Advances in Complex Systems

Emergence of regulatory networks in simulated evolutionary processes

Dirk Drasdo*

Interdisciplinary Center for Bioinformatics (IZBI), University of Leipzig, Haertelstr. 16/18 D-04107 Leipzig, Phone: +49 341 97-16686, Fax: +49 341 97-16709, Email: drasdo@izbi.uni-leipzig.de

Matthias Kruspe

Bioinformatics Group, Department of Computer Science, University of Leipzig, Haertelstr. 16/18 D-04107 Leipzig, Phone: +49 341 97-16675, Fax: +49 341 97-16709, Email: matthias@bioinf.uni-leipzig.de

Received 08.04.2005

*Corresponding author

Despite the spectacular progress in biophysics, molecular biology and biochemistry our ability to predict the dynamic behavior of multicellular systems under different conditions is very limited. An important reason for this is that still not enough is known about how cells change their physical and biological properties by genetic or metabolic regulation, and which of these changes affect the cell behavior. For this reason it is difficult to predict the system behavior of multicellular systems in case the cell behavior changes for example as a consequence of regulation or differentiation. The rules that underly the regulation processes have been determined on the time scale of evolution, by selection on the phenotypic level of cells or cell populations. We illustrate by detailed computer simulations in a multi-scale approach how cell behavior controlled by regulatory networks may emerge as a consequence of an evolutionary process, if either the cells, or populations of cells are subject to selection on particular features. We consider two examples, migration strategies of single cells searching a signal source, or aggregation of two or more cells. Both can for example be found in the life cycle of *Dictyostelium discoideum*. However, phenotypic changes that can lead to completely different modes of migration have also been observed in cells of multi-cellular organisms for example as a consequence of a specialization in stem cells or the de-differentiation in tumor cells. The regulatory networks are represented by Boolean networks and encoded by binary strings. The latter may be considered as encoding the genetic information (the genotype) and are subject to mutations and crossovers. The cell behavior reflects the phenotype. We find that cells adopt naturally observed migration strategies, controlled by networks that show neutrality, robustness, and redundancy. We carefully analyse the regulatory networks and the resulting phenotypes by different measures and by knockouts of regulatory elements. We illustrate that in order to maintain a cells' phenotype in case of a knockout, the cell may have to be able to deal with contradictory information. In summary both, the cell phenotype as well as the regulatory network emerged behave as their biological counterparts observed in nature.

Keywords: Single-cell based model, cell migration and aggregation, artificial evolution, boolean network, genotype-phenotype relationship

1. Introduction

The understanding of the principles underlying the complex organization processes during development and the life cycle of organisms requires the identification and the understanding of the well orchestrated interplay of the functional building blocks such as genetic and metabolic networks, whole cells, or organs on many times and length scales. For example, the behavior of cells cannot completely be understood without understanding the principles underlying their intracellular regulation processes that determine their behavior and their properties. On the other hand the intracellular regulation processes have emerged as a consequence of selection on the phenotypic level of cells and organisms, so an understanding of the principles underlying the intracellular regulation processes should also involve the evolutionary time scale. Biophysical and mathematical models have been proofed useful to explain growth and migration phenomena of cells in many different biological systems such as the aggregation and pattern formation in bacteria populations ([35], [34], [3]), and *Dictyostelium discoideum* (e.g., [5], [24], [29]), convergent extension ([38]), blastulation and gastrulation ([28], [6], [9]), and even the growth of cell populations and tumors *in vitro* (e.g. [10], [8], [27], [16], [17]). They often fail to make predictions, however, for situations which have not been experimentally studied. This may

reflect the lack of models to consider the changes of cell behavior dictated by the intracellular regulation (and differentiation) machinery. Most models of multicellular organization consider cells as "simple" physical particles, and usually fail if cells actively change their behavior, or their biophysical or cell-biological properties either as a response to environmental changes, or interactions among each other, or due to an internal clock. For example, a change of the nutrient conditions of bacteria can completely modify the way a bacterium moves or divides [3]. An animal cell can completely change the way it moves as a consequence of physiologic differentiation or neoplastic de-differentiation, i.e., show a transition of its phenotype for example from a mesenchymal to an amoeboid movement [15]. The migration phenotype is observed to depend on the cells' ability to acquire polarity, pseudopod protrusion, cortical stiffness, cytoskeletal contractibility and dynamics, as well as physical contact with ECM environment. Further factors that play an important role in determining the cells' migration phenotype are the down-or up-regulation of cell-cell and cell-surface receptors, of proteases, and the ability of a cell to decipher external signals that direct cell movement ([15] and refs. therein). This change of behavior must be a consequence of internal decisions due to "rules" that are encoded in the genetic and metabolic information units of the cell, which have been selected on the time scale of evolution. Hence a model that permits prediction of individual cell or multicellular behavior should combine a description of a cell with a description of the rules that dictate the change of its behavior or parameters. A step into this direction has already been suggested by Hogeweg ([19], [20]). In these references the generation of morphogenetic (multicellular) structures on the time scale of development are studied that arise from cell-cell interactions and cell differentiations due to rules that appear as a consequence of an (artificial) evolutionary process. Within computer simulations with this model generic mechanisms could be identified that have been found in nature such as engulfment, growth meristem, intercalary growth, and intercalate and stretch.

