Towards shape and structure optimization with evolutionary development

Till Steiner, Markus Olhofer, Bernhard Sendhoff

2006

Preprint:

This is an accepted article published in Proceedings of the ALife X: The Tenth International Conference on the Simulation and Synthesis of Living Systems. The final authenticated version is available online at: https://doi.org/[DOI not available]

Towards shape and structure optimization with evolutionary development

Till Steiner, Markus Olhofer and Bernhard Sendhoff
Honda Research Institute GmbH, Offenbach/Main, Germany
{markus.olhofer,bs}@honda-ri.de

Abstract

We present an approach to model an artificial developmental system based on cells that interact through gene regulatory networks for shape or structure optimization.Cell differentiation is accomplished by positional information provided by transcription factors that diffuse inside a simulated environment. Different actions such as cell division, apoptosis and cell-cell communication are implemented. The system is applied to a simple target shape optimization problem. We show that our model has the ability to form stable patterns of the transcription factors, comparable to patterns found in biological systems. We believe that stable patterns of the transcription factors are a prerequisite for stable and controlled growth which in turn is required for shape or structure optimization.

Introduction

Evolutionary algorithms have been successfully used in many areas like e.g. technical optimization, operations research and design optimization. In particular, in the later field, in combination with appropriate simulation tools (like computational fluid dynamics) innovative technical design solutions have been obtained, see e.g. (Sonoda et al., 2004).

At the same time, the complexity of the reachable design is limited, because in most cases (e.g. spline representation) it directly relates to the dimension of the search space. Alternative representations such as free form deformation, allow unrestricted complexity of the shape, however not of the changes of the shape (Menzel et al., 2005).

Besides the limited complexity, it is also difficult to endow shapes and structures with certain properties that might be desirable in some cases like symmetry, self-similarity or properties that reflect constraints of the physical world. Whereas it is unclear whether such properties could play an important role for purely technical designs (like e.g. turbine blades), they seem more intuitively useful for structures like brain-like information processing systems or for aesthetic design.

Of course in biological evolution, the representation, i.e., the genotype – phenotype map is a very complex nonlinear dynamical process, which has received increasing attention over the last decade and which is now firmly established as the field of evolutionary development or evo-devo (Coyne, 2005). A number of simulation environments of evolutionary development for varying purposes have been put forward. In most cases, cellular growth conducted by genes that are regulated by spatio-temporal signals, and the ability of genes to produce these signals themselves lead to genetic regulatory networks, which define the central part of these simulations.

In this paper, we investigate an approach toward modeling cellular development for evolutionary shape or structure optimization. We start with a very simple target shape. This shape is chosen in a way that its development needs mechanisms, which we believe are necessary for the evolution of highly complex shapes. Especially, we regard the activation and deactivation of genes at specified spatial positions as crucial.

In order to generate the shape, it is important to deactivate genes which are necessary for the growth in one direction at the same place where other genes which generate the growth in a different direction have to be activated. Finally, the termination of growth is necessary.

At the same time, we analyze our system with regard to pattern formation of transcription factors (TFs) which play a central role in the regulatory process.

In the next section, we provide a short biological motivation and review some existing models of developmental biology and pattern formation. In particular, the interaction between pattern formation of the regulating substances and the dynamics of the regulation network itself constitutes the core part of our system, which we will describe in detail in Section 3. We will observe the pattern formation ability of our model system and the consequent optimization behavior for a simple target shape optimization problem in Section 4 and discuss our approach and our results in the last section.

Modeling developmental biology

The development of mammalian life begins with one fertilized egg cell (zygote) that divides several times so that every new cell contains the same genetic information as the first one. Initially, cell differentiation depends on specific

chemicals (transcription factors) that provide positional information inside the growing organism. This information leads to a selective expression of genes, because the presence of specific transcription factors controls the binding of RNA-polymerase to promotor regions of genes on the DNA. Transcription factors can also bind to enhancers and silencers on the DNA that are not necessarily close to the gene, therefore, modifying gene expression in many ways. Because transcription factors are gene products themselves, spatio-temporally coupled genetic regulatory networks can emerge. These networks create stable spatial expression patterns of transcription factors, which are e.g. necessary for axis formation in early embryonic development. A comprehensive introduction to developmental biology can be found in (Gilbert, 2003).

