The Growing Canvas of Biological Development:
Multiscale Pattern Generation on an Expanding Lattice of Gene Regulatory Nets

René Doursat

Department of Computer Science and Engineering
University of Nevada, Reno
http://www.cse.unr.edu/~doursat

The spontaneous generation of an entire organism from a single cell is the epitome of a self-organizing, decentralized complex system. How do nonspatial gene interactions extend in 3-D space? In this work I present a simple model that simulates some biological developmental principles using an expanding lattice of cells. Each cell contains a genetic regulatory network (GRN), modeled as a feedforward hierarchy of switches that can settle in various on/off expression states. Local morphogen gradients provide positional information in input, which is integrated by each GRN to produce differential expression of identity genes in output. Similarly to striping in the *Drosophila* embryo, the lattice becomes segmented into spatial domains of homogeneous genetic expression that resemble stained glass motifs. Meanwhile, it also expands by cell proliferation, creating new local gradients of positional information within former single-identity domains. Analogous to a "growing canvas" painting itself, the alternation of growth and patterning results in the creation of a form. This preliminary study attempts to reproduce pattern formation through a *multiscale, recursive and modular process*. It explores the elusive relationship between nonspatial GRN weights (genotype) and spatial patterns (phenotype). Abstracting from biology in the same spirit as neural networks or swarm optimization, I hope to be contributing to a novel engineering paradigm of system construction that could complement or replace omniscient architects with decentralized collectivities of agents.

1 Introduction

The spontaneous generation of an entire organism from a single cell is the epitome of a self-organizing, decentralized complex system. Through a precise spatiotemporal interplay of genetic switches and chemical signaling, a detailed architecture is created without explicit blueprint or external intervention. Recent dramatic advances in the genetics and evolution of biological development, or "evo-devo" [e.g., Carroll et al. 2001], have started the foundations of a future discipline of *generative* development. The goal is to unify organisms beyond their seemingly endless diversity of form and describe them as variations around a common theme. The variations are the specifics of the genetic code; the theme is the generic elementary laws by which this code controls its own expression, whether triggering cell division, differentiation, self-

assembly or death, towards form generation. On this stage, evolution is the player.

"How does the one-dimensional genetic code specify a three-dimensional animal?" [Edelman 1988] How does a static linear code dynamically unfold in time (metabolic dynamics) and space (cell assembling)? How do nonspatial gene interactions extend spatially? The missing genotype-phenotype link in biology's Modern Synthesis constitutes the main question in evo-devo. These issues also potentially open entirely new perspectives in engineering disciplines, including software, robotic, electrical, mechanical or even civil engineering. Could an architecture, device or building, construct itself? Would it be possible for a swarm of software agents or small robotic components, each containing a "genetic code", to self-assemble? Although themselves emergent, our cognitive faculties are strongly biased towards identifying central causes and require great effort to comprehend massively parallel processes. We spontaneously tend to ascribe the generation of order to one or a few highly informed agents, following anthropomorphic stereotypes (designer, architect, manager, etc.). Yet, heteronomous human-designed order is the most sophisticated of all forms of organization. In living systems, autonomous decentralized order is the natural norm because it is the most cost-effective. Information is distributed over a large number of relatively ignorant agents, making it easier to create new states of order by evolving and recombining their local interactions. To imitate Ulam's famous quip, self-organized systems are the "nonelephant" species of systems science—yet they are the least familiar of them. Since natural systems are *not* engineered: what can human-made systems learn from them? [Braha et al. 2006, Chap. 1]

In this preliminary study I present a simple prototype simulating some aspects of biological development with dynamical systems coupled as cellular automata. Following the artistic metaphor of a growing canvas that paints itself [Coen 2000], I attempt to reproduce developmental pattern formation by a *multiscale recursive process*. Starting with a few broad positional and identity domains of cells, details are gradually added by cell proliferation and differentiation using only local information. A "shape" emerges from self-assembling agents that all carry the same code. Parts 2 and 3 lay out the main ideas behind the model, which is further discussed in Part 4.

2   Stained Glass Patterning

This part describes the principles of developmental patterning based on spatial unfolding of genetic code [Carroll et al. 2001]—a process that could be nicknamed "shape from switching"—and gathers them into an abstract 2-D model. Section 2.1 briefly summarizes current knowledge about gene regulatory networks and positional integration. In Section 2.2 I offer a model of "stained glass" patterning based on an array of multitier networks. This model is then generalized in Part 3 to a *multiscale* system that incorporates the growth of the organism in a *recursive* and *modular* way.

