A morphogenesis model for multiagent embryogeny

Gregory Beurier\textsuperscript{1}, Fabien Michel\textsuperscript{2}, and Jacques Ferber\textsuperscript{1}

\textsuperscript{1}Laboratoire d’Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM) 161 rue Ada, 34392 Montpellier Cedex 5 - France 
\{beurier,ferber\}@lirmm.fr

\textsuperscript{2}Centre de Recherche en Sciences et Technologies de l’Information et de la Communication (CReSTIC/LERI) Université de Reims, Moulin de la Housse BP 1039 51687 Reims Cedex 2 France 
fabien.michel@univ-reims.fr

Abstract
This paper describes a bio-inspired method that enables the production of artificial multiagent organisms starting from a single agent. This method relies on two complementary computing approaches: (1) mimicking the functioning of segmentation genes and homeotic genes in the development of natural embryos and (2) using techniques from the evolutionary computing and the artificial embryogeny fields. The paper first details these mechanisms and then studies some possibilities of such a bio-inspired approach developing multiagent systems that evolve and self-organize in various flag patterns.

Introduction
Living systems outperform in many ways complex systems that man has produced. Seeking inspirations from these systems, several established and active research communities are organized around themes such as self-organization (Camazine et al., 2001), DNA computing, artificial neural networks (Boers et al., 1993), morphogenesis (Roggen et al., 2003), evolutionary computation (Koza, 1994), etc.

The problems tackled with evolutionary computation become increasingly complex and traditional approaches are becoming less efficient. In the nascent field of artificial embryogeny (Stanley and Miikkulainen, 2003), researchers explore new ways to outperform the classical direct mapping of evolutionary methods (from genotype to phenotype) by proposing various approaches to evolve and grow artificial organisms (indirect encoding).

This paper proposes an artificial embryogeny to explore possibilities of evolving a multiagent system (MAS) (Ferber, 1999). In our model, the agents behavior mimics the functioning of the segmentation genes and homeotic genes involved in the animals morphogenesis, based on the drosophila larva development model. Agents replicate and interact to construct organisms following morphogenetic rules.

The outlines of the paper are as follows: A review of some relevant works on artificial embryogeny is first presented. Secondly, a biological background is given to explain the role of segmentation and homeotic genes in the morphogenetic process. The MAS model is given in the third part: we first define the cell-agents and then describe the way to make them evolve to construct multiagent organisms. Simulations of the model are developed and then discussed. Finally, we draw conclusions from this work.

Artificial embryogeny related works
Considering artificial embryogeny research, a clear distinction can be made between grammatical approaches and cell chemistry approaches (Stanley and Miikkulainen, 2003).

Using grammatical approaches, a number of researchers have studied the potential of Lindenmeyer systems (L-systems) (Prusinkiewicz and Lindenmayer, 1990) for generating Artificial Neural Networks (ANNs) which are a major challenge for artificial embryogeny. Boers and Kuiper used evolutionary algorithms to evolve the rules of L-systems that generate ANNs (Boers et al., 1993). Gruau abandoned L-systems to develop a language called cellular programming based on the transformation of graphs (Gruau, 1993). Cellular programming enables the controlling of the division of cells that grow into ANNs.

Grammatical approaches produce good results but do not seem to be well fitted for evolutionary processes (Stanley and Miikkulainen, 2003). So, a number of researchers have managed to simulate the low-level aspect of natural embryogeny. One of the earliest works on cell chemistry approaches returns to Turing (Reaction/Diffusion) (Turing, 1952). Turing proposed linear equations to achieve spatial differentiation, which is done by postulating two substances with mutual interaction and different spatial distribution, namely the morphogens. De Garis was one of the first to present a work based on cellular automata, cell chemistry and genetic programming (de Garis, 1999). De Garis’s work achieved to grow simple non-convex organisms. Moreover, de Garis extended his work to obtain convex organisms thanks to the addition of external sources of morphogens. In fact, exogenous sources of morphogens are commonly used in artificial embryogenies. For instance, Eggenberger used exogenous substances to induce a symmetry break in his digital organisms (Eggenberger, 1997).

