

# COMPLEXITY OF THE SEARCH SPACE IN A MODEL OF ARTIFICIAL EVOLUTION OF GENE REGULATORY NETWORKS CONTROLLING 3-DIMENSIONAL MULTICELLULAR MORPHOGENESIS

MICHAŁ JOACHIMCZAK<sup>1</sup>

BORYS WRÓBEL<sup>2</sup>

*Computational Biology Group, Department of Genetics and Marine Biotechnology  
Institute of Oceanology, Polish Academy of Sciences  
Powstańców Warszawy 55, 81-820 Sopot, Poland*

<sup>1</sup>*mjoach@iopan.gda.pl*

<sup>2</sup>*bwrobel@iopan.gda.pl*

The question of what properties of biological systems allow for efficient evolutionary search in complex fitness landscapes (evolvability) is one of the central interests both for the research in the field of evolutionary biology and artificial life. In this paper we attempt to address this issue by using a model of 3-dimensional multicellular development in which cell fate is determined by differential gene expression in each cell. In our model cells can vary in size and can move freely in 3-dimensional space, affected by forces of adhesion and repulsion. The development relies on an indirect mapping between the genotype and the morphology (the phenotype). Cell differentiation is allowed by positional information provided by diffusible factors. The state of the gene regulatory network coded by the genome determines the cell fate (such as division, death, growth). The genetic elements in our systems define points in  $N$ -dimensional space. The connectivity in the gene regulatory network is determined by the proximity of these points; one can imagine the evolutionary process as their movement in space. Changing the number of dimensions of this space allows to ask directly the questions about the effect of the complexity of the search space on the efficiency of the evolutionary search. Higher dimensionality results in a larger search space, but in our model this search space can still be explored thanks to the action of genetic operators that allow for duplications of genetic elements, a mutational mechanism that allows for regulatory innovations in the network.

*Keywords:* artificial embryogeny; gene regulatory network; evo-devo; complexity of the fitness landscape; neutrality; evolvability.

## 1. Introduction

Artificial embryogeny (AE), a subdiscipline of the field of artificial life (AL), investigates the development of multicellular bodies in simulation. The main concern of AE is to understand how complex systems (phenotypes, such as multicellular bodies) can be represented using compact codes (genotypes) [27].

Complex multicellular organisms have evolved on Earth on several independent occasions. Their lineages led to today's animals, fungi, and several lineages of multicellular eukaryotic algae, one of which led to higher plants. The development of organisms with functionally differentiated cells has several common features [26]. Perhaps most importantly, biological multicellular development is not possible without differential expression

of genes (switching on/off different genes in different cells). This is because each cell (with some exceptions) has the same repertoire of genes (the genome, the genetic material).

The representation of multicellular development in biological genomes is possible thanks to the fact that physical (mechanical, chemical) features of the cells and their components result from the laws of physics. The shape of the embryo is the result of the interplay between the information encoded in the genome and the rules of physicochemical interactions. These interactions shape biological objects at various levels. First of all, they shape the 3-dimensional structure (folding) of the (macro)molecules of which biological objects are composed. 3D structure determines the chemical and mechanical properties of the molecules. At a higher level, interactions between the molecules determine what molecules are synthesised in the cell. In part, this determination is thanks to a subsystem of the cell that is known as the gene regulatory network (GRN). This subsystem consists of two parts. One is the genome, built of one or several long molecules of nucleic acid (DNA for all known cellular organisms). The second is the collection of gene products (biological macromolecules: RNA and proteins). Segments of the genome that code specific products are called “genes” [13]. Gene products are constructed based on the information coded in the genes in a process of gene expression. Gene expression begins with transcription (from the genome to RNA) and may end in translation (from RNA, built of nucleotides, to protein, built of amino acids), protein folding and post-translational processing.

Gene products (RNAs, proteins) can have various functions: structural/mechanical, catalytic or regulatory. Regulatory products interact chemically with each other and with genomic DNA to determine which genes are expressed and thus which products are formed in the cell. These regulatory relations can be represented as a network of regulatory connections between the genes (hence the term “gene regulatory network”). Computations using GRNs have similarities to computations with artificial neural networks, which thus provide a better metaphor for the cellular computation than a Turing machine (with, say, the genome as the tape). The state of the network can be viewed as the basis for the memory about the differentiation state of the cell, inherited at cell division.

Catalytical and structural products (and their concentration) can be viewed as the output of this subsystem. Structural products (and their chemical interactions) determine the shape of the cell and its mechanical properties. Other structural products, often locating physically at the cell boundary (the membrane), determine the adhesion properties of the cells and formation of cell assemblies. The processes that shape the embryo (such as cell differentiation, cell migration, reshaping of multicellular structures) thus depend not only on the state of the gene regulatory networks but also on the physicochemical properties of gene products and the cell itself. In other words, the organism’s phenotype is the joint product of its genotype (the genetic information) and the rules of physics (and chemistry). The emergent phenomena produced by the laws of physics allow for large increases in complexity of phenotypes without great increases in the complexity of genomes, lowering the required dimensionality of the search space [9].

The first step of gene expression (transcription) starts at specific sites (stretches of the nucleotide sequence) in the genome which are called promoters. The term “promoter” may signify in a narrow sense the site directly involved in the start of transcription (where the

enzymes that form the transcript bind). Coding regions may be preceded by many such sites (resulting in transcripts of different length). In a broad sense, “promoter” may mean the whole “promoter region”, containing potential transcription starts intermingled with so-called transcriptional factor binding sites (which bind activatory and inhibitory macromolecules). A still broader term is “cis regulatory region” (“cis” signifies the region must lie on the same molecule that the regulated coding sequences). This term includes sites not necessarily close to the start of transcription: enhancers (if activatory) and silencers (if inhibitory). Such sites may regulate several neighbouring genes.

Transcriptional activity of promoters is a prime determiner of the concentration of functional molecules in the cell, but not the only one. Gene expression is regulated also during the subsequent steps (post-transcriptional and post-translational). This is determined by the affinity of the regulating factors (RNAs, proteins) to the sequences of transcripts and polypeptides (which ultimately correspond to the sequences in the genome). Technically, these are also “cis” regulators from the point of view of genetics (although they are usually not referred to as such in molecular biology). We will use the term “promoter” throughout the paper to refer to (an abstraction of) one of several cis regulatory regions which together regulate one or several coding regions.

Why “one or several”? Clustering of several genes in so-called operons under control of a combination of cis regulatory regions is not, as originally thought, restricted to bacteria, but occurs also in eukaryotes (for a recent review, see [4]). Operons allow for co-regulation (co-transcription) of functionally related coding regions (for example, involved in the same biochemical process). Similar many-to-many relationship between regulatory and functional elements exists when multiple transcripts share common regulatory regions (enhancers and silencers [13]), when large products (e.g. proteins), once formed, are cut into smaller pieces with different functions or when gene products are multifunctional. In particular, many proteins consist of several parts (protein domains) that are functionally and structurally independent, even though they form one macromolecule.

When building biologically-inspired complex artificial systems, it is not necessary to faithfully simulate low-level processes [27], such as those described above. However, less abstract approaches make it easier to design experiments that test particular biological hypotheses. More abstract implementation (such as “grammatical approaches”; see [27]) are more limited in this respect. Such implementations include so-called “cellular encoding” in which a grammar tree is used to encode steps in the development. The instructions in the tree are actions performed by single cells (such as division, cell death, forming connections between the cells). Different cells in the embryo read different parts of the tree at the same developmental step. Such grammar trees can be evolved using the genetic programming methodology [14, 19]. In contrast, “chemical approaches” are less abstract, and attempt to model production and diffusion of regulatory substances in the embryo. These biologically-inspired systems include systems that use artificial gene regulatory networks (aGRNs).