A number of simulation systems for evolutionary development have been suggested. Usually, these systems are either driven by biological questions, where research focuses on modeling physio-chemical processes to get a better insight into the underlying principles or by the aim to translate such principles into problem solving strategies for a variety of tasks. The degree of abstraction that is necessary for both purposes can vary and depends on the actual question that is to be understood or problem that is to be solved.

An example for a detailed simulation driven by biological questions is the work by (Mendoza et al., 1999). The authors model genetic regulatory networks to predict and simulate morphogenesis of various mutated phenotypes of arabidopsis thaliana. According to (Reil, 2003), Eggenberger-Hotz for the first time combined a simulation of development with genetic regulation. He simulated cellular growth conducted by a functional genome, which is divided into regulatory elements and structural elements. Cells interact physically, communicate, divide and die. He was able to simulate biologic observations, like invagination of cell sheets (Eggenberger-Hotz, 2004b) and genetically regulated movement (Eggenberger-Hotz, 2004a).

An example for research aiming at the principles of developmental biology is the work by (Bowers, 2005), who investigates modularity. He presents a model in which genes are linked in a chain of expression. Cell-cell communication is achieved by approximating diffusion via Gaussian distributions. Two predefined chemical gradients provide directional information for the developmental process. Bowers uses a French-flag like target for cell differentiation and achieves individual solutions with high fitness. However, no stable cell distribution is reached, which results in overgrowth. (Federici, 2004) introduces embryonic stages that cope with the general problem of evolvability of development. Phenotypic simulations are computed on a coarse discretization level. The French flag problem and related target patterns are investigated. He finds that the embryonic stages have positive effects on the performance of the optimization algorithms.

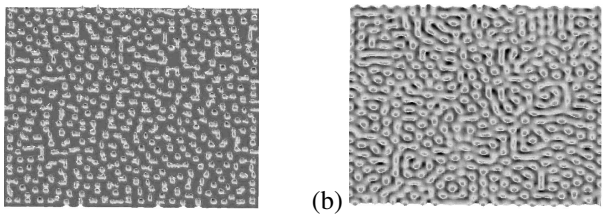

Figure 1: Simulation of the Gray-Scott Model. Dark gray denotes low and bright gray high u concentration.
(a) $f = 0.024$, $k = 0.085$; (b) $f = 0.029$, $k = 0.072$.

Some general models exist that can be applied to both research areas. E.g. (Fleischer, 1995) presents a multiple-mechanism developmental model, that includes intracellular activities and reaction diffusion equations as well as chemical and mechanical interactions between cells.

More abstract models ignore the detailed cellular structure and use systems of differential equations to simulate cellular behavior. Mostly, models are chosen to produce macroscopic effects, so that the desired results determine the level of abstraction. For example, the phenomenon of differential gene expression, which is visible in the color-patterns on the skin of many animals can be investigated by reaction-diffusion equations. (Turing, 1952) proposed interacting systems of diffusing chemicals (hence the name reaction-diffusion) that produce patterns like those observed in biology. He formulated a set of equations that are able to produce striped and spotted patterns. (Meinhardt, 1998) investigated a variety of reaction-diffusion equations to simulate biological pattern formation. (Witkin and Kass, 1991) created a great variety of artificial textures that can be used in computer graphics, partly resembling biological patterns. A straightforward example is the Gray Scott Model (Pearson, 1993), which is formulated in equations (1,2). It describes the dynamics of the concentrations of two interacting substances (u, v), with diffusion coefficients D_u and D_v.

$$\frac{\partial u}{\partial t} = D_u \frac{\partial^2 u}{\partial x^2} - uv^2 + f(1-u) \quad (1)$$

$$\frac{\partial v}{\partial t} = D_v \frac{\partial^2 v}{\partial x^2} + uv^2 - (f+k)v; \quad (2)$$

$$\frac{D_u}{D_v} = 7. \quad (3)$$

Figure 1 shows stable distributions of the chemical concentration u for different values of the parameters f and k. Note that a fundamental difference in the resulting pattern can be observed, although the absolute differences of the chosen values are very small.

