Integrated Cellular and Gene Interaction Model for Cell Migration in Embryonic Development

Hien Nguyen and Mingzhou (Joe) Song
Department of Computer Science
New Mexico State University
Las Cruces, NM 88003, USA
hinguyen@cs.nmsu.edu, joemsong@cs.nmsu.edu

Abstract—**The relationship between cellular behaviors and protein concentrations is central for embryonic development. An integrated cellular and gene interaction model is proposed to reveal this relationship. Protein concentrations vary spatiotemporally based on locations of cell, gene-gene interactions and the diffusion mechanism. On the other hand, cellular behaviors differentiate spatially, driven by cell-cell communication and protein concentrations in each cell. The model integrates a variation of the reaction-diffusion equation at the gene expression level and a particle-based cellular model based on the differential adhesion hypothesis for cell sorting at the cellular level. Cell sorting based on the adhesion hypothesis and patterns of inhibitory genes were simulated to illustrate the model. This model provides a more comprehensive basis to explain pattern formation during embryogenesis, than existing approaches that ignore the cellular components**

Keywords- embryonic multi-scale modelling; differential adhesion; reaction-diffusion equations

I. INTRODUCTION

Embryonic development is a complex process in which the life form progresses from a single homogeneous cell to an embryo composed of heterogeneous cells. To describe such an intricate process, many models [2, 3, 5, 7, 11] have been developed to investigate gene interactions at the molecular level, to account for spatiotemporal patterns of gene expression at the cellular level during embryogenesis. One category of widely used mathematical models is based on the reaction-diffusion mechanism. These models have been established for a variety of biological phenomena, including animal coats [12], human brain development [1], and gene regulatory interactions [5]. A well-studied example is stripe formation in the fruit fly embryo, where the reaction-diffusion mechanism is used to describe the dynamic behaviors of maternal genes [2, 5, 10, 11].

In developmental biology, the reaction-diffusion mechanism is applicable thanks to the assumption that gene expressions, regulating each other, can diffuse freely in space. In the case of *Drosophila*, the assumption is valid at the blastoderm stage when the cell membrane has not formed. Thus the corresponding cleavage cycles from 10 to 13 are well investigated [2,3,5,10,11]. Unfortunately, the assumption becomes problematic when the spatial boundary formed by a cell is considered. Firstly, gene regulation, for transcription, translation, and post-translation, mostly occurs inside the cell. Hence, genes can only regulate each other where a cell exists and conditions in the cell are right. Secondly, expressed gene products cannot freely diffuse because of the enclosing cell membranes. Therefore, cells have a critical impact on gene interactions. In the mean while, cells have their own operations, including mitosis, migration, communication with other cells, and death. Not surprisingly, these behaviors are heavily regulated by gene expressions in the cell. How cells and gene expressions behave regarding each other as an integrated system is a profound question in biology.

Intuitively, models that consider both the cell and gene expression at the same time are better to capture their dynamic behaviors. This kind of models can be found in the field of artificial life. Bentley [13] developed an Evolutionary Developmental System (EDS), which considered a cell as an agent with receptors to sense the environment and effectors to perturb the environment. The molecular parameters, including synthesis rate, diffusion rate, decay rate, transcription factors, etc., are all contained in the artificial genome. The EDS was successful in studying evolved mechanisms to construct some basic patterns. Fleischer [16] developed a complicated system called simulation testbed, in which physical, chemical and electrical factors were modeled in cell-cell communication. Agarwal [15] provided a Cell Programming Language (CPL) to specify and simulate a variety of biological phenomena. In his thesis, he also provided an implementation of cell motility using differential adhesion with fixed cell affinity. More recently, Jiang et al. [9] proposed a multi-scale model which uses the Cellular Potts Model for intercellular interactions, and PDE for intracellular and extracellular interactions. Although these mentioned models study both cells and proteins, Fleischer [16] and Agarwal [15] aimed to discover evolutionary mechanisms to apply to computer systems and/or algorithms, while Jiang [9] applied the model for the growth of (avascular) tumor, where cells occupied all of the space and cell affinity is not affected by gene regulations (only cyclic proteins are considered in intracellular PDE).

