Development of Virtual Embryos with Emergent Self-Repair

Tim Andersen  Richard Newman  Tim Otter
Crowley Davis Research, Inc.
280 South Academy Ave., Suite 140
Eagle, Idaho, US 83616
(tim, richard, timmo)@cdres.com

Abstract
We have constructed a computational platform that incorporates principles of biological development and primitives derived from cell biology into a rule-based architecture. When coupled to an evolutionary search, this platform can evolve 3-dimensional embryos with robustness to damage and high capacity for self-repair, even though repair is not explicitly encoded or part of the fitness criteria. One such embryo, a 117-cell cube, achieves this robustness through cell signaling and gene regulatory networks (SGRNs) that curtail cell division when the embryo is intact and support replacement of lost cells when damage shifts the balance between growth and repression. These results demonstrate that relatively simple SGRN architectures can generate complex and robust emergent properties.

Introduction
In the past several years, developmental computation has proven to be a versatile, powerful approach for studying the relationship between genetically encoded components and the properties of the systems in which they operate (Kumar and Bentley 2003, Stanley and Miikkulainen 2003). To examine emergence of higher order processes in simple cellular systems we have constructed a computational platform where genes, gene expression, and gene products are resources controlled by virtual cells, which carry out the core processes of development: cell division (duplication), signaling (communication), and metabolism. While this shift to a cellular rather than gene-based architecture may seem trivial, we have evolved various 3-dimensional clusters of cells, “virtual embryos” that display robustness to injury and high capacity for self-repair, even though repair is not explicitly encoded or included in the fitness criteria driving the evolutionary process. This paper analyzes the gene regulatory networks and signaling pathways of one of these self-repairing embryos and provides insight regarding mechanisms that control its organization, stability, and robustness to injury.

Fundamentally, development (or ontogeny, “coming into being”) results from processes carried out by cells – metabolism, signaling, division, growth, differentiation, etc.– as they become coordinated components of a multicellular body (Gerhart and Kirschner 1997). Developing organisms exhibit almost uncanny regularity and reliability, behaving as if a program with computer-like precision is guiding ontogeny, and so the term “genetic program” has become standard parlance among developmental biologists (Keller 2002). Taken literally, though, the “genetic program” metaphor is misleading, because it implies that there actually is a prescribed sequence of operations by or upon genes to control development. But genes specify no more than the potential composition of an organism, and genetic code contains no explicit information regarding metabolism, homeostasis, adaptability, or other higher-order properties (Brenner 1999, Nijhout 1990). For instance, early-stage vertebrate embryos have a remarkable ability to detect and correct for loss of one or more cells produced by early divisions, a capacity known as regulative development or regulation (Carlson 1996), and it is unreasonable to suppose that such compensatory powers are specified in the genetic code. Instead, regulation is an emergent property of cellular systems built from genetically encoded elements.

Reduced to its bare essentials, a living cell has four basic capabilities: gene expression, metabolism, cell division, and growth. During development, these four capabilities enable cells to replicate, communicate with one another, and differentiate. The central question for the present studies is whether faithful implementation of fundamental cellular processes can produce virtual organisms with realistic behaviors. In addition, we believe that this approach can provide insight regarding the underlying mechanisms of higher order properties such as self-repair, and it may prove advantageous for modeling living tissues.

Computational platform
We have chosen to focus on a cellular level of granularity. The computational platform\(^1\) includes a developmental engine, with a rule-based architecture incorporating cellular processes such as division, cell death, signaling, and gene expression. Each virtual cell is a compartment

\(^{1}\) CellSim\textsuperscript{TM} is proprietary software developed by Crowley Davis Research, Inc.
with a genome and internal molecules that mediate exchange and communication with its surroundings, as shown in Figure 1. Signaling and control of gene expression are treated computationally as molecular processes, where a single virtual molecule may subsume the functions carried out by several molecules in a living cell. Virtual cells may be killed by injury or by apoptosis, but cell senescence is not represented in the computational platform.

![Diagram of a virtual cell](image)

*Figure 1: Organization of a virtual cell.*

Phenotype is modeled as the result of a developmental process, starting from a single cell and its genome. Properties such as embryo shape and self-repair arise as the multicellular virtual embryo develops. An evolutionary process selects virtual embryos with a desired shape.

