Multicellular Development, Self-Organization, and Differentiation

Sanjeev Kumar

Krasnow Institute for Advanced Study
George Mason University
Fairfax, VA, 22030, USA
kumars@cs.gmu.edu

Abstract. This paper describes experiments performed using a multicellular test-bed, the Evolutionary Developmental System (EDS). The EDS is an object-oriented model of biological development designed for the study of 3D multicellular development and differentiation. Genetic regulatory networks are used to specify and control the construction of varied multicellular morphologies. The construction of these morphologies, from a single cell, using the symmetry-breaking mechanisms of cell signaling and asymmetric division is investigated. In addition, analysis reveals that multicellular differentiation emerges during the process of development.

1 Introduction

Artificial life and developmental biology overlap on some quite important topics. One obvious topic is that of construction. Constructing robust complex adaptive systems in a self-organizing manner is a notoriously difficult problem that highlights fundamental issues of scalability, modularity, self-organisation, and self-repair. But Nature has solved these problems by evolving development—the process or set of processes responsible for constructing organisms [8]. Exactly how does development construct? How does the genome code for and control the transformation of a single cell into a complex multicellular system with well defined structures and form? How is symmetry broken in an otherwise homogenous system leading to heterogeneity [7] and specialised cell identities and structures?

Inspired by biological development, computational development is seen as a potential solution to such problems. This paper reports on a small subset of experimental results summarised from a doctoral thesis. The work addresses the problem of understanding the self-organising mechanisms and principles of development. The application chosen was that of constructing primitive 3D, geometric shapes, which have proved useful in areas such as computer graphics. The long term view, however, is that self-organisation through developmental principles may help us to construct complex systems that exhibit advantageous properties such as self-construction and self-healing. Section 2 presents a short overview of the biologically plausible model of development. Section 3 details two experiments

performed to study the construction of two morphologies using symmetry-breaking mechanisms; it also details the analysis of the construction process revealing that multicellular differentiation emerges.

2 The Evolutionary Developmental System (EDS)

The Evolutionary Developmental System is an object oriented computer model of many of the natural processes of development [5]. At the heart of the EDS lies the developmental core. This implements concepts such as embryos, cells, cell cytoplasm, cell wall, proteins, receptors, transcription factors (TFs), genes, and cis-regulatory regions. Genes and proteins form the atomic elements of the system. A cell stores proteins within its cytoplasm and its genome (which comprises rules that collectively define the developmental program) in the nucleus. The overall embryo is the entire collection of cells (and proteins emitted by them) in some final conformation attained after a period of development. A genetic algorithm is wrapped around the developmental core. This provides the system with the ability to evolve genomes for the developmental machinery to execute. The following sections describe the main components of the developmental model: proteins, genes and cells.

Proteins

In the EDS, the concept of a protein is captured as an object. In total there are forty proteins (see [5] for more details), each protein having five member variables:

- an ID tag (simply an integer number denoting one of forty six predefined proteins the EDS uses to control cellular behaviour)
- source concentration (storing the concentration of the protein)
- two sets of co-ordinates (isospatial [3] see fig. 1, and Cartesian)
- a bound variable (storing whether or not a receptor has bound a protein).

(The latter is only used in receptor proteins.)

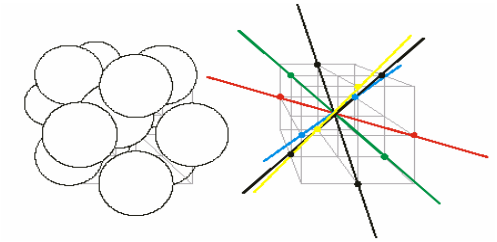


Fig. 1. Isopatial coordinates permit twelve equidistant neighbours for each cell.

A protein's source concentration variable is responsible for storing the protein's current concentration. In order to calculate concentration levels for a protein at a distance (during cell signaling, for example) or the creation of a new protein, the appropriate diffusion, production and decay rates are required. Proteins are able to diffuse within an embryo through an implementation that uses a Gaussian function centred on the protein source [5]. All coefficients are evolved and in order to access them the protein's ID tag serves as an index into the genome (which acts as a lookup table).