A developmental model should be able to produce stable patterns of the transcription factors (TFs) during the microscopic simulation similar to the ones obtained from macroscopic simulations of reaction-diffusion equations. Thus, we can regard the spatio-temporal stability of the transcription

factors both as an observable as well as an indicator for a stable development process.

The model

We choose an abstraction level comparable to that of (Eggenberger-Hotz, 2004a) with a functional genome that is suitable for evolutionary computation.

We simulate a developmental process that uses cells as phenotypic representation. The genotype is described by a virtual DNA which is subject to the process of evolution and describes the developmental process of the phenotype rather than its final appearance. Generally, individuals grow in discrete steps (developmental time steps) from one cell to the final shape by using simulated transcription factors.

The Developmental model

The environment We define a "virtual egg" that provides the physical environment for the simulation of development. It has the form of a two dimensional grid (directions x and y) with fixed resolution and boundaries. In the beginning of development, the virtual egg contains one single cell that is a ball with a fixed radius placed in the center of the grid. Generally, cell positions are determined by floating point values for x and y and do not necessarily correspond to grid points.

The cell Cells are the entities that represent the phenotype, because cell positions are evaluated for fitness computation. All cells inside one egg contain the same DNA. They have the ability to divide, which means that a new cell is placed next to the initial one. The exact position of the new cell depends on genetic information, but the distance between the initial and the new cell is always twice the radius of a cell.

The cells possess the ability to produce a transcription factor that diffuses inside the egg. Since the position of the cell's center is generally not a grid point, the release of transcription factors is simulated by an increase of the concentration of that transcription factor at the nearest grid point.

After apoptosis (cell death) the cell is removed from the grid.

The DNA The simulated functional DNA is a vector of genes. Each gene consists of one structural unit (SU) and several regulatory units (RU). Whereas structural units provide the information for the cell's actions, regulatory units act as the controlling entities. The activation function of the regulatory units are evaluated in every developmental time step. Specifically depending on the presence of transcription factors, all regulatory units of one gene contribute to its overall activity, i.e., they determine whether a gene is active or inactive at the position where the cell is located.

Transcription factors Transcription factors consist of a type a^{TF}, a distribution of concentrations that is associated with every point of the grid δ, a diffusion constant D and a decay rate γ. The type is used to compute a chemical distance to the regulatory units. Therefore, a transcription factor with a small chemical distance to a regulatory unit has a greater influence on that unit than TFs with larger distances. This ensures that regulatory units will react specifically to certain transcription factors, which is explained in more detail in the following.

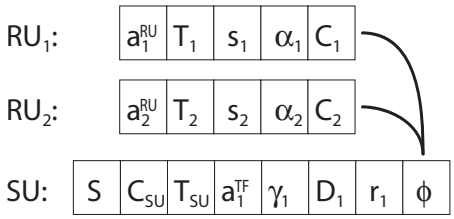

Figure 2: The layout of one gene with two regulatory units.

The translation process A typical gene with two regulatory units is depicted in Figure 2. All parameters belong to the set of real numbers and are subject to optimization. With the exception of T_n, which is constrained to $[0,\infty[$, they belong to $[0,1]$.

The structural unit consists of eight variables. The first variable S denotes the type of the action that will be performed. C_{SU} and T_{SU} are used for regulation, see below. If S codes for the release of a transcription factor, the variables a_k^{TF}, γ_k, D_k and r_k will be used to characterize this transcription factor. a_k^{TF} denotes the chemical type of the transcription factor; γ_k is the decay rate. D_k is the diffusion constant and r_k is the expression rate with which the cell produces the transcription factor. If S codes for cell division, the variable ϕ denotes the division angle. If S codes for apoptosis, no further variables of the gene are used.

Whether the translation of a gene is carried out or not depends on the regulatory process described below.