2.1   Genetic switches integrate positional information

An important amount of genetic information critical for development is contained in the nonexpressed parts of DNA. Stretches of nucleotide sequences, called genetic

regulatory sites or *switches*, control gene expression at multiple stages in the developing organism. Switches are generally located upstream of the genes they regulate and are bound by specific proteins, whose effect is to interfere with the enzymatic processes responsible for genetic transcription (Fig. 1a). These proteins selectively attach to sequences of the DNA strand, as keys to locks, and can either impede ("repress") or accelerate ("promote") transcription. As genes are often reused at different stages and places during development, one genetic switch can take the form of a complex of multiple DNA segments separated by gaps and potentially bound by various combinations of promoter and repressor proteins.
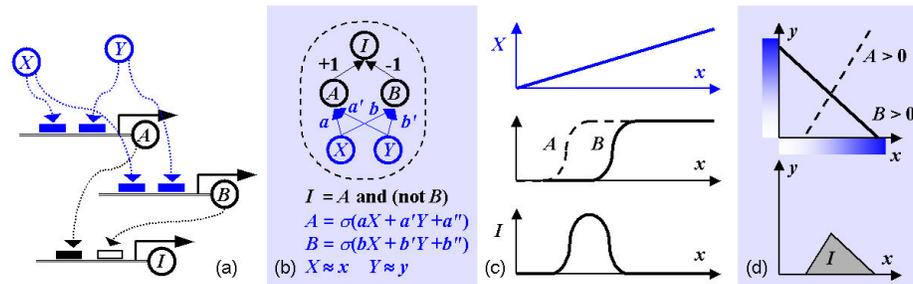

Fig. 1: Principles of spatial patterning from a genetic network. (a) Schematic view of gene regulatory interactions on DNA strands: proteins X and Y combine to promote the transcription of genes A and B by binding to their upstream regulatory sites, which leads to proteins A and B (assuming a simple one gene-one protein relationship); thereafter, A promotes, but B represses, the synthesis of I. (b) Formal view of the same GRN. (c) Variation of expression levels on one spatial axis, construed as a chain of GRNs: the concentration of X follows a gradient created by diffusion; this gradient triggers a gain response in A and B at two different thresholds, thus creates boundaries at two different x coordinates (for a given Y level); these domains in turn define the domain of identity gene I, where A levels are high but B levels are low. (d) Same spatial view in 2-D: the domain of I covers the intersection between high A and low B.

As regulatory proteins are themselves synthesized by the expression of other genes, the developmental genetic toolkit can be globally described as a complex web of regulatory influences, or *gene regulatory network* (GRN) (Fig. 1b). Although the structural and functional properties of GRNs are not fully understood, it seems that they are broadly organized into functional *modules* [Schlosser & Wagner 2004, Callebaut & Rasskin-Gutman 2005] that reflect the successive stages and anatomical modules of organismal development. Studies of stripe formation in the *Drosophila* embryo have identified a cascade of morphological refinements that start with global molecular gradients and progressively lead to the precise positioning of appendages [e.g., Coen 2000, Carroll et al. 2001]. Each period is characterized by the activation of a group of genes that respond to regulatory signals from the previous group, and trigger the next group (Fig. 1c). As an example, initial protein deposits of maternal origin located asymmetrically in the egg start diffusing across the syncytium. Then, depending on their concentration on each point, these initial molecules regulate the

expression a first set of genes at different levels and positions on the embryo's axes (antero-posterior, dorso-ventral and proximo-distal). The regulatory sites for these genes are sensitive to the concentration of maternal proteins at various thresholds, creating staggered domains of expression in the form of "stripes". These stripes in turn intersect in various ways to give rise to the next generation of gene domains (Fig. 1d). In particular, they form *identity domains* (*Drosophila*'s "imaginal discs"), where cells are characterized by a common signature of genetic activity, setting the basis for differentiated limb growth (leg, antenna, wing, etc.) and organogenesis.