Genes

In nature, genes can be viewed as comprising two main regions: the cis-regulatory region [2] and the coding region [1, 6]. Cis-regulatory regions are located just before (upstream of) their associated coding regions and effectively serve as switches that integrate signals received (in the form of proteins) from both the extracellular environment and the cytoplasm. Coding regions specify a protein to be transcribed upon successful occupation of the cis-regulatory region by assembling transcription machinery.

The EDS uses a novel genetic representation termed the cis-trans architecture (fig. 2), based on new empirical genetics data emerging from experimental biology labs [2].

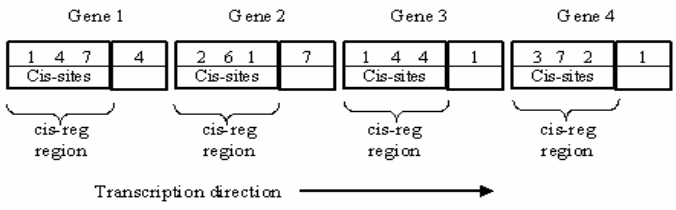


Fig. 2. An arbitrary genome created by hand. Genes consist of two objects: a cis-regulatory region and a coding region. Each number denotes a protein.

Equation	*Explanation*
$input_j = \sum_{i=1}^{d} conc_i * \text{int}eraction\_strength_i$ **(1)**	Where $input_j$ = total input of all TFs assembling upon the jth gene's cis regulatory region; *i*, = current TF; D = total number of TF proteins visible to the current gene; $conc_i$ = concentration of the ith TF at the centre of the current cell; $weight_{ij}$ = interaction strength between TF *i* and gene *j*.
$activity_j = \frac{input_j - THRESHOLD\_CONSTANT}{SHARPNESS\_CONSTANT}$ **(2)**	Where $activity_j$ = total activity of the jth gene; $input_j$ = total input to the jth gene; SHARPNESS_CONSTANT = a constant taken from the range 0.1-0.001 and is typically set to 0.01.
$activation\_probability_j = \frac{1+\tanh(activity_j)}{2}$ **(3)**	Gene activation probability. Where $activation\_probability_j$ = activation probability for the jth gene; $activity_j$ = total activity of the jth gene.

Table 1. Equations used to calculate the activity of a single gene by summing the weighted product of all transcription factors regulating a single structural gene.

The first portion of the genome contains protein specific values (e.g., protein production, decay, diffusion rates). These are encoded as floating-point numbers. The

remaining portion of the genome describes the architecture of the genome to be used for development; it describes which proteins are to play a part in the regulation of different genes. It is this latter portion of the genome that is employed by each cell for development.

Currently, the EDS's underlying genetic model assumes a "one gene, one protein" simplification rule [1, 6] to aid in the analysis of resulting genetic regulatory networks. The genome is represented as an array of Gene objects (fig. 2). Each gene objects contains two members: a cis-regulatory region and a protein coding region. The cis-regulatory region contains an array of TF target sites; these sites bind TFs in order to regulate the activity of the gene. The gene then integrates these TFs and either switches the gene 'on' or 'off'. Integration is performed by summing the products of the concentration and interaction strength (weight) of each TF to find the total activity of all TFs occupying a single gene's cis-regulatory region, see eqn. 1 table 1. This sum provides the input to eqn. 3, yielding a probability between 0 and 1 of the gene firing [4].

Cells

Cell objects in the EDS have two state objects: current and new. During development, the system examines the current state of each cell, depositing the results of the protein interactions on the cell's genome in that time step into the new state of the cell. After each developmental cycle, the current and new state of each cell is swapped ready for the next cycle.

The EDS supports a range of different cell behaviours, triggered by the expression of certain genes. The behaviours used for the experiments described in this work are:

- division (when an existing cell "divides", a new cell object is created and placed in a neighbouring position)
- the creation and destruction of cell surface receptors
- and apoptosis (programmed cell death).

The EDS uses an n-ary tree data structure to store the cells of the embryo, the root of which is the zygote (initial cell). As development proceeds, cell multiplication occurs. The resulting cells are stored as child nodes of parents nodes in the tree. Proteins are

stored within each cell. When a cell needs to examine its local environment to determine which signals it is receiving, it traverses the tree, checks the state of the proteins in each cell against its own and integrates the information.

The decision for a cell to divide in the EDS is governed by the ratio of division activator protein to repressor; the direction (or isospatial axis) the daughter cell is to be placed is non-random and is specified by the position of the mitotic spindle within the cell see [5] for more details.