In this paper we illustrate by *in silico* simulations the emergence of cell migration strategies, if either the cells individually, or populations of cells are subject to selection on particular properties in an artificial evolution process, namely the ability (1.) to locate a signal source, and (2.) to form aggregates of two or more cells. The migration of our model cells are controlled by artificial regulation networks that are encoded by binary sequences (the genotype). We carefully analyze the emerged boolean networks and find they show many features that have been observed in biological networks such as neutrality and robustness. The studied biological situations (1.) and (2.) occur for example during the life cycle of *Dictyostelium discoideum*. The Dictyostelium cells aggregate towards several distinct aggregation sites by walking up concentration gradients of cAMP [32], [18]. This type of behavior is known as "chemotaxis". The aggregation sites are determined by the distribution of amoeba. Neighboring cells respond to cAMP in two ways. They initiate a movement towards the cAMP pulse and they release cAMP on their own. After this the cell is unresponsive to further cAMP pulses for several minutes. Eventually, all cells aggregates to

a single center. On the other hand, in the absence of a morphogen, the cells perform a random movement. Directed migration and aggregation of cells are central events and occur for example also during wound healing, where fibroblasts migrate into the wound and form a network which is then gradually filled by cell proliferation, during the guidance of monocytes (and macrophages) and neutrophils as a response on chemokines secreted, for example, by T_H1 cells at infection sites. Macrophage development is a typical representative for a cell that undergoes a change in its migration phenotype during its development. It transforms from an amoeboid stem cell in the bone marrow to a macrophage by differentiation which is characterized by a significant up-regulation of certain integrins and by a re-arrangement of the cytoskeleton from diffuse to strongly focalized [15]. The inverse transformation is also observed and is accompanied by the abrogation of pericellular proteolysis, the strengthening of RHO/ROCK signal pathways, and the weakening of integrin-ECM interactions. Our main motivation was to develop a model concept that allows the assessment of which cell function (here: migration strategy) would be best adapted to a specific situation (here: detection of a signal source and formation of multicellular aggregates) either from a set of pre-given alternatives, or as an arbitrary combination from a set of pre-given alternatives, or by exploration of a novel strategy. This also includes the switching between different modes of behavior (migration strategies) controlled by rules of the regulatory network if the environmental condition change. For this purpose we apply the concept of an evolutionary reactor on *in silico*-single cell-and multicellular systems. The same concept has been frequently considered for the evolution of molecular systems (e.g. [12], [13], [14], [31]). We start with a defined number of system copies ("species") each consisting of one or many cells. A species is characterized by a number of binary strings that corresponds to the number of cell types permitted in each system copy (for a detailed explanation, see next section), We distinguish between different situations. Most of our simulations assume that initially the binary strings of different species (i.e. the cells in different system copies) are chosen independently and at random, equally distributed in sequence space. Hence initially the simulation starts with "species" from many points in sequence space. We find that this choice of initial conditions accelerates the convergence of our algorithm. We find the same species if we start with a population of cells with the same networks and study their Darwinian evolution. Thereby we believe that our approach firstly represents a modeling strategy which does not determine a specific cell function within a mathematical model *a priori* but allows for an investigation of potential cell functions, and secondly permits to study Darwinian evolution with only minor modifications of the parameters and initial conditions within the same modeling scheme.

What we also hope to illustrate in this paper how morphogenetic and evolutionary processes may be related in an conceptional approach. For this purpose we also studied an example where all system copies in the evolutionary reactor initially are identical and study how the species evolve. However, we like to emphasize that it is not our objective to "build an animal" and to reproduce minute details of cell

aggregation, or to provide a 1:1 picture of the different hierarchies of intracellular molecular organization including chromosomes, exon/intron structures, genes, proteins, and protein networks, or of the evolutionary process that has lead to it. Despite the spectacular progress in molecular biology, biochemistry and biophysics quantitative information on these processes is scarce. So a model reproducing every aspect of a specific developmental process and how it emerges from evolution cannot be realistic and would imply too many unknown parameters.
Our objective is to link processes on many lengths and time scales in a (largely simplified) multi-scale approach. However, the framework we use may be readily applied to each of the above individual-cell based models. In addition our model concept may also provide a potential concept to understand how building blocks and principles underlying gene regulatory networks may have emerged during evolution which is a topic of current interest [25].
Similar approaches are presently under consideration for biomimic systems. Here the main scope is to regulate movement and perception of robots by circuits that are built on biological principles (e.g. [37], [11], [30]). In these approaches neural networks form the control unit of robots. In this paper we study instead Boolean network with discrete states as a simple model for genetic regulation. This has two advantages. Firstly Boolean networks can directly be encoded by binary sequences which may be considered as direct representation of the genes. Hence mutations or crossovers can directly be linked to the genetic level. Secondly Boolean networks permit an unambiguous classification due to the discrete states of their elements. The simplicity of the model thus allows for a detailed analysis of all functional hierarchies that appear in the model.
This paper is organized as follows. In the next section we briefly summarize important technical details of our model. Then we present results to the detection of signal sources by single cells, and for cell-cell aggregation.

