

Biological Development of Cell Patterns: Characterizing the Space of Cell Chemistry Genetic Regulatory Networks

Nicholas Flann¹, Jing Hu¹, Mayank Bansal¹,
Vinay Patel¹, and Greg Podgorski²

¹ Computer Science Department, Utah State University, Logan UT, USA

² Biology Department, Utah State University, Logan UT, USA

nick.flann@usu.edu

<http://www.cs.usu.edu/~flann/index.html>

Abstract. Genetic regulatory networks (GRNs) control gene expression and are responsible for establishing the regular cellular patterns that constitute an organism. This paper introduces a model of biological development that generates cellular patterns via chemical interactions. GRNs for protein expression are generated and evaluated for their effectiveness in constructing 2D patterns of cells such as borders, patches, and mosaics. Three types of searches were performed: (a) a Monte Carlo search of the GRN space using a utility function based on spatial interestingness; (b) a hill climbing search to identify GRNs that solve specific pattern problems; (c) a search for combinatorial codes that solve difficult target patterns by running multiple disjoint GRNs in parallel. We show that simple biologically realistic GRNs can construct many complex cellular patterns. Our model provides an avenue to explore the evolution of complex GRNs that drive development.

Keywords: GRN, cascading GRNs, recurrent GRNs, artificial embryology, development, developmental programs, cellular differentiation, differential equations, hill-climbing.

1 Introduction

Development from a single cell, the zygote, into the adult organism is a remarkably complex and poorly understood process. One common way to create patterns in the early embryo is through the use of diffusible morphogens that form gradients to provide positional information to embryonic cells. Cells acquire different identities within a developing field according to the levels of morphogen they detect. Broad divisions within the embryo can be established this way. More elaborate and finely resolved patterns are often created later when cells that have acquired coarse positional identities interact across membranes. These interactions generate different cell types in arrangements that include borders, patches and mosaics. Development in *Drosophila* follows this pattern, beginning with morphogens that specify broad divisions along the anterior-posterior and

dorsal-ventral axes, followed by cellular interactions that create sharply bounded divisions [18].

Many investigators have developed computational models that reproduce elements of development [20][19][13][15][5]. This is a relatively young field with great potential to provide insights that complement and extend those obtained by classical biological investigations. An important approach to modeling development is Artificial Embryogeny (AE). One path in AE follows a grammatical approach where sets of rules in the form of grammatical rewrite systems are evolved [1][2]. The other path of AE research is a cell chemistry-based approach that simulates the way structures emerge in biology e.g. [1][8][9][10][6].

Inspired by nature's elegance and precision in solving the problem of embryogenesis, we have developed a cell chemistry-based AE system to search for biologically realistic GRNs capable of generating patterns found in embryos. We study the effectiveness of GRNs in solving target patterns of cell types such as borders, patches, and mosaics that are observed in biological development. The power of this GRN search is that it allows unrestricted exploration of ways to solve embryological patterning problems. This stands in contrast to the actual embryo which, while solving complex problems of patterning, is a prisoner of its evolutionary history.

2 Approach

2.1 Cell Pattern Problems

Our focus is on how GRNs can build 2D spatial patterns similar or identical to those generated during biological development. A sample of the 43 patterns used in this study is shown in Fig.1. The individual patterns are grouped into three broad types: mosaic, border and patch. A mosaic pattern is defined as a periodic pattern in one or two dimensions repeated across the sheet of cells. A border pattern identifies single cells or single lines of cells that divides one group of cells