Instead of designing biologically inspired computing systems, our ultimate goal is to develop modeling to explain accurately biological patterns. In this paper, we illustrate our model by simulating cell migration at the cellular level and the inhibitory gene interaction at the molecular level. The cell migration is simulated based on the differential adhesion

978-1-4244-8302-0/10/$26.00 ©2010 IEEE

hypothesis; and the inhibitory interaction is simulated using the reaction-diffusion mechanism. Our simulation shows that, in terms of pattern formation, the integrated model gives more dynamic results from non-cell-based gene expression models. Broadly speaking, such a model will be more comprehensive than the typical non-cell specific models used for gene networks and pattern formation.

II. The Integrated Cellular And Molecular Model

We define an integrated cellular and molecular model, whose system diagram is shown in Figure 1.

The system consists of two interconnected subsystems, one at the gene expression level and the other at the cellular level. The gene expressions level subsystem is responsible for dynamics of gene regulation. This lower-level subsystem accepts cell positions from the cellular level subsystem as an input. In turn, the higher-level subsystem, responsible for dynamics of cell movements, takes the spatiotemporal data of protein concentrations as the input. In general, the integrated system can be learned by parameter estimation in gene regulatory network reconstruction, as well as in pattern formation.

In our model, we make the following main assumptions. Firstly, the proteins can diffuse one place to another place in the neighborhood. In the presence of cells, the diffusion rate would be much lower than when no cells appear to block. Secondly, cells move based on the differential adhesion hypothesis. This hypothesis was proposed by Malcom Steinberg in 1964 [6], stating cells sort in the way that maximizes the adhesion energy. Thirdly, we assume that the adhesion of a cell depends on a particular kind of protein (which is the cadherins family in reality). And most importantly, space and time are discrete.

This proposed model allows us to explore gene regulatory interactions in a broader embryonic developmental stage, e.g., not restricted to the blastoderm stage. In addition, it also allows us to investigate cell movement in a more dynamical way.

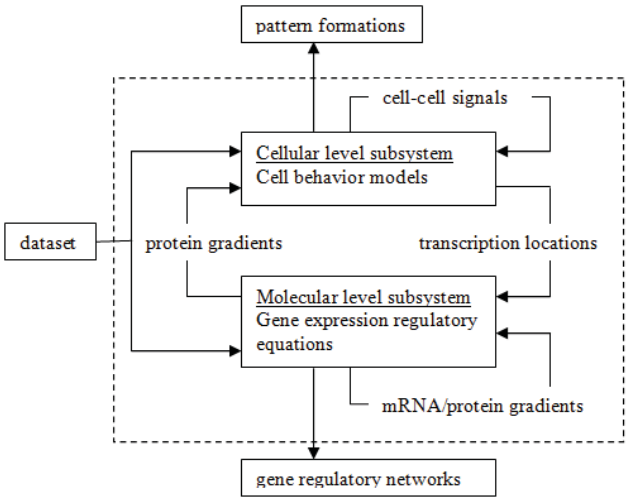

Figure 1. System diagram of the integrated cellular and molecular model. The model includes the cellular subsystem and the molecular subsystem.

A. Cellular and Molecular States of a System

The space of the embryo is divided into a grid of **sites**. Each site is a location where at most one cell can be present. A site has several neighboring sites with which it has contact. Given a site s and a time step t, we define the following functions.

The function C(s,t) to indicate the presence of a cell at site s and time t is defined as

$$C : \text{Site X Time} \rightarrow \{0,1\}$$

$$C(s,t) = \begin{cases} 1, & \text{if a cell is present} \\ 0, & \text{if no cell is found} \end{cases}$$

The function P(s,t,i) to specify the concentration of i^{th} protein at site s and time t is defined as

$$P : \text{Site X Time X } \aleph \rightarrow \Re,$$

$\aleph$: natural numbers ; $\Re$: real numbers

$$P(s,t,i) \geq 0 \quad \forall s,t,i$$

B. Cellular and Molecular Interactions

According to the differential adhesion hypothesis, different cells have different adhesion. We assign to each cell an adhesive factor, which is proportionate to the concentration of the k^{th} protein. The adhesion energy between two cells is the product of their adhesive factors.