2. Model

In order to study the evolution of migration strategies of cells we need model representations of 1. space, 2. an individual cell, 3. intracellular regulatory network, 4. the evolution process during which the network evolves. We model cells as point objects on a d-dimensional square lattice with boundaries. In order to take into account excluded volume interactions between cells we assume that one lattice point can be occupied by only one cell. Cells can move to each of the $3^d - 1$ surrounding neighbor sites (Moore neighborhood). The intracellular regulatory network is represented by a Boolean network (BN). This certainly is an oversimplification in many biological situations but it is noteworthy that despite of their simplicity and shortcomings, Boolean networks have been successfully used to model the gene regulatory network in a number of biological systems as e.g. *Drosophila melanogaster* [1]. A Boolean network consists of elements (e.g. genes) which can be either ON (expressed) or OFF (not expressed) [21]. We assume our Boolean networks have three types of

elements, a set of input elements $\{I\}$ which allow the cell to sense its environment, a set of internal elements $\{E\}$, which allow to establish a memory, or directly affect the migration behavior of the cells, and a set of the output elements $\{O\}$ which do not influence the internal or input elements. Note, that the internal and output elements together determine the migration action of a cell. In Dictyostelium or neutrophils the input elements may summarize the cell surface receptors that permit the cell to sense the local strength of a signal and the machinery that permits the cell to translate this into a gradient information [36], [22]. On the models' level of abstraction, the output elements summarize the RHO-family GTPase that controls the actin polymerization in the protrusion, further modules of the protusion machinery, and the traction-generating machinery of the cell. The direction of cell migration can also be influenced by non-diffusible chemical cues attached to the ECM or to the surface of cells [2]. The internal elements allow the cell to build up a memory such as a positive feedback loop observed during microtubule orientation and actin polymerization in the protrusive region of the cell, being responsible for persistence of cell movement.

A rule table determines how the network state in the next point of time, $t + \Delta t$ develops as a function of the network state at the present point of time t. Δt is a fixed number, denoting the time period between two successive network updates and can be set to $\Delta t = 1$ without loss of generality. The network is updated synchronously. For the output element j, $O_j = f(\{I\}, \{E\})$, for the internal element r, $E_r = h(\{I\}, \{E\})$ $(j = 1, ..., o,\ r = 1, ..., e)$. Here o is the number of output and e the number of internal elements. Be further i the number of input elements. Then, O_j $(\forall j)$ and E_r $(\forall r)$ are both binary strings of length 2^{i+e}. One boolean rule table in this case is characterized by $L = (e + o)2^{i+e}$ states. By concatenating the states of internal and output elements in the form $S = \{E_1(I_1, I_2, ..., E_1, E_o, ...)$, $E_2(I_1, I_2, ..., E_1, E_o, ...)$, ..., $O_1(I_1, I_2, ..., E_1, E_o, ...)$, $O_2(I_1, I_2, ..., E_1, E_o, ...), ...\}$ the rule table can be easily encoded in a linear bit-string of length L. Modifications in the bit-strings correspond to changes of the network rules. The total number of different rules is $N = 2^L$.

Our basic simulation schemes are designed to obtain a fast convergence of the artificial evolutionary process to an optimal solution. However, Darwinian evolution is a special case of our algorithm by appropriate parameter settings as is explained below.

Basic simulation schemes: (see Fig. 1; for the sake of clarity we formulate the algorithmic schemes for the evolutionary process (I.) and the fitness evaluation (II.) for one cell as they are used in the studied situations of subsection 4.1; however they are readily extended to many cells):

Scheme I: evolutionary process (Fig. 1a):

(1) Define a fitness function F which encodes the conditions to which a cell has to adapt. F depends on the studied biological situation (example: eqn. (1)).
(2) Consider μM copies of the biological system and subdivide them into M sub-

populations of size μ (in our simulations, $\mu = 20 - 80$, $M = 4$). Each copy consists of a cell and its environment. A cell is characterized by its binary string of length L that encodes the rules and topology of its regulatory network (see above). The evolutionary process operates on the binary string of a cell only. Therefore the fitness of a binary string can be identified with the fitness of the cell which is characterized by that binary string. Generate a random binary string of length L independently for each of the $M\mu$ cells with a "0" or "1" with equal probability at each of the L positions of the binary string. The $M\mu$ binary strings determine the parental population P_p.

(3) Evaluate the binary strings of all cells in the parental population by investigating the fitness of the cells (see Scheme of the evaluation process below; for this only the binary string has to be transfered to the subroutine that investigates a cells' fitness). The evaluation process is based on the behavior of the cell on the developmental time scale as depicted in Fig. 1b and explained separately below.

(4) Exchange randomly selected binary sequences between the M subpopulations with probability p_e. (This step serves to accelerate the convergence of the algorithm and in principle may be omitted.)

(5) Select $M\lambda/2$ pairs of binary strings from the parental population P_p randomly such that the selection probability for a binary string is taken proportional to the fitness value of the corresponding cell (roulette-wheel). Between each pair of binary strings cross-overs occur with probability p_c. The resulting $M\lambda$ binary strings determine the filial population P_f. (Note that due to the selection process binary strings with higher fitness in the parental population have a larger chance to have offspring in the filial population.)

(6) Mutate all binary strings in P_f with probability p_m. (We have also performed simulations where we allowed for cross-overs in addition to mutations but did not find noteworthy differences to the case without cross-overs.)

(7) A new parental population is formed according to the following rules:

- Select the Mn best binary strings (the elite) from the parental population P_p, mutate them with probability p_m, and store them in P_e ($n \leq \mu$).
- Select $M(\mu - n)$ binary strings from the filial population P_f according to the roulette wheel procedure and add them to the Mn binary strings of P_e. Now P_e consists of $M\mu$ binary strings.
- Replace the strings in P_p by the strings in P_e.