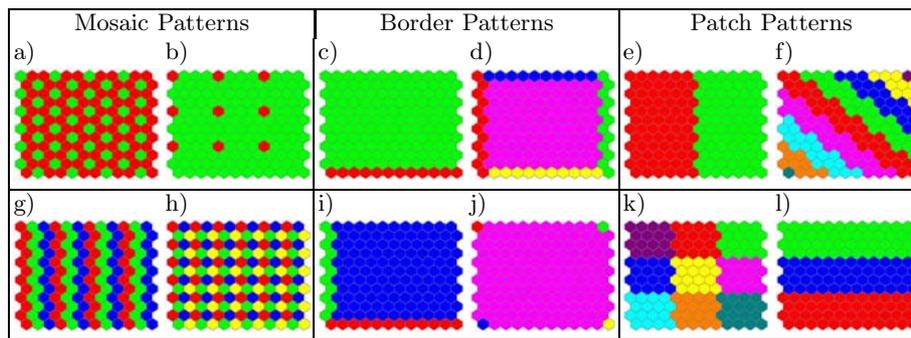


Fig. 1. A sample of the 43 target patterns used in this computational study. Each problem is represented as a 2D sheet of hexagonal cells (12 by 12, or 24 by 24). The complete set of target patterns is given in supplementary material [27].

from another or from the medium. A patch pattern is an aperiodic partitioning of the sheet of cells into contiguous groups of distinct cells types. Segments are a biologically significant subset of the patch pattern. Each of the pattern types is common in natural development. For example, regular mosaics occur in *Drosophila* epidermal neurogenic clusters where a single neuroblast becomes surrounded by non-neuronal supporting cells [18] and in the vertebrate retina where ganglion cells are encircled by cells of other types [24]. Patches in the form of segmental repeats are seen in the obvious segmental divisions along the insect anterior-posterior axis [18] and in the segmental array of vertebrate somites [26]. Borders are seen frequently, such as in the strip of cells that becomes a signaling center at the boundary between the anterior and posterior compartments of the *Drosophila* wing imaginal disc [18] and in the apical ectodermal ridge at the division between the dorsal and ventral ectoderm of the vertebrate limb bud [25].

2.2 Modeling Biological GRNs

Space of GRNs. Biological models of GRNs can be described as a graph, where each node represents a distinct protein's expression level and each edge represents influences among proteins. A protein is influenced when its production or inhibition is controlled as a function of other protein expression levels. Since production and inhibition are defined as rates of change, the GRN is naturally modeled as a set of coupled differential equations. Fig.2 shows an example of a 3 protein, 5 edge GRN and its input and output protein expression pattern.

Table 1. The edges composed to form a GRN. The rate of change of protein P_0 for cell σ is defined in terms of the weighted *expression* of protein P_i in σ and neighboring cells, where ω_j is the strength of edge j ($0.0 \leq \omega_j \leq 1.0$), $f(x) = \frac{x^2}{(1+x^2)}$; $g(x) = 1 - f(x)$; $h(x) = \frac{2}{(1+e^{-x})} - 1$; $n(\sigma)$ returns the set of directly neighboring cells (i.e., that share a common membrane); n_S, n_W, n_N, n_E return the cell directly neighboring to the South, West, North and East respectively.

Label	Description	Name	Definition
∇ A	P_0 Diffusion with zero boundary conditions	Diff0	$\frac{dP_0(\sigma)}{dt} = \omega_j \frac{\partial^2 P_1(\sigma)}{\partial x^2} = \nabla P_1(\sigma)$
B	P_0 Diffusion with fixed boundary conditions	DiffX	$\frac{dP_0(\sigma)}{dt} = \omega_j \frac{\partial^2 P_1(\sigma)}{\partial x^2} = \nabla P_1(\sigma)$
C	P_0 direct expression by P_1	ExprDirect	$\frac{dP_0(\sigma)}{dt} = \omega_j f(P_1(\sigma))$
D	P_0 direct inhibition by P_1	InhDirect	$\frac{dP_0(\sigma)}{dt} = -\omega_j f(P_1(\sigma))$
E	P_0 driven to same as P_1	Same	$\frac{dP_0(\sigma)}{dt} = \omega_j (f(P_1(\sigma)) - P_0(\sigma))$
F	P_0 driven to opposite of P_1	Oppos	$\frac{dP_0(\sigma)}{dt} = \omega_j (g(P_1(\sigma)) - P_0(\sigma))$
G	P_0 driven to difference in values between P_1 and P_2	Error	$\frac{dP_0(\sigma)}{dt} = \omega_j (h(P_1(\sigma) - P_2(\sigma)) - P_0(\sigma))$
H	P_0 autocatalysis and reciprocal control by P_1	ExprQuad	$\frac{dP_0(\sigma)}{dt} = \omega_j (f(\frac{P_0(\sigma)^2}{P_1(\sigma)} + \psi_j))$
I	P_0 quadratic inhibition by P_1	InhQuad	$\frac{dP_0(\sigma)}{dt} = \omega_j (g(P_1(\sigma)^2) - \psi_j)$
\diamond J	P_0 driven to same as cell neighbors values of P_1	SameNeig	$\frac{dP_0(\sigma)}{dt} = \omega_j (f(\sum_{\rho \in n(\sigma)} \frac{P_1(\rho)}{6}) - P_0(\sigma))$
K	P_0 driven to opposite of cell neighbors values of P_1	OpposNeig	$\frac{dP_0(\sigma)}{dt} = \omega_j (g(\sum_{\rho \in n(\sigma)} \frac{P_1(\rho)}{6}) - P_0(\sigma))$
L	P_0 driven to difference in opposing cell neighbor values of P_1	ErrorNeig	$\frac{dP_0(\sigma)}{dt} = \omega_j (f(\sum_{\rho \in n(\sigma)} \frac{P_1(\rho) - P_1(op(\sigma, \rho))}{6}) - P_0(\sigma))$
M:P	P_0 driven to same as geometric neighbor value of P_1 ; with $i \in N, W, S, E$	SameNeig _i	$\frac{dP_0(\sigma)}{dt} = \omega_j (f(P_1(n_i(\sigma))) - P_0(\sigma))$

