Evo-devo *in silico*: a Model of a Gene Network Regulating Multicellular Development in 3D Space with Artificial Physics

Michał Joachimczak, Borys Wróbel
Computational Biology Group, Department of Genetics and Marine Biotechnology
Institute of Oceanology, Polish Academy of Sciences
Powstańców Warszawy 55, 81-820 Sopot, Poland
mjoach@gmail.com

Abstract

We present a model of multicellular development controlled by a gene network in which the connectivity is determined by the proximity of sequences in N-dimensional space. Thus the sequences of individual genes can be visualised as points in space which approach or move away from one another as the genomes evolve. The genotype-phenotype (morphology) mapping in our model is indirect, relies on artificial physics, and allows cell adhesion and free movement in 3D space. Cell differentiation is allowed by positional information provided by factors that diffuse in this space, and the differential gene expression in each cell determines the cell fate (such us division, death, growth and movement). We apply a genetic algorithm to find genotypes that can direct morphogenesis of non-trivial asymmetrical shapes. We then investigate the mechanism of such developmental process and the features of gene regulatory network that direct the embryogenesis.

Introduction

Generation of two-dimensional (2D) patterns, such as the French flag (Wolpert, 1969, see also Steiner et al., 2006), is much simpler then the problem of 3D morphogenesis. Recent progress in this field was facilitated by the introduction of gene regulatory networks (GRNs) as an indirect regulatory mechanism in 3D artificial embryogenesis that relies in part on simulated physics (e.g. Eggenberger-Hotz, 2003b,a, 2004, see also Bongard and Pfeifer, 2003), in contrast to the initial more abstract approaches, such as generative encodings (see e.g. Prusinkiewicz and Lindenmayer, 1996).

In biological systems, the structure of GRNs is encoded indirectly in nucleic acid-based genomes. Coding sequences are accompanied (usually, preceded) by regulatory areas (cis-regulators, promoters) which regulate the level of gene expression. The coding sequences code for functional products: catalysers of biochemical reactions (enzymes, ribozymes), proteins that have structural/mechanical roles, and finally, regulatory products that bind other products or bind to regulatory areas in the genome to control the production of other products (gene expression). The gene products are the nodes in the biological GRNs while the edges are defined by regulatory interactions. The amino acid sequence of a protein product (or a nucleotide sequence of an RNA molecule) defines its 3D structure in a way that is still far from being fully understood (this is the so-called folding problem). Interactions between three-dimensional molecules (between proteins, proteins and nucleic acids, and between RNA molecules, etc.) are even more difficult to model.

Several approaches to encode the structure of the artificial GRNs in the genome using an abstraction of this "lock and key" mechanism of molecular recognition have been proposed. For example, product-promoter affinity can be determined in an all-or-none manner by a direct match between numbers assembled from the digits in the genomic sequence (Quayle and Bullock, 2006) or coded directly in the genome, possibly with real-number rounding (Bongard and Pfeifer, 2003). Jakobi (1995) used a different approach, with promoter affinity (a discretised value from 0 to 1) determined by the match between triplets of "chemicals": characters in a regulatory protein sequence (from a 64-letter alphabet) and in the genome (from a four-letter alphabet). The triplets are found indirectly in a metaphor of genome scanning by the RNA polymerase, folding of the regulatory protein, and protein-protein interaction between them. The method proposed by Eggenberger-Hotz (2003a) is much simpler and relies on direct proximity of real numbers encoded in the genome. Bit-by-bit comparison of 32-bit integers is another method of similar complexity (Banzhaf, 2003; Kuo et al., 2004). Bentley (2003) proposed a much more indirect approach based on encoding the coordinates for subsets of Mandelbrot set and matching their similarity.

We extend the approach that uses the proximity of real numbers (a 1D approach) by introducing a model of GRN in which product-promoter affinity depends on the Euclidean distance between points in N-dimensional gene sequence space. As the genomes evolve, these points approach or move away from one another.

The Model

Outline

In the model of embryogenesis proposed here, multicellular development starts from a single cell. Each cell of an

individual has the same linear genome: a list of genetic elements. Each element is characterized N+1 real values (N coordinates in the gene sequence space and a gene modifier, see below) and an integer (element type), with "gene" (a metaphor of a coding sequence) and "promoter" (an abstraction of a regulatory region) as the main types.