(8) Continue with step 3. until a certain stop criterion is fulfilled. The number of iterations ($= g$) determines the evolutionary time.

Scheme II: evaluation process of a binary string:

(1) Initialize the fitness to $F = 0$.

(2) Initialize the state of each element of the Boolean network to 0 (in order to start from a well defined initial state; however, we confirmed for selected parameter

settings that the results of the evolutionary process do not depend on the precise choice of the initial network state).
(3) Place the cell on one of k possible initial positions on the lattice (see below).
(4) Initialize the lattice properties (e.g., set the morphogen concentration on the lattice to zero etc., see below.)
(5) Perform the following sequence of steps T times:

- Determine the state of all input elements of the BN of the cell.
- Apply the Boolean rules (represented by the binary string under evaluation) to update the state of the internal and the output elements.
- Perform cell actions according to the state of the output elements.
- Update the fitness F (for example according to eqn. (1)).
- Update lattice properties (for example according to eqn. (2)).

(6) Continue with step 2. until the cell has been set to each of the k initial position exactly once.
(7) Return the fitness value F for the cell to evolutionary program (see step 2. of the scheme of the evolutionary process).

The motivation for the initial choice of random strings in step 2 of the scheme of the evolutionary process (which is always used throughout this paper if the opposite is not stated explicitly) and of the selection steps 4 and 6 of the scheme of the evolutionary process was to find the "fittest" cells (with the best adapted strategies) efficiently, avoiding the algorithm to stick in local fitness optima. In case the main motivation is to model a gradual evolution process from one point in sequence space it can be favorable to start instead with an initial cell population where all Bit-strings are the same and choose the parameters of the evolutionary algorithm in such a way that Darwinian evolution is modeled. For this purpose we also performed simulations with the parameter settings $\lambda = \mu$ and $n = 0$ which corresponds to a general replacement of the ancestor population by the progeny of the fittest individuals from the ancestor population. In subsection 4.1 we also give an example for the outcome of our algorithm if initially all $M\mu$ cells start with the same binary string which encodes a random walk movement.
In case the biological system consists of a population of many cells (as in subsection 4.2, i.e., if each copy of the biological system in step 2 of scheme I consists of many cells) then the number of bit-strings that are considered to be subject to the evolutionary process within each copy depends on the number of permitted cell types. If one looks for a population of one cell type only, still only one bit-string needs to be considered as for the above example of one cell. If one looks for as many cell types as cells in the cell population, then for each cell of the population one bit-string has to be considered as being subject to evolution.

3. Analysis tools

Analysis of Networks: We analyzed the connectivity of the boolean elements by M-analysis [33] in the population obtained at convergence of the evolutionary process. This permits an identification of the key elements and key links within a network for each (functional) phenotype.
Analysis of Robustness: Usually the final population consists of different species. Some of them encode the same function with a different network. In order to study the robustness of the networks that have emerged in the evolutionary process, we performed in-silico knock-outs and analyzed the resulting network phenotype. Here, we define a "knock-out" by setting the state of an element to zero.
Characterization of selected species: In order to characterize the population of species on the evolutionary time scale we measure the average, minimum, and maximum fitness of the evolving population as well as the skewness and kurtosis of the fitness value distribution. In order to classify the migration strategies used by the cells of the population at different stages we analyze the regulation networks and the phenotype of the species.

4. Results

We consider different evolution strategies (Darwinian evolution, accelerated artificial evolution), different biological situations (one cell searching for a signal source, many cells searching for a signal source, cell-cell aggregation) and different degrees of network complexity in different spatial dimensions ($d = 1, 2$).
In the first part we consider the most simple problem, the searching of a single cell for a signal source in $d = 1$ dimensions since this problem already shows most of the features that we found for the more complex problems and in higher dimensions. Then we consider migration in two space dimensions which is natural for cells such as for example Dictyostelium cells. Note, however, that the one dimensional situation could already be studied in an appropriate experimental setting. Finally we study the aggregation of many cells and situations in three dimensions.

4.1. *Cells searching for a signal source*

Known different types of cell movement include chemotaxis, haptotaxis, galvanotaxis, contact guidance, thermodynamical types of interactions, or random walk-like movements [18]. The usual way is to include them into mathematical models is to directly specify the type of movement in a certain biological situation in the equations or rules *a priori*. In contrast to this approach we assume cells have to establish a migration strategy in an artificial evolutionary process which enables them to find a signal source located at $\mathbf{x} = \mathbf{0}$. The fitness function F is given by

$$F = \frac{1}{aT} \sum_{i=1}^{a} \sum_{t=0}^{T-1} \left\{ \delta(\mathbf{x}^{(i)}(t)) \right\} + F_2, \tag{1}$$