3. Experiments: Multicellular Development and Differentiation

The application chosen was that of constructing primitive 3D, geometric shapes that have proved useful in areas such as computer graphics. This section details the development of two morphologies, a plane and a cube, using symmetry-breaking mechanisms to study multicellular development and differentiation.

3.1. Experiment 2: System Setup

The experiments used the parameter settings shown in table 2.

Evolution	
Parameter	Value
Runs	100
Population Size	70
Generations	100
Tournament Size	30
Number of Genes (N)	10
1-Point Crossover	100%
Gaussian mutation rate per gene	0.01
Gene Sharpness Constant	0.001

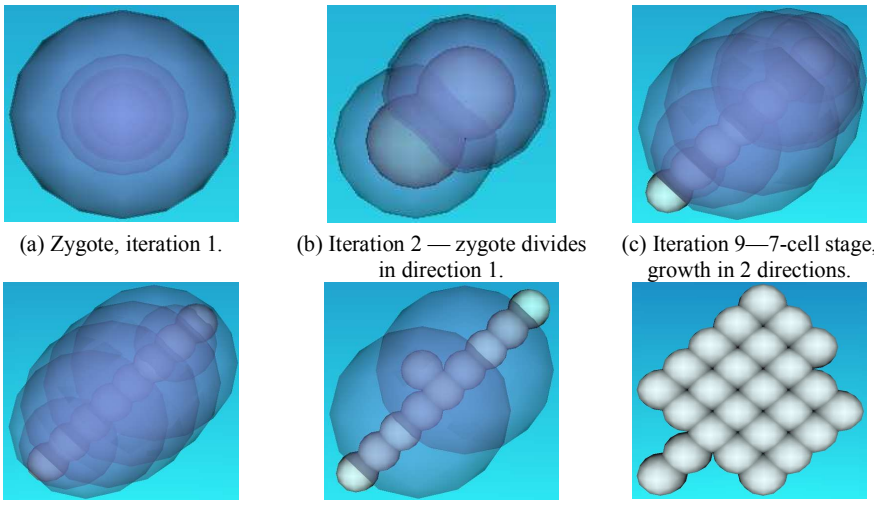
Development	
Parameter	Value
Developmental iterations	30
Cis-Sites (K)	2
Division Threshold	0.2
Apoptosis Threshold	0.2
Symmetric Division	Evolved
Asymmetric Division	Evolved
Cell Signaling	On
Cell Division Strategy	Evolved

Table 2. List of parameters and values.

Experiments were performed on a 1.47GHz AMD Athlon XP+ processor with 1Gbyte of 333MHz DRAM.

3.2 Results and Analysis

3.2.1 Plane



(a) Zygote, iteration 1.

(b) Iteration 2 — zygote divides in direction 1.

(c) Iteration 9—7-cell stage, growth in 2 directions.

(d) Iteration 10 — 8-cells, growth occurs in upper right corner, this cell is fated to die.

(e) Iteration 11 — 10-cell T-shape embryo, with division in directions 0 and 11.

(f) Final state of plane — with all proteins removed.

Fig. 3. The development of the best plane.

[27, 11 | 36] [15, 4 | 20] [15, 36 | 17] [7, 37 | 6] [11, 27 | 33] [7, 15 | 30]
[17, 36 | 2] [22, 5 | 22] [8, 13 | 10] [28, 7 | 0]

Fig. 4. Evolved genome for the best plane.

The first morphology was that of a plane (see [5] for fitness function, which was closeness to the target shape, defined by its mathematical equation). The evolved genome is shown in fig. 4. Figure 5 provides an illustration of how the plane developed over time through a series of snap-shots taken at intervals during development. The last image shows the final plane with all proteins removed for clarity. Gene 10 evolves to emit protein 0, and it is protein 0 that is used as a local-acting signal during cell signaling (snapshot a shows the build up of protein 0, depicted as the sphere with the largest diameter). Development continues with the zygote dividing in direction 0 (snapshot b) in the 2nd iteration. Development proceeds in this manner until iteration 9.

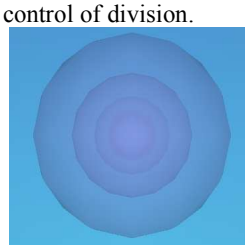
Snapshot (c) shows a crucial change in development from the 5-cell stage (not shown)—two divisions occur in iteration 9. These divisions are due to the zygote and the fifth cell. The zygote manages to increase its level of protein 6 above a threshold and change division direction to 1, while the fifth cell (shown in the lower left corner of snapshot d) also divides but in direction 0, giving rise to this seven cell embryo. The new cell in the lower left corner is fated to die.