One of the key interests in our research is the gradual increase in complexity of regulatory networks and inherent mathematical properties of the corresponding graphs. We thus allow the genomes of arbitrary size and regulatory units that have no upper limits on the number of nodes they directly interact with. Furthermore, we provide multiple layers of possible interactions in developing artificial organisms for evolution to tinker with. On the most basic level, inside a single cell, the products that we call "internal" have affinity to promoters and control the expression of other genes. On the level of a whole organism, special class of products ("external products") may diffuse from the source cell. External products may have affinity to promoters but also to products belonging to another class: receptors. Thus external products are a metaphor for morphogens in natural embryogenesis, and their interactions provide a mechanism for cell differentiation and differential growth. Furthermore, cells develop in an environment with simple simulated physics: overlapping cells repel each other while daughter cells are attached to the mother cells with simulated springs (a metaphor of adhesive forces between cells).

Regulatory Units

Regulatory units are formed from a series of promoters followed by a number of genes and are the basic building blocks for the structure of the GRN. There are a metaphor of regulatory units in nucleic acid-based genomes, in which several protein- or RNA-coding regions can be under the control of several regions that affect gene expression at different levels: pre-transcriptional, post-transcriptional (stability, transport, and translation of transcripts), and post-translational (stability, transport and activity of proteins).

The majority of existing models of artificial embryogeny follows the scheme of multiple promoters and one product (e.g. Beurier et al., 2006; Steiner et al., 2006; Eggenberger-Hotz, 1997, 2003b,a, 2004). However, many-to-many relationship between the promoters and regulated genes is common both in prokaryotes and eukaryotes (Gerstein et al., 2007). Indeed, clustering of several genes in so-called operons is not, as originally thought, restricted to bacteria, but common also in eukaryotes (for a recent review, see Blumenthal, 2004; Gerstein et al., 2007). Such arrangement allows for co-regulation of co-transcribed genes that are closely related functionally (for example, involved in the same biochemical process). A similar logic applies to multiple transcripts sharing common regulatory regions (promoters, enhancers, silencers; Gerstein et al., 2007), polyproteins, and ineed, multidomain proteins (with domains responsible for separate functions).

To locate regulatory units in our model, each genome is scanned linearly. Whenever a sequence of elements consisting of at least one promoter followed by at least one gene is detected, it is treated as one unit that extends until the next promoter. Promoters and genes outside regulatory units are ignored. For example, in a genome "GGGppGpGGpGpp" (where each p is a promoter and each G a gene), three regulatory units (square brackets) exist: GGG [ppG] [pGG] [pG] pp.

Two types of promoters are introduced: additive and multiplicative. To compute the level of activation of a regulatory unit (the expression level of all of the products), we first compute the activity of all of its promoters:

$$p_i = \sum_{k=1}^{K} L_k w_{k,i} \,. \tag{1}$$

where p_i is the activity of a given promoter, K is the total number of regulatory factors in the genome (e.g. internal or external products, see below), L_k denotes the perceived level of the factor k, and $w_{k,i}$ is the promoter-factor affinity:

$$w_{k,i} = \begin{cases} d_{k,i} \leq 5 : \mathrm{sgn}(m_k m_i) \frac{2|m_k m_i|(5-d_{k,i})}{10d+|m_k m_i|} \,; \\ \\ d_{k,i} > 5 : 0 \,. \end{cases} \tag{2}$$

where $d_{k,i}$ is the Euclidean distance between the sequences of the promoter i and the gene of factor k, while m_k and m_i are the values of their modifier fields. In other words, affinity is 0 (no interaction) when the distance is larger than 5, and at a maximum (10) when the distance is 0. For intermediate distances, the affinity falls hyperbolically, rapidly for small $|m_k m_i|$ and approximately linearly for large $|m_k m_i|$. The signs of the modifiers determine if the effect is inhibitory or excitatory.

All the genes belonging to a given regulatory unit have the same level of expression:

$$L_\Omega = f(\prod_{i=0}^{I} p_{m,i} \sum_{j=0}^{J} p_{a,j}) \,. \tag{3}$$

where I and J denote the number of multiplicative and additive promoters (respectively) and $p_{m,1..i}$ and $p_{a,1..j}$ describe their activations. It is possible to allow for the bias in activation ($p_{m,0}$, $p_{a,0}$), but we use the identity element of, respectively, multiplication and addition ($p_{m,0} = 1$,$p_{a,0} = 0$). f is a sigmoidal function returning value from (0,1) with the threshold at 0.5 (the steepness of the sigmoid was kept constant in all the experiments).

The presence of a multiplicative promoter in a regulatory unit results in a strict requirement for the expression of the associated product, otherwise the whole unit remains inactive. This feature (where a subset of transcription factors is

necessary to initiate gene expression in an 'all-or-nothing' fashion) is known to be common in gene regulation but would not be easily captured by a purely additive function. Additionally, introducing multiplicative promoters provides evolution with a mechanism to switch off the whole regulatory unit with a single mutation.