In aGRN systems the behaviour of cells depends on the presence or concentration of products of artificial genes which can have specific functions. Artificial genomes consist of parts that code for products (“genes”) and parts that allow for regulation of the production

of these products.

The models of GRNs published so far differ in (at least) two important aspects. One concerns the fact that some capture only the on/off state of particular genes [24, 25], while other attempt to model varying concentration of gene products [7, 9], often including the process of formation and decay [1, 2, 5, 10, 11, 17, 21]. The last approach gives more justification for the term “chemical approach” [27]). Other differences lie in the method of determination of interaction between system components (between products, between products and regulatory regions). Some systems rely on distances between integers/real numbers e.g. [1, 8, 21], some include an abstraction of the processes of translation and/or folding (such as incrementing the digits in the sequence, e.g. [24]). Some aGRN systems introduce more sophisticated models of these two processes, motivated by biology (such as RNA folding in the work of [11]) or mathematical elegance (such as fractal proteins of Bentley [2]).

The main objective of this paper is to present a model built to investigate the relations between evolution and development in biological systems. Our focus is thus to create a model which incorporates as many features of biological systems relevant to the problem as possible [27]. These features include: determination of cell fate through differential gene expression, allowing for changes of timing and ordering of gene expression (and thus of the events in embryogeny: heterochrony) and allowing for complexification of the aGRN. We design mutational changes so that they do not result in abrupt changes in phenotypes. Our model allows in principle for stochasticity during development, which is important for investigation of canalization and robustness planned in our further work. An important feature of our model is that the cells can assume freely positions in 3D space. Most previous approaches to 3D development rely on putting the cells on a grid [7, 8, 10, 20]. We also provide a way for our aGRN to receive information about cell surroundings, including “mechanical” information (limited so far to information about local cell density). This kind of information is important in biological development [22].

However, high biological realism presents a serious challenge not only in terms of computational resources, but primarily in terms of an increased search space. Higher number of parameters in more realistic models implies search spaces that are much larger, multi-dimensional and rugged. In our model, genetic elements correspond to points in  $N$  dimensional space. Connections in the aGRN depend on distances between these points. Thus the second aim of our paper is to investigate the effect of changing dimensionality of this space. On the one hand, when the number of dimensions is large, the search space is increased. On the other hand, it is easier for the points associated with genetic elements to move ‘neutrally’ (without forming/destroying connections unnecessarily).

The paper is organized as follows. In section 2 the model is described in detail. In Section 3 we investigate which genetic operators affect evolvability when the search space is increased gradually by increasing the size of the set of real numbers associated with each genetic element (‘dimensionality’). Since in principle higher dimensionality corresponds in our model to more possibilities for neutral mutational moves, we ask a question if this has an effect on evolvability. The last section provides a discussion of the results and of the planned future work.

## 2. Our model of 3-dimensional multicellular development controlled by an artificial gene regulatory network

In our model, the development starts from a single cell. Cells die, divide and move in 3D space (no grid is used). Each cell has the same genome, so the topology of the GRN is the same in each cell, but the expression of genes may vary (Fig. 1a; please note that this is an abstract diagram: single molecules are not individually modelled, only their concentrations in each cell are calculated). Cell fate depends on cell history and position in 3D space. Both affect the inputs to the GRN (Fig. 1a): the repertoire of perceived diffusible products (produced by other cells) and perceived external factors, present in the environment (Table 1), for example, information about the nutritional state in the embryo or local cell density (congestion). The GRN's outputs correspond to specific changes in cell fate (Table 2).

In biological systems regulation occurs at different levels. In our model three types of interactions are used: internal product-promoter, morphogen-promoter and morphogen-receptor (Fig. 1a). Direct interaction between morphogens and effectors is not allowed. In other words, diffusible products or factors can influence cell fate only indirectly, modifying the expression of internal products. Differential expression of receptors with affinity to different morphogens allows for directing cellular growth toward or against morphogen sources (Fig. 1c). Morphogens can be produced by the cells of the developing embryo. Additionally, gradients of some morphogens (external factors) are a feature of the environment throughout the development. This is the symmetry breaking mechanism in our model (it allows cell differentiation after the first division), inspired by insect embryogenesis (for a popular introduction, see [6]).

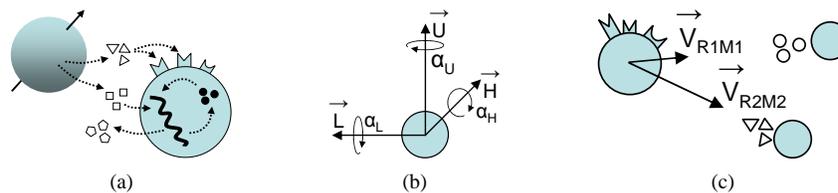


Fig. 1. The cell in our model is a sphere with particular mechanical properties and an associated division vector, which determines the direction where new cell is placed after division (a). Cell fate depends on cell history and position in 3D space, which affect what diffusible products (produced by other cells) and external factors (Table 1) it perceives. Connectivity in the GRN is established using three types of interactions: internal product-promoter, morphogen-promoter and morphogen-receptor. The direction of the division vector is the sum of the internal division vector (b), rotated and multiplied by dedicated effectors (Table 2) and the external division vector, which direction is determined by the gradients of perceived morphogens (c).

### 2.1. The genome: genetic elements and regulatory units

The genome in our model is a list of genetic elements. The genome is parsed to determine which promoters/genes belong to which regulatory units (Fig. 2): the basic building blocks of our aGRN. The size of the genome is not fixed: element deletions and inser-

Table 1. External factors in our model (inputs of the GRN).

External factor	Description
$P_1, P_2, P_3, P_4$	morphogens emitted from 4 fixed locations in space, perceived level depends on the distance and can be used to orient cells in space
“1”	a product that is perceived at the constant level of “1” and can be used as a threshold by the nodes of the network
Time	during development increases linearly from 0 to 1
Generation	increased after each cell division
Congestion	proportional to the number of surrounding cells
Energy	decreases in a cell after division
Total energy	a global value for the embryo, decreases after each division

Table 2. Effectors in our model (outputs of the GRN).

Effector	Description
Divide	if above a preset threshold, the cell divides
Die	if above a preset threshold, the cell dies and is removed from the embryo
Freeze	if above a threshold, the cell stops updating its regulatory network, but remains in the embryo
Change radius	increases the radius of the cell beyond the default value
Division vector	controls how much the division direction is determined by internal cell orientation (vs gradient-based cell orientation)
$\alpha_H, \alpha_U, \alpha_L$	3 effectors controlling the amount by which internal cell orientation is rotated in the daughter cell after division (Fig. 1b)

tions/duplications are allowed during evolution. Each element in the genome is characterized by a set of  $2+N$  real values (Fig. 2).  $N$  numbers define a point in space associated with each genetic element. Its position affects the connectivity of the genetic element in the GRN (Fig. 3). The figure shows how a connection between a gene coding an internal product (G2) and a promoter (a cis regulating element P21) is deduced, but the same method is used to establish all possible connections in our aGRN (Fig. 1). This includes in particular the outputs of the GRN (Table 2), which are treated as if they were promoters permanently linked to products with a specific function (for example, triggering cell division). The inputs (Table 1) are treated exactly the same as the products coded by the genes in the regulatory units, but their expression levels are externally driven.