Snapshot (d) shows the state at iteration 10 in which the new cell in the lower left corner of the image ceases dividing; instead, the upper right portion of the image shows division has occurred again in direction 1. The cell in the lower left and upper right corners of snapshot (d) are ephemeral additions and are ultimately fated to commit suicide (or apoptosis). Snapshot (e) shows the zygote's first daughter cell has managed to divide in direction 11 resulting in a T-shape. Thereafter, cells eventually start to divide in direction 11 and others in direction 10, thus giving rise to the main body of the plane. Genetic regulatory networks are thus able to specify developmental programs that successfully code for, and control, the construction of a restricted morphology, such as a plane, from a single cell.

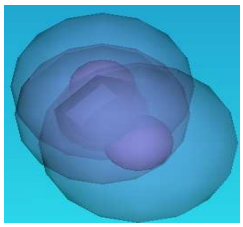
3.2.2 Cube

Having examined the restricted morphology of a plane, this section studies a larger, symmetrical morphology: the cube. The second morphology was a cube, (see [5] for

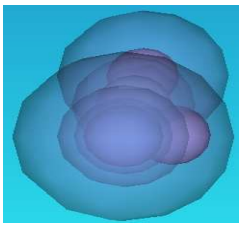
fitness function). Evolution has evolved gene 7 (which emits protein 4) for directional control of division.



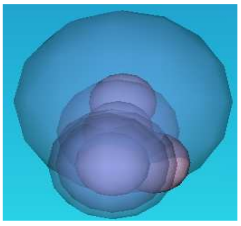
(a) Zygote



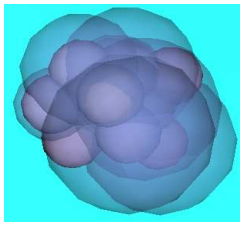
(b) Iteration 4 — 3-cell stage.



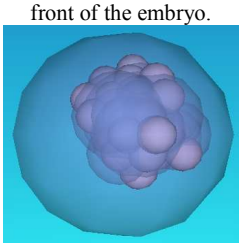
(c) Iteration 5 — 4-cell stage. New cell placed in the lower front of the embryo.



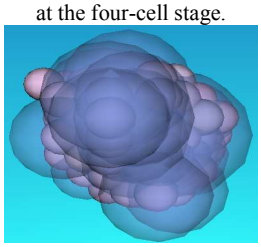
(d) Iteration 10 — 5 iterations later and the embryo remains at the four-cell stage.



(e) Iteration 15 — 11-cell stage with many long-range proteins removed for clarity.

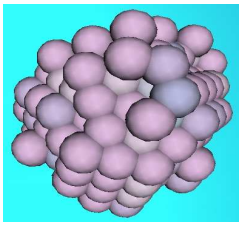


(f) Iteration 20 — the core of the cube begins to take form.



(g) Iteration 25 — cube structure almost established.

(h) Final state of cube with all long-range proteins present.



(i) Final state of cube with proteins removed.

Fig. 5. The development of the best cube.

[32, 3 | 28] [0, 35 | 20] [4, 30 | 28] [26, 28 | 0] [26, 36 | 30] [33, 32 | 31]
[25, 8 | 4] [27, 23 | 21] [22, 1 | 27] [37, 8 | 14]

Fig. 6. Evolved genome for the best cube.

The gene expression plots (see fig. 7) reveal important differential gene expression patterns between the two cells. Noticeably, genes 3 and 9 are expressed, albeit sparingly, in the zygote, but not at all in the daughter cell. Other important differences in gene expression between the two cells are the expression of genes 7 and 8, which are both increasingly activated, in the zygote, over time, but are seldom activated in the daughter cell.

In the zygote, two proteins control (more or less) the activation of gene 3: proteins 0 and 4, conferring inhibitory and excitatory stimuli, respectively. In the daughter cell, levels of both proteins 0 and 4 are low due to division, and so do not provide sufficient inhibition or activation. Instead, it falls, to proteins 24 and 37 to provide inhibition, and to cell signaling. Signaling initially delivers large inhibitory stimuli through the receptor 13-proteins 4 and 31 signal transduction pathways from the first division in iteration 3, see [5] for the full analysis and for details on receptors.