Products

We introduce three types of products that can be coded by genes in a regulatory unit: internal, external and receptors. Internal products affect the expression of the regulatory units only in the cell in which they are produced. External products (morphogens) diffuse from the producer cell and bind to promoters and receptors in other cells. Receptors, on the other hand, interact with external products and influence the axis of cell division (the division vector) by shifting it toward or away from the source. Since the cells may differ in the pattern of gene expression, also their set of active receptors may be different. This allows each cell to orient its division in relation to the pattern of morphogens that are available at a given moment.

The affinity between a morphogen and a promoter or a receptor is defined by Eq. 2, but to simulate simple diffusion, the perceived concentration of the morphogen depends not only on the level of production but also on the distance from the producing cells (the sources). The level of morphogen m perceived by the cell c is:

$$L_{c,m} = \sum_{i=1, i \neq c}^{I} l_{i,m} \frac{1}{1 + D_{i,c}} \,. \tag{4}$$

where I denotes the number of cells at the current developmental stage, $D_{i,c}$ is the distance from the cell c to the source i in the 3D space of the developing organism and $l_{i,m}$ is the level of morphogen m this source produces (the expression level of the morphogen, Eq. 3).

Additionally, to simulate some of spatiotemporal effects generated by diffusion, the actual value of $l_{i,m}$ is delayed in time, depending on the distance from the source. This adds some additional memory cost of storing morphogen levels from previous time steps for each morphogen in each cell, but at basically no additional computational cost. Alternatively, one could obtain more realistic diffusion by simulating a 3D grid in which morphogens would diffuse (e.g. Beurier et al., 2006; Steiner et al., 2006). However, computational cost of updating fine-grained diffusion levels in 3D would be considerable.

Other Types of Genetic Elements

Three additional element types exist: pseudogenes, external factors and effectors. They are ignored when the genome is scanned for regulatory units.

Pseudogenes. If any genetic element is mutated to a pseudogene, its sequence will be shielded from the selective pressures until another mutation changes the element type.

External factors. External factors act in exactly the same way as external products (that is, they may interact with regulators and receptors), however their levels of expression are not regulated by the cell: the factors are provided externally at predefined levels. Also, for the positional factors (see below) the source locations are predefined. External factors can be thus viewed as inputs of the GRN and their possible interactions with receptors is an initial mechanism for breaking the symmetry of cell divisions.

Two subclasses of external factors are introduced. The first one consists of positional morphogens, emitted from four different points in three-dimensional space. They are a metaphor for maternal factors in natural embryogenesis. Thus each cell is provided with enough positional information to locate itself in 3D (four points is the minimal number that allows 3D trilateration). The perceived level of the positional external factors in a particular cell depends on the Euclidean distance to the source, in the same way as the level of external products (Eq. 4).

The perceived level of external factors in the second class does not depend on cell location. The level of one is constant throughout the development (a "1" signal), so it can be used as a simple threshold for any of regulatory unit in the GRN. The other provides a time signal, its expression increasing linearly from 0 to 1 during the developmental process. The next two are somewhat related: a generation counter (incremented in each daughter cell after division), and the energy depletion level, which increases from 0 to 1 (each cell division has some cost for both the daughter and the mother cell). The level of the last factor in this class depends on the number of neighbours (saturating at 1 for 8 cells in close proximity) and thus allows the cell to detect when it lies in a densely packed cell structure.

Effectors. Effectors can be viewed as outputs of the GRNs. They either correspond to actions each cell can take during the development or allow to adjust the parameters of the developmental process. Each effector is defined by its sequence in the N-dimensional gene space, and products that have their sequences close enough will add to its activation (using Eq. 2). In a way, this parallels the promoters, and indeed one can consider each effector as a special regulatory unit with a single promoter.

Cell actions consists of all-or-none responses when activation levels of corresponding effectors reach a certain threshold. These are: division, apoptosis (programmed cell death) and freezing (after which expression levels in the cell are no longer updated). In the second group of effectors, the following parameters are updated by a value corresponding to the activation level defined by Eq. 3: cell radius, spring

length, internal division vector length and internal division vector angles.

Since the list of possible effectors and external factors is predefined and we prefer to avoid defining a separate element type for each, the assignment of a given element coding an effector or an external factor to a particular function depends on its order in the genome. After all the functions are assigned, the rest is treated as pseudogenes.

Development

Development starts from a single cell and proceeds over discrete time steps. It stops either when a maximum time step is reached or when an individual embryo exhausts its initial energy.

The state of each cell is determined by the expression levels of all of the products in the genome, cell coordinates in 3D space (these are real values, no grid is used), cell radius, the orientation of the cell division vector, energy level, and several parameters related to the physical model. As the fate of the cells depends on the differential gene expression, it is essential to provide a mechanism that will break the initial symmetry of cell divisions. This is the function of the initial gradients of positional external factors. A similar mechanism is known to direct the initial stages of insect embryogenesis (for a popular introduction, see Carroll, 2005).