In summary, molecular gradients of morphogenetic factors (the "keys") provide positional information [Wolpert 1969] that is integrated in each cell nucleus by its genetic switches (the "locks") along several spatial dimensions. Developmental genes are regulated by differential levels of lock-key fitness and expressed in specific patterns or territories in the geography of the embryo. A territory represents a combination of multiple gene expression values, i.e., morphogenetic protein concentrations. In this study I model the territories of embryo partitioning similarly to the colorful compartments created by the intersection of lead cames in stained glass works.
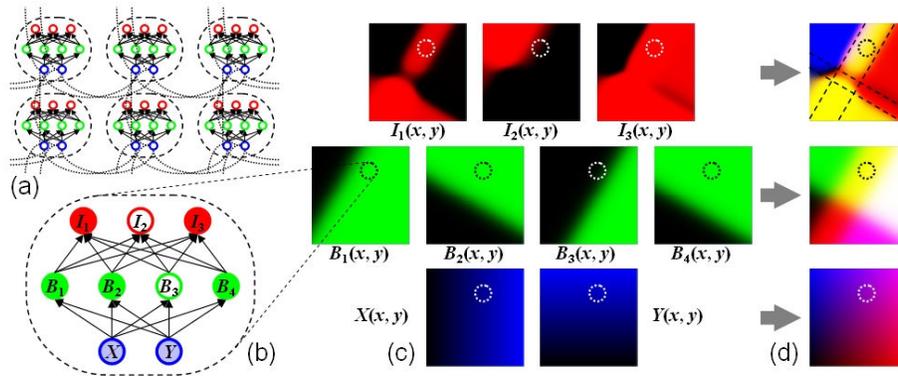

Fig. 2: Numerical simulation of stained glass domain formation from a simple feedforward GRN. (a) Subset of the 100×100 lattice of coupled identical GRNs: bottom nodes X and Y are linked to their neighbor counterparts to create diffusion. (b) This GRN contains 4 "boundary" genes and 3 "identity" genes: colors represent expression levels at a sample location (dashed circles). (c) 2-D domain maps created by each node on the lattice, under random weights: X and Y form horizontal and vertical gradients; $B_{1...4}$ nodes form diagonal boundaries; $I_{1...3}$ are expressed in various intersection segments of the $B_{1...4}$ domains. (d) Same 2-D maps, superimposed (using a different color scheme); bottom: X and Y; middle: B_1, B_3 and B_4; top: $I_{1...3}$.

2.2   A feedforward GRN model of stained glass segmentation

The core architecture of the virtual organism is a network of networks, construed as a 2-D lattice of identical GRNs (Fig. 2a). Each GRN represents a cell and is connected to neighbor cells via some of their GRN nodes [Mjolsness et al. 1991, Salazar-Ciudad et al. 2000, von Dassow et al. 2000]. In the present work, my first attempt at modeling stained glass patterning uses a simplistic feedforward GRN template (Fig. 2b)

containing three layers: (1) a bottom layer with two *positional* nodes, X and Y, (2) a middle layer of *boundary* nodes $\{B_i(t)\}_{i=1...n}$ and (3) a top layer of *identity* nodes $\{I_k(t)\}_{k=1...m}$ (Fig. 2b). Activity variables X, Y, B_i and I_k denote gene expression levels, assumed equivalent to the concentrations of proteins that these genes synthesize. In a first stage, the boundary nodes compute discriminant functions of the positional nodes via weighted sums and sigmoid thresholding: $B_i = \sigma(w_{ix} X + w_{iy} Y - \theta_i)$, where $\{w_{ix}, w_{iy}\}_{i=1...n}$ are the connection weights from layer (1) to layer (2), θ_i is B_i's threshold value and $\sigma(z) = 1 / (1 + \exp(-\lambda z))$. Given a set of weight values, the effect of a boundary node is to segment the plane into two half-planes of binary expression levels, weak and strong (Fig. 2c-d, middle). Parameter λ controls the slope of the logistic function σ, i.e., the sharpness of the boundaries (Fig. 1c). In a second stage, the identity nodes are activated by particular combinations of 'on' and 'off' boundary nodes, following the same thresholded sum rule: $I_k(t) = \sigma(\sum_i w'_{ki} B_i - \theta'_k)$, where $\{w'_{ki}\}$ are the weights from layer (2) to layer (3), and θ'_k is I_k's threshold value. Therefore, territories of high identity gene expression are comprised of polygon-shaped regions, at the intersection of multiple boundary lines (Fig. 2c-d, top).
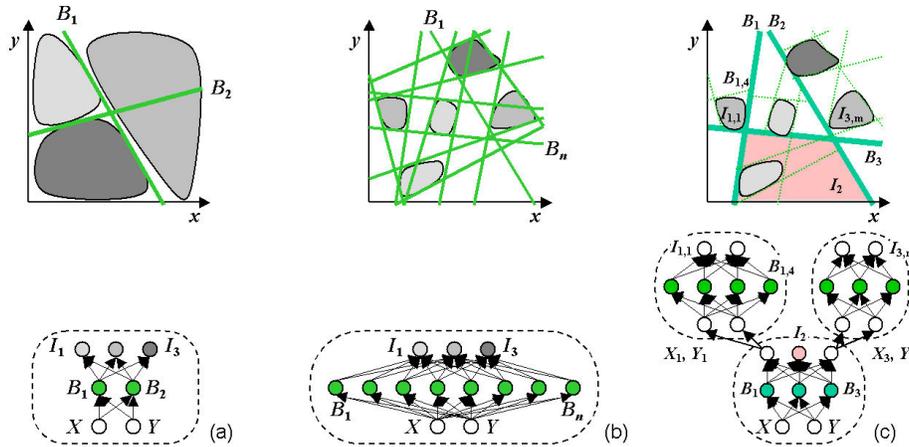