The adhesive factor AF(s,t) of site s at time t is defined as follows:

$$AF : \text{Site X Time} \rightarrow \Re$$

$$AF(s,t) = P(s,t,k)^* C(s,t)$$

where k indicates the k^{th} protein that controls the adhesion.

The adhesive energy $AE(s_1,s_2,t)$ between two sites s_1 and s_2 at a time t is defined as

$$AE : \text{Site X Site X Time} \rightarrow \Re$$

$$AE(s_1,s_2,t) = AF(s_1,t)^*AF(s_2,t)$$

The swap operation between two sites s_1 and s_2 is described as follows

Case $C(s_1,t) = 0$ AND $C(s_2,t) = 0$: Skip

Case $C(s_1,t) = 1$ AND $C(s_2,t) = 0$:

Exchange $C(s_1,t)$ and $C(s_2,t)$

$P(s_2,t,i) \leftarrow P(s_2,t,i) + P(s_1,t,i)\ \forall i$

$P(s_1,t,i) \leftarrow 0\ \forall i$

Case $C(s_1,t) = 0$ AND $C(s_2,t) = 1$:

Exchange C(s1,t) and $C(s_2,t)$

$P(s_1,t,i) \leftarrow P(s_1,t,i) + P(s_2,t,i)\ \forall i$

$P(s_2,t,i) \leftarrow 0\ \forall i$

Case $C(s_1,t) = 1$ AND $C(s_2,t) = 1$:

Exchange $P(s_1,t,i)$ and $P(s_2,t,i)\ \forall i$

The energy $E(s_1, s_2)$ obtained when swapping two sites s1 and s2 is calculated as follows:

$$E: \text{Site X Site} \rightarrow \Re$$

$$E(s_1,s_2) = \sum AE(s_1,s_n) + \sum AE(s_2,s_m) - \sum AE(s_1,s_m) - \sum AE(s_2,s_n)$$

where s_m represents the neighbor sites of s_2 except for s_1, and s_n represents the neighbor sites of s_1 except for s_2.

Cell movement is performed at every time step, by a cell site updating algorithm described as follows:

1: set all sites unvisited
2: repeat until all sites are visited
3: randomly select an unvisited site **s**
4: select the neighbor site **s'** such that
$E(s,s') = \max_{\forall s_k \text{ in the neighborhood of } s} E(s,s_k)$
5: swap **s** and **s'**
6: mark **s** as visited

Also, at each time step, the protein concentrations at a site are also updated by the following differential equations:

$$\frac{dP(s,t,i)}{dt} = C(s,t) * \sigma * S(\sum m_{ij} * P(s,t,j) - h_i) + \sum D(s,s_k,t) * (P(s_k,t,i) - P(s,t,i)) + \chi * P(s,t,i)$$

where

σ : synthesis rate
S : sigmoid function $S(x) = 1/(1+e^{-x})$
m_{ij}: coefficients representing the relationship between i^{th} and j^{th} proteins
h_i: threshold for protein synthesis
$D(s,s_k,t)$: diffusion rate between *s* and s_k at time *t*
χ : degradation rate.

The first term represents protein synthesis; the second term diffusion of protein molecules; the third term protein degradation. This differential equation is implicit for mRNA to protein transcription, but captures protein-protein interactions.

The diffusion rate $D(s_1,s_2,t)$ depends on whether cells are present at sites s_1 and s_2. It is defined as follows:

D: Site X Site X Time $\rightarrow \Re$

$$D(s_1,s_2,t) = (1-C(s_1,t))*(1-C(s2,t))*d_1 + [C(s_1,t)*(1-C(s_2,t)) + C(s_2,t)*(1-C(s_1,t))]*d_2 + C(s_1,t)*C(s_2,t)*d_3$$

where d_1, d_2, d_3 are diffusion rates for the three cases respectively. The first term represents the case when cells exist at neither site; the second term when one cell is present; the third case for existence of both cells.