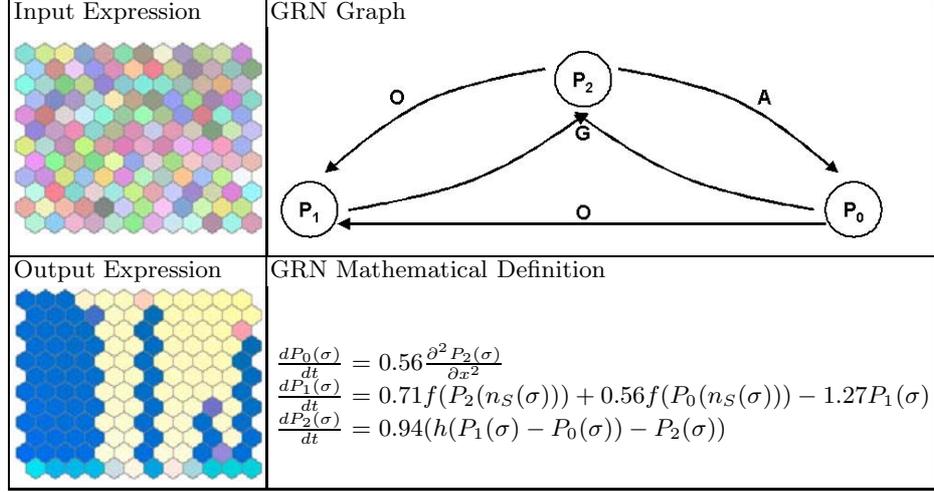


Fig. 2. Example of a GRN. The left column shows the input and output protein expression patterns, where the color of each cell is computed by mapping protein expression directly onto the *RGB* values. The right column shows the GRN as a graph and the associated set of coupled differential equations. The letters indicate edge types shown in Table 1.

Table 1 illustrates the protein influences considered in this study. Within an individual cell, protein expression can be controlled by a single protein (the direct control edges *C-F*, *I*) or some function of multiple proteins (the combinatorial control edges *G*, *H*). Over the sheet of biological cells, proteins influence each other through both long-range and short-range signaling. Edges *A* and *B* implement long-range signaling through diffusion under different boundary conditions. Edges *J-P* implement short-range signaling, where a cell can sense protein expression levels in directly neighboring cells across contacting membranes [20]. Notice edges *M-P* enable a cell to signal to a specific geometric neighbor cell. Such capabilities are known to be utilized to build internal segment borders [13] and rely on morphogenic gradients constructed earlier that establish anterior-posterior and dorsal-ventral axes. There are $d^m p^m p^{2p(m-p)}$ possible GRNs with p proteins and m edges (with d edge types). In all search studies, the GRN space is constrained to be connected, have at least one cycle, and with each protein having at least one *in edge*. A GRN is solved by first setting the protein values of each cell from a uniform random distribution, then numerically solving the differential equations using the Runge-Kutta method (with $dt = 0.05$) until either a fixed point (where the average update error $\leq 10^{-8}$ per cell; see Fig.2), or an instability is heuristically detected.