Fig. 3: Illustration of recursive morphological refinement based on a hierarchical GRN. (a) These 3 domains can be separated using only 2 boundary nodes. (b) The boundary layer increases rapidly with the amount of morphological detail. (c) A more economical and realistic approach uses a hierarchy of sub-GRNs: the identity nodes of the lower (earlier) modules define broad domains that subsequently trigger the higher (later) modules, via new local gradients. (Admittedly, this simplified diagram does not show a decrease in the total number of B nodes, but this is due in part to the absence of homologous modules; see text for a discussion.)

This feedforward network architecture is analogous to the multilayered perceptron, an artificial neural model applied to input-output classification in high-dimensional space. In the perceptron model, the bottom nodes code for input examples, the top nodes correspond to output classes and the middle layer (or "hidden units") provides elementary discriminant functions used by the output nodes to clas-

sify. A perceptron generates class domains similar to the patterns observed here (Fig. 2), also in more than 2-D. The main property of a multilayered perceptron is its ability to be programmed to reproduce a great variety of class distributions. During a training phase, or "supervised learning", input examples are presented to the network and weights are iteratively modified to minimize the discrepancy between the observed and expected output. Thus, the network can progressively adapt to a desired configuration of class domains partitioning the input space. Whereas a two-layer perceptron can only achieve linear separations (the half-planes created by the boundary nodes), a perceptron containing three or more layers is theoretically versatile and can reproduce class domains of almost all shapes and forms by combining these linear separations at a higher level. However, a necessary condition for this versatility is that the class domains remain reasonable connected sets, or piecewise connected sets. A major drawback is that the number of hidden units (linear boundaries) increases rapidly as the class domains become more fractured and scattered in small pieces (Fig. 3), eventually tending to infinity in the limit of discontinuous points.

3 Multiscale Segmentation: The Growing Canvas

In the previous section we saw that a lattice of coupled perceptron-like GRNs was able to generate a stained glass pattern of identity gene domains in two phases: a first phase establishing half-planes and a second phase combining these half-planes into polygonoid domains. Naturally, the basic boundaries need not be linear; other types of discriminant functions can be used on the middle layer nodes, e.g., polynomial, radial, Boolean, etc. Moreover, GRNs are not strictly feedforward as genes can be interacting within a layer in cooperative or competitive ways to refine each other's response (Fig. 4b). In other words, boundary lines and identity domains could be displaced by mutual attraction or repulsion instead of arising independently. Each layer with its recurrent connections might also correspond to an "attractor state" in the sense of Kauffman's genetic nets [Kauffman 1969]. In this case the feedforward propagation would correspond to a chain of scheduled phase transitions between transient attractors. Across all these variants, however, two main properties remain: (i) intermediate genes provide elementary domains that are further combined by downstream genes in nontrivial, morphology-specific ways; (ii) the intermediate tier scales rapidly with the amount of morphological detail (Fig. 3b).