III. Simulation Studies

We carried out three simulation studies to illustrate various aspects and capabilities of our integrated model. In the first study, we present the cell sorting simulation in an isolated cellular level subsystem. The result is conformable to the biological phenomenon that led to the adhesion. The second study shows the inhibitory interaction of two genes in an isolated gene expression level subsystem. The third study presents the simulation result for an integrated system that includes both components in the previous two studies.

A. Cell Sorting based on the Adhesion Hypothesis at the Cellular Level

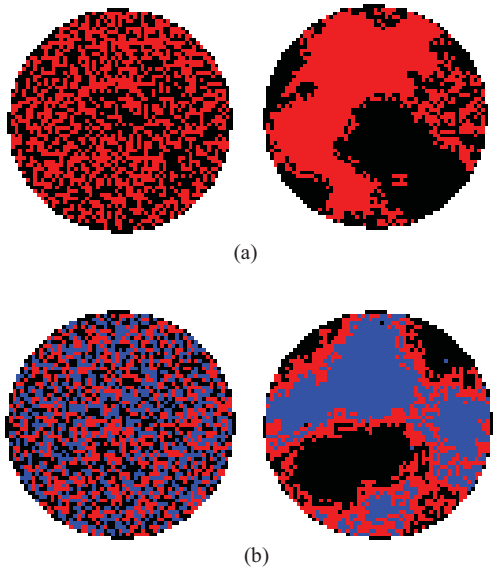

Figure 2. Cellular pattern formation during cell sorting**.** Dots represent cells. Blue cells have the strongest adhesion factor; red ones have intermediate adhesion factors; and the black ones have the weakest adhesion factors. (a) Cells sorting with two types of cells. Left: initial configuration. Right: sorted configuration. (b) Cell sorting with three types of cells. Left: initial configuration. Right: sorted configuration.

The adhesion hypothesis originated from the experiments of Townes and Holtfreter in 1955 [6]. In their experiments, they extracted cells from different layers of amphibian embryos after the neural tube had formed and mixed them together. Then they observed that the cells aggregated to form layers as in the original embryos.

We artificially created multiple types of cells with different adhesion factors to test if our model can conform to the adhesion hypothesis. The number of cells of each type is assigned randomly, and the adhesion factors satisfy the aggregation condition of the adhesion hypothesis.

Global pattern formation due to the cell adhesion is illustrated in Figure 2. Although the cells were randomly distributed spatially, consistent global patterns are formed after cell adhesion is modeled. This study demonstrates the capability of the proposed model in capturing cell adhesion behaviors at the cellular level.

B. Expression and Regulation among Inhibitory Genes at the Molecular Level

Now we inspect our model at the gene expression level without considering the cellular level. We assume that cells are distributed all over the embryos. Hence, genes can interact at all sites. C(s,t) is set to 1 for all s and t. Cell motion is switched off.

Figure 3 is the gene expression pattern formation over time of the two inhibitory genes. Until time step 5, the inhibited gene product (the blue one) can grow up at the middle of embryo because the regulating gene product (the maroon one) is not present in this region. After that, the regulating gene products spread all over the embryo, because

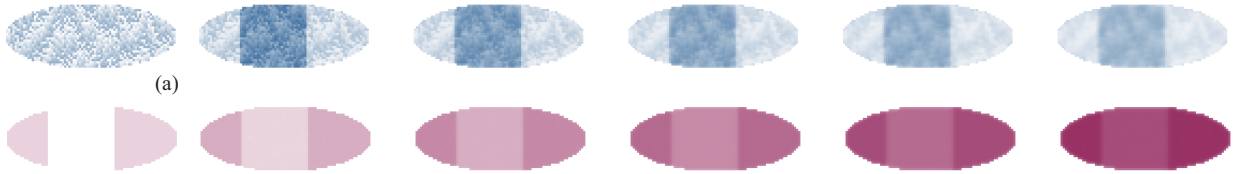

Figure 3. Expression pattern formation due to inhibitory gene interactions without cellular information**.** The inhibited protein is represented by light blue, the more concentrated the darker. The following parameters are used in generating the patterns: σ=1, d_1= 0.05, d_2=0.03, d_3=0.01, χ = 0.05, h_i = 0, m = -20 (inhibition coefficient). (a) The inhibited gene expression. (b) The regulating gene expression. From left to right: the change of the gene expression concentration over time.

of the diffusion as well as the transcription. Therefore, the inhibited gene product is suppressed but fades away because of the degradation.