When a new cell is formed, it is attached to its mother with an elastic spring and placed at a very small distance, so the two cells initially overlap in space. The default length of the spring is equal to the sum of cells' radii and can be increased by designated effector gene, if present. During subsequent time steps after division, the elasticity of the spring will push away the cell in the direction of the mother's division vector. The spring ensures that the new cell moves away in the desired direction while remaining close to the mother, simulating simple adhesion. Repulsion between any two cells (at a certain distance) ensures that the cells do not overlap in space. To prevent brusque movements of the cells, their motion is slowed down with simulated viscosity. In a manner similar to spring length, the radius of the daughter cell is controlled by a dedicated effector. No activity means default value, maximum activity translates into a twofold increase.

The position of the daughter cell after division is influenced by two mechanisms, each corresponding to one of two auxiliary vectors maintained for each cell: the internal and external division vector. The direction of the sum of these two vectors (the cell division vector) gives the direction of the spring that attaches the daughter to the mother cell.

The first mechanism is directly based on the mechanism used in 3D L-systems (Prusinkiewicz and Lindenmayer, 1996). A daughter cell inherits the internal vector from the mother. At this point the vector is rotated in the daughter cell. Each of the three angles of rotation is affected by the expression of one of three effectors in the mother cell (Prusinkiewicz and Lindenmayer, 1996). If the effector is activated (Eq. 1), the rotation is positive, repression by inhibitory regulators results in negative rotation. An additional effector is used to determine the length of the internal vector. The default vector length is 0, which means that if there are no products acting on this effector (or the element corresponding to it is not present in the genome), the direction of the cell division vector will not be influenced by this mechanism.

The second mechanism allows to orient the vector towards or away from morphogen sources. High positive affinity between the sequence of an active receptor and the sequence of a morphogen perceived in the cell shifts the direction of the external vector toward the source of this morphogen. Negative affinity shifts the vector in the opposite direction. The overall effect is a sum of interactions of all receptors in the given cell with all morphogens produced by every source:

$$\vec{V_c} = \sum_{r=1}^{R} \sum_{m=1}^{M} \sum_{s=1, s \neq c}^{S} l_r w_{r,m} L_{c,m} \vec{\delta}_{s,c} \,. \qquad (5)$$

where R denotes the total number of receptors in the genome, M the total number of external products and external factors defined in the genome, S is the number of sources (cells and four positional external factors), l_r is the expression level of the receptor r in the cell (Eq. 3), $w_{r,m}$ is the morphogen-receptor affinity (Eq. 2), $L_{c,m}$ is the perceived level of the morphogen (Eq. 4), and $\vec{\delta}_{s,c}$ is the normalized vector from the given cell to the source.

To allow for a control of cell divisions, we provide an input to the GRNs that is a metaphor of the nutritional/energetical state of the cell. Since in our model each cell division has some energetical cost, the cell energy can be exhausted by rapid divisions. The same applies to the whole developing individual: there is a limit on the total energy that can be used during the development. As mutations causing uncontrolled cell divisions put a high drain on computational resources during evolution, early exhaustion of such individual's energy can help keep the problem in check. Additional biological realism is introduced by requiring a brief (10 simulation time steps) period of division arrest right after a division, both in the mother and in the daughter cell. Arrested cells update the state of their GRN normally but cannot divide no matter how high the expression of the corresponding effector. This has an additional advantage of giving the simulated physics the time to adjust the position of the new cell.

Fitness evaluation

The most obvious way to assess the fitness in simulations of morphological development is to count how many cells fit inside the desired shape, penalising for each cell outside the shape (e.g. Kumar, 2004). This approach works well when cells can only take certain locations on the grid, but leads to

undesired results when cells take arbitrary positions in space and can temporarily overlap. The possibility to reach high fitness by producing densely packed and highly overlapping cells would allow to exploit the simulated physics and other features of the model in an unintended way. We propose an alternative: a cuboid in 3D space that contains the target shape is divided into cubical voxels and each voxel is marked either as internal or external to the shape. To compute fitness, we iterate over each cell and check whether they occupy internal voxels, and if so, those voxels are marked as occupied. This approach has several advantages. First of all, it is efficient and allows to avoid repeated scoring of voxels occupied by overlapping cells. Secondly, it allows the cells to adopt different sizes (and even shapes, although this is not explored here). Finally, it is possible to give higher weights for some of the voxels to assist the evolution of morphologies that otherwise do not evolve easily.

Implementation

The computations are simplified by first transforming the genome into a GRN graph, in which only if the distance between the sequences is smaller than the threshold (5), an edge is drawn (see Eq. 2). During the development, it allows to update the state of the GRN using a list of factors that affect each promoter or receptor.

