expression at high concentration. However, at low concentration, it works as an inhibitor of dpp. The concentration level of Ci is high near the A-P boundary, where it is exposed to the Hh signal. The concentration is low in the other cells in anterior compartment [Dominguez, et al., 96]. In this simulation, although the Ci concentration levels are same in the whole anterior compartment, it is appropriate that a strong dpp stripe expression appears near the A-P boundary, due to the Hh activation which is higher than Ci inhibition. This phenomena in the wing disc supports the simulation results, and suggests the possibility that Ci represses the dpptranscription not only in the wing disc but also in the leg disc near the A-P boundary.

### 6.2 The Coaxial Patterns of Gene Expressions

The most interesting point of leg formation is how the P-D axis and coaxial patterns form. One intuitively appealing hypothesis, as postulated in the boundary model, is that a diffusive factor is produced at the central region where both Wg and Dpp exist, and according the concentration level of this diffusive factor, different genes are expressed. However, a recent study by Lecuit and Cohen [Lecuit and Cohen, 97] casts doubt on such a possibility. Lecuit and Cohen argue that dll is activated when both Wg and Dpp are above a certain threshold. They also indicate that dac is directly activated by Wg and Dpp, not mediated through dll. In addition, Campbell *et al* [Campbell *et al.*, 93] reports that *aristaless*(*al*) is activated when both Wg and Dpp exist at a certain threshold, where this threshold is higher than that of dll. Together with related experimental results, it seems that all rings forming the coaxial pattern of the leg imaginal disc are directly formed by Wg and Dpp activation.

This hypothesis was validated by our simulation. As shown in Figure 3, a coaxial pattern can be formed by changing the activation thresholds for each gene (*al*, *dll*, *dac*). The gene *al* expresses in the most distal tip of the leg disc, because Wg and Dpp exist in high concentration in the center of the disc. The *dll* expression requires the a high concentration of Wg and Dpp, but not higher than *al* does. The *dac* expresses at a low concentration of Wg and Dpp, but does not express at a high concentration of Wg and Dpp.

#### 6.3 The Bi-Polar Model

From a series of simulations, it became clear that in the leg imaginal disc, the P-D axis is directly determined by the diffusion of wq and dpp products. We confirm that Lecuit and Cohen's model can actually create a coaxial shape. Combined with the expression patterns of other related genes, such as al, and several experiments on regeneration, as discussed in [French et al., 76] and [Meinhardt, 83], we now believe that a new model can be formulated which we call the Bi-Polar Model. The Bi-Polar model assumes that there are two sources of morphogen, and possibly one fate determinant factor for the A-P axis. In the Bi-Polar Model, Wg and Dpp act as morphogens, and determine their own positional value according to their concentration level. Figure 5 shows simplified view of the concentration-level contour for wg and *dpp* products. An interesting point of this diagram is that, within anterior or posterior hemispheres, the two-dimensional position of a cell can be uniquely determined by the level of concentration of Wg and Dpp. The A-P axis is determined by En and Ci, thus only one hemisphere need to be specified to determine a given cell.

In addition, for cells with the same concentration level of one of two gene products, there is a gradient crossing the equi-concentration line created by the other gene. For example, even if two cells at distance are exposed to the



Figure 5: The gradients of Wg and Dpp concentration. (a). Equi-concentration contour for Wg and Dpp. Two contours determine the expression of the target genes at different thresholds, and two dimensional position of a cell. (b). The thresholds of the genes which form the coaxial expression patterns.

same level of Wg, the Dpp levels are different. This enables cells to identify their position even if they are exposed to same level of concentration of one of two gene products.

It has been argued that the polar coordinate model is only empirical rules, rather than being a model describing underlying mechanism. The basic problem is how two positional value systems, the circular value and circumferential value, are established. In particular, no feasible molecular mechanisms have been proposed to explain how circumferential values are established. The bi-polar model provides a feasible molecular solution to this problem. As described in the previous paragraph, all cell positions can be uniquely identified. While it is a major question whether or not cells are actually using fine resolution circumferential values, instead of using coarse-grain sectors based on positional values, as postulated in the boundary model, the important point is that the bi-polar model can, if necessary, provide fine grained positional values.

The problem with the boundary model is that it assumes diffusive factors from the intersection of boundaries to form chemical gradients to establish the P-D axis. The bi-polar model is more flexible, and general explanation, since bi-polar sources of diffusive factors are likely to be established adjacent to the intersection of the boundary. In the boundary model, circumferential values are assigned using coarse Anterior-Ventral, Anterior-Dorsal, and Posterior sectors, instead of fine-grained values. In the bi-polar model, a sector-based positional value can be easily created if we assume that cells are using coarse decision logic. Therefore we argue that the bi-polar model is superior to existing models.