Fleischer and Barr (Fleischer and Barr, 1992) developed
a simulation framework for multi-cellular pattern formation including chemical diffusion. This framework also includes mechanical factors such as cell adhesion, genetic factors, etc. Fleischer and Barr showed the difficulty of maintaining the size and the shape of a multi-cellular organism and pointed out the necessity of combining multiple mechanisms in the pattern formation to guarantee the robustness of the organism. COMPUCELL 3D has been developed to simulate morphogenetic processes in different organisms and also implements phenomena such as cell growth, diffusion of chemical gradients, etc (Izaguirre et al., 2004). COMPUCELL 3D highlighted the necessity of modeling each cell state of an organism to implement differentiation processes. Still, the morphogenesis model of COMPUCELL 3D is limited by the fact that it a priori assumes the shape of the organism as a function of time.

Inspired by the cell adhesion process, Hogeweg (Hogeweg, 2000) developed a model to simulate morphogenetic processes such as cell migration or engulfing. Using Boolean networks of genetic regulation, Hogeweg achieved to evolve complex artificial organisms. However, Hogeweg’s model does not use morphogen gradients and cannot handle the simulation of phenomena such as chemotaxis or chemical genes regulation.

Miller developed artificial organisms based on Cartesian Genetic Programming which is an extension of Boolean networks (Miller, 2003). Miller’s goal is to evolve a developmental program inside a cell to create multicellular organisms. The obtained organism can be then considered as a program. Still, this method seems to be expensive in computing resources for simple problems for which developing a single program is easier.

**Biological background**

**Morphogenesis**

The proposed chemical approach is a multiagent model directly inspired by the functioning of genes involved in the morphogenesis process of animals. To better describe this model, we now briefly introduce the biological concepts we will use later on.

Morphogenesis has been first studied via the observation of the alteration of the developmental process in species. At the end of the 19th century, Bateson discovered few mutants drosophila (a fruit fly) which parts of the body were transformed into another ones. At the end of the 20th century, Lewis discovered the genes involved in these alterations. These genes are known to be the same involved in the process of morphogenesis, the homeotic genes.

Genes are regions of DNA inheritance of living creatures. A gene basically encodes the chemical structure of a protein and the way this protein (the transcript) is produced. A protein is produced by a cell when the coding part (CP) of the gene is transcribed and translated. The transcription/translation process depends on interactions between regulatory elements (REs), which are other parts of the gene, and transcription factors which are cells internal or external molecules. The interaction process that takes place between a transcription factor and a gene is called the regulation process.

The homeotic genes encode transcription factors that control the expression of genes responsible for particular anatomical structures, such as wings, legs, and antennae. Basically, during the early phases of the development of animals, these genes are expressed within the organism to assign an identity to regions which are already established. In the case of the insects, these regions are called segments and are created by a cascade of regulations of genes, namely the segmentation genes.

**Drosophila larva development model**

To better understand the role of homeotic genes and segmentation genes, we will now briefly describe the early phases of drosophila larva development. For further information on this subject see (Li et al., 2001). The genes that priorly act in the developmental process are already present in the egg before the embryo starts to transcribe its own genes. These maternal genes are expressed during oogenesis (production of female gamete) and they produce the mRNAs (maternal ribonucleic acids) which are stored within the egg. Initial asymmetries in the egg along the anterior-posterior axis (from head to tail) are set up by localization of mRNAs at the anterior and posterior poles of the embryo. These mRNAs are translated and diffuse through the embryo, forming two opposite long-range protein gradients (see Fig. 1.a.). These two protein gradients are transcription factors that regulate a first set of segmentation genes, the gap genes. A gene X may be active only at high concentrations of transcription factors, whereas a gene Y could also be active at lower concentrations, and therefore at a greater distance from the source of the transcription factors. It is the case for gap genes which roughly subdivide the embryo along the anterior/posterior axis (see Fig. 1.b.). The gap genes encode transcription factors that regulate the expression of a second set of segmentation genes, the pair-rule genes. The broad domains of gap gene expression are translated into a striped pattern and

![Figure 1: A cascade of gene activation in the drosophila larva development.](image-url)
the embryo is divided into pair of segments (see Fig. 1.c.). Finally, the pair-rule genes encode transcription factors that regulate the expression of a third set of segmentation genes, the segment polarity genes. The segment polarity genes set the anterior/posterior axis of each segment (see Fig. 1.d.). Once established, the homeotic genes are regulated to assign each segment its own identity by regulating groups of functional/structural genes.