The dynamics of cell movement is simulated with simple Newtonian physics, using Runge-Kutta 4th-order integration. Springs behave according to Hooke's law and additional repulsive force is introduced between any two cells that overlap.

For complex GRNs it takes considerably less time to compute the new location of the cells compared to the time taken to update the state of the GRN. It is thus possible to update the GRN state, for example, only every 10 steps of simulated physics.

Genetic algorithm

All the results obtained in this work were obtained using a generational genetic algorithm with constant-size population of 300 individuals. A new generation was formed by copying 5 genomes without mutation (elitism), 150 with mutations and crossover, and 145 by mutation only. We allowed for multi-point crossover between genomes of different sizes. The candidate genomes for the next generation were chosen using tournament selection (which is not susceptible to the scaling of the fitness function). Elite individuals replaced the elite individuals from the previous generation if their fitness was equal. This allows elite genomes to wander through the neutral regions in the sequence space (which may allow for a more efficient evolutionary search, Shipman et al., 2000).

The elements in the genome are the lowest level of abstraction in our model, so the genetic operators were designed to work on the level of the genetic elements (rather than single bits or real values): each had a predefined probability of occurrence per element in a genome and per generation. The first operator results in a modification of the modifier or of the coordinates in the N-dimensional sequence space. The coordinates are modified by addition of a small value drawn from a Gaussian distribution. This operator corresponds to simple mutations in nucleic acid sequence (such as point mutations, short deletions and insertions in the coding or regulatory sequences). The second mutational operator, on the other hand, does not have any obvious biological interpretation, and allows to change the sign of the modifier. Another mutational operator allows for a change in the element type (with unequal probabilities for each type), in particular, a change of any element to a pseudogene and vice versa, with an obvious biological parallel. However, we allow any type change, which includes a direct change of a receptor into a morphogen or a promoter to a product (and vice versa), while conserving the sequence. In further work, we plan to explore if this feature helps evolution, at any rate, it does not have an obvious natural counterpart.

The remaining mutational operators act on the level of whole elements (element deletion, duplication, and insertion of a randomly created element) and the whole genome: deletion of a segment of the genome with random start and end point and a duplication of such a segment to a random position in the genome.

In the experiments described below, we set the probabilities of deletions to be around twice as high as probabilities of element duplications and insertions. Such deletion pressure restricts the accumulation of elements whose presence does not affect fitness (i.e. in which mutations are neutral) and so prevents the unnecessary growth of genome size. This particular solution to the genome size issue was partly motivated by biological realism (Charles et al., 1999), and partly by difficulties in properly balancing the fitness function faced by an alternative: a fitness cost to larger genomes. However, some level of neutral elements (which include pseudogenes) is beneficial. In natural genomes the presence of regions in which mutations do not affect the phenotype (neutral regions, junk DNA) allows for the appearance of innovations beneficial from the point of view of natural selection (Shipman et al., 2000). Such regions are shielded from the selective pressures which allows for bolder movements in the sequence space.

Results

In all our experiments, the evolution started from the same simple genome (Fig. 1A) designed by hand and containing four regulatory units, all regulated by external factors. The products in two units have effect on the division effector, two other induce rotation of the internal division vector and its length. The remaining external factors and effectors are defined but the nodes are not connected to the others in the GRN (and are not shown in Fig. 1B). The number of dimen-

sions of the gene sequence space was set to two.

In a way, the presented model attempts to trade simplicity for biological realism. In further work we plan to address the question whether the model can be simplified (or, indeed, complicated). Before it is possible, we need to ask if this initial version allows for efficient search in non-trivial fitness landscapes, by challenging the genetic algorithm with target shapes of different difficulty. While highly symmetric structures (spheres, ellipsoids) or slightly more demanding half-ellipsoids evolved quite easily (not shown), asymmetric morphologies, shown on Fig. 2 are a difficult task, and usually over 500 generations were needed to find a solution. Interestingly, the solutions found by the genetic algorithm did not rely on cell death (a mechanism that we observed to be used often in the development of half-ellipsoids; not shown) but rather on differential cell division and cell growth (changing cell radius) in different regions.