The regulatory process During one time step, the activity G of every gene is evaluated by solving

$$G(F) = \frac{1}{1+e^{-(\tan(C_{SU}\cdot\frac{\pi}{2})\cdot F)}} - T_{SU}. \qquad (4)$$

Equation (4) describes a threshold function that results in an active gene if G is positive. C_{SU} influences the slope of the function and T_{SU} the actual threshold. F is the sum of the activity values from all regulatory units of the gene:

$$F(f_{kn}) = \sum_{k=1}^{K}\sum_{n=1}^{N} f_{kn}, \qquad (5)$$

where K denotes the number of transcription factors and N the number of regulatory units. Function f_{kn} is given by

$$f_{kn}(\delta_k, d_{kn}) = s_n \cdot \left(\frac{2}{1+e^{-(\tan(C_n\cdot\frac{\pi}{2})\cdot(\delta_k\cdot d_{kn}-T_n))}} - 1\right), \qquad (6)$$

where C_n and T_n belong to the regulatory unit and describe slope and threshold for f_{kn} respectively; $s_n \in \{-1,1\}$. The concentration of the k-th transcription factor at the position of the cell is given by δ_k. The function d_{kn} determines the chemical distance:

$$d_{kn} = e^{-\alpha_n \cdot \tan(\sqrt{(a_k^{TF} - a_n^{RU})^2} \cdot \frac{\pi}{2})}, \quad (7)$$

where a_k is the type of the transcription factor and a_n the corresponding value of the regulatory unit. α_n is an additional weight, which is also part of the regulatory unit. The tangent function is used in equation (7) to enable an easy scaling of the Euclidean distance between a_k^{TF} and a_n^{RU} to $[0,\infty[$, so that $d_{kn} = 0$ if the distance between a_k^{TF} and a_n^{RU} is 1.

Note that some parameters (e.g. the type S) are converted to discrete values for the developmental process. Also, not all parameters of a gene are used depending on S. Therefore, not all parameters are always subject to selection pressure.

Diffusion After a transcription factor has been released, its diffusion on the 2D grid is governed by the simulation of the diffusion equation (8) using an explicit Euler scheme.

$$\frac{\partial \delta_k}{\partial t} = D_k \frac{\partial^2 \delta_k}{\partial x^2} - \gamma_k \cdot \delta_k \quad (8)$$

$$\delta_k(t=0,x) = \begin{cases} r_k + \delta_k^0(x), & x = x_0 \\ \delta_k^0(x), & x \neq x_0 \end{cases}, \quad (9)$$

where δ_k denotes the concentration of transcription factor a_k^{TF} and x_0 is the closest grid position to the cell. If transcription factor a_k^{TF} has been released before, there will be a residual increase in concentration on the grid which is denoted by $\delta_k^0(x)$.

The Evolutionary model

A ($\mu = 30$, $\lambda = 100$) evolution strategy with individual mutative self-adaptation of strategy parameters without recombination has been used in all simulations.

The fitness of each individual that consists of the virtual DNA and the egg is determined either by evaluation of the cell positions after development or by the distribution of the transcription factors.

In addition to mutation, we include gene duplication as a specific operator in our evolutionary model. Gene duplication is believed to be a major driving force in biological evolution. It consists of the insertion of a duplicate of genes into the DNA. This allows one of the two identical copies to mutate while none of the original functionality is lost. In our model, genes that were not activated during the whole developmental process are replaced with a probability of 1% by a randomly chosen gene that was active during development.

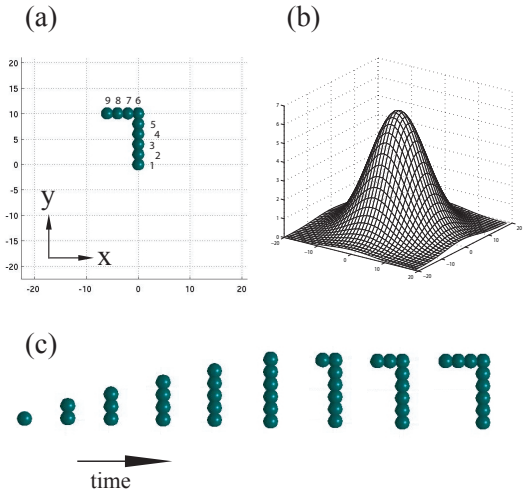

Figure 3: a) The target shape inside the environment. b) The distribution of one predefined transcription factor. c) The development of the best individual after the evolutionary process.