**Incorporation of Biological Features**

Two primary actions, division and death, are available to each virtual cell. These actions are supported by a rudimentary metabolism that includes stimulation by growth factors, sending signals, receiving signals, and control of gene expression. Metabolic components can be linked to form gene regulatory networks within cells, or signaling networks between cells, or combinations thereof (SGRNs, Figure 5). Network architecture is not determined a priori or part of the fitness criteria, but instead it arises during evolution of virtual embryos and their genomes. Linkage between network components depends on the extent of match or overlap between numerical signatures of components. If two signatures are identical, they match precisely and interaction of the components is maximal, while components with vastly different signatures interact weakly, if at all (see Virtual Molecules, below, for details).

**Configuration of Virtual Cell and Environment**

The parameters necessary for development are configurable. These include placement and concentration of growth factors or other molecules, defining space for cells to divide, sequencing of actions, neighborhood definition, and rules that govern calculation of molecular affinity.

After configuration, development is initiated by placing a single cell into the environment. The cell’s genome then interacts with molecules in the environment and molecules that the cell produces. Depending upon these interactions, each gene may be promoted or inhibited. When a gene is promoted, the transcription apparatus produces the molecules defined by the gene’s structural region. These newly produced molecules may in turn interact with the cell’s genome, affecting rates of transcription at the next step. Development is thus governed by inputs from the environment, neighboring cells, and by internal feedback mechanisms.

The main development process uses two inner loops that are executed with each iteration. The first loop executes cell signaling. The second loop determines the factors that affect transcription for each cell, performs transcription, and then executes any resulting actions, division or death.

**Division, Growth, and Death**

A cell may have genes that encode division or death molecules, and their intracellular concentration increases as the corresponding genes are transcribed. Division or death is then a function of the concentration of these two types of molecules. When a cell dies, it is removed immediately. Alternately, if a cell divides, a new equal-sized cell with identical properties is placed in an unoccupied adjacent location in the 3-dimensional cell grid. The initial cell has 26 placement options, while subsequent divisions are usually more restricted by neighboring cells. Growth is represented as a simple two-state function that is tied to division: prior to division a single cell occupies one unit of size, and upon division, size doubles as another identical cell appears next to the first cell. Although a cell may be increasing its potential to divide, intermediate stages of growth are not represented, nor does the model accurately represent the progressive decrease in cell size typical of early embryonic development in vivo. Division resets the division potential to zero, similar to the resetting of the cell cycle clock. A cell that is ready to divide may not do so if all 26 adjacent positions are already occupied, similar to contact inhibition seen in noncancerous living cells.

**Virtual molecules**

The numerical signature of a virtual molecule includes three primary components: indicant, sensitivity, and type. In addition, molecules may be associated with other parameters. For example, sources of molecules in the environment have a defined location and an exponent that determines the shape of the gradient. Molecules in the environment may enter the cell and influence its operations. Transcription of genes within a cell can create molecules, which in turn may influence its own operations.
or those of neighboring cells. There are 5 types of molecules:
- Cell division factor – determines when a cell divides
- Death factor – determines cell death
- Transcription factor (TF) – interacts with control elements
- Receiver molecule – accepts signals from neighboring cells
- Sender molecule – broadcasts signals to neighboring cells.

Although division and death are complex processes that involve many molecules in vivo, each is represented by a single molecule type in the computational model. Virtual receivers, senders, and TFs correspond to receptors and their associated signal transduction molecules, hormones or other secreted molecules, and transcription factors. The concentration of molecules is decayed at each step by a set percentage, configured to 10.0% for the experiments reported in this paper.

Indicant and sensitivity values are single-precision floating point numbers that determine the strength of interaction between a transcription factor and a control element, or between one cell’s sender molecule and a neighboring cell’s receiver molecule, as defined by equation (1).

\[
affinity (m, n) = e^{-m \text{indicant} - n \text{indicant} \sqrt{m \text{sensitivity} + n \text{sensitivity}}} \]  

where \( m, n \) are molecules

**Transcription Apparatus**

The genome of an organism is organized into a set of “gene assemblies” (Figures 1 and 2). Each gene assembly has a regulatory region that controls the activity of the structural genes in the gene assembly, and a structural region that defines the molecules produced by transcription.