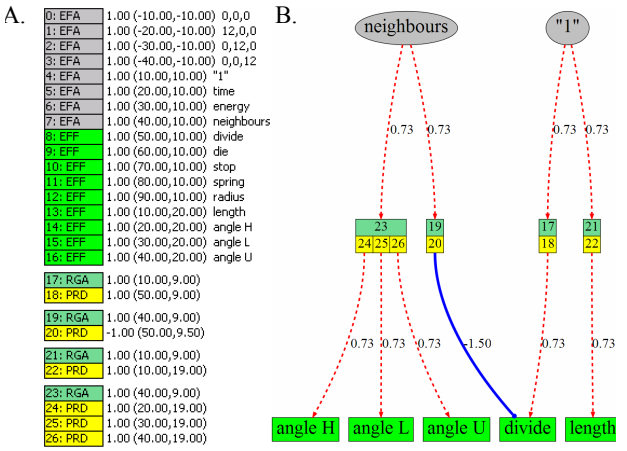

Figure 1: The seed genome (A) and the corresponding gene regulatory network (GRN; B). The genome consists of 27 elements (the value of the modifier, the coordinates in 2D sequence space are listed on the right): 8 external factors (the first 4 are positional factors, with 3 coordinates in 3D developmental space), only 2 of which are connected to the GRN, and 9 effectors, of which only 5 are connected, and 6 genes in 4 regulatory units.

The genome of the best stem-cap individual (Fig. 2A) codes for only three external products. To investigate their role, we have used a standard procedure used in molecular biology: knock-out experiments. Only the deletion of one of 3 morphogens (mpg3 or *capless*) had a large effect on fitness (Fig. 3A): no cap formation. Adding the third dimension to the sequence space allows to incrementally move the position of *capless* away from the plane in which all the other genes lie. This is a simple way to decrease the weight of a connection between a given gene and all the other nodes in the GRN, and corresponds to introducing point mutations as opposed to gene deletion. It can be seen (Fig. 3B) that

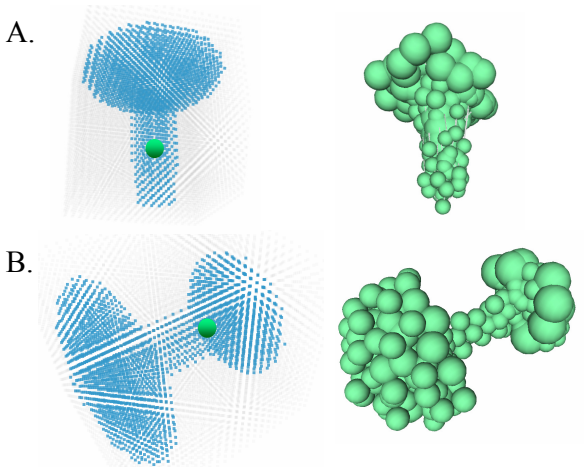

Figure 2: Difficult target shapes (left; the sphere marks the position of the first cell, dots represent target voxels): stem-cap (A) and asymmetrical dumb-bell (B), and best evolved solution phenotypes (right).

the effect of such operation on the development is "dose-dependent": the more the sequence was disturbed, the larger the effect. We can conclude that expression of *capless* allows for the development of a defined morphological structure.

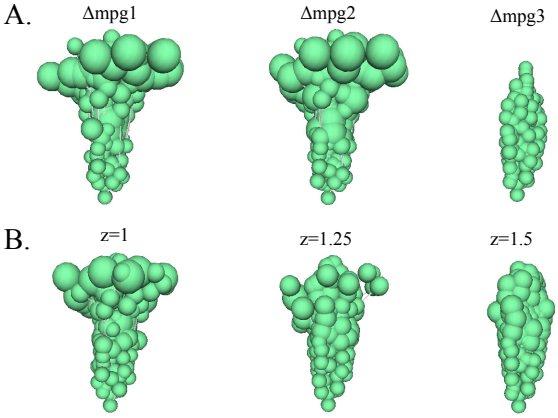

Figure 3: Mutational analysis of the best-solution individual to the stem-cap target. Only the deletion of one of 3 morphogens (mpg3 or *capless*) has a large effect on fitness (A). Shifting the location of this morphogen away from the XY 2D plane in which all the other genes lie (along the Z axis) has an incremental effect on fitness (B).

Interestingly, both the production and the perceived level of this morphogen in the developing structure is not asymmetrical along the main axis of development (Fig. 4A). It seems that all the cells produce this factor and its perceived concentration increases dramatically after the developmental step 60 when the cell number doubles (cell divisions are synchronised in the development of this individual, taking

the advantage of the division arrest mechanism). This means that it is not the asymmetry of *capless* expression that allows for the cap formation. Rather, the increase in concentration of this morphogen at step 60 causes asymmetric cell division and cell growth when the cell number doubles again after step 70. In other words, another mechanism must be used for cell differentiation along the embryo axis. We confirm this conclusion by creating an embryo in which all the cells express *capless* at a constant level (Fig. 4B).

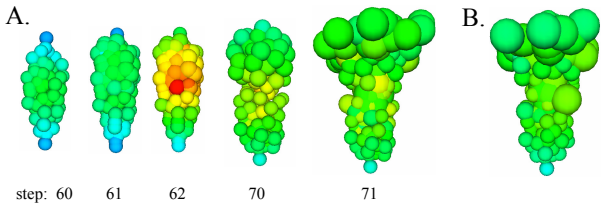

Figure 4: The perceived level of *capless* in the development of a stem-cap structure. Panel A shows the level of this morphogen in the best-solution individual at different developmental steps. Panel B: the phenotype of an individual in which all the cells produce *capless* at the same level throughout the development.