Results

Gene activity

Cell differentiation is the result of the activation of different genes in different cells. We investigated the ability of the system to reproduce the shape depicted in Figure 3a), which is one of the simplest shapes that needs different kinds of transitions in gene activity depending on cellular position: the gene coding for division in y-direction must be active in the first 5 cells only, so the transition is 'initially on' to 'off', division in x-direction must be active only in the cells 6 to 8, so the transistion is 'initially off' to 'on' and then again to 'off'. In this way we are able to observe a simple kind of cell differentiation. A Gaussian distribution of one transcription factor centered around the first cell is predefined to provide a minimum of positional information, see Figure 3b). The fitness, which is minimized during evolution, is evaluated using a modified Hausdorff-distance d_H between the cell centers of the target $\mathcal{Z}_1$ and those of the individuals $\mathcal{Z}_2$ after the developmental process:

$$d_H = \frac{1}{2}\left(\sum_{i=1}^{|\mathcal{Z}_1|} \min\{|\vec{a}_i - \vec{b}|^2; \vec{b} \in \mathcal{Z}_2\} + \sum_{j=1}^{|\mathcal{Z}_2|} \min\{|\vec{b}_j - \vec{a}|^2; \vec{a} \in \mathcal{Z}_1\} \right). \quad (10)$$

Theoretically, two genes should be sufficient to evolve the target shape. However, practically, successful individuals always consist of at least five genes. We varied the number

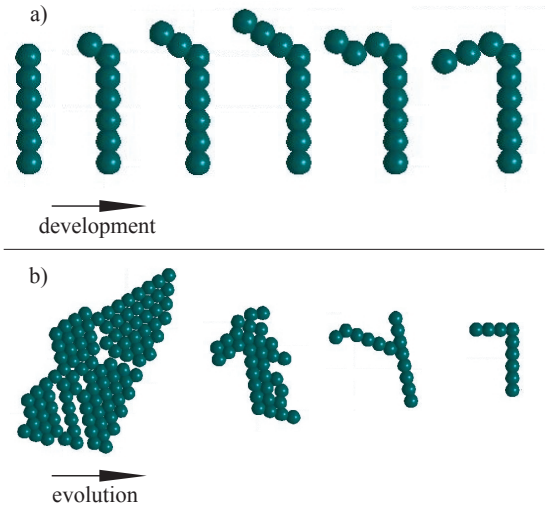

Figure 4: a) An exemplary development of another good result. b) The best individuals after their development at ascending developmental stages.

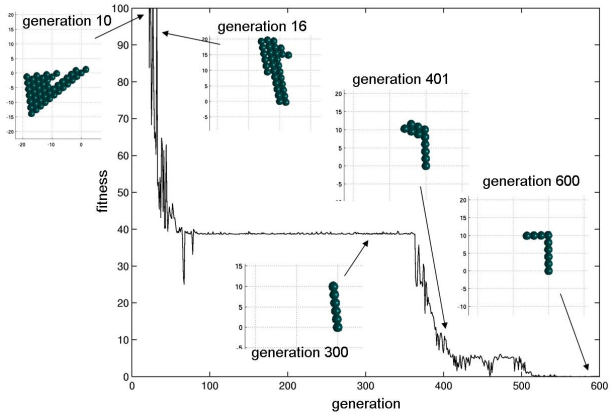

Figure 5: Fitness versus generations of a typical evolutionary run. Some example patterns after development are shown.

of genes between 2 and 12. Although the reason is not clear yet, we might speculate that five genes are necessary to make use of gene duplication, which seems to play a large role in our evolutionary model. Indeed, if we switch gene duplication off, we are not able to reproduce the target shape[1].

An example of a different developmental process encoded by the evolved gene regulatory network is shown in Figure 4a). Compared to the best individuals (Figure 3c), we observe a different growth strategy which leads to an individual with relatively high fitness including the important bend of the target shape. Figures 4b) and 5 show selected best individuals at different generations. We note that at the early stages of evolution we can observe a rather "wild" growth process whereas at later stages the growth is more controlled. Thus, the stable control of growth is the issue; growth itself is easily accomplished.