C. Integrated Gene Interactions and Cell Sorting

In this simulation study, we integrate the cellular and the molecular level subsystems. As in the study in 3.2, the inhibited gene product is initially randomly distributed while the regulating one is initially concentrated at the ends of the embryo. To consider also the cellular subsystem, gene products can only be enhanced at locations where a cell is present. The cell adhesion factor is controlled by a third gene product, which takes part in regulatory interactions by the following coefficients: m_{31} = 5, m_{32} = 3, and m_{33}=0 (no self regulation).

Figure 4 shows the spatiotemporal progression of pattern formation. Considering the cellular constraints caused the proteins not to concentrate as in Figure 3. The simulation also shows that the stripe pattern is not present anymore. This study demonstrates the challenge of considering both cellular and molecular components in a model.

The pattern of the cell distribution shows local engulfment of cells. That is an expected pattern under the adhesion hypothesis. However, cell movements also require other conditional factors than adhesion, such as neighbor density or cell shape. Therefore, adhesion-only models may produce unrealistic patterns.

IV. DISCUSSION

We have proposed an integrated spatiotemporal model for gene interaction and cell sorting using reaction-diffusion mechanism and differential adhesion hypothesis. The core idea we explored is to couple two levels of abstraction, i.e., molecular and cellular, to model diffusion considering cell boundaries and adhesion alteration based on protein concentrations. We have created an integrated model to generate spatiotemporal patterns to mimic some aspects of embryo development. Each subsystem is consistent with basic biological principles; the entire model can account for some complex spatial pattern formation. More importantly, as both molecular and cellular levels are considered, the model can be applied to a variety of biological processes. We have demonstrated its utility as a basis for pattern formation during embryogenesis. Particularly, it can be applied in those developmental stages when proteins do not freely diffuse, providing a first step to study the embryo after the blastoderm stage.

However, spatiotemporal data for both gene products and cells are necessary to train this proposed model. This has posed challenges in biological data acquisition, as well as computational modeling. In addition to adhesion, other biological processes may also influence cell movement. Although integrating other cell behaviors such as division and differentiation can lead to more realistic models, it also makes system modeling challenging. These issues are

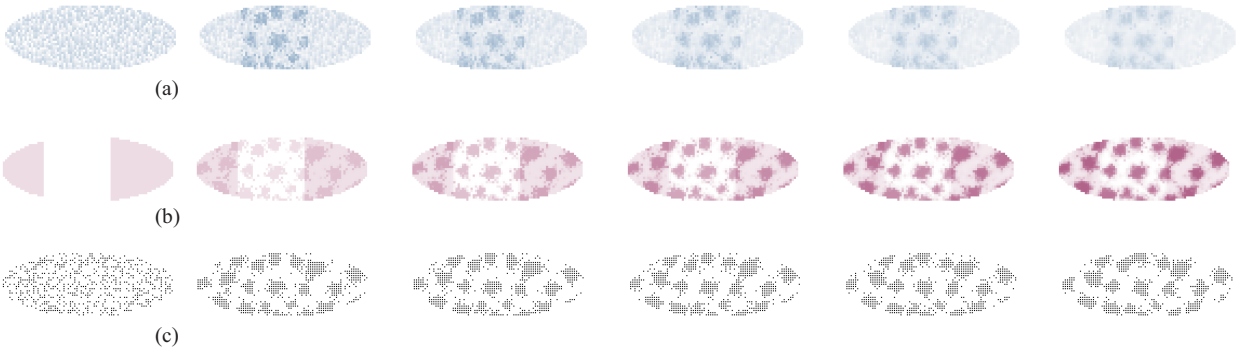

Figure 4. Pattern formation due to both inhibitory gene interaction at the expression level and cell sorting at the cellular level. The following parameters were used in generating the patterns: σ=1, d_1= 0.05, d_2=0.03, d_3=0.01, χ = 0.05, h_i = 0, m = -20 (inhibition coefficient). (a) The inhibited gene expression. (b) The regulating gene expression. (c) The cell distribution.

generally recognized and have received attention from research community.