Point (ii) is the main focus of this section. The simple feedforward GRN presented so far, although theoretically universal in its ability to generate segmented patterns, is limited in practice by scaling. Moreover, an organism's shape is not fully generated in every detail in just two phases, but rather *grows* in incremental stages. Biological observations indicate that transcription regulation "cascades" down from site to site in a broadly directed fashion, so that a GRN structure consists of many more than three tiers. In general, early developmental genes seem to pave the way for later sets of genes and are generally not reused. A frequent exception to this principle are multivalent genes that reappear in different organs and at different times during development. Yet, their capacity for repeated expression relies on independent combinations of switches (Fig. 4d), which can still be represented by a feedforward network

with duplicate nodes in different layers (red nodes in Fig. 4). Each node carries out a different switch logic under a GRN architecture that remains globally directed.
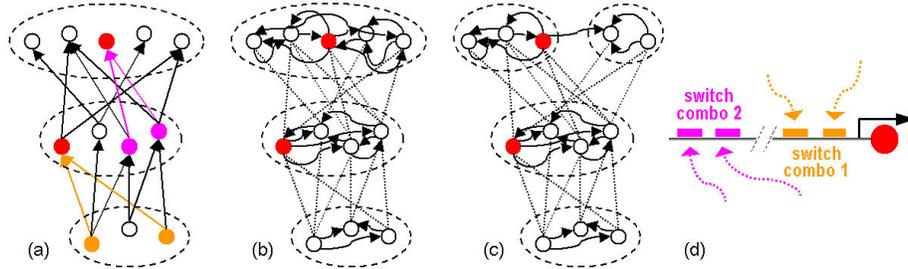

Fig. 4: Variations on the simple feedforward multitier GRN architecture (a). (b) Recurrent connections are added inside each tier. (c) Tiers are subdivided into subtiers to create a hierarchy (compare Fig. 3c). (d) A multivalent gene (red node in all schemas), i.e., a gene that is repeatedly expressed at different periods and locations of the developing organism, can be formally represented by duplicate nodes in separate tiers: these nodes are activated by different protein combinations in input (here, pink and orange) via different regulatory functions.

Therefore, to account for progressive growth and morphological refinement, a more adequate GRN model would consist of a *hierarchy* of subnetworks, where each subnetwork performs a function similar to the simple feedforward network presented above. Instead of relying on a single group of boundary nodes to cover fine-grain details (Fig. 3a-b), the image can be iteratively refined by the action of several mapping modules and submodules (Fig. 3c). At first, the network at the bottom of the hierarchy only establishes broad identity domains; then these identity domains in turn trigger specialized subnetworks that further create local partitioning at a finer spatial scale, etc. Morphological details are added in a hierarchical fashion, analogous to inclusions of small stained glass motifs into bigger ones (Fig. 5a). Fractal patterning has also been explored in "map L-systems" [Siero et al. 1982]; however, these use symbolic rules and explicit geometrical features instead of coupled dynamical units.

In parallel to this hierarchical refinement, the medium *expands*, i.e., cells multiply and expression domains enlarge. Thus, morphological refinement basically proceeds by alternation of two fundamental steps: (1) subdivision of a uniform identity domain into finer identity domains; (2) enlargement of the new identity domains, which receive further subdivisions during step (1). The transition from (1) to (2) is carried out by cell proliferation, whereby daughter cells inherit mother cell types (i.e., their current state of genetic expression: RNA, protein and metabolic concentrations) and temporarily preserve the local identity of an expanding domain. The reciprocal transition from (2) to (1) corresponds to the creation of new local gradients of positional information within former single-identity domains. Expanding domains are mapped by new local coordinate systems that activate the entry points of a regional subnetwork of the GRN in the next layer of the hierarchy. These new local gradients emerge from a diffusion process similar to the original diffusion of global coordinates *X*, *Y*. For example, during proliferation a small number of mother cells retain signaling