The target of *stable growth* naturally leads to the question of evolving development models with stable patterns of transcription factors providing sufficiently stable positional information for cell differentiation. Therefore, we investigated transcription factor distributions inside our system and compared them to patterns found by the macroscopic simulations of the Gray Scott Model.

Concentration Patterns of Transcription Factors

The expression rate of transcription factors depends on the activity of the corresponding gene. The expression rate r_k is scaled by $G(F)$, if the gene codes for a transcription factor. Therefore, the concentration of the released transcription factor is proportional to the amount by which the threshold T_{SU} in equation 4 is exceeded.

The following process was used for our investigation. We evaluate a row of 30 identical cells, each cell consisting of two genes with two regulatory units. Gene type S is limited to the release of a transcription factor. Patterns should form as a response to random finite amplitude perturbation. The fitness is determined by maximizing the variance of the transcription factor concentration after convergence. Oscillatory solutions receive a low fitness. This should lead to stable distributions of the concentrations of transcription factors.

The genes that produce satisfactory solutions are then positioned on an array of 20x20 cells. Two examples for the resulting patterns are shown in Figure 6. The pattern formation resembles a transient process. Before a stable solution is reached, we observe damped oscillations. In some patterns, oscillations persist. Furthermore, graded distributions can emerge that always orientate along the longest extension of the simulation area; a phenomenon observed by Meinhard in reaction-diffusion systems. Therefore, this simple model with 2 transcription factors showed the ability to form stable patterns, which should in principle be sufficient to give positional information for cell differentiation.

It is nice to observe the emergence of stable patterns for the transcription factors because on the one hand we are able to reproduce characteristics of macroscopic reaction-diffusion systems with our simple cellular model and on the other hand we believe stable TF patterns are a necessary ingredient for stable shape formation using developmental growth processes.

Discussion

In this paper, we have shown that a complex genotype-phenotype mapping based on a simple model for development with gene regulatory networks and differential gene

[1] So far, we have not run any experiments with recombination instead of gene duplication.

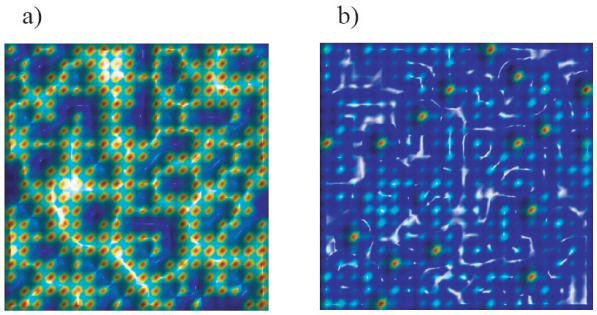

Figure 6: Transcription factor-patterns emerge during simulation of cellular behavior: a) stripe-like patterns, b) spot-like patterns.

expression can in principle be used for shape optimization. Of course, the chosen shape is very simple (although the sharp bend is not so easy to realize with other – inherently smooth – representations like splines) and it is necessary to show that the approach scales with larger shape complexity and interesting constraints that we mentioned in the introduction.

One major problem is the continuity of the complex mapping: evolution strategies become trapped very early in local optima. This indicates that the fitness landscape is rather rugged with small, steep optimal peaks. One reason might be the rather simple implementation of cell division, without taking into account any cell adhesion or cell movement, which would certainly dampen the extreme effects of a variation in e.g. the division angles. Other elements of the mapping add to the rugged nature of the landscape like the complex chain of regulatory elements that is vulnerable to mutation. It is interesting to note that in nature on the contrary such chains are often relatively stable and inherently robust. It is worthwhile to explore how this could be achieved in an artificial system.

Attempts to use good, low complexity features in the beginning of evolution and to increase complexity during evolution by gene duplication used in our simulations or by approaches suggested in (Federici, 2004) and (Stanley and Miikkulainen, 2003) seem promising to overcome this problem. Note that without gene duplication, the target shape in Figure 3 was not reached in our simulations.