One additional value associated with each element determines its type (listed in Table 3). The other allows to modify the weight of the connections in the GRN. The connectivity of a given element in the GRN depends on the Euclidean distance of the corresponding point to other points (Fig. 3). One can imagine (or actually visualise if  $N \leq 3$ ) that as genomes evolve, these points approach one another or move away.

The main purpose of modifier fields is to determine if an interaction is activatory or inhibitory (this occurs when the modifiers of two elements have different signs). It means

that a particular gene product can have positive and negative effects on different regulatory units (many biological factors also act as activators or inhibitors depending on context). Additionally,  $|m_k m_i|$  affects the shape of the (hyperbolic) affinity curve (Fig. 3): the function becomes more convex as  $|m_k m_i|$  decreases and is approximately linear for large  $|m_k m_i|$ . While the position of the point associated with an element can be understood as an abstraction of protein or nucleic acid structure, both the distance and the product of modifiers can be understood as an abstraction of the chemistry at the surfaces of the two macromolecules, say a transcriptional factor (a protein) and DNA. Both determine the binding strength between the two and the effect of this binding on binding of the enzymes necessary for transcription. Distance is more related to binding strength and the product to the specificity of binding. A low value of the product of modifiers (shown next to the curves) means that strong weights are possible only at short distance.

The chemical affinity between structures associated to two elements is calculated as follows:

$$w_{k,i} = \begin{cases} d_{k,i} \leq 5 : \text{sgn}(m_k m_i) \frac{2|m_k m_i|(5-d_{k,i})}{10d_{k,i}+|m_k m_i|} ; \\ d_{k,i} > 5 : 0 . \end{cases} \quad (1)$$

where  $d_{k,i}$  is the Euclidean distance between the points  $i$  and  $k$  associated with the elements, while  $m_k$  and  $m_i$  are the values of their modifier fields. Constants are chosen so the affinity is 0 (no interaction) when the distance is larger than 5 and at a maximum (10) when the distance is 0. The maximum distance for interaction is an element of biological realism. Without it, elements far away from each other could interact in principle. Its introduction allows also to simplify the calculations (otherwise all the possible connections in the GRN would exist). The actual value can be viewed as a factor of scale in our system. A maximum value for affinity prevents biologically unrealistic effects. Without it, for a very high affinity connection even very small concentrations of modelled substances could have huge effects.

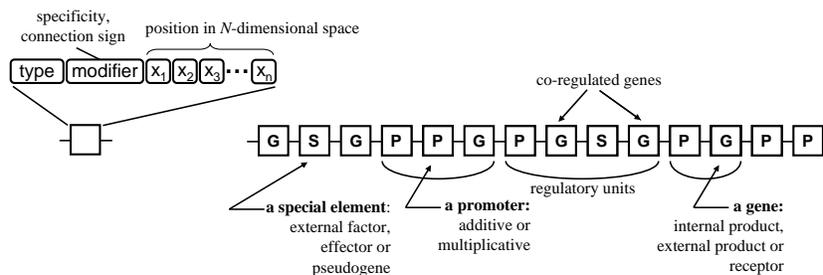


Fig. 2. The genome and the genetic elements. Each genetic element is a collection of  $2+N$  numbers. The first number determines the type of the element (described in detail in Table 3). There are three basic classes of genetic elements: promoters, genes and special elements. The promoters and genes define regulatory units: one or more promoter followed by one or more gene. Special elements are ignored at this parsing step, but effectors and external factors are assigned to specific inputs/outputs based on their order in the genome.

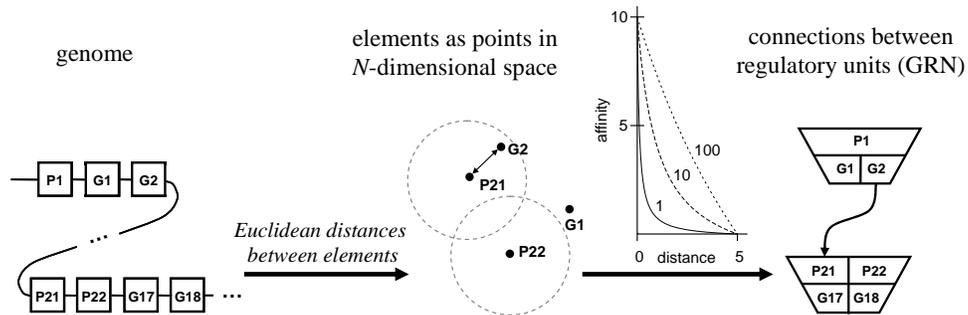


Fig. 3. Coding the gene regulatory network (GRN) in the linear genome. The genome is the collection of genetic elements, each of which is associated with a  $N$ -dimensional point in space. Distances between points define which elements in the GRN are connected. The weight of the connection between two elements depends not only on distance, but also on the product of modifiers associated with each other (Eq. 1). The connection is inhibitory when this product is negative.

Table 3. The types of genetic elements defined in our model and the associated structural features of the cell. Promoters and genes define regulatory units, special genes are ignored in this process.

Class	Type	Description
Promoters	Additive,	Inputs of regulatory units, their effects sum up or multiply, respectively (see Eq. 3)
	Multiplicative	
Genes	Internal product	Binds to promoters inside the same cell
	External product	Diffuses from the cell and binds to promoters and receptors, possibly of other cells
	Cell receptor	Orients cell division according to the gradient of external product(s)
Special	External factor	Input for the GRN (signals from the environment)
	Effector	Output of the GRN (actions that cells can perform)
	Pseudogene	Ignored, but mutation in the type field may turn it (back) into other element, maintaining the values of other fields

## 2.2. Synthesis and degradation of products

To calculate the activation level of a regulatory unit, we first compute the activity of all its promoters:

$$p_i = \sum_{k=1}^K L_k w_{k,i} . \quad (2)$$

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 chemical affinity, corresponding for example to the chemical affinity of a transcription factor to a stretch of DNA sequence (a promoter understood as a cis regulatory region). The rate of synthesis of the set of  $\Omega$  products coded in a regulatory

unit is established as follows:

$$\frac{dL_{\Omega}}{dt} = a_c f'_A \left( \prod_{i=0}^I p_{m,i} \sum_{j=1}^J p_{a,j} \right) - d_c L_{\Omega}. \quad (3)$$

where  $I$  and  $J$  denote the number of multiplicative and additive promoters (respectively), while  $p_{m,1..i}$  and  $p_{a,1..j}$  describe their activations (Eq. 2). The presence of a multiplicative promoter in a regulatory unit results in a strict requirement for the presence of a product with an affinity to it, otherwise the unit is not expressed. ‘‘All-or-nothing’’ regulation is quite common in biological systems and is difficult to incorporate when only additive units are used (such as in classical perceptron neural networks). The actual indexing of the promoters starts from 1,  $p_{m,0}$  is reserved for the identity element of multiplication ( $p_{m,0} = 1$ ).  $a_c$  and  $d_c$  are synthesis and degradation constants, respectively, both set to 1 for the experiments described in this work. This equation can be understood as describing the synthesis/degradation of a transcript coding for several proteins or for a multifunctional protein. In our previous work [15, 16], in contrast, gene expression was calculated in a less realistic manner, in the same way the activation levels in a perceptron-like artificial neural network, allowing in principle the gene expressions to change between zero and maximum value instantaneously.  $f'_A$  in Eq. 3 is a sigmoidal function with range  $(-1, 1)$ :

$$f'_A(x) = \frac{2}{1 + e^{-\omega(x-1)}}. \quad (4)$$

where  $\omega$  describes the steepness of the sigmoid and does not change during the evolution or development ( $\omega = 10$  was used). A sigmoid shape of the function relating gene expression to the expression of regulatory proteins is a common feature in biological genetic switches, reflecting e.g. cooperativity of binding of proteins to DNA.