where $\mathbf{x}^{(i)}(t)$ denotes the position of the cell at time t in the i-th run, where $i = 1, 2, ...a$ determines the different initial positions, and $\delta(\mathbf{x}(t))$ denotes the Kronecker symbol (i.e., $\delta(\mathbf{x}(t)) = 1$ if $\mathbf{x} = \mathbf{0}$, and zero otherwise). F measures the fraction of time a cells stays on the signal source. I.e., the larger is the time a cell spends on the signal source, the larger is its fitness F. $F_2 = 1$ if the cell is able to find the signal source from all initial positions and zero otherwise. This insures that cells that finally find the signal source from all initial positions get a benefit.
We classify the migration strategies in a two step process: Firstly we verify that a network encodes a potentially successful strategy, i.e. stops on the signal source once it arrives at it. For those networks that encode potentially successful strategies we classify the network rules that the cell uses on its search process (the migration towards the signal source) with respect to the strategy that they encode. (Hence mutations that affect elements which are not used by a cell do not affect the phenotype and hence are neutral.) Often we find cells that use network elements which are characteristic for more than one strategy (e.g., random walk and chemotaxis, see below) although their migration is dominated by a specific strategy so the phenotype appears to be (almost) the same as for a pure strategy and consequently has almost the same fitness. We introduced a threshold value $\Theta \in [0.5, 1)$ that allows to determine the relative contributions of each strategy that appears in cells that use a mixture of different strategies. If a cell uses a certain strategy for more than $(\Theta \times 100)\%$ of its steps towards the signal source it is attributed to this strategy (note that only for $\Theta = 1$ a cell uses a strategy on the whole way to the signal source, i.e. uses a pure strategy.) In case no or more than one migration strategy could be identified for a cell by the described procedure we classify the behavior strategy of the cell as "mixed". However, the assignment of a cell to more than one strategy is very rare and occurs only for Θ close to 0.5.
In order to give a clear illustration of the underlying model concept we in this subsection mainly focus on a one-dimensional lattice. At the end of this subsection and in the next subsection we consider a $d = 2$-dimensional lattice.

4.1.1. *One space dimension:*

We study $K = 3$ initial positions, and varied T, the mutation rate p, the initial distance x_0 of the cell from the signal source, and Θ. The other parameters were kept constant (see tables 5, 6 in Appendix). $x_1(0) = -x_0$, $x_2(0) = 0$, and $x_3(0) = x_0$, with $x_0 \in \{4, 7\}$, and choose $T \in \{30, 200\}$. The optimal fitness F_{opt} can be calculated analytically for each parameter set (x_0, T) to $F_{opt} = (2(T-(x_0-1))+T)/(3T)+F_2$. For $x_0 = 7, T = 100$ $F_{opt} = 0.96 + F_2$ as confirmed in table 2 (note that $F_2 = 1$ this the value for F_{opt} implies that the cell stops on the signal source). This corresponds to a cell that on one hand directly walks from the starting points $x_1(0) = -x_0 = -7$, $x_3(0) = x_0 = 7$ to the signal source and stays on it, and on the other hand does not move at all, if its starting point was the position of the signal source, i.e., for $x_2(0) = 0$.

Table 1 shows the network we included into a cell which allows the cell to establish a number of different optional strategies: (1.) deterministic straight movement independent of any signal molecule concentration (*SD*: straight deterministic), (2.) return strategy (*RP*: establishes one or more return points), (3.) chemotactic movement, i.e., a deterministic movement into the direction of the local morphogen gradient (*CHT*), (4.) random walk (*RW*), or (5) other, mixed strategies (*mixed*).
We have also studied species which contain only sub-networks of table 1 which permitted only a subset of the strategies (1.-4.). Before we report on the results of the network shown in table 1 we briefly summarize some major results on these simulations.
The SD-strategy only allows the cell to access the signal source either form $+x_0$, or $-x_0$, but not both. All other strategies are in principle suited to ensure that a cell is able to detect the signal source from all initial positions, but the time the cells need to find the signal source differs for the different strategies. For the random walk-strategy this time scales $\propto x_0^2$. In general the RW-strategy works only with probability one in $d \leq 2$ dimensions (since the return probability is one only in $d \leq 2$ dimensions). An alternative successful strategy invented during the evolutionary process is to introduce a return point (*RP*) (Fig. 2). Here a cell uses the internal elements to count the number of steps it has performed into a certain direction, and returns if the signal source has not been found. The maximum distance a cell is able to travel before it returns is $2^e - 1$. However, for large distances from the signal source, and in particular in $d > 1$ dimensions the RP-strategy (or any other deterministic search strategy) requires a large network with many correctly linked elements in order to permit complex strategies and hence is unlikely to emerge.
Cells that are attracted by signal sources or form aggregates often use long range informations encoded in a morphogen gradient. We assume the signal source secretes signal molecules which are able to spread by diffusion with rate D and decay with a rate γ. The equation for the local morphogen concentration reads

$$\frac{\partial c(\mathbf{x})}{\partial t} = D\frac{\partial^2 c(\mathbf{x})}{\partial \mathbf{x}^2} + \zeta \cdot \delta(\mathbf{x} - \mathbf{x_0}) - \gamma \cdot c(\mathbf{x}), \tag{2}$$

where $\mathbf{x}_0 = 0$. In our simulations we numerically integrate the equation using the explicit Euler method.

Phenotypes: Fig. 3 shows the proportion of species that adopt one of the five types of behavior explained above if the networks (cf. table 1) permit to adopt all the five above defined strategies for two different initial choices of the binary strings in the evolutionary process.

In Fig. 3a, the binary string for each of the $M\mu$ cells was initially chosen randomly and independently for each cell (see step 2. of scheme I in section 2). In Fig. 3b all

$M\mu$ cells initially start with the same binary string where the binary string encodes a pure random walk movement. In the latter case, after ~ 10 generations the initial fraction of random walkers decreases from $F_{RW} = 1 \rightarrow F_{RW} \approx 0$ while at the same time the species that perform mixed strategies increase to a value of almost one. At large g the dominating strategy is chemotaxis followed by mixed strategies for both initial choices of the bit-string distribution, compare Figs. 3a,b.