Each control element in a gene assembly has three characteristics: indicant, sensitivity, and effect. The promotion of a gene assembly is based on the total affinity of all transcription factors for each control element, from equation (2), and the amount this contributes to promotion of the gene assembly calculated from equation (3).

\[
totalAffinity (c) = \sum_{m \in TFs} \text{concentration}(m) \times affinity (m, c) 
\]

where \( c \) is a control element, and \( m \) is a TF \( (2) \)

\[
promotion (c) = c \cdot \text{effect} \times (1 / (1 + e^{\text{totalAffinity}(c) - p})) \]  

where \( p \) is a constant that shifts the promotion threshold.

The promotion value for each control element in the gene assembly is then summed to determine the total activity level for the gene assembly, and this value directly determines the concentration of gene products defined by each of the structural genes in the gene assembly. Thus, at each step and for each structural gene in the assembly, the concentration of its product molecule is increased by an amount given by equation (4).

\[
totalPromotion (g) = \sum_{c \in A} promotion (c) \]  

where \( g \) is a structural gene, \( c \) is a control element, and \( A \) is the gene assembly such that \( g \in A \)

**Cell Signaling**

A cell may also receive information from neighboring cells. The simplest neighborhood consists of cells adjacent to the reference cell. In these studies, neighborhood was specified to be any other cells within a defined radius. All signals are broadcast: a cell cannot send a signal in a particular direction, nor can a cell determine from which cell or which direction a signal originates.

For a signal to be transmitted the sender and the receiver must each create specific molecules. To send a signal, a cell must create molecules of type ‘signal’ from an appropriate gene. At each step, each cell identifies the cells in its neighborhood and presents those signals to its neighbors. For a cell to receive a signal, it must have produced receiver molecules tuned to that signal. The interaction of sender and receiver is calculated using equation (1), and then the total signal strength for each receiver is determined by equation (5).

\[\text{Figure 2: Gene assembly organization. See text for details.}\]
Here is the text from the document:

\[
total\text{Affinity}(r) = \sum_{m \in \text{sender}} \text{concentration}(m) \times \text{affinity}(m, r)
\]

where \( r \) is a receiver molecule \( (5) \)

When a receiver senses a signal, it generates an internal signal defined by that receiver molecule in the cell’s genome, not by the particular signal detected. Accordingly, different cells can respond in different ways to the same external stimulus.

The strength of the internal response is determined using equation (3), and for the experiments reported in this paper, the effect was set to 1. Internal signals do not decay at the same rate (10% per step) as for other molecule types; instead, they decay completely in the same step. A cell’s response thus depends directly on the interaction of sender and receiver molecules, and so signaling is sensitive to changes in its neighbors.

**Evolutionary strategy**

An evolutionary engine automates the search for embryos with the desired target shape. An individual is defined by both its genome and the conditions in which it develops; genetic operators may modify the genome, or the context for development. This design is akin to in vitro culture of embryos, where the culture conditions or the genetic makeup may be adjusted.

A seed population of cells, with varying genotypes, develops in each generation to yield a population of multicellular embryos with different properties. Three basic steps are required to process an individual in the population. First, the configuration specified by the individual is instantiated and a single cell with the defined genome is placed in that environment. The computational engine then runs until a stable shape or a limited number of steps (1000 in these studies) is reached. Stability is determined by comparing the number of contiguous steps that have occurred without cell division or death. If this number surpasses a user-defined threshold then the embryo is assumed to have stabilized. Finally, an individual’s fitness is determined by a shape matching metric based on the shape distributions method (Osada et al. 2002), with the fittest individuals selected to seed the population of the next generation. If stability is not reached, the fitness of the individual at the step limit is assessed and penalized by a small constant factor to favor those individuals that achieved stability.