Cell differentiation in biological systems is guided by transcription factors (TFs). TFs form stable spatio-temporal patterns to provide positional information to the cells. We observed pattern formation ability in our simulations that produces TF distributions resembling biologic patterns. Emerging patterns and how their characteristics vary with changes of the parameters are comparable to those of reaction-diffusion systems. Therefore, the system presented exhibits a wide variety of possible patterns, which seems to be an important property for exploiting cell differentiation for shape optimization.

Patterns strongly depend on boundary and initial conditions, and react very sensitively to parameter variation. Therefore, mutations have a substantial influence on the pattern formation ability of TFs. Nevertheless, stable patterns emerge under the pressure of selection. This supports our interpretation that stable patterns are required for stable and controlled growth.

Future research

Firstly, it is important to demonstrate the scalability of the approach for more complex shapes or structures. Secondly, it is necessary to get a better understanding of the evolution of development in itself by observing e.g. the role of duplication (vs. standard recombination) or the role of stability against mutations (and on a more biological note also against environmental fluctuations.) We will investigate whether a local behavior guided by information exchange between neighboring cells leads to more stable patterns as suggested by (Salazar-Ciudad et al., 2000).

References

Bowers, C. P. (2005). Formation of modules in a computational model of embryogeny. In *Proceedings of the Congress on Evolutionary Computation (CEC)*, pages 537–542. Piscataway: IEEE Press.

Coyne, J. (2005). Switching on evolution. *Nature*, 435:1029–1030.

Eggenberger-Hotz, P. (2004a). Asymmetric cell division and its integration with other developmental processes for artificial evolutionary systems. In *Proceedings of the Nineth International Conference on Artificial Life IX*.

Eggenberger-Hotz, P. (2004b). Genome-physics interaction as a new concept to reduce the number of genetic parameters in artificial evolution. In *Proceedings of the Congress on Evolutionary Computation (CEC)*.

Federici, D. (2004). Using embryonic stages to increase the evolvability of development. In *Proceedings of the Workshop on Regeneration and Learning in Developmental Systems, WORLDS*.

Fleischer, K. W. (1995). *A Multiple-Mechanism Developmental Model for Defining Self-Organizin Geometric Structures*. PhD thesis, California Institute of Technology, Pasadena, California.

Gilbert, S. F. (2003). *Developmental Biology, Seventh Edition*. Sunderland: Sinauer Associates, Inc.

Meinhardt, H. (1998). *The Algorithmic Beauty of Sea Shells*. Heidelberg: Springer-Verlag.

Mendoza, L., Thieffry, D., and Alvarez-Buylla, E. R. (1999). Genetic control of flower morphogenesis in arabidopsis thaliana: a logical analysis. *Bioinformatics*, 15:593–606.

Menzel, S., Olhofer, M., and Sendhoff, B. (2005). Application of free form deformation techniques in evolutionary design optimisation. In *Sixth World Congress on Structural and Multidisciplinary Optimization*.

Pearson, J. E. (1993). Complex patterns in a simple system. *Science*, 261.

Reil, T. (2003). Artificial genomes as models of gene regulation. In Kumar, S. and Bentley, P. J., editors, *On Growth, Form and Computers*, pages 256–277. London: Elsevier Academic Press.

Salazar-Ciudad, I., Garcia-Fernández, J., and Solé, R. V. (2000). Gene networks capable of pattern formation: From induction to reaction-diffusion. *J. theor. Biol.*, 205:587–603.

Sonoda, T., Yamaguchi, Y., Arima, T., Olhofer, M., Sendhoff, B., and Schreiber, H.-A. (2004). Advanced high turning compressor airfoils for low reynolds number condition, Part 1: Design and optimization. *Journal of Turbomachinery*, 126:350–359.

Stanley, K. O. and Miikkulainen, R. (2003). Achieving high-level functionality through complexification. In *Proceedings of the AAAI-2003 Spring Symposium on Computational Synthesis. Stanford, CA*. AAAI Press.

Turing, A. M. (1952). The chemical basis of morphogenesis. *Phil. Trans. R. Soc. Lond.*, 237:37–72.

Witkin, A. and Kass, M. (1991). Reaction-diffusion textures. *Computer Graphics (SIGGRAPH 91 Conference Proceedings)*, 25:299–308.