**Proposed MAS organism model**

The proposed model is a MAS based on reactive agents able to evolve and replicate. Our approach aims to closely mimic the role of segmentation genes and homeotic genes in the process of morphogenesis:

- Environment takes the role of egg/embryo, in which maternal gradients of morphogens/mRNAs are diffused.
- Agents take the role of cells and their behavior is an interpretation of a genome composed by segmentation genes and homeotic genes (behavioral genes). They are called cell-agents.

**Environment and information**

The environment is a discretized rectangular surface (8-connectivity) that primarily acts as interaction medium. Gradients of maternal morphogens are initially present to initiate the developmental process. These gradients are placed randomly or in an ad-hoc manner, depending on the organism adaptation objectives (see the evolutionary experiments section). Maternal gradients assume the role of biological mRNAs as in natural morphogenesis. An alternative to the use of an exogenous source of morphogens is discussed in the end of the paper. Nevertheless, maternal gradients are a fundamental feature of natural embryogenetic processes and must be considered when designing bio-inspired embryogenies.

The information management is comparable to a biological plain diffusion and involves equalizing the concentrations of emitted proteins. The proteins diffuse at each time step according to a diffusion coefficient \( \delta \) and a decay rate \( \rho \). The physical mechanism of diffusion, when mapped onto a discrete system (space and time), may be expressed quite simply: at each time step, squares give out a fraction of the proteins they hold, in all "directions" (to each neighbor). The fraction is computed by using the diffusion coefficient \( \delta \). At the same time, the proteins also decay according to their decay rate \( \rho \). Hence, for the simplest version of the plain diffusion algorithm, a square \( i \) that holds a quantity \( \Phi_i \) of a protein \( P \), characterized by a diffusion coefficient \( \delta_p \) and a decay rate \( \rho_p \), will give a small fraction of \( P \), \( \Delta \Phi \), at each time step to each of its neighbors, independently of squares, of neighbors, and of time. Each transfer of protein to a neighbor \( j \) can then be captured by the following equation:

\[
\Delta \Phi_{i \rightarrow j} = (\Phi_i - (\Phi_i \times \rho_p)) \times (\delta_p/8)
\]  

**Cell-agents**

The proposed approach relies on extending classical regulatory networks (Stanley and Miikkulainen, 2003) by distinguishing morphogenetic interactions (segmentation genes) and functional interactions (homeotic genes). Each cell-agent behavior is an interpretation of a genome composed by two types of genes: segmentation genes, which are involved in the patterning process of the organism, and homeotic/behavioral genes that encode the behavior of the agents and give the identity of the regions of the formed organism. Both homeotic/behavioral terms are used since some behaviors (such as the replicating one) are not dedicated to a particular region/segment of the organism.

A gene is composed by two parts: (1) the coding part (CP) and (2) the regulatory elements (REs). The CP encodes the gene primary function. The REs encode the interaction properties that enable the gene regulation (as in classical artificial regulatory networks).

The CP of segmentation genes encodes protein properties. When a segmentation gene is activated, a parameterized diffusing protein is produced. The segmentation gene CP is implemented in a set of binary digits, namely a bitset. The diffusion coefficient and the decay rate of the protein are encoded in 7 bits (values ranged from 0.00 to 1.00 with a precision of one-hundredth). The CP can be easily extended to add extra properties to the produced proteins encoded in the same bitset (e.g. type, color, etc.).

In the case of behavioral genes, the CP type is related to the implemented function. Indeed, the CP works as an evolutionary computing standard tool. For instance, the CP can be a tree encoding a part of the cell-agent program (like in genetic programming (Koza, 1994)), a bitset encoding a specific parameter/value of the cell-agent behavior (like in genetic algorithm (Holland, 1987)), etc. For instance, we can imagine encoding the color of a cell-agent by using a set of 24 bits representing the RGB canal, or a moving behavior with simple commands encoded within a tree.