### 2.3. Modelling diffusion

Diffusible substances in our model include external products and external factors, coded by different genetic elements (Table 3). Internal and external products affect the regulated regulatory units in exactly the same way, but only external products can bind to promoters in other cells than the one that produces them. In addition, external products and diffusible external factors can interact with receptors. Both diffusible factors and diffusible (external) products are called morphogens, the difference between them is that the perceived levels of the factors do not ultimately depend on the production by some cell and are provided as a part of the environment of the developing embryo. All external factors (diffusible and not) can interact with promoters. Four of the factors are emitted from four specific points in 3-dimensional space in which the embryo develops (Table 1). This is the symmetry-breaking mechanism in our model. Four points are used because this is, in principle, the minimal number of signal sources necessary to perform self-localization in 3D space (e.g. via 3D trilateration).

The interactions between the receptors and morphogens influence the vector of cell division by shifting it toward or away from the morphogen source (Fig. 1). Since the cells

may differ in the pattern of gene expression, also their set of active receptors may be different, allowing each cell to orient its division in relation to the pattern of morphogens that are available at a given moment. The affinity between morphogens and promoters or receptors is defined by Eq. 1. In addition, we simulate simple diffusion: the perceived concentration of the morphogen in a given cell depends on the distance to the source and the level of the morphogen in the source (level of production). The level of morphogen  $m$  perceived by the cell  $c$  is:

$$L_{c,m} = \sum_{i=1, i \neq c}^I l_{i,m} \frac{1}{1 + D_{i,c}}. \quad (5)$$

where  $I$  denotes the number of cells of a developing embryo (potential producers),  $D_{i,c}$  is the distance from the cell  $c$  to a source cell  $i$  in the 3D space of the developing organism, and  $l_{i,m}$  is the level of morphogen  $m$  this source cell produces. The actual value of  $l_{i,m}$  is delayed in time, depending on the distance from the source. This allows to simulate some of spatiotemporal effects of diffusion.

#### 2.4. Multicellular development

Daughter cells after division inherit the state of the GRN (the concentration of products) from the mother, but the concentration of the effector of division is reset to zero. Daughter cells are placed at a small distance from the mother cells initially (indeed, they overlap). Throughout the development, cells that overlap repulse each other, and cells closer than a specified distance will also experience an adhesive force, which is necessary to maintain the integrity of the cellular structure. The dynamics of cell movement is simulated with simple Newtonian physics, using Runge-Kutta 4th order integration. This is an extension of our previous work in which cells were simply placed at a specific distance after division [16], and possibly then pushed each other away, remaining connected with springs [15].

The direction in which new cell is placed is determined by the division vector (Fig. 1b), which is controlled by two mutually supporting mechanisms. The evolutionary process can take advantage of any or both mechanisms to shape the development, and a dedicated effector regulates their relative importance (Table 2). Each mechanism uses one of two auxiliary vectors that are maintained for each cell. One is the internal division vector inherited from the mother (together with two associated perpendicular vectors) and possibly rotated at division (Fig. 1b). The other is the external division vector (Fig. 1c), named so because it is oriented towards or away from the sources of external products and diffusible external factors.

Internal division vector draws from the mechanism used in 3D L-systems (see e.g. [23] for a detailed explanation). Its direction is inherited from the parent cell and rotated depending on the expression of three angular effectors (Table 2). A somewhat similar, yet simpler concept (limited to 12 possible directions) was earlier used for 3D embryogenesis by Kumar [20]. The length of the internal vector depends on the activation of a separate designated effector, so if products binding to it are expressed at a low level, the direction is determined in larger extent by the external division vector, which relies on the morphogen-receptor interactions.

The external division vector is determined by a sum of interactions of all receptors in the given cell with all morphogens produced by every source:

$$\vec{V}_c = \sum_{r=1}^R \sum_{m=1}^M \sum_{s=1, s \neq c}^S l_r w_{r,m} L_{c,m} \vec{\delta}_{s,c}. \quad (6)$$

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 sources for positional external factors),  $l_r$  is the expression level of the receptor  $r$  in the cell (established using Eq. 3),  $w_{r,m}$  is the morphogen-receptor affinity (Eq. 1),  $L_{c,m}$  is the perceived level of the morphogen, and  $\vec{\delta}_{s,c}$  is the normalised vector from the given cell to the source.

Non-diffusible external factors (Table 1) that give historic/temporal (time, generation counter, energy level) and “mechanical” (congestion) information provide a way to control cell divisions. In particular, each division has some energetic cost. Cell energy can be exhausted by rapid divisions. There is also a limit on the total energy that can be used during the development of the whole individual. Uncontrolled cell divisions cause the energy to be exhausted at early developmental stages, allowing to terminate costly simulation of unfit individuals.

## 2.5. Genetic algorithm and the fitness function

We use a genetic algorithm in which a new generation is formed by copying 5 genomes without mutation (elitism), 245 with mutation, and 50 with mutations and multi-point crossover between genomes of different sizes (when the recombination operator was enabled). Tournament selection is used to choose candidate genomes for the next generation. Elite genomes can wander through the neutral regions in the search space (which may allow for a more efficient evolutionary search [12, 25]): after a neutral mutation an elite genome always overwrites the previous version.

Initial generation was formed from a population of randomly created individuals. Each was created by inserting elements representing all possible effectors and external factors at the beginning of the genome and 6 regulatory units with random size are formed by sampling the number of regulatory elements from a normal distribution  $N(3, 3)$ , and the number of gene products from  $N(2, 2)$ . The probability of choosing an additive promoter was 2.5 times more than for choosing a multiplicative promoter, the ratio for products was 3:3:1 (internal:external:receptor). The values of modifiers were initially set at 10 with the sign determined randomly. In order to investigate the effects of changing the number of coordinates used in establishing the position of points corresponding to genetic elements (which affects the complexity of the search space) only the first coordinate was drawn from  $N(0, 20)$  for the genomes in the initial populations. If the dimensionality was higher than 1, the other coordinates were set to zero. In other words, points corresponding to all genetic elements in the genome initially lie on one line.

The genetic operators are designed to work on the level of genetic elements rather than single bits or real values, because a genetic element is the basic unit of heredity in

our model. Each operator (see Table 4) had a predefined probability of occurrence per element in a genome and per generation. In this work the genetic algorithm is used as

Table 4. Genetic operators used in the experiments and their symbols.