Fig. 5 presents a graph representation of a GRN controlling the development of asymmetric dumb-bell shown on Fig. 2B. It can be observed that majority of high-weight connections are inhibitory. It is also interesting that only some of the inputs (constant signal, only one spatial external factors out of four) and possible effectors are used. In other words, the development takes advantage of the changes in cell radius, internal division vector length and its rotations in two directions out of three allowed by the model. The developmental mechanism in this particular GRN does not use cell death, freezing or changes in spring length. The analysis of other GRNs evolved in our experiments also showed that only a subset of developmental mechanisms is actually needed to enable the morphogenesis of non-trivial shapes. However, quantitative analysis of hundreds of evolved GRNs would be necessary to infer any general properties of evolved networks.

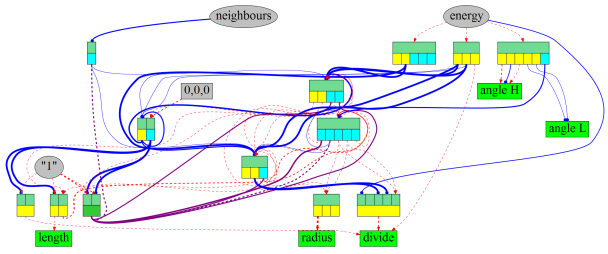

Figure 5: The GRN controlling the development of a asymmetrical dumb-bell shape in Fig. 2B. Dashed lines correspond to excitatory connections.

Discussion and Future Work

Our model extends the ideas first presented in several seminal papers by Eggenberger-Hotz who introduced GRNs with the affinity based on the similarity of real numbers (albeit in one dimension, 2003b; 2003a; 2004) and physics based on springs (2003b; 2003a). However, in this previous work cells grow on a grid and the embryo structure is further reshaped by controlling the forces between neighbouring cells. In contrast, in our model the development relies on cells of different size dividing freely in 3D space, and pushing each other away.

At the present stage, many features can be seen as unnecessary complications. Future work will show how many can be removed without compromising the ability of the evolutionary process to solve non-trivial tasks. Some parameters were included to allow future analysis of their effect on evolvability (this influenced, for example, our choice to use the insertion/deletion ratio as a way to control the genome size). At this preliminary stage, only a perfunctory analysis of the values of many parameters was possible (this applies, for example, to the thresholds in Eq. 2). Moreover, some parameters are related (for example the optimal value of the thresholds depends on the average size of mutational steps). However, the analysis of the effects of the inclusion of some parameters and their particular values on evolvability requires systematic experiments that need considerable time.

The main drain on computational resources in our experiments is frequent apparition of individuals which develop by uncontrolled cell divisions, and the energy depletion mechanism plus a limit on cell number, while keeping this problem in check, limits the potential of the artificial embryology. One of the possible solutions is to allow for a slow increase of energy in a manner that will reward slow controlled growth.

On the other hand, some features can be viewed as unnecessary simplifications. Perhaps tethering the daughter cells just to the mother cells is one of them, and since we already introduce the concept of cellular neighbourhood, attaching the cells to the neighbours (possibly taking advantage of the receptor compatibility) may provide a fruitful direction of further development (Bongard and Pfeifer, 2003; Eggenberger-Hotz, 2003b,a, 2004). Another direction is to allow dynamic changes in cell size and spring forces (in the present version both remain set after division; cf. Eggenberger-Hotz, 2004 where a similar feature allows for simple locomotion).

It would be also interesting to increase the realism of dynamics of gene expressions by introducing finite rates of change in product concentrations, represented with differential equations (e.g. as in Banzhaf, 2003; Kuo et al., 2004).

We might, however, argue that the shape of the parameter space in our model is not as complex as in the models presented previously. For example in the Eggenberger's model

each product is specified by as much as 7 different parameters. Since our model allows to compare the efficiency of the evolutionary search in gene sequence spaces of different dimensionalities, we will be able to investigate this issue in future work.

Primarily, however, we plan to go beyond the genetic algorithm as an approach to investigate the interplay between evolution and ontogeny using our model. The development of an actual artificial life setting, with competition for limited resources in a simulated world with a spatial structure (that would allow for at least temporal separation of subpopulations) will be the main objective of our further work. A genetic algorithm is a search method that allows only for a preliminary assessment of evolvability, but has obvious limitations: features like elitism, tournament selection, constant population size, fixed values of the parameters of the evolutionary process are not features of natural selection. Only an artificial life setting will allow to properly investigate such issues like the effects of the episodes of low population sizes, and the ability of the self-adapting systems that can tune their mutation rates to reach new adaptive peaks. We believe that only in such a setting some evolutionary questions considering the robustness of the network (and the related question epistasis), the role of mutations in regulatory regions for the evolutionary innovations, and the statistical properties of the evolved GRNs can be meaningfully explored.