Our current and future work develops along the following directions. Since the model is applicable to three-dimensional spaces in principle, a future task is to extend the current two-dimensional implementation of the model to three dimensions in space. Another direction is to extend the modeling work in gene networks to incorporate spatial constraints due to the existence of cells. With these efforts and increasingly available spatiotemporal data of gene expression at cellular resolution, we look forward to gaining more insight on the cellular and molecular mechanisms underlying diverse cell behaviors.

REFERENCES

[1] J. Lefèvre, J.F. Mangin, "A Reaction-Diffusion Model of Human Brain Development", PLoS Computational Biology, April 2010.

[2] D. Papatsenko, "Stripe formation in the early fly embryo: principles, models, and networks", BioEssays Volume 31 Issue 11, 2009, pp. 1172 – 1180.

[3] C.W. Li, B.S. Chen, "Stochastic Spatio-Temporal Dynamic Model for Gene/Protein Interaction Network in Early Drosophila Development", Gene Regulation and Systems Biology, 2009:3, pp. 191-210.

[4] C. Fowlkes et al. "A Quantitative Spatiotemporal Atlas of Gene Expression in the Drosophila Blastoderm," Cell, 133(2), 2008, pp. 364-374.

[5] Y. Fomekong-Nanfack, J.A. Kaandorp, J.G. Blom, "Efficient parameter estimation for spatio-temporal models of pattern formation: Case study of Drosophila melanogaster", Bioinformatics, vol. 23, 2007, pp. 3356-3363.

[6] S. F. Gilbert, Developmental Biology, 8th ed., Sinauer Associates, Inc., USA, 2006.

[7] A. Deutsch, S. Dormann, Cellular automaton modeling of biological pattern formation : characterization, applications, and analysis. Birkhauser Boston, USA, 2005.

[8] E. Klipp, R. Herwig, A. Kowald, C. Wierling, Systems Biology in Practice: Concepts, Implementation and Application, Wiley-VCH, Germany, 2005.

[9] Y. Jiang, J. Pjesivac-Grbovic, C. Cantrell, J.P. Freyer, "A Multiscale Model for Avascular Tumor Growth", Biophysical Journal Vol. 89, 2005, pp. 3884–3894.

[10] J. Jaeger et al., "Dynamical analysis of regulatory interactions in the gap gene system of Drosophila melanogaster", Genetics 167, 2004, pp. 1721-1737.

[11] J. Jaeger et al., "Dynamic control of positional information in the early Drosophila blastoderm", Nature 430, 2004, pp. 368-371.

[12] J.D. Murray, Mathematical Biology, Vol 2, Springer-Verlag, USA, 2004.

[13] S. Kumar, P.J. Bentley, "Biologically inspired evolutionary development", Proceedings of the International Conference on Evolvable Systems: from biology to hardware, Trondheim, Norway, 2003.

[14] H. De Jong, "Modeling and Simulation of Genetic Regulatory Systems: A Literature Review", Journal of Computational Biology, Volume 9, Number 1, 2002, pp.67-103.

[15] P. Agarwal, "Cell-based Computer Models in Developmental Biology", Ph.D dissertation, Dept of Computer Science, New York University, 1993.

[16] K. Fleischer, A. Barr, "A simulation testbed for the study of multicellular development: The multiple mechanisms of morphogenesis", Artificial Life III, 1993, pp. 389-416.

[17] F. Graner, J.A. Glazier, "Simulation of Biological Cell Sorting Using a Two-Dimensional Extended Potts Model", Physical Review Letters, Vol 69, 1992, pp.2013-2016.

[18] J. Campos-Ortega, et al., The embryonic development of Drosophila melanogaster, Springer, 1985.