Range	Operator	Description	
Part of element	⊗	Position change	moves the element-associated point in a random direction by a distance drawn from $N(0, 10)$ or $N(0, 0.1)$
	⊗, ⊙	Type change	modifies the element type; in some experiments only changes to and from pseudogenes (pseudogenization) were allowed (⊙)
	⊙	Modifier value change	adds a value drawn from $N(0, 1)$ to the modifier field; this affects the specificity of the interaction
	⊖	Modifier sign change	switches all the interactions of this element from inhibition to activation or vice versa
Element	□	Insertion	creates a new random element
	△	Deletion	removes a single element
	▽	Duplication	inserts a copy of the element after the original
Whole genome	▲	Large deletion	removes a group of genetic elements, starting and ending element is chosen randomly
	▼	Large duplication	duplicates a genome fragment at a random position, the starting and ending element is chosen randomly
	⋈	Recombination	multipoint crossover between two individuals

an imperfect model of biological evolution, with fitness determined by the similarity of evolved morphology of artificial multicellular organisms to a target (Fig. 4). A voxel-based approach was used to determine the level of similarity: cubical voxels are marked as either inside or outside the shape and voxels occupied by cells are counted [15]. In this way, no fitness advantage is provided by densely packing overlapping cells.

The morphological process was allowed to last for 300 iterations to give time for the forces of repulsion and adhesion to act. However, the fitness was set to zero if the cell number changed in the last 50 iterations. It was also set to zero if the number of cells passed the limit of 100 or the embryo exhausted its energy. In other words, the embryogenesis was required to self-terminate and maintain this state. In contrast, in our previous experiments [16] embryogenesis was stopped after 50 iterations or when the number of cells reached 250, and most embryos reached either of the limits. Such limits are conceptually equivalent to the existence of some required external factor whose depletion stops the development. Self-termination of the development is a more difficult task as it not only requires the embryo to develop the correct morphology, but also to switch regulatory network of each cell into the stable state that will prevent further divisions.

### 3. Results

In our model each genetic element determines a point in  $N$ -dimensional space. As the genomes evolve, these points approach and move away from each other. It is natural to

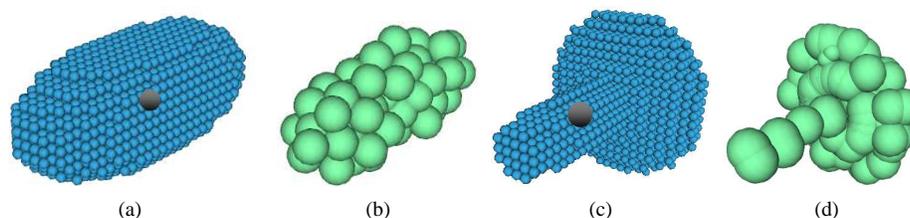


Fig. 4. Evolution of multicellular embryos in 3D. Panel (a) shows the target shape and (b) the best-fit individual obtained in the experiments described in this paper. The grey sphere indicates the location (inside the target shape) where the initial cell from which development starts is placed. The target and the result of experiments with the previously described version of the model [16] are shown for comparison (c,d).

ask if the dimensionality of this space affects the ability to evolve target shapes (evolvability). Mutations affecting the position of points associated with each element in our experiments did not affect single coordinates: all coordinates change when a vector with random length and direction in the  $N$ -dimensional space is added (Table 4). The actual mutational mechanism ensures that the average disturbance of the position does not depend on the dimensionality and no particular direction is given preference. It also allows for the whole  $N$ -dimensional space to be accessible with a single mutation, although smaller moves have higher probability. This operator is a metaphor of point mutations or short deletions/insertions in the coding or regulatory sequences in biological genomes. Changes in these sequence affect, respectively, the structure of a gene product or the local structure of the nucleic acid from which the genome is built.

In principle, one could expect that higher dimensionality will have a negative effect on evolvability as it increases the size of the search space. Still, there are reasons to expect beneficial effects of higher dimensionality. More dimensions allow for more neutrality in the mutational process. Neutrality can be beneficial as populations stuck in a local sub-optimum may still wander around in neutral directions in the fitness landscape searching for paths that allow for increase in fitness [12, 25]. The exact impact of the neutrality on evolvability is subject of intensive research and multiple studies confirm both positive and negative effects (see [12] for a recent review).

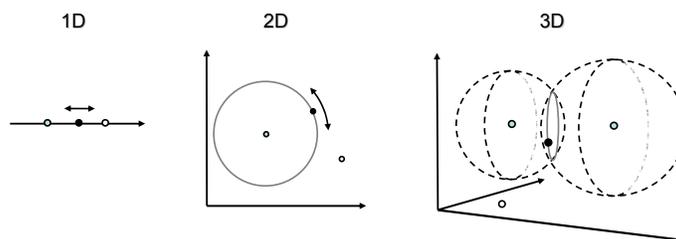


Fig. 5. The effect of dimensionality on the neutrality of changing position of points associated with genetic elements. As the dimensionality increases, there are more trajectories that let the points remain equidistant to other points after mutation.

It is easy to visualise the way in which the dimensionality could affect neutrality. In 1 dimension, it is impossible for a point corresponding to a genetic element to move closer to some other point using mutational changes in a single coordinate without moving away from the points laying in the other direction (Fig. 5). As the dimensionality increases, there are more possible trajectories that allow approaching some points while remaining equidistant to the other (Fig. 5).

In this paper, we investigated the effect of increasing the dimensionality by challenging the genetic algorithm with the task of self-terminating embryogenesis with a symmetrical shape as a target (Fig. 4). Fig. 6 shows the results. As the number of dimensions increases, so does the search space, but we hoped to observe some positive effect of dimensionality (perhaps an increase of evolvability followed by a decrease). No such effect was observed in our experiments. We did observe detrimental effects of the increased search space on evolvability, but only when the repertoire of genetic operators was limited to just one: position change.

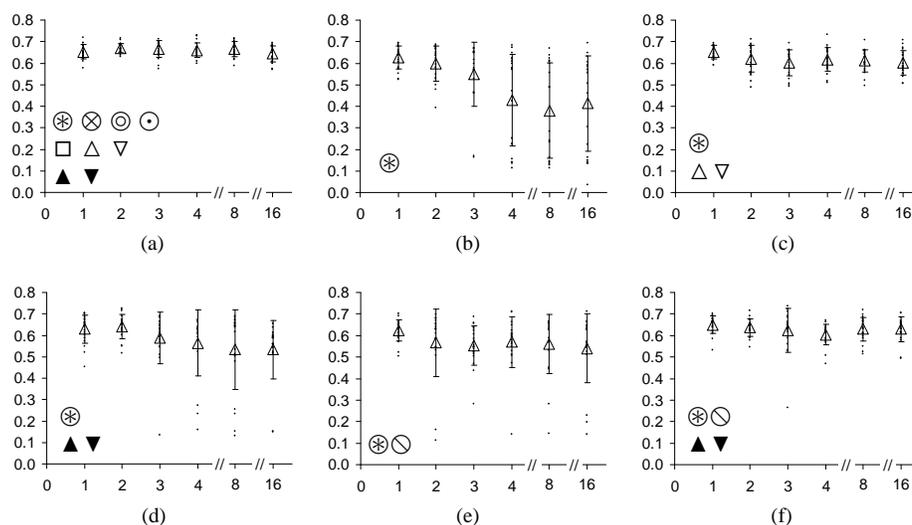


Fig. 6. The efficiency of the evolutionary search. The fitness of the highest-scoring individuals in each of the 20 independent runs of the genetic algorithm after 250 generations is shown for different complexity of the search space (number of dimensions), with their means (triangles) and standard deviation (bars). Different panels show the results obtained with different combinations of genetic operators, using the symbols in Table 4): all operators without recombination (a), only position change (b) or position change with element duplication/deletion (c), large duplication/deletions (d), pseudogenization (e) or with pseudogenization together with large duplications/deletions.