For the experiments reported in this paper, mutation operators were applied to modify location, strength, and affinity of environmental gradients, and the 'effect' parameter of gene control elements. After each generation, individuals were selected for mutation with a 10.0% probability, and the set of mutations operators was applied to them. The evolutionary algorithm did not utilize any operators for crossover, duplication, deletion, or creation of genes. In biological terms, this is equivalent to asexual reproduction without major genome rearrangements. The initial population was derived from an individual randomized by 10 successive rounds of mutations applied to gradients and gene regulatory parameters as described above, except that mutation was certain and no selection was performed between rounds of initializing mutation.

The starting genome was minimal, only the basic set of genes encoding molecules essential for survival: a rudimentary sensory system with one receiver and one sender, one “regulatory” gene encoding a TF that could control the activity of other genes or its own, and one gene for cell division. There were two sources of growth factor in the environment. After several hundred generations these sources had been repositioned so they were nearly superimposed, with slightly altered gradient strengths and shapes. The fittest individual’s regulatory gene had mutated to a nonfunctional form, and the strengths of promotion or inhibition of the three remaining genes had been changed substantially by mutation.

**Results**

Most individuals of the starting population (64 individuals with average fitness = 0.4, ~ 4-6 cells/embryo) develop little, if at all. After many generations, more complex and complete patterns of development emerge. At generation 574 the fittest individual develops from a single cell to yield a stable phenotype with fitness 0.938 (1.0 = perfect cube), missing only the 8 corner cells of the target 125-cell cube (Figure 3A). Development of this individual is complete in 325 steps, but even when the development engine continues to run for 2500 additional steps, the embryo does not change shape, and there are no visible fluctuations in internal state (Figure 3B).

Stability of embryo shape can also be demonstrated by damaging the embryo during development (not shown), or after development has reached the stable 117-cell form (Figure 3C). In both cases, this embryo has a remarkable capacity for self-repair and produces nearly the same shape regardless of the extent or timing of damage. These results indicate that development is robust to damage: it can arrive at the same end result from different starting points, and the paths of repair from various damaged states converge on the same stable form, a cube missing its corners. The embryo exhibits a type of “shape homoeostasis”.

A more careful analysis of the response to minor damage reveals some small deviations from this pattern of perfect, convergent repair. Although the surface cells of the 117-cell cube appear to be in the same state of repressed growth (light color, left image, Figure 4), in fact these cells occupy slightly different local environments that differ in their capacity for repair, and in the pattern of repair. For example, ablation of the center cell of a cube face does not repair (C, Fig 4), while damage at the lower edge (A, A’) repairs perfectly, and ablations in several other positions (B,D,E) repair and generate extra cells on an adjacent...
surface. If the extra cells are then killed, they do not regrow. In aggregate, these results show that the cube is one of several metastable shapes, and that more than one phenotype may result from a single genotype, though the variations in form are minor.

Development and repair rely on signaling to curtail growth (Figures 4 and 5). The cube develops in an environment that is nearly equipotential for growth: the gradient of growth factor across the entire developmental field varies by less than 18%, and at the edge of the cube embryo the concentration of growth factor is ~88% of the peak value near the center of the cube. That is, all 117 cells occupy positions that have enough growth factor available to divide, but division of any cell is inhibited by signals from its neighbors. The effect of removing a cell is to disinhibit its former neighbors (Figure 4A, 4B) which then accumulate growth factor and divide into an adjacent unoccupied position. In Figure 4A, the cell marked by a white asterisk (*) repaired the damage. The other two nearby cells accumulated growth factor but did not divide and subsequently became repressed. Thus, the stable shape of this embryo represents a balance between growth and repression by neighbors. When the receiver gene (Gene 2) was deleted, the embryo grew unabated, filling the entire development space (12 x 12 x 12 = 1728 cells) in 35 steps, underscoring the importance of signaling and the potential of the environment to support growth.