material that is not inherited by daughter cells, then later on start diffusing these molecules asymmetrically from the borders of the expanding domain toward which they were pushed. Naturally, steps (1) and (2) are not strictly separate; they can overlap and unfold at different rates in different body parts without global synchrony.
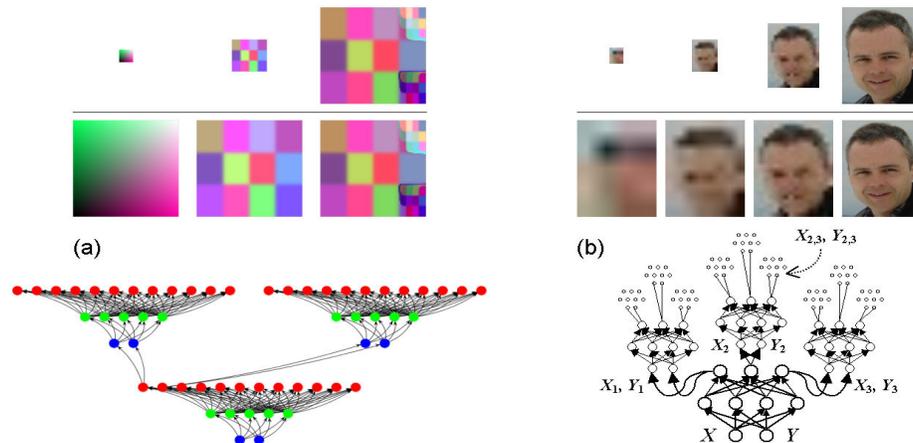

Fig. 5: The growing canvas of morphogenesis: pattern formation proceeds by successive refinement steps on an expanding medium. (a) Simulations with a 2-layer hierarchy of 3-tier GRNs; each GRN contains 2 horizontal + 3 vertical boundary nodes that give rise to 12 rectangular identity domains; 2 early domains become subdivided. (b) Same general idea illustrated by a portrait at multiple scales of resolution [after Coen 2000] and a generic network diagram.

4 Discussion and Future Work

In this article I have shown the possibility of multiscale pattern formation based on an expanding lattice of hierarchical, feedforward gene regulatory networks. The alternation of growth and patterning ultimately results in the creation of a "form", represented by an image or shape on the lattice. The hierarchical GRN structure adequately supports both the temporal sequence of developmental stages and the spatial accumulation of details. Most importantly, compared to a single layer of discriminant functions (Fig. 3a-b), a hierarchy allows the reuse of *modules* and thus can greatly reduce the number of intermediate gene nodes needed to generate a full mapping. In the fictitious example of Fig. 3 the number of *B* nodes is actually not smaller in the hierarchical network than in the flat network because the pattern does not contain repeated or "homologous" parts. However, in biological development a key property is *modularity*: organisms are made of repeated segments, most apparent in arthropods or the vertebrates' column and digits. Genetic sequencing has revealed many identical or highly overlapping stretches of DNA code not only within individuals but also across species, which indicates that during evolution segments might have duplicated then differentiated [Carroll et al. 2001]. In the present model, this would correspond to reusing the same subnetwork multiple times within one hierarchical layer (as in Fig. 5a, second layer), then mutating these copies to create variants.

In this study I attempted to clarify the elusive relationship between genotype and phenotype, specifically how a nonspatial genetic network can unfold in space to become a 2-D or 3-D shape. The problem of the quantity of information or "genetic cost" is of central importance here: What is the minimal number of gene nodes needed to cover a given amount of morphological details? Organisms obviously contain far less genes than cells or even local regions of homogeneous cell identity. This means that a GRN is actually extremely sparse and development cannot be entirely specified by positional information and switch-based logic. Modularity and component reuse are also not sufficient to explain the relative paucity of genetic instructions. Therefore, other "epigenetic" factors must contribute to the generation of morphological details: cell-to-cell signaling interactions (differential adhesion, Turing patterns), shape-sculpting programmed cell death (e.g., between digits), differential growth rates and topological transformations (gastrulation, mechanistic folding, etc.) [Webster & Goodwin 1996, Nagpal 2002]. A more comprehensive model of self-organized development should contain all of the above mechanisms (Fig. 6).
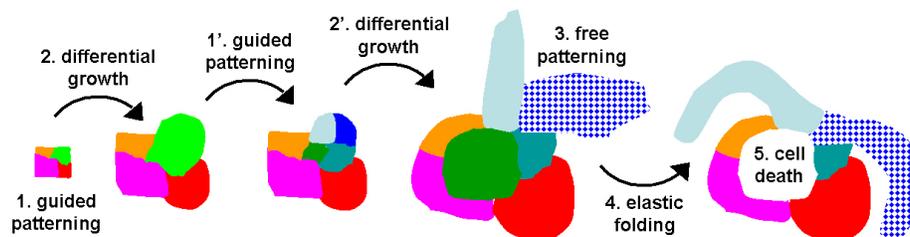

Fig. 6: The big picture of morphogenesis. A comprehensive model of form development would integrate several elementary genetic and epigenetic mechanisms. (1) Guided patterning: GRN-controlled establishment of expression maps (this article). (2) Differential growth: deformations (bulges, offshoots, limbs) created by domain-specific proliferation rates. (3) Free patterning: texture (stripes, spots) emerging from Turing instabilities. (4) Elastic folding: transformations from mechanistic cellular forces. (5) Cell death: detail-sculpting by domain removal.