In 1 dimension evolution with only position change resulted in similar fitness for all 20 runs (average: 0.63, standard deviation: 0.05; Fig. 6b) to the fitness reached when all the genetic operators were permitted excluding recombination (av. 0.65, st. dev. 0.05; Fig. 6a). However, in 16 dimensions a considerable decrease in average fitness and increase in variance was observed (position change only: av. 0.41, st. dev. 0.22; all operators without re-

combination: av. 0.65, st. dev. 0.03). When also the inversion of the modifier sign was permitted during evolution, the detrimental effect of high dimensionality was less pronounced, but still observable (for 16 dimensions: av. 0.54, st. dev. 0.15; not shown).

It could also be observed (for 16 dimensions: av. 0.54, st. dev. 0.16; Fig. 6e) when mutations could change position and element type, but only to pseudogene and back to the original type (this is the mechanism of pseudogenization used in all our experiments). On the other hand, when all element type changes were added to position changes, the effects of high dimensionality disappeared (for 16 dimensions: av. 0.60, st. dev. 0.05; not shown). They also disappeared when element duplication and deletion was permitted in addition to position change (for 16 dimensions: av. 0.60, st. dev. 0.06; Fig. 6c). However, when large duplications and deletions were allowed instead of element duplication/deletion, some detrimental effect persisted (for 16 dimensions: av. 0.54, st. dev. 0.16; Fig. 6d); it disappeared when pseudogenization was added to the combination of operators (so position change, large duplication/deletions and pseudogenization was possible; for 16 dimensions: av. 0.60, st. dev. 0.06; Fig. 6f). Also, when recombination (unequal crossing-over) was added as a second operator to position change, the results (for 16 dimensions: av. 0.60, st. dev. 0.06; not shown) were intermediate between single and large deletions/duplications.

For 2 dimensions, it is possible to visualise the trajectories of points corresponding to individual genetic elements during evolution (Fig. 7). In this visualisation lines start at the positions coded by the elements of the best genome at the end of a run and then go backwards in time (generations) towards the original position coded in one of the genomes in the initial population. Each segment of each line corresponds to a single mutational step (a change in position accepted by the evolutionary process).

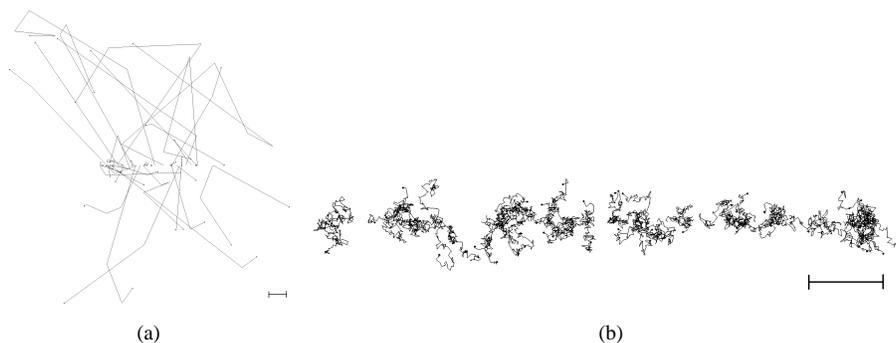


Fig. 7. The actual trajectories of points corresponding to genetic elements in 6 regulatory units in a genetic regulatory network which allows for the development of an ellipsoid shape (Fig. 4b with the self-controlled termination of development as an objective). Panel (a) corresponds to large mutational steps with small frequency (mutation rate per element: 0.01), panel (b) shows the results for 100-times lower mutational steps with all the elements changing their coordinates in each generation (mutation rate per element: 1). The scale bars correspond to the maximum distance of interaction (Eq. 1).

When mutations affects the position of points in  $N$ -dimensional moving them over large random distances, mutations that are accepted by the evolutionary process correspond to

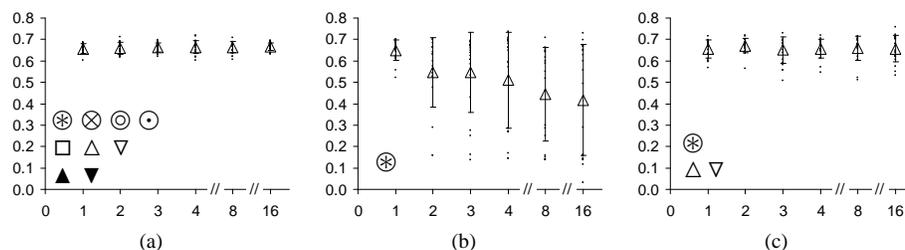


Fig. 8. The efficiency of the evolutionary search with 100 times smaller step than in the experiments shown in Fig. 6. The fitness of the highest-scoring individuals in each of the 20 independent runs of the genetic algorithm after 250 generations is shown for different complexity of the search space (number of dimensions), with their means (triangles) and standard deviation (bars). Panel (a) corresponds to all genetic operators, (b) to coordinate mutations only, and (c) to coordinate mutations and single element duplication/deletion.

very large changes in position (Fig. 7a), very often larger than the maximum distance of interaction (Eq. 1). Yet the argument illustrated in Fig. 5 is only applicable to a situation when the changes in point positions are small. Small movements require high probability of mutation in order for the genomes to evolve. We have thus run three experiments in which the mutational step (the distance of point movement in a random direction) was 100 times lower than in the experiments shown in Fig. 6, but the probability of changing the coordinates of each element in each generation was set to 1 (100 times higher). Such mutation rate may seem high, but lower rates (e.g. 0.1) after the same number of generations were found to yield solutions with much lower fitness.

All the same, the results obtained for small mutational steps and high rate of mutation were similar to the results for large mutational steps and low mutational rate. When all operators without recombination were used (Fig. 8a), no detrimental effect of high dimensionality was observed, but it was present in the experiment in which only position could be changed (Fig. 8b), and ameliorated by the addition of duplications and deletions of single genetic elements (Fig. 8c). The averages for 20 runs and 16 dimensions were, respectively: 0.67 (st. dev. 0.02), 0.42 (st. dev. 0.26) and 0.66 (st. dev. 0.06). When the changes in point position are visualised for the experiment in which small mutational steps occur for each element in each generation, the effect on the trajectories can be easily observed (Fig. 7b). Interestingly, the lines in Fig. 7b form 6 clusters corresponding to 6 regulatory units in the gene regulatory network (Fig. 9b). This may stem from the fact that the networks obtained in this way are much more densely connected than the networks obtained through the evolutionary process with 100 times larger mutational steps and 100 times lower mutational rate (Fig. 9a).