2.3 Determining GRN Computational Adequacy

Our goal is to characterize the space of biological GRNs and determine their computational adequacy to solve spatial design problems found in nature. Three

searches of the space of GRNs were performed. First, an open-ended Monte Carlo search [13] identifies simple GRNs that form interesting or useful spatial patterns. Second, a randomized hill-climbing search identifies simple GRNs that solve the specific spatial patterns of Fig.1. Finally, we searched for combinatorial codes that solve the difficult target patterns of Fig.1 by running multiple disjoint GRNs in parallel.

When searching the space of GRNs we attempt to identify the simplest GRN by systematically considering GRNs of increasing number of proteins and connectivity. In addition, we consider the adequacy of spatial signaling mechanisms by limiting signaling to diffusion or just cell-cell contact, or both. We define diffusion only signaling as ∇ , where edges are drawn from equations A – I of Table 1, and contact only signaling as \diamond , with edges from equations C – P of the same table.

3 Methods and Results

Monte Carlo Search: The Monte Carlo search samples the GRN space by constructing a GRN as a random graph. We vary the number of proteins p , $2 \leq p \leq 5$ and the number of edges m , $p \leq m \leq p + 3$. Each GRN is evaluated for its *interestingness* by first solving the GRN with an initial random protein expression then evaluating the spatial regularity of the resulting expression pattern. The algorithm is provided in [27]. The spatial regularity is evaluated by a heuristic measure of interestingness [14] over the frequency terms of the 2D FFT of the protein expression pattern.

Monte Carlo Results: Fig.3 shows some high scoring GRNs and their patterns of expression. Results suggest that the space of GRNs is indeed dense, with weak

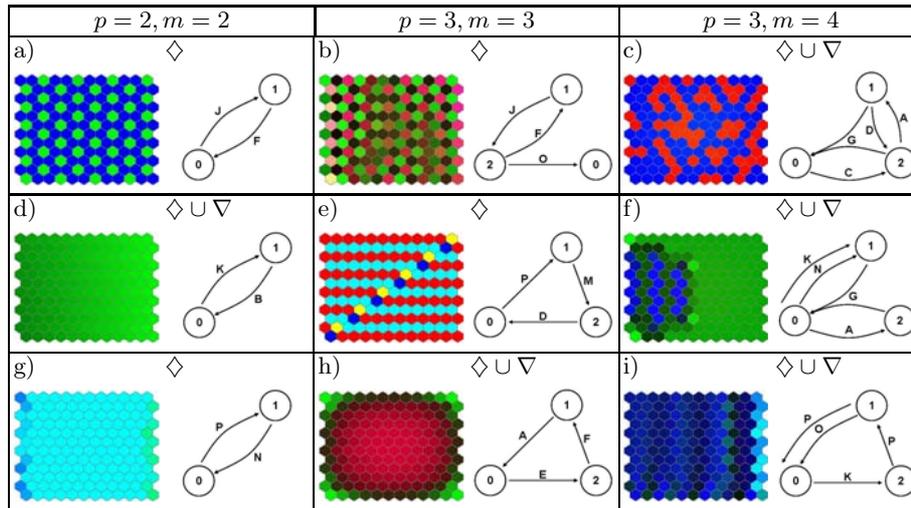


Fig. 3. Examples of high scoring Monte Carlo results and underlying GRN

random sampling identifying many useful GRN's. Small GRN's were found that identify N,S,E,W borders, corners, diagonal partitions, alternating stripes of period 2, 3, 4, and 6, both vertically and horizontally, and mosaics of varying periodicity and regularity. Most interesting were the irregular patterns found but not included in Fig.1, such as mosaic patterned patches (e.g., Fig.3f) and irregular cell clusters (e.g., Fig.3c)

Regular mosaics, patches and borders were all found when edges were restricted to direct contact signaling, demonstrating that this short-range, local form of signaling is a powerful mechanism. Borders and simple segments were found when edges were restricted to diffusible signals, while the irregular patterns on the right of Fig.3 required both direct contact and diffusible signaling.