The REs encode the interaction between genes and protein gradients to enable the regulation process. The formal description of REs is inspired by the Reaction/Diffusion
model (Turing, 1952) and encodes a chemical reaction involving genes transcription factors (the diffusing proteins). But, instead of encoding the chemical result of a chemical reaction, the REs encode the possibility of this reaction to appear. First of all, all the transcription factors of the gene are encoded within a bitset defining a set of transcription identifiers. The transcription factors of the gene are randomly chosen (C, B, D, A in Fig. 2). The REs second part is called the reaction part (RP) and encodes the chemical reaction itself. The RP is decomposed into equal parts representing reaction factors with each of the transcription factors. The last part encodes the reaction unit quantity, an integer that matches the perceived quantity of proteins with the reaction (as the mole unit in chemistry).

A cell-agent perceives all proteins in its square and applies the chemical reaction for all of its genes. For a given gene, the result of the reaction depends on the restricting chemical factor:

\[
    r = \min \left\{ \forall (h \in H), \frac{Q_h}{u + F_h} \right\}
\]  

The reaction result \( r \) can be interpreted in various ways. In **standard** reactions, if \( r > 1 \), the chemical reaction intervenes and the regulatory elements activate \( G \): the cell-agent perceives a sufficient amount of transcription factors. If \( r = 0 \), the gene is inhibited.

The REs have been extended to perform more sophisticated tasks. For instance, a second RP can be added to inhibit the gene when \( r > 1 \). This permits to greatly increase the way a gene can be regulated but slows down the evolutionary process. Another extension deals with activation/inhibition thresholds. The quantities of perceived transcription factors may be high enough to react several times (e.g. \( r = 8 \)). So, a part can be added to the REs to limit the regulation process to a certain number of reactions: under or over (or both) a threshold encoded in the gene, the CP is inhibited.

### Genetic model

The evolutionary principles of the model are inspired by classical approaches of evolutionary computing. A standard evolutionary algorithm has been extended to encompass the notion of organism. Every cell-agent possesses a **behavioral gene** encoding the replication function (mitosis).

The evolutionary algorithm works as follows:

1. Generate maternal gradients in the environment \( E \).
2. Generate a population \( P \) of cell-agents with random genomes.
3. While (criterion not reached) {}  
4. Place each cell-agent of \( P \) in \( E \) and let it grow.
5. Assign fitness to each genome of the cell-agents according to an objective function with respect to the formed organism.
6. Select \( n \) cell-agents for reproduction with respect to the fitness of their genome.
7. Reproduce cell-agents by taking two parents at a time and use reproduction operators on the set of genes of their genomes.
8. Apply mutation operators on the genes of the offspring genomes.
9. Replace old population by offspring according to replacement strategies. 

**Step 1** is trivial: the generation of the maternal gradients can be done either randomly or by the simulation designer to accelerate the population convergence. In most of the cases, gradients are deposited according to symmetry axes. This phase intervenes at first in order to correctly initialize the regulatory elements of the generated genomes (transcription factors).

**Step 2** consists in randomly generating coding parts and regulatory elements for every gene of the cell-agents. This generation can be different according to the nature of the gene (i.e. bitsets, tree of instructions, graphs of values, etc.). The fitness is calculated on the formed organism but based on the designer desiderata (task to achieve, size, color, etc.). The fitness is calculated on the formed organism but assigned to the genome of the single cell-agent that has given birth to the whole organism.

**Step 3** is used to alter genetic information on the genes of the genomes of the newly created cell-agents. This operation is used to prevent the evolutionary algorithm from stagnating at local optima.

**Step 4** and **5** are crucial ones. They consist in simulating each single cell-agent and evaluating the quality of the formed organism. Fitness is given by external(s) observer(s) based on the designer desiderata (task to achieve, size, color, etc.). Again, various techniques can be used depending on the type of the genes: Simple Crossing-over, Double Crossing-over, etc. (Koza, 1994).

**Step 7** consists in creating new cell-agents (with new genomes) starting from pre-selected cell-agents. The offspring are generated in the hope that they will be better than their parents (in the sense of fitness). Again, various techniques can be used depending on the type of the genes: Simple Crossing-over, Double Crossing-over, etc. (Koza, 1994).

**Step 8** is used to alter genetic information on the genes of the genomes of the newly created cell-agents. This operation is used to prevent the evolutionary algorithm from stagnating at local optima.