#### 4. Discussion

The main purpose of our work is to build a biologically realistic model of artificial embryogeny that can be easily adapted to ask specific questions in the field of theoretical biology, questions concerning evolution of gene regulatory networks (such as GRN modularity) and evolution of form and of functions related to form (embryo patterning, symmetry, brain-

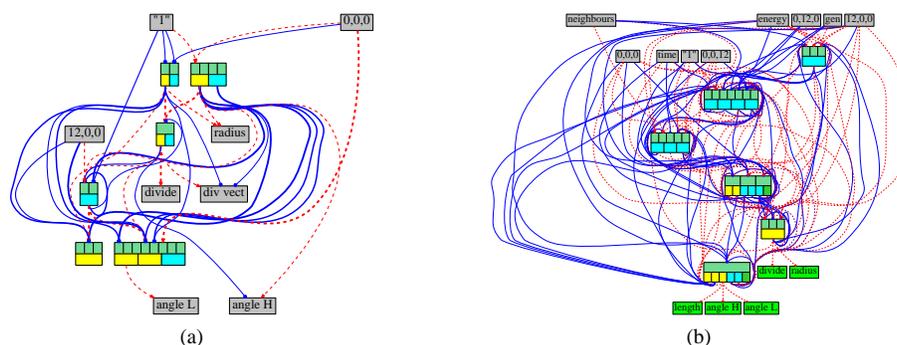


Fig. 9. The gene regulatory network with 6 regulatory units formed by a mutational process with large mutational steps (random distance drawn from  $N(0, 10)$  with small frequency (mutation rate per element: 0.01), panel (b) shows the results for 100-times lower mutational steps with all the elements changing their coordinates in each generation (mutation rate per element: 1).

body co-evolution). This is why we have consciously taken the decision to directly test our model in a task that involves morphogenesis (rather than, e.g. the ability to reach a specific gene expression pattern in a single cell, as in [1, 18]). The implementation of our model is modular. For example, different ways to define chemical interactions in the system are possible, although so far only a representation of each genetic element as a collection of real numbers was explored. This representation was designed in a way that permits to graphically visualize the mutational steps during evolution as changes in the position of points that correspond to genetic elements.

In this and previous papers [15, 16] we show that our model allows for the development of complex shapes, visually similar to the shapes that can be grown using the model presented by Kumar [20] and Eggenberger [7–10]. Our design of the GRN shares some similarities with Eggenberger’s design, but several important differences exist between our model and previous approaches. First of all, in our model cells divide freely in 3D space and gene expression can influence cell size. This is an important development, and we hope it will allow for further progress in modelling biologically realistic embryogenesis. Secondly, we allow for many-to-many relationship between *cis* regulating regions (which we call promoters) and the coding regions (genes), that is, the same combination of promoters regulates a set number of genes. Previously presented models of artificial embryogeny (e.g. [3, 7–10, 17, 20, 28]) and artificial GRNs [1, 11, 18, 21, 24] follow a many-to-one scheme of multiple regulatory regions (promoters) and one gene/product.

In terms of engineering, simple models have a clear advantage of being more efficient computationally. But simple models have drawbacks when one is interested in asking questions about the role of specific processes in biology. If cell migration is not included in the model, its role in the multicellular development cannot be investigated. If the fixed length genome is a feature of the model, it makes it impossible to address, say, the role of gene duplication/deletion in the emergence of modularity in the GRN and development. Allowing for a many-to-many relationship between the promoters and genes in our system will hopefully allow to ask questions about the importance of co-regulation. To repeat, our mo-

tivation is to build a simulation platform that could be used in research on many biological phenomena without the need to drastically change the model to address a specific question. On the other hand, the process of model building has to be, to some extent, incremental, slowly pushing the limits of what is possible to simulate.

This paper introduces three important extensions of our previous work [15, 16]. The first is a chemically realistic determination of product concentration (production vs. decay, this was previously modelled in [1, 2, 5, 10, 11, 17, 21]). In our previous models, activation or inhibition of a regulatory unit resulted in setting the product level directly to a different value (a similar approach was used e.g. by Eggenberger [7, 9]). The second new feature is that the development of the embryo is now shaped not only by differential cell size, but also by forces of adhesion and repulsion between the cells (in a more realistic manner than in our previous work [15]). The third modification lies in establishing a more difficult morphological task: in the experiments presented in this paper the embryos were required to self-terminate the development.

To the extent tested so far, these changes affected negatively the efficiency of evolutionary exploration when it comes to development of a predefined target shape. The fitness values obtained for morphologies more complex than ellipsoids were usually lower compared to those achieved with simpler versions of the model (Fig. 4 and [15, 16]).

The search for optimal conditions involves careful investigation of what elements of the model are necessary for maintaining evolvability. The current design is, in many ways, redundant. Duplication of elements may occur through the duplication of single genetic elements, several of them or through recombination. A single genetic element may become a pseudogene directly, through a change of gene type, or indirectly. A simple example of indirect pseudogenization would be a mutation in a regulatory region to include a multiplicative promoter with which no product coded by the genome interacts. However, explicit definition of pseudogenes makes it easier to ask questions about the effects of pseudogen formation on evolvability.

In the experiments reported here the difficulty of the task could be varied gradually. This is thanks to the element of our model which was originally designed to allow for better graphical visualisation of the evolutionary process: in our model each element corresponds to a point in  $N$ -dimensional space. The number of dimensions can be set arbitrarily high (higher number of coordinates cannot be visualised and require more computational resources, both in terms of memory and time; we have reached up to 128 dimensions in our experiments). Increased dimensionality of this space results in an increased search space. On the other hand, higher dimensionality allows for more neutrality. This is because in our system  $N$  coordinates coded in a genetic element are not mutated separately. Instead, a mutational step moves the corresponding point in a random direction at a random distance (in other words, one step affects all the coordinates). Higher dimensions allow that when a point A moves closer to point B, its distance to point C may not be affected.

The results of our experiments lead us to conclude that the possibility of such neutral moves does not increase evolvability in our system. The only effect of high dimensionality that could be observed was detrimental, but it was present only when the genetic algorithm was crippled and mutational changes were restricted to position moves. The effect consisted

of lowering the average fitness of the best individual in 20 experiments and higher variation (Fig. 6 and 8), meaning that the genetic algorithm more often got stuck in local sub-optima. When any of the possible genetic operators that allow for duplications of genetic elements was allowed, the detrimental effect was greatly reduced, with the largest reduction in the experiment in which deletion/duplication of single elements was allowed.

Presence of a genetic operator that allowed the type of an element to change also increased the efficiency of the evolutionary search. Not all such changes are biologically relevant (for example, a change from a promoter to a gene coding for a product). Others are more so. A change from an additive to a multiplicative promoter can be viewed as a change that results in a strict requirement for a transcription factor. A change from an external product to an internal product or vice versa corresponds to a removal (or addition) of a protein motif that causes the protein to be exported from the cell. Perhaps the most interesting is a change from an element to a pseudogene and vice versa. The availability of this move in the search space did not result in an improvement on par with the introduction of single element duplications/deletions. This agrees with our knowledge of biological evolution. The current biological view is that new functions arise by first duplication, which allows for neutral mutations, a process helped by pseudogenization of one of the copies. This is followed by subsequent recovery of gene expression and further functionalization. It is the duplication that is more central here. When an element is (perhaps temporarily) silenced by mutation without being removed from the genome (“pseudogenized”), it is removed from the GRN, with possibly detrimental effects on fitness. In contrast, when a previously duplicated element becomes pseudogenized, it is more likely that the move is neutral. This is why the evolvability of a system with pseudogenization can be still improved by addition of duplication/deletion of large genome fragments (and vice versa, Fig. 6def).

Recombination between genomes is commonly used to allow populations to escape from local maxima. In other words, the expected effects of allowing recombination should go in the same direction as the increased neutrality. This is indeed what was observed in our experiments (not shown). Further experiments are needed to find out if this is the result of forming a common gene pool or rather stems directly from the fact that unequal crossing-over may result in large duplications (and deletions), and if so, if such duplications lead to the emergence of modularity in the network.

Other mutational events, such as the change of the modifier sign or its value, when present in combination with position change, but without the mechanism leading to gene duplication, did not help much in the evolutionary search when the number of coordinates was high (not shown).