Table 2 shows the maximum, average and minimum fitness of the different strategies in Fig. 3a. The average fitness of the chemotaxis strategy is $F_{CHT} \approx 1.96 = F_{opt}$. The mixed strategy which is the second frequent strategy has the second largest average fitness. For the straight movement (SD) the average fitness is close to the optimum fitness value for straight movement strategy which is $F_{opt}^{SD} = (T - (x_0 - 1) + T)/3T = 0.64\overline{6}...$ for the parameters of Fig. 3. Nevertheless it is even transiently a rarely used strategy probably since the number of different networks that encode mixed strategies is much larger (compare table 3). Hence, the fraction of strategies does not directly reflect the fitness of the strategy but also how probable it is, that a species with this strategy emerges. As we will see below, a further important aspect is robustness: strategies that require a high degree of organization within the Boolean network, such that closely related networks in sequence space do not encode the same strategy anymore are less likely to occur than it would be expected from their fitness. Fig. 4(a) shows how the maximum, average and minimum fitness for the population of Fig. 3(a) evolves. The figure suggests that the convergence of $f(g)$ is determined by the average fitness which converges at $g \approx 400$; the maximum fitness converges slightly faster than $f(g)$. Fig. 4(b) shows the variance, skewness and kurtosis for the parameters of Fig. 3(a). All values converge become stationary at $g \approx 200 - 400$. The skewness changes from a positive to a negative sign at small $g \approx 1$. This becomes immediately obvious if one looks at the fitness distribution (Fig. 5) which is peaked as small fitness values for small g and at large fitness values for large g. The kurtosis has a minimum at $g \approx 1$. The positive sign at large g reflects the leptokurtic character of the distribution. The distribution at large $g \approx 20$ is bi-modal with a higher peak at $F - F_2 \approx 0.96$ and a smaller peak at $F - F_2 \approx 0.68$.

In Fig. 6 we vary the parameters x_0, T, p and Θ for initially random bit strings (i.e., the initial condition is the same as in Fig. 3(a)). The typical scenario in Fig. 6(a-d) is the same as in Fig. 3(a). With increasing generation g the fraction of species which use 'mixed' strategies decreases while the proportion of species that perform a random walk is (partly intermittently) increased (at small g in Fig. 3(a), Fig. 6(a),(c),(d); in 6(b) a detectable RW-peak did not form). Eventually pure chemotaxis which has the largest fitness becomes the most adopted strategy. The underlying mechanisms for the latter are twofold. Firstly cells performing mixed strategies may perform chemotaxis after a small number of mutations that either result in the use of network elements that are characteristic for chemotaxis or eliminate the use of those network elements that are not characteristic for chemotaxis. This line of argument is supported by the observation that some of the cells that

use a mixed strategy have a fitness close to the fitness of a chemotactic strategy (see table 2) and the fraction of networks that encode mixed strategies is by far the largest (see table 3). Secondly, since the probability that cells have offspring is proportional to their fitness, cells that already perform chemotaxis have a larger probability to form offspring that also perform chemotaxis. However, as a consequence of the large fraction of networks that encode mixed strategies the proportion of cells that perform mixed strategies remains large. An example for a mixed strategy are cells that transiently perform a random walk, depending on the states of their internal elements and eventually drift to the signal source by chemotaxis where they stop. The reason for the intermittent peak of random walkers which switch to a deterministic walk and stop if they reach the signal source is, that the sub-network they use is relatively simple, i.e., does neither require many elements, nor a high degree of connectivity. The network for random walkers only needs $D = 1$ if $A = 0$, and $D = 0$ and $F = 0$ if $A = 1$ (see table 1). However, their fitness is small so for large g the RW strategy almost disappears (Fig. 3a, Fig. 6a-d). Decreasing x_0 (Fig. 6a) and increasing T (Fig. 6b) both increase the intermittent peak of random walkers and decrease the asymptotic fraction of cells doing chemotaxis. In both cases the relative difference in fitness between the different strategies for which the cell eventually stop on the signal source becomes smaller, since the cells then spend more time on the signal source than to find the signal source. The main effect of a decrease in x_0 is an increase of the cells that use the RW-strategy since the fitness of random walkers depend stronger than linear on x_0 (the time a cell needs to walk a distance of x_0 is $\propto x_0^2$ for the RW strategy). A reduction of Θ from 0.9 (Fig. 3a) to 0.5 (Fig. 6c) results in a classification of most mixed strategies into either the chemotaxis or RW-strategy, hence both fractions increase.
A decrease of the mutation rate p increases the number of generations until chemotaxis can be established, but at the same time increases the fraction of species doing eventually chemotaxis (Fig. 6d). The reason is that a mutation of states that phenotypically result in mixed strategies of slightly smaller fitness than for chemotaxis becomes rare. On the other hand, increasing p reduces the fraction of cells that eventually perform chemotaxis largely since chemotaxis requires a well-orchestrated interplay of network elements and hence is sensitive to perturbations by mutations in the bit-string that encodes the regulatory network.