**Step 9** determines which offspring will replace old cell-agents in order to generate the new population. Elitism, Steady State Replacement and CHC Selection are commonly used techniques for this purpose (Koza, 1994).

Although this algorithm is a classical model of evolutionary programming, it has been designed to fit our embryoge-
netic/organism point of view: the genomes of the cell-agents neither encode problem solutions nor methods for solving a problem, but rather encode the chemical interactions and the growth of an organism which is expected to have properties that will eventually resolve a problem from an external point of view.

**Evolutionary experiments**

To validate our model and evaluate its capabilities, classical artificial life experiments have been conducted creating organisms that construct flag patterns. The model has been implemented on the MAS platform MadKit (Gutknecht et al., 2001) using the TurtleKit framework (Michel et al., 2005).

Some modifications have been done to the model to ensure a faster convergence of the system. Firstly, four maternal gradients are deposited according to symmetry axes in order to bootstrap the morphogenetic process. The maternal gradients consist in linear gradients of proteins deposited from one side to the other side of the environment in arbitrary quantities (from 0 to 5000 in the presented simulations). In the drosophila larva model, these gradients are emitted by the mother during the early phase of the development according to the anteroposterior axis. Secondly, the coding parts of some genes are simplified by defining, in an ad-hoc manner, the encoded behaviors. When such genes are regulated, they directly activate the cell-agent behaviors instead of extracting these behaviors from an interpretation of the coding part of the gene.

The fitness of the organisms is calculated by an external observer-agent when the growing process stops or when the size of the organism reaches the size of the environment. This fitness is given by the observer-agent to the genome of the cell-agent which has formed the whole organism. The fitness consists in a percentage of similarity between the graphical representation of the formed organism and a predefined flag pattern. The simulation ends when an organism is at least similar at 98% to the predefined pattern. The cell-agents which are eligible for reproduction are chosen via a roulette wheel selection with respect to the computed fitness. The mutation operator consists in randomly switching bits in the genes. At each new generation, the probability for a bit to mutate is 0.15%. The reproduction operator is a two-point crossover.

**The French flag experiment**

The French flag model of Wolpert (Wolpert, 1968) has been the inspiration for the first task the model has to achieve. This model has already been studied by Miller (Miller, 2003) using CGP. The evolved organism has to grow a recognizable French flag.

To exhibit the role of maternal gradients, the first system has been implemented without segmentation genes. All the genes of the cell-agents are directly regulated by maternal gradients that are initially present in the environment. The genome of the cell-agents consists in four genes: mitosis, blue, white and red. All genes have been implemented in bit-sets and the coding part simply activates a particular behavior of the cell-agents. The mitosis gene controls the replicating behavior. When the mitosis gene is activated, cell-agents replicate in the free neighbor spaces. When a color gene is activated, cell-agents take the corresponding color. By default, the cell-agents are green. Each transcription identifier of the transcription factors of the genes has been encoded in a set of 2 bits, the reaction factors of the reaction parts have been encoded in a set of 5 bits and the reaction unit quantity has been encoded in a set of 6 bits.

One hundred simulations have been made using 30 individuals as population, in a 100x100 environment. We obtained the growth of a French flag organism in all cases. We obtained the convergence of the population in an average of 250 generations. Figure 3 shows the growth of a mature French flag organism.

**The Japanese flag experiment**

This second experiment exhibits the role of segmentation genes in the patterning process. The genome of the cell-agents consists in three genes: mitosis, segmentation and red. By default, the cell-agents are white. Each transcription identifier of the transcription factors of the genes has been encoded in a set of 2 bits, the reaction factors of the reaction parts have been encoded in a set of 5 bits and the reaction unit quantity has been encoded in a set of 6 bits. The same environment (size and maternal gradients) than in the French flag experiments has been used. It permitted us to reuse the mitosis gene of evolved French flag agents. Indeed, the mitosis gene specifically controls the size of the formed organism. So, we used a predefined Japanese flag pattern of the same size as the French flag pattern. It is a fundamental feature of the model: by reusing evolved genes, the performance of the evolution algorithm can be significantly increased and a step by step evolution is possible. Figure 4 shows the growth of a Japanese flag organism.