Acknowledgements

The computational resources used in this work were obtained through the support of the Polish Ministry of Science and Education (project N303 291234), the Tri-city Academic Computer Centre (TASK) and the Interdisciplinary Center for Molecular and Mathematical Modeling (ICM, University of Warsaw; project G33-8).

References

Banzhaf, W. (2003). On the dynamics of an artificial regulatory network. In *Advances in Artificial Life*, volume 2801 of *Lecture Notes in Computer Science*, pages 217–227. Springer Berlin / Heidelberg.

Bentley, P. (2003). Evolving fractal proteins. In *ICES 2003, Evolvable Systems: From Biology to Hardware*, volume 2606 of *Lecture Notes in Computer Science*, pages 81–92. Springer Berlin / Heidelberg.

Beurier, G., Michel, F., and Ferber, J. (2006). A morphogenesis model for multiagent embryogeny. In *Proceedings of the ALife X: The Tenth International Conference on the Simulation and Synthesis of Living Systems*, pages 84–90. MIT Press.

Blumenthal, T. (2004). Operons in eukaryotes. *Briefings in Functional Genomics and Proteomics*, 3(3):199–211.

Bongard, J. and Pfeifer, R. (2003). Evolving complete agents using artificial ontogeny. In *Morpho-functional Machines: The New Species (Designing Embodied Intelligence)*, pages 237–258. Springer-Verlag, Berlin.

Carroll, S. (2005). *Endless Forms Most Beautiful: The New Science Of Evo Devo And The Making Of The Animal Kingdom*. WW Norton & Company.

Charles, H., Mouchiroud, D., Lobry, J., Gonalves, I., and Rahbe, Y. (1999). Gene size reduction in the bacterial aphid endosymbiont, *Buchnera*. *Molecular Biology and Evolution*, 16(12):1820–1822.

Eggenberger-Hotz, P. (1997). Evolving morphologies of simulated 3D organisms based on differential gene expression. In *Proceedings of the Fourth European Conference on Artificial Life*, pages 205–213. MIT Press.

Eggenberger-Hotz, P. (2003a). Exploring regenerative mechanisms found in flatworms by artificial evolutionary techniques using genetic regulatory networks. In *The Congress on Evolutionary Computation, CEC '03*, volume 3, pages 2026–2033.

Eggenberger-Hotz, P. (2003b). Genome-physics interaction as a new concept to reduce the number of genetic parameters in artificial evolution. In *The Congress on Evolutionary Computation, CEC '03*, volume 1, pages 191–198.

Eggenberger-Hotz, P. (2004). Asymmetric cell division and its integration with other developmental processes for artificial evolutionary systems. In *Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems*, pages 387–393. MIT Press.

Gerstein, M., Bruce, C., Rozowsky, J., Zheng, D., Du, J., Korbel, J., Emanuelsson, O., Zhang, Z., Weissman, S., and Snyder, M. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Research*, 17(6):669.

Jakobi, N. (1995). Harnessing morphogenesis. *Proceedings of Information Processing in Cells and Tissues*, pages 29–41.

Kumar, S. (2004). *Investigating Computational Models of Development for the Construction of Shape and Form*. PhD thesis, Department of Computer Science, University College London.

Kuo, D. P., Leier, A., and Banzhaf, W. (2004). *Evolving Dynamics in an Artificial Regulatory Network Model*, volume 3242 of *Lecture Notes in Computer Science*, pages 571–580. Springer Berlin / Heidelberg.

Prusinkiewicz, P. and Lindenmayer, A. (1996). *The algorithmic beauty of plants*. Springer-Verlag New York, Inc., New York, NY, USA.

Quayle, A. P. and Bullock, S. (2006). Modelling the evolution of genetic regulatory networks. *Journal of Theoretical Biology*, 238(4):737–753.

Shipman, R., Shackleton, M., and Harvey, I. (2000). The Use of Neutral Genotype-Phenotype Mappings for Improved Evolutionary Search. *BT Technology Journal*, 18(4):103–111.

Steiner, T., Olhofer, M., and Sendhoff, B. (2006). Towards shape and structure optimization with evolutionary development. In *Proceedings of the ALife X: The Tenth International Conference on the Simulation and Synthesis of Living Systems*, pages 70–76. MIT Press.

Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. *Journal of Theoretical Biology*, 25(1):1–47.