Galván-López and Poli [12] note that the controversy on whether neutrality helps or hinders evolutionary search may stem from the considerable variability in the problems and the lack of a common view on how neutrality can be increased. This makes the comparison of results obtained with different models rather tricky. It is even more tricky to compare the results obtained with artificial life models to that obtained using biological systems. We believe that our model allows for that, but also that these issues should be investigated in a setting more biologically realistic than a genetic algorithm. It would be much more realistic to create a system in which the evolution is fuelled by competition for

limited resources in a simulated world with a spatial structure. In biological evolution, particular morphologies are not a goal in themselves, they only indirectly influence the ability to produce progeny. An artificial life platform would permit the investigation of morphological innovations more relevant to biology. We hope that our model can be used in such a platform to experimentally test hypotheses concerning, for example, the robustness of the gene regulatory networks, the issue of epistasis, and the statistical properties of the evolved GRNs. These are some of the problems we plan to address in future work.

## 5. Acknowledgements

The computational resources used in this work were obtained through the support of the Polish Ministry of Science and Education (project N303 291234 and N N519 384236), the Tri-city Academic Computer Centre (TASK) and the Interdisciplinary Centre for Molecular and Mathematical Modelling (ICM, University of Warsaw; project G33-8).

## References

- [1] Banzhaf, W., On the dynamics of an artificial regulatory network, in *Proceedings of the 7th European Conference (ECAL 2003), Dortmund, Germany, September 14-17, 2003*, Lecture Notes in Artificial Intelligence (Springer Berlin / Heidelberg, 2003), pp. 217–227.
- [2] Bentley, P. J., Fractal proteins, *Genetic Programming and Evolvable Machines* (2004) 71–101.
- [3] Beurier, G., Michel, F., and Ferber, J., A morphogenesis model for multiagent embryogeny, in *Proceedings of the ALife X: The Tenth International Conference on the Simulation and Synthesis of Living Systems* (MIT Press, Cambridge, MA, 2006), pp. 84–90.
- [4] Blumenthal, T., Operons in eukaryotes, *Briefings in Functional Genomics and Proteomics* **3** (2004) 199–211.
- [5] Bongard, J., Evolving modular genetic regulatory networks, in *Proceedings of the 2002 Congress on Evolutionary Computation, CEC '02*, Vol. 2 (2002), pp. 1872–1877.
- [6] Carroll, S., *Endless Forms Most Beautiful: The New Science Of Evo Devo And The Making Of The Animal Kingdom* (WW Norton & Company, 2005).
- [7] Eggenberger, P., Evolving morphologies of simulated 3D organisms based on differential gene expression, in *Proceedings of the Fourth European Conference on Artificial Life* (MIT Press, 1997), pp. 205–213.
- [8] Eggenberger, P., Exploring regenerative mechanisms found in flatworms by artificial evolutionary techniques using genetic regulatory networks, in *The Congress on Evolutionary Computation, CEC '03*, Vol. 3 (2003), pp. 2026–2033.
- [9] Eggenberger, P., 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*, Vol. 1 (2003), pp. 191–198.
- [10] Eggenberger, P., 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* (MIT Press, Cambridge, MA, 2004), pp. 387–393.
- [11] Flamm, C., Endler, L., Müller, S., Widder, S., and Schuster, P., A minimal and self-consistent in silico cell model based on macromolecular interactions, *Philosophical Transactions of the Royal Society B: Biological Sciences* **362** (2007) 1831–1839.
- [12] Galván-López, E. and Poli, R., An empirical investigation of how and why neutrality affects evolutionary search, in *GECCO '06: Proceedings of the 8th annual conference on Genetic and Evolutionary Computation* (ACM, New York, NY, 2006), pp. 1149–1156.

- [13] Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., Emanuelsson, O., Zhang, Z. D., Weissman, S., and Snyder, M., What is a gene, post-ENCODE? History and updated definition., *Genome Research* **17** (2007) 669–681.
- [14] Gruau, F., Whitley, D., and Pyeatt, L., A comparison between cellular encoding and direct encoding for genetic neural networks, in *Genetic Programming 1996: Proceedings of the First Annual Conference*, eds. Koza, J. R., Goldberg, D. E., Fogel, D. B., and Riolo, R. L. (MIT Press, Stanford University, CA, USA, 1996), pp. 81–89.
- [15] Joachimczak, M. and Wróbel, B., Evo-devo *in silico*: a model of a gene network regulating multicellular development in 3D space with artificial physics, in *Artificial Life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems*, eds. Bullock, S., Noble, J., Watson, R., and Bedau, M. A. (MIT Press, Cambridge, MA, 2008), pp. 297–304.
- [16] Joachimczak, M. and Wróbel, B., Evolution of 3D development controlled by a gene regulatory network: The complexity of the search space and evolvability, in *Proceedings of the 8th German Workshop on Artificial Life 2008 (GWAL-8)* (2008).
- [17] Knabe, J. F., Nehaniv, C. L., and Schilstra, M. J., Evolution and morphogenesis of differentiated multicellular organisms: autonomously generated diffusion gradients for positional information, in *Artificial Life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems*, eds. Bullock, S., Noble, J., Watson, R., and Bedau, M. A. (MIT Press, Cambridge, MA, 2008), pp. 321–328.
- [18] Knabe, J. F., Nehaniv, C. L., Schilstra, M. J., and Quick, T., Evolving biological clocks using genetic regulatory networks, in *Artificial Life X: Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems*, eds. Rocha, L. M., Yaeger, L. S., Bedau, M. A., Floreano, D., Goldstone, R. L., and Vespignani, A. (MIT Press/Bradford Books, 2006), pp. 15–21.
- [19] Komosinski, M. and Rotaru-Varga, A., Comparison of different genotype encodings for simulated three-dimensional agents, *Artificial Life* **7** (2002) 395–418.
- [20] Kumar, S., *Investigating Computational Models of Development for the Construction of Shape and Form*, Ph.D. thesis, Ph. D. Thesis, Department of Computer Science, University College London (2004).
- [21] Kuo, D. P., Leier, A., and Banzhaf, W., Evolving dynamics in an artificial regulatory network model, in *Parallel Problem Solving from Nature - PPSN VIII, Lecture Notes in Computer Science*, Vol. 3242 (Springer Berlin / Heidelberg, 2004), pp. 571–580.
- [22] Mulder, B., Developmental biology: On growth and force, *Science* **322** (2008) 1643–1644.
- [23] Prusinkiewicz, P. and Lindenmayer, A., *The algorithmic beauty of plants* (Springer-Verlag, New York, NY, 1996).
- [24] Reil, T., Dynamics of gene expression in an artificial genome - implications for biological and artificial ontogeny, in *Proceedings of the 5th European Conference on Artificial Life, Lecture Notes In Computer Science*, Vol. 1674 (Springer-Verlag, London, UK, 1999), pp. 457–466.
- [25] Shipman, R., Shackleton, M., and Harvey, I., The Use of Neutral Genotype-Phenotype Mappings for Improved Evolutionary Search, *BT Technology Journal* **18** (2000) 103–111.
- [26] Smith, J. and Szathmáry, E., *The Major Transitions in Evolution* (WH Freeman & Co Ltd, 1995).
- [27] Stanley, K. and Miikkulainen, R., A taxonomy for artificial embryogeny, *Artificial Life* **9** (2003) 93–130.
- [28] Steiner, T., Olhofer, M., and Sendhoff, B., 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* (MIT Press, Cambridge, MA, 2006), pp. 70–76.