The formation of the red circle intervenes when (1) the
reusing the pre-evolved mitosis gene in all cases with an average of 200 generations. Without mention. The convergence of the population has been obtained using 30 individuals as population in a 100x100 environment factor. One hundred simulations have been made (3) the red gene factor of the red gene encodes a protein which is a transcription segmentation gene to a mature Japanese flag organism.

Figure 4: Growth of fittest program from a single cell-agent to a mature Japanese flag organism.

segmentation gene encodes a protein which is a transcription factor of the red gene, (2) this protein has fittest parameters, (3) the red gene encodes an adequate reaction with this transcription factor. One hundred simulations have been made using 30 individuals as population in a 100x100 environment. The convergence of the population has been obtained in all cases with an average of 200 generations. Without reusing the pre-evolved mitosis gene, for one hundred simulations, the convergence of the population has been obtained with an average of 290 generations, showing (1) the modularity of the model and (2) the added value provided by reusing previously evolved genes.

Experiments and model evaluation

Compared to other similar experiments which have been published in the literature, the obtained results are very promising. Firstly, the evolved organisms are almost equal to each predefined pattern. Secondly, these results have been achieved in a number of generations which was less than expected. Finally, the Japanese flag experiment showed that it is possible to successfully reuse evolved genes within different organisms. This result has urged us to work on establishing a kind of gene library that could be used to ease the design of new organisms based on the reuse of predefined/pre-evolved genes either (1) during the a priori design phase or (2) at runtime, directly within the evolutionary engine (manually or automatically).

A model limitation relies on the ad-hoc use of predefined maternal gradients. In fact, even if maternal gradients most likely are a fundamental feature of natural processes of morphogenesis, it would be very interesting to design organisms that do not require exogenous sources of morphogens prior to the emission of the cell-agents themselves. Considering this issue, we are currently working on applying Reaction/Diffusion techniques to define a better local control: substituting maternal gradients by the behavior of the agents. Indeed, Gierer and Meinhardt (Gierer and Meinhardt, 1972) have shown that interactions between heterogeneous gradients of morphogens can lead to symmetry breaks and polarity gradients formations.

Additionally, we are exploring some modifications of the reaction part of the genes: (1) we are extending the chemical reaction to transform the classical reaction factors into a polynomial function of the transcription factors, (2) this polynomial function is encoded and evolved within a node tree instead of a simple bitset. Preliminary experiments showed great improvements in the pattern refining and in the symmetry breaking of the simulated organisms.

Another way to improve the expressiveness of the model is to act on the mutation mechanisms. So, we plan to modify the mutation operators to permit modifications of the genome structure of the cell-agents. In fact, following the establishment of new sources of gradients in a pattern as a consequence of the regulation of segmentation genes, new gradients may be fired leading to more refined patterns. However, this refining may be insufficient and lead to local minima. So, we are exploring the introduction of new segmentation genes during the evolution process to ensure a better refining of complex organisms.

Conclusion

We have presented and discussed a bio-inspired developmental model of multiagent organisms. This model is an artificial embryogeny closely inspired by the morphogenetic process and is implemented using the multiagent paradigm. The evolved entities are reactive agents sensitive to protein gradients. The system has been simulated to exhibit its capabilities in evolving complex organisms.

Perspectives of such a work are numerous, but we are actually interested in three issues. The first one deals with multiagent systems design. We plan to construct swarm organisms giving the cell-agents low-level functions, enabling them to resolve simple tasks which require distributed resolution. Another interesting long-term perspective deals with data encryption and data compression. It can be interesting to consider predefined patterns of organisms as a data to compress or to encrypt. Evolved genomes would be thus considered as compressed data which have to evolve to be readable. In the same way, we can imagine to encrypt data in a genome. Then, the relevant key to decrypt data could be either the environment properties or the behavior of the agents.

Finally, we are very interested in refining our inspirations from biological processes. In a general way, the artificial embryogeny field can help in increasing knowledge on the mechanisms of complexity appearance in natural organisms. Seeking inspirations in these mechanisms can greatly contribute in designing complex systems. Indeed, Bentley and Kumar demonstrated that an evolutionary approach of embryogeny provides significant benefits to evolve complex solutions in evolutionary computation (Kumar and Bentley, 2003).
References