Network analysis: Fig. 7 reflects the contributions of the individual network elements to the cell phenotypes after the fitness has saturated with a fitness distribution as in Fig. 5. Fig. 7(a) shows the fraction of individuals that were still able to detect the signal source after knockouts of $1, 2, 3, 4, 5$ elements, Fig. 7(b) the wiring diagram generated based on an analysis of the mutual information (see methods) that measures the information transfer between network elements at successive points of time. Fig. 7 represents an average over $N = 16000$ individuals. In almost all of the $N = 16000$ individuals, A influences F. If $A = 1$ (source found), then $F = 0$ (cell stops migrating). The states of B, C encode the gradient and

determine the state of element E which determines the direction of movement. If the cell is on the signal source, the gradient vanishes and $B = 0$, $C = 0$. Hence, the information on whether the signal source is found in the presence of a signal molecule is encoded in the elements B, C as well. Accordingly, some of the species do still maintain their strategy at constant fitness after element A is knocked out (Fig. 8a). In our networks, $B = 1, C = 0$ ($B = 0, C = 1$) corresponds to $\partial c/\partial x < 0$ ($\partial c/\partial x > 0$); $c(x, t)$ is the local morphogen concentration. Note that cells that use element A to detect whether they have found the signal source or not, only need to determine positive or negative gradients (but not a zero gradient since this is equivalent to being on the signal source) and hence in principle need only either element B or element C in addition to A. Only those cells which do not use element A need both, elements B and C to determine the gradient information and whether they have arrived at the signal source. Many cells use A and both, B and C, or only B if they approach the signal source from one, and only C if they approach it from the other direction. In the presence of a gradient neither the internal elements nor the elements that permit a random movement are crucial and their influence on the direction of the movement or on halting is negligible. If B (or C) is knocked out, those cells that use A and C but not B (or A and B but not C) may still able to maintain a successful phenotype which is able to detect the signal source and remain on it. However, the phenotype is in general not the same as before the knockout has been performed. To see this, consider the case where B is knocked out, i.e., $B = 0$. Then, in case $\partial c/\partial x < 0$, $B = 0$ and $C = 0$ while $A = 0$. I.e,, $B = 0$ and $C = 0$ indicates no gradient when $A = 0$ indicates that the signal source has not been found; hence a contradiction in the model setting in which the gradient vanishes only if the cell is positioned on the signal source. The cell cannot distiguish whether the information that $B = 0$ is a consequence of a dys-function (here due to the knockout) or if $B = 0$ encodes a true information hence the cell cannot resolve the conflicting information between $B = 0$, $C = 0$ on one hand and $A = 0$ on the other hand. So the cell must make its moves independent of the state of element B. The cell may, however, use element C to determine the gradient in case the gradient is not zero and ignore the state of elements B, C in case $A = 0$. This requires a particular arrangment within the boolean transition rule table. Hence whether after a knockout a cell is able to maintain its phenotype largely depends on the very particular organization of the other elements. The cell may also ignore the state of the elements B, C in case $C = 0$ (which without a knockout would correspond to $B = 1$) and use instead a random walk if $\partial c/\partial x < 0$ (which is usually encoded by $B = 1$, $C = 0$), while for $\partial c/\partial x > 0$ it may still use the states of elements B, C to determine its direction of motion. In the former case the phenotype is as the unperturbed phenotype, in the latter case the phenotype corresponds to a random walk if the cells starts on the rhs. of the signal source, and a directed movement if it starts from the lhs. (compare peaks in Fig. 8(c)).

We find that the ultimate test of whether an element is indispensable or not is given by assessing the phenotype after the respective element is knocked out. The

network analysis by the mutual information score if applied to the observed network transitions does not permit to assess whether an element is indispensable. The condition that the mutual information MI transfered from one, or a set of elements to another element is larger than zero ($MI(...) > 0$) tends to over-estimation the influence (Fig. 7(b)). The reason is that many hypothetical wirings, although detected by the mutual information, do not affect the phenotype and hence are not subject to selection. If they are knocked out, there is no effect on the phenotype. Part of this ambiguity between the assessment of the phenotype by the MI and knockout experiments is also a consequence of the fact that not all network transitions are observed hence the data are incomplete. In this case, the criterion that an influence exists if the mutual information score is larger than zero is too weak and should be replaced by a χ^2-test (for a detailed discussion of reconstructing network from incomplete data including this issue, see [26]). However, even in case a χ^2-test is applied it cannot be excluded that the significance level is chosen not properly to detect relevant wirings and ignore irrelevant ones. In order to obtain an qualitative estimate for the influence of a non-zero threshold we analyzed our networks assuming that wirings between network elements are accepted only if $MI(...) > s$ for several $s \in [0, 1]$. We found that increasing s from $s = 0 \rightarrow s = 1$ firstly results in a slight decrease of the detected wirings in the network until a threshold value at $s \approx 0.25$, above which the number of wirings immediately become very sparse. The wirings at $s < 0.25$ represent phenotypes that are close to the picture obtained from simulated knockout experiments. Note that applying REVEAL in its classical form (see [23]) underestimates the influences since the classical criterion of REVEAL to determine the complete influence (which is (mutual information/entropy)$= 1$ [23]) does not suffice to identify all influences (Fig. 7(c)) since it works properly only if all transitions can be observed.

In order to assess the relevance of each element to the phenotype we knocked out all combinations of the input and internal elements and study whether after the knockout the cells were still able to detect the signal source (Fig. 7(a)). For knockouts of one, two, three, four and five elements we find that element A is the most indispensable element for a functioning phenotype, followed by elements B, C. Surprisingly, even the internal elements sometimes represent cell function; in $\sim 50\%$ of the single-element-knockouts they turned out to be crucial. However, in $2-9\%$, even 4-element-knockouts are tolerated, revealing a surprisingly large robustness of the phenotype. The preferred strategy in case of many-element knockouts is a random walk which stops as far as the cell is on the signal source. This strategy even works in case of five elements are knocked out, for example, if all internal elements and the elements B, C are knocked out.