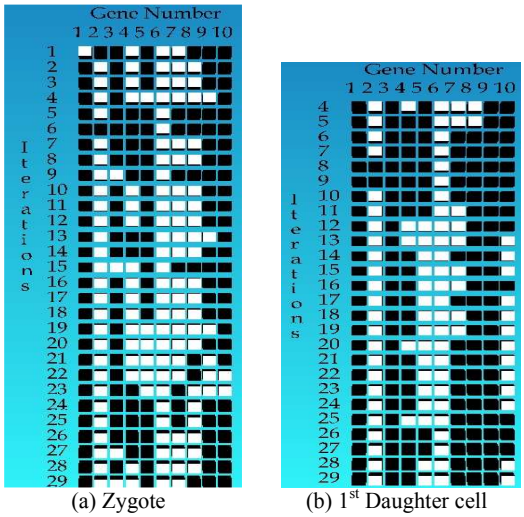
Over time, as receptor 13 decays, so too does the inhibitory stimulus received through that pathway. However, receptor 14 is synthesised more frequently in the daughter cell due to the expression of gene 10. It must be noted that both cells by virtue of expressing a different subset of genes also have a different subset of active receptors. The zygote begins development with an assortment of receptors, while the daughter cell (and later progeny) inherit their state including receptors from their parent, and then begin to express different genes and consequently different receptors. For example, at iteration 21 the zygote has receptors 10 and 9 while the daughter cell has receptors 14, 10 and 9.

The zygote manages to activate gene 9, occasionally, due to proteins 0 and 4, which confer activating stimuli. After cell division, cell signaling also contributes to the occasional activation of gene 9 in the zygote. When the zygote divides, protein 31 is symmetrically distributed to the daughter cell in iteration 4, after which the daughter cell continues to activate gene 6 resulting in more protein 31 synthesis. Protein 31 is a single large inhibitory influence on the expression of gene 9 in the daughter cell.

The symmetric and asymmetric division of two important activator proteins, 0 and 4, result in the low distributed concentration of these proteins in the daughter cell. The

low concentration of these activating proteins means the large inhibition, due to protein 31, becomes insurmountable ensuring the permanent deactivation of gene 9 in the daughter cell.

Consequently, differential gene expression emerges as a result of the symmetry-breaking mechanisms of cell signaling and asymmetric cell division (which is used sparingly).



(a) Zygote (b) 1st Daughter cell

Fig. 7. Gene activation plot for the zygote (a) and the 1st daughter cell (b) of the best cube. Where white and black cells represent activated and inactivated cells, respectively.

4. Conclusions

This paper has described experiments performed using a multicellular test-bed, the Evolutionary Developmental System (EDS), an object-oriented model of biological development. Development with symmetry-breaking mechanisms was used to

successfully construct 3D multicellular morphologies from a single cell. During the development of these morphologies phenomenena resembling biological differentiation emerged with different cells exhibiting differential gene expression. An analysis of how the morphologies were constructed, including the translation of genetic states into spatial instructions, and how differentiation emerged was presented.

Acknowledgments

Many thanks to Lewis Wolpert, Michel Kerszberg, and Jeff Bassett for helpful advice and criticism. This research was funded by Science Applications International Corporation (SAIC).

References

[1] Alberts, B., Dennis, B., et al. (1994). *Molecular Biology of the Cell*. Garland Publishing
[2] Davidson, E.H. (2001). *Genomic Regulatory Systems: Development and Evolution.* Academic Press.
[3] Frazer, J. (1995) *An Evolutionary Architecture*. Architectural Assoc. London.
[4] Kerszberg, M. and Changeux, J-P. (1998). A Simple Molecular Model of Neurulation. In BioEssays 20: 758-770.
[5] Kumar, S. (2004). *Investigating Computational Models of Development for the Construction of Shape and Form*. Ph.D. Dissertation, Department of Computer Science, University College London, UK.
[6] Lewin, B. (1999). *Genes VI*. Oxford University Press, Oxford, UK.
[7] Turing, A. (1952). *The chemical basis of morphogenesis*. Phil. Trans. R. Soc. London B, 237:37-72.
[8] Wolpert, L. (2001). The Principles of Development. Oxford University Press, Oxford, UK.