## 7 Conclusion

We have developed a biologically faithful simulation system for *Drosophila* leg formation. We simulate major eight genes and their interactions, and successfully reproduce the patterns of gene expressions observed in actual leg discs. Our results support the theory that the coaxial patterns of gene expressions of *al*, *dll* and *dac* in the leg disc are formed by two chemical gradients of *wg* and *dpp* products. In order for the simulated expression patterns to be consistent with actual expression patterns, we had to introduce some unknown interactions. Thus, we propose the hypothesis that Ci is required to repress *dpp* transcription in the anterior compartment in the leg disc. Ci repressing *dpp* has been found in wing imaginal disc, but it has not been clearly shown in the leg disc.

Detailed analysis and simulation of P-D axis formation led us to formalize the findings by Lecuit and Cohen on direct activation of *Dll* and *Dac*, combined with other related research results. The Bi-Polar Model postulates that the positional information along the P-D axis is determined by two morphogen gradients, which form bi-polar patterns of concentration gradients. We feel the bi-polar model is a superior alternative to existing models, because it can subsume existing models and is well supported from the molecular level.

## 8 Acknowledgement

This work was carried out under the Virtual *Drosophila* Project at Sony Computer Science Laboratory Inc. We would like to thank to Dr. Mario Tokoro at Keio University for his useful advice. We also want to express our thanks to Dr. Tetsuya Kojima at Tokyo University for providing useful comments and suggestions. Deep thanks also go to Sean Luke and Shugo Hamahashi, whose advice, suggestions and criticism have contributed substantially to this paper.

- H. Kitano, S. Hamahashi, J. Kitazawa, K. Takao, and S. Imai. Virtual Biology Laboratories: A New Approach of Computational Biology. In Proceedings of Fourth European Conference on Artificial Life, 274-283, 1997.
- 2. M. Morohashi and H. Kitano. A Method to Reconstruct Genetic Networks Applied to the Development of Drosophila's Eye. In Proceedings of the Sixth International Conference on Artificial Life, 72-80, 1998.
- S. Hamahashi and H. Kitano. Simulation of Drosophila Embryogenesis. In Proceedings of the Sixth International Conference on Artificial Life, 151-160, 1998.
- 4. H. Ueda and H. Kitano. A Generalized Reaction-Diffusion Simulator for Studying the Molecular Basis of Pattern Formation in Biological Systems.

In Proceedings of the Sixth International Conference on Artificial Life, 462-466, 1998.

- 5. V. French, P. J. Bryant, and S. V. Bryant. Pattern Regulation in Epimorphic Fields. *Science*, 969-981, 1976.
- 6. H. Meinhardt. A Boundary Model for Pattern Formation in Vertebrate Limbs. J. Embryol. Exp. Morphol., 115-137, 1983.
- 7. S. Kondo. A Mechanistic Model for Morphogenesis and Regeneration of Limb and Imaginal Discs. *Mechanism of Development*, 161-170, 1992.
- A. M. Turing. The Chemical Basis of Morphogenesis. Philos. Trans. R. Soc. Lond B 237, 37-72, 1952.
- 9. M. Bate and A. M. Arias. The Development of Drosophila melanogaster, vol II. CSHL Press, 1993.
- T. Lecuit and S. M. Cohen. Proximal-distal Axis Formation in the Drosophila Leg. Nature, 139-145, 1997.
- M. Dominguez, M. Brunner, E. Hafen, and K. Basler. Sending and Recieving the Hedgehog Signal: Control by the *Drosophila* Gli Protein Cubitus interruptus. *Science*, 1621-1625, 1996.
- 12. K. Basler and G. Struhl. Compartment Boundaries and the Control of Drosophila Limb Pattern by Hedgehog Protein. Nature, 208-214, 1994.
- P. J. Bryant. The Polar Coordinate Model Goes Molecular. Science, 471-472, 1993.
- S. Eaton and T. B. Kornberg. Repression of *ci-D* in Posterior Compartments of *Drosophila* by *engrailed*. *Genes and Development*, 1068-1077, 1990.
- G. Campbell, T. Weaver, and A. Tomlinson. Axis Specification in the Developing Drosophila Appendage: The Role of wingless, decapentaplegic and the Homeobox Gene aristaless. Mechanism of Development, 1113-1123, 1993.
- W. J. Brook and S. M. Cohen. Antagonistic Interactions Between Wingless and Decapentaplegic Responsible for Dorsal-Ventral Pattern in the Drosophila Leg. Science, 1373-1377, 1996.
- M. Zecca, K. Basler, and G. Struhl. Direct and Long-Range Action of a Wingless Morphogen Gradient. *Cell*, 833-844, 1996.