Fig. 8 shows the fitness distribution in case of knockouts.

4.1.2. *Two space dimensions:*

Table 4 explains the network elements in $d = 2$. Figs. 9, 10 show typical examples of chemotaxis in $d = 2$, which constitutes the dominant strategy in $d = 2$ (as for $d = 1$). In Fig. 9 the cell does not move, in Fig. 10 the cell performs a random walk until it is able to sense the morphogen. In both cases the cell move straight towards the position of the signal source as far as it is able to sense the morphogen concentration. The fitness of both strategies is the same as long as the expansion of the morphogen concentration (by diffusion) is fast compared to the typical time a cell needs from the initial cell position to the signal source by a random walk strategy. In $d = 2$ our algorithm didn't find a return (or, any other successful deterministic search) strategy with a reasonable number of elements. Although such a strategy certainly exists (e.g. walking in a spiral-formed pattern around the starting position) it is unlikely to emerge since it needs many internal elements (so the search space is large) and a high degree of internal network organization (so the fraction of networks that encode a RP strategy is small). For example, in $d = 1$ for the network in table 1 the search space already contained $2^{384} \approx 10^{120}$ states and the fraction of networks that perform a RP strategy was very small (see table 3).
In two dimensions the search space is much larger than in one dimension which is why the fraction of individuals that are still able to detect the signal source after knockouts is significantly smaller than in one dimension (Fig.a 11). However, if knockouts are performed prior to evolutionary optimization, again many individuals can be identified that are able to detect the signal source and have a comparable fitness to those in which no element was knocked out (Fig. 12). As shown in Fig. 13 a knockout in $d = 2$ can affect in many ways in how far a cell is able to sense its environment. Still, the cell can find complex strategies to largely replace the function of the knocked-out element. Often, the phenotype of sub-optimal species (after knock-outs) is almost indistinguishable from the phenotype of optimal species due to including a complex mixture of stategies that involve internal elements, random moves and directed moves. For example, on a larger scale, the zick-zack-movement of the cell in Fig. 13 on a coarser scale would be almost indistinguishable from a purely random component in the cell movement. The phenotypic differences are expected to become even less pronounced if the cell movement is not limited to the sites of a lattice (i.e., in an off-lattice model).

4.2. *Cell-cell aggregation*

Our strategy also allows to study multicellular phenomena, as cell-cell aggregation. In this subsection we study a selection criterion which favors the formation of many cell-cell contacts. As observed for example in *Dictyostelium*, we assume that in principle each cell is able to secrete a morphogen (by an additional network element that, if it is ON, leads to the secretion of a signal molecule). The simulations were performed in $d = 2$ dimensions on a 2-torus, i.e., with period boundary conditions in x-and y-direction (Fig. 14). Our fitness criterion here is the number of cell-cell

contacts integrated over time, i.e.,

$$F = \frac{1}{aT} \sum_{i=1}^{a} \sum_{t=0}^{T-1} N_{cc}(t), \tag{3}$$

where $N_{cc}(t)$ is the number of cell-cell contacts at time t. In the absence of chemoattractants cells perform a random walk and stop as far as they have formed small aggregates (Fig. 14a). In the presence of morphogens cells adopt a chemotaxis strategy after a number of generations and walk up the local morphogen gradient. For sufficiently large D a single morphogen maximum forms to which all cells migrate (Fig. 14b) while a small D favors many small cell aggregates (Fig. 14c).
We were also able to find situations in which one cell secreted a morphogen while the other cells didn't but were attracted by the morphogen. Fig. 15 shows a pair of cells in which one cell secretes a morphogen while the other cell has learned to sense the morphogen and to walk up the morphogen gradient. All these scenarios are reminiscent to the situation found in *Dictyostelium*. Thereby our modeling strategy is able to generate different types of migration behavior that occur in different stages during cell-cell aggregation in *Dictyostelium* without having implemented them *a priori*.

5. Discussion

In this paper we studied simple examples for the formation of migration strategies of individual cells or groups of cells that are controlled by cell-internal regulation networks in a multi-scale approach that takes into account different time and length scales. We represented the regulation networks by Boolean networks. The networks rules were not determined *a priori*. Instead we used an artificial evolutionary process in order to allow for a selection for those network rules for which the cells best adapt to given selection criteria on the phenotypic level, such as the finding of a signal source or the formation of cell-cell aggregates. We first encoded the information of the boolean rule tables by binary strings, which encode the "genetic information". Then, within an artificial evolution process on the level of a cell or a population of cells, we used the machinery of genetic algorithms to exchange genetic information between the strings by crossovers, modify strings by random mutations, and reinvestigate the composition of a new population. Here, we mainly studied the case where each network initially had a random set of rules, but showed for an example that the asymptotic frequency distribution of different strategies were not affected by whether we started the evolution process with a random set of rules or with the same rule for all species. Due to the simplicity of our approach, a detailed analysis of the emerged networks and phenotypes were possible. We find, that our modeling strategy selects for species which migration strategies correspond to those found in natural biological systems. Moreover, we find the formation of simple strategies (