Just as interesting were the patterns not found. No complex orthogonally segmented patterns such as the checkerboard pattern of *Drosophila* proneural clusters [22] were found (e.g. Fig.1k). Neither were complex orthogonal mosaic patterns, seen in the zebra fish retina [21] (e.g. Fig.1h). These results suggest that some pattern problems may be too difficult to solve using a single simple ($p \leq 5$, $m \leq 8$) GRN starting from a random distribution.

Hill Climbing Search: The hill climbing search takes target patterns from those given in Fig.1 and attempts to find simple GRN's that best solve each problem. The algorithm is provided in [27]. The mutation operator adjusts the strength parameters ω_k of each edge k by a small random amount. At a lower frequency, the edge *type* (Column 1 Table 1), *in-protein* or *out-protein* are randomly modified. Mutation preserves the connectedness constraint and signaling constraints of the GRN as described above.

To evaluate the fitness of a GRN, its fixed point expression distribution is matched against the provided target pattern. In natural systems, combinatorial codes of protein expression are used to discriminate cell types. The matching function is designed so that combinatorial protein expression codes are discovered, then evaluated as to how they identify cells of the same type, while discriminating cells of different types. Let $D(\sigma, \rho)$ be the Euclidian distance between the protein values of two cells σ and ρ . First, the cells are *k-means* clustered using $D(\sigma, \rho)$ into bins, then the bins are matched against the types assigned in the given target pattern. Cells clustered into the same bin are assigned a unique identifier, $B(\sigma)$. Good clusterings are identified that minimize the distance $D(\sigma, \rho)$ when $B(\sigma) = B(\rho)$ and maximize the distance when $B(\sigma) \neq B(\rho)$. The match is measured by comparing $B(\sigma)$ with the assigned types in the target pattern $T(\sigma)$ (as in [13]) corresponding to distinct colors in Fig.1. To evaluate robustness, each GRN is run over multiple initial random protein distributions and over two cell array sizes (12×12 and 24×24 .)

Hill Climbing Results: Fig.4 shows a selection of results executing hill climbing over the patterns in Fig.1. In Fig.4, combinatorial codes are denoted as Boolean expressions over the proteins of cell $P_i(\sigma)$, where H denotes a high expression, L denotes a low expression, and M denotes an intermediate expression corresponding to cluster values found by *k-means*. As suggested by the Monte Carlo stud-

	Signaling	GRN Graph	Code	Target	Solution
a)	\diamond		$\begin{array}{c cc} P_0(\sigma) & L & H \\ P_1(\sigma) & H & L \\ \hline B(\sigma) & 0 & 1 \end{array}$		
b)	$\diamond \cup \nabla$		$\begin{array}{c cc} P_0(\sigma) & H & L \\ P_1(\sigma) & L & H \\ P_2(\sigma) & L & H \\ \hline B(\sigma) & 0 & 1 \end{array}$		
c)	\diamond		$\begin{array}{c cccc} P_0(\sigma) & L & L & H & L \\ P_1(\sigma) & M & L & L & H \\ P_2(\sigma) & H & H & L & L \\ \hline B(\sigma) & 0 & 1 & 2 & 3 \end{array}$		
d)	\diamond		$\begin{array}{c cc} P_0(\sigma) & L & H \\ P_1(\sigma) & L & L \\ P_2(\sigma) & L & H \\ P_3(\sigma) & H & L \\ \hline B(\sigma) & 0 & 1 \end{array}$		
e)	$\diamond \cup \nabla$		$\begin{array}{c cc} P_0(\sigma) & L & L \\ P_1(\sigma) & H & H \\ P_2(\sigma) & H & L \\ P_3(\sigma) & L & L \\ \hline B(\sigma) & 0 & 1 \end{array}$		

Fig. 4. Examples of high scoring Hill Climbing runs. Each row is the best run found for the target pattern. The signaling constraint, simplest GRN, the combinatorial code discovered, the target pattern and expression pattern are shown in each row.

ies, many target patterns were easily solved from initially random distributions. For example, the GRN and combinatorial code shown in Fig.4a is the *dual* of the classic solution for lateral inhibition with feedback for mosaic construction [20]. Significantly, like the Monte Carlo simulation, simple hill climbing from an initially random pattern failed to create many biologically relevant patterns, including complex orthogonal mosaics (*e.g.* Fig.1h) and patches (*e.g.* Fig.1k).