In sum, the complete morphogenetic program (the "theme") consists of a set of developmental laws at the cellular level, while the parameters of this program (the "variations") are represented by specific GRNs. Fed into the same program, different GRNs will give rise to different shapes. At this point, evolution enters the stage (the "player") and two complementary issues arise: given specific GRN weights, what pattern will the growing canvas create? Conversely, given a desired pattern as a target, what values should the GRN weights take to produce this pattern? Beyond the biological challenge of unraveling real-world molecular pathways, these questions also raise a technological challenge: dynamic self-assembly and autonomous design in the absence of a global symbolic blueprint—e.g., as in swarm robotics or distributed software agents. Previous artificial models of development have mainly followed a bottom-up approach by observing, classifying and selecting patterns emerging from given GRNs, whether randomly wired or biologically detailed. For example, aposteriori statistical analyses try to identify prototypical wiring patterns in GRNs

[Salazar et al. 2000]. My intention is to explore the top-down, "reverse-engineering" approach and suggest through this preliminary study that potentially any given spatially-explicit blueprint could be encoded in the weights of a nonspatial GRN. While I have not proposed a specific algorithm to compute the weights in this article, known methods could be investigated and adapted, whether direct calculation (back-propagation, or back-compilation of global effects into local rules, similarly to [Nagpal 2002]), fitness-based evolutionary algorithms, or a combination of both.

In the same spirit as artificial neural networks or ant colony optimization, my goal is less a faithful reproduction of biological mechanisms than their abstraction and potential application to computational and technological problems. Drawing from biological development, I hope to be contributing to a novel engineering paradigm of system construction (virtual or physical) [Braha et al. 2006] in which the emergence of complex structures does not exclusively rest on one omniscient architect, but partly or fully on a decentralized collectivity of simple agents, each endowed with a low-cost, incomplete network of instructions.

Bibliography

[1]   Callebaut, W., & Rasskin-Gutman, D. (ed.), 2005, *Modularity*, The MIT Press.

[2]   Carroll, S. B., Grenier, J. K., & Weatherbee, S. D., 2001, *From DNA to Diversity*, Blackwell Scientific (Malden, MA).

[3]   Coen, E., 2000, *The Art of Genes*, Oxford University Press.

[4]   Edelman, G. M., 1988, *Topobiology*, Basic Books.

[5]   Kauffman, S. A., 1969, Metabolic stability and epigenesis in randomly constructed genetic nets, *Journal of Theoretical Biology*, **22**: 437–467.

[6]   Braha, D., Bar-Yam, Y., & Minai, A. A. (ed.), 2006, *Complex Engineered Systems: Science Meets Technology*, Springer Verlag.

[7]   Mjolsness, E., Sharp D. H., & Reinitz, J., 1991, A connectionist model of development, *Journal of Theoretical Biology*, **152**: 429–453.

[8]   Nagpal, R., 2002, Programmable self-assembly using biologically-inspired multi-agent control, *1st Int Conf on Autonomous Agents*, July 15-19, Bologna, Italy.

[9]   Salazar-Ciudad, I., Garcia-Fernández, J., & Solé, R., 2000, Gene networks capable of pattern formation, *Journal of Theoretical Biology*, **205**: 587–603.

[10] Schlosser, G., & Wagner, G. P. (ed.), 2004, *Modularity in Development and Evolution*, The University of Chicago Press.

[11] Siero, P., Rozenberg, G., & Lindenmayer, A., 1982, Cell division patterns: syntactical description and implementation. *Comp Graph Image Proc*, **18**:329-346.

[12] von Dassow, G., Meir, E., Munro, E. M., & Odell, G. M., 2000, The segment polarity network is a robust developmental module, *Nature*, **406**: 188–192.

[13] Webster, G., & Goodwin, B., 1996, *Form and Transformation*, Cambridge.

[14] Wolpert, L., 1969, Positional information and the spatial pattern of cellular differentiation development, *Journal of Theoretical Biology*, **25**: 1–47.