More comprehensive mapping of the repair capabilities was carried out in two series of experiments summarized in Figure 6. In the first series, totipotency (the ability of one cell to form a complete embryo) was assessed by deleting all possible combinations of 116 cells, leaving one cell remaining. Following repair, the resulting embryo's shape and cell count were measured. In the second series, each of the 117 cells was deleted. Again, following repair, each embryo's shape and cell count were measured. In all cases,
Figure 5: The signaling and gene regulatory network for the 117-cell cube embryo. Solid arrows, promotion; dotted lines, inhibition; dashed lines, production or interaction (sender/receiver). Numbers indicate strength of promotion (+values) or inhibition (-values). Two sources of growth factor promote genes 0, 1 and 2. Gene 1 produces sender, gene 2 produces receiver and interaction of the sender/receiver pair ultimately inhibits gene 0, which stops cell division. The network is drawn with an affinity threshold of 0.8; weaker interactions are not represented.

repair was perfect or deviated by only a few missing or extra cells (range, 111-123 cells). Perfect repair occurred in 53 of 117 cases (45%) with single cell deletions (Figure 6B) and 83 of the 117 embryo cells (71%) were totipotent.

In parallel experiments we have evolved embryos with other stable shapes, including solid or hollow balls of cells, hollow cubes, and a rectangular solid “brick”. Embryos become hollow by apoptosis of the inner cells of an initially solid mass, similar to formation of primitive ducts in vivo. Some of these evolved hollow forms have the desired shape but they cannot repair damage because the remaining outer cells lack the ability to divide. The solid brick repairs minor or major damage nearly perfectly (100% of minor damage/ 98% of cells totipotent), while other shapes, such as a hollow ball, exhibit moderate capabilities for self-repair (30% of minor, 16% of major damage repaired perfectly). These results indicate that different genomes and networks may produce embryos with similar shapes, but only some support self-repair.

Discussion

Our goal in these studies was to devise a computational approach that faithfully incorporates principles derived from cell and developmental biology. If successful, this system should be able to grow multicellular virtual organisms with higher order properties such as homeostasis, which could be measured by shape stability.

Figure 6: Repair of individuals sustaining major damage (A) or minor injury (B). Each panel depicts the results of “embryoblast” experiments involving loss of 116 cells (A) or loss of 1 cell (B). See text for details.
and self-repair. To be consistent with early development of vertebrate embryos, implementation of cell division, signaling and gene expression would require close mapping from in vivo to in silico to capture the character of each process and its controls. Thus, cell division was implemented as simple division as one cell into two daughters with the same state as the original cell, subject to inhibition by contact or signaling from neighboring cells. Signals are sent by non-directional broadcast, and signal reception depends on a cell having produced properly tuned receiver molecules. In addition, internal signals are decoupled from the external signal so different cells can respond in different ways to the same broadcast. Our construct for gene expression permits cross-talk among TFs and control elements: a TF interacts with more than one control element and each control element is regulated by several TFs. Furthermore, expression is proportional to the strength of promotion over a range of values (equations 2-4), not simply “on” or “off”, and gene products may accumulate from step to step. These constructs are elements of a deliberate strategy supported by Harris’ proposal that body form arises from a delicate balance of the properties of cells and their interactions rather than from imposed constraints (Harris 1987).

Such features may bias our virtual embryos toward shape homeostasis, including stability of shape and repair. As far as we know, ours is the first report of an evolved 3-dimensional, self-repairing virtual embryo that also attempts to quantify repair and analyze its mechanisms. Analysis is enhanced by composite SGRN diagrams (Figure 5) that summarize interactions between cells and control of gene expression within cells throughout the embryo’s life, during ontogeny and in response to injury. Our results show that the embryo's stability emerges from interactions between neighboring cells, and not from trivial mechanisms such as timer-limited growth or careful placement of chemical gradients in the environment. Furthermore, the same processes that build the embryo also can maintain and repair it.

Finally, our evolutionary strategy—with its “stable steps” criterion, fitness penalty for failure to stabilize, and variation of the environment from one generation to the next—may favor individuals that can buffer changes in growth conditions using mechanisms that make them stable and robust to damage. These features are quite reasonable, as any organism that exhibits unrestrained growth is unlikely to survive, much less be competitive against species with better-regulated development. If we view our experiments as virtual tissue culture, variation of the environment is tantamount to changing culture conditions to find suitable ones. Furthermore, while some species are adapted to nearly constant environments, others must cope with conditions that fluctuate daily, seasonally, or erratically. Modifying the virtual environment from generation to generation introduces a modest amount of unpredictability, well within the range of plausible variation. Biological fitness is a measure of the match between an organism's phenotype and its environment, living and non-living. The goal is to evolve genotypes that support development of high fitness phenotypes in a range of environments; successful schemes of development can buffer environmental variation, at least to some degree.