Table 5 summarizes a systematic computational exploration of adequacy, where the simplest GRN was identified using only direct contact signaling \diamond , or only diffusible signaling ∇ , or both signaling types. Target patterns were selected from each problem class (mosaic, patch and border).

Mosaics: Diffusion signaling alone is inadequate to create mosaics, while direct contact signaling is both necessary and sufficient. Furthermore, GRNs for complex mosaics, such as those in the zebra fish retina [21] (*e.g.* Fig.1h, where h denotes the Code) are difficult to find when searching from random GRNs. This suggests an alternative incremental evolutionary strategy of mutating GRNs that solve related mosaics. An example of related GRNs is shown in Fig.3. The GRN in Fig.3b that solves this complex pattern is a mutant version of the simpler GRN in Fig.3a, the difference being the addition of one protein and one edge. We are currently exploring the power of this evolutionary strategy.

Combining GRNs Results: The results of running two GRNs in parallel are shown for each target pattern in the last column of Fig.5. Complex orthogonal target patterns (*e.g.* Fig.1k), which were hard to solve using a single GRN, can easily be solved by this method. For example, a single GRN of 5 proteins and 7 edges (both contact and diffusion) is needed to identify the nested circular pattern (Fig.5h), whereas two small (2 protein, 2 edge) GRNs can solve it easily.

4 Conclusions

In this work we show that simple biologically realistic GRNs are very powerful and capable of constructing many complex cellular patterns. One significant and unanticipated result is that cell-cell contact signaling is sufficient to form many global patch patterns. The complexity of the cellular patterns formed by the simple GRNs modeled here poses the question of why biological GRNs are so complex. This study revealed some patterns that could not be solved by a single small GRN. These patterns, however, were solvable when disjoint GRNs were run in parallel and their protein expression levels combined. Our results support the view that complex GRNs may have evolved in nature by combining simpler modules. Our model provides an avenue to explore the evolution of complex GRNs that drive development.

Acknowledgements

The authors wish to thank Dharmesh Shar, Ranjitha Dhanasekaran, Swapna Challa and Prashant Rai for help in preparing the paper and writing the code.

References

1. Stanley, K., Mikkulainen, R.: A Taxonomy For Artificial Embryogeny. *Artificial Life* 9(2): (2003) 93–130
2. Lindenmayer, A.: Mathematical Models For Cellular Interaction In Development: Parts I and II. *Journal of Theoretical Biology*. **18** (1968) 280–315
3. Gruau, F.: Neural Network Synthesis Using Cellular Encoding And The Genetic Algorithm. Doctoral dissertation, Ecole Normale Supérieure de Lyon, France (1994)
4. Gruau, F., Whitley, D., & Pyeatt, L.: A Comparison Between Cellular Encoding And Direct Encoding For Genetic Neural Networks. In J. R. Koza, D.E. Goldberg, D. B. Fogel, & R. L. Riolo (Eds.), *Genetic Programming 1996: Proceedings of the First Annual Conference*, Cambridge, MA: MIT Press (1996) 81–89
5. Hogeweg, P. Computing an organism: on the interface between informatic and dynamic processes. *Biosystems*, (2002) 64: 97-109
6. Astor J.S. & Chris Adami C. A Developmental Model for the Evolution of Artificial Neural Networks. *Journal of Artificial Life*. (2000) **6:3** 189-218
7. Turing, A.: The Chemical Basis Of Morphogenesis. *Philosophical Transactions of the Royal Society B*. **237** (1952) 37–72
8. Fleischer, K., Barr, A.H.: A Simulation Testbed For The Study Of Multicellular Development: The Multiple Mechanisms Of Morphogenesis. In C. G. Langton (Ed.), *Artificial life III*. Reading, MA. Addison-Wesley (1993) 389–416