Developmental computation has been promoted as a powerful approach for a wide range of engineering problems (reviewed by Stanley and Miikkulainen 2003). The choice of cellular attributes and strategies for their implementation depend on the problem. For instance, GRNs have been used to control development (Bongard 2002, Federici 2005) or to mediate behavioral responses to stimuli (Quick et al. 2003, Taylor 2004), but not both. Other studies have focused on developmental processes per se rather than applying principles of cell biology to solve a particular engineering problem (e.g., Andersen et al. 2005, Eggengerger 1997, 2003, Miller 2004, and the present study). Many of these approaches incorporate features derived from living cells and so resemble another superficially, but closer inspection reveals subtle but important differences that are reflected in the relationships among encoded elements, cellular primitives, and higher order properties such as shape, metabolism, and self-repair.

Multicellular bodies (phenotypes) developed by cell division and death have been created using 2-dimensional, non-toroidal cellular automata using control mechanisms based on reaction-diffusion systems (Furusawa and Kaneko 1998), Boolean circuits (Miller 2004), or neural nets (Federici 2005). Many of these virtual organisms are clearly capable of self-repair, but because these studies do not quantify the extent of repair or the range of conditions where repair occurs, nor analyze its underlying mechanisms, it is impossible to make definitive comparisons. However, it is fair to say perfect repair is rare, and that most of these organisms repair incompletely or become unstable upon damage.

Instability may stem from rules for cell division and growth that are inconsistent with natural systems: one cell can produce four progeny at once, displace any adjacent cells, and the state of differentiation of the daughter cells is specified by the original (mother) cell (Federici 2005), or one cell can produce eight progeny at once, overdwiting the state of any adjacent cell (Miller 2004); such rules and abrupt changes in state may bias these systems toward unrestrained growth. On the other hand, a more biologically faithful division of one cell into two nearly identical daughter cells coupled with diffusional limitations of nutrients leads to a balance between growth by cell division and death by starvation (Furusawa and Kaneko 1998), or coupled with contact inhibition but not death it produces self-repairing organisms of stable size (Streichert et al. 2003). While we are only just beginning to comprehend how local rules and information can generate global behaviors like self repair, this analysis suggests that faithful implementation of biological concepts does matter.
Shape stability and self-repair have been addressed directly by two groups of researchers using random Boolean networks (Streichert et al. 2003) or a dynamical system of differential equations (Fleischer and Barr 1993, Fleischer 1996) to develop 2-dimensional multicellular bodies. Fleischer’s system also can produce 3-dimensional bodies with stable shapes, but their capabilities for self-repair were not reported. Streichert et al. evolved artificial embryos with RBNs that support limited growth and self-repair. In the best cases when several cells are killed after an embryo stabilizes, it reestablishes its original number of cells, but probably not the original shape, because of stochastic variations in the topology of cell neighborhoods. Both of these approaches use a simple division of one virtual cell into two daughter cells that either inherit the same properties as their progenitor (Streichert et al. 2003), or simultaneously divide and differentiate (Fleischer 1996).

To extend our studies, we have begun to explore several possibilities: adding primitives such as cell adhesions, movement, shape, and senescence; using more complex fitness functions to guide the evolutionary search; enabling mutation operators such as gene duplication and deletion; implementing rules for cell placement that are more flexible than the cell grid presented in this paper; and using gene knockout experiments and targeted mutations to analyze network interactions of SGRNs. The expanded platform has produced more complex virtual tissues that exhibit self-repair. By examining their repair capabilities and analyzing their regulatory networks we hope to gain further insight regarding the mechanisms that support self-repair and homeostasis.

Acknowledgments

We thank the staff at Crowley Davis Research for their hard work, creativity, and thoughtful input. Support: U.S. Army Medical Research and Materiel Command Cooperative Agreement DAMD17-02-2-0049.

References