9. Mjolsness, E., Sharp, D. H., & Reintz, J.: A Connectionist Model Of Development. *Journal of Theoretical Biology.* **152** (1991) 429–453
10. Kaneko, K., Furusawa, C.: Emergence Of Multicellular Organisms With Dynamic Differentiation And Spatial Pattern. *Artificial Life.* **4** (1998) 77–93
11. Federici, D.: Using Embryonic Stages To Increase The Evolvability Of Development. In proceedings of WORLDS Workshop on Regeneration and Learning in Developmental Systems, hosted by GECCO (2004)
12. Roggen, D., Federici, D.: Multi-Cellular Development: Is There Scalability And Robustness To Gain? In proceedings of PPSN VIII 2004 The 8th International Conference on Parallel Problem Solving from Nature (2004) 391-400
13. von Dassow, G., Meir, E., Munro, E.M., & Odell, G.M.: The Segment Polarity Network is a Robust Developmental Module. *Nature.* **406** (2000) 188–192
14. Hilderman, R.J., Hamilton, H.J.: Heuristics For Ranking the Interestingness Of Discovered Knowledge. In N. Zhong and L. Zhou, editors, Proceedings of the Third Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'99), Beijing, China, (1999) 204–209
15. Izaguirre, J.A., Chaturvedi, R., Huang, C., Cickovski, T., Coffland, J., Thomas, G., Forgacs, G., Alber, M., Hentschel, G., Newman, S.A., & Glazier J.A.: CompuCell, a Multi-Model Framework For Simulation Of Morphogenesis. *Bioinformatics.* **20** (2004) 1129-37
16. Bongard, J. C.: Evolving Modular Genetic Regulatory Networks. In Proceedings of the IEEE 2002 Congress on Evolutionary Computation (CEC2002), IEEE Press, (2002) 1872–1877
17. Taylor, T.: A Genetic Regulatory Network-Inspired Real-Time Controller For a Group Of Underwater Robots. In Intelligent Autonomous Systems 8 (Proceedings of IAS8), F. Groen, N. Amato, A. Bonarini, E. Yoshida and B. Krse (eds.), IOS Press, Amsterdam, (2004) 403–412.
18. Lawrence, P.A.: *The Making Of a Fly.* Blackwell Scientific Publications, Oxford (1992)
19. Shvartsman, S.Y., Muratov, C.B., & Lauffenburger, D.A.: Modeling And Computational Analysis Of EGF Receptor-Mediated Cell Communication In *Drosophila* Oogenesis. **129** *Development* (2002) 2577–2589
20. Collier, J.R., Monk, N.A.M., Maini, P.K., & Lewis, J.H.: Pattern Formation By Lateral Inhibition With Feedback: A Mathematical model Of Delta-Notch Inter-cellular Signalling. **183** *J. Theor. Biol.* (1996) 429–446
21. Mochizuki, A.: Pattern Formation Of the Cone Mosaic In the Zebrafish Retina: A Cell Rearrangement Model. *J. Theor. Biol.* **215** (2002) 345–361
22. Skeath, J.B.: At the Nexus Between Pattern Formation And Cell-Type Specification: the Generation Of Individual Neuroblast Fates In the *Drosophila* Embryonic Central Nervous System. *Bioessays.* **212** (1999) 922–931
23. Jessell, T.M.: Neuronal Specification In the Spinal Cord: Inductive Signals And Transcriptional Codes. *Nat. Rev. Genet.* **1** (2000) 20–29
24. Cook, J.E. and Chalupa, L.M.: Retinal Mosaics: New Insights Into An Old Concept. *TINS* **23** (2000) 26–35
25. Atabef, M., Clarke, J.D.W., Ticle, C.: Dorso-Ventral Ectodermal Compartments and the Origin of Apical Ectodermal Ridge in Developing Chick Limb. *Development* **124** (1997) 4547–4556
26. Gossler, A., de Angelis, M.H.: Somitogenesis. *Curr. Top. Dev. Biol.* **38** (1998) 225–287
27. <http://www.cs.usu.edu/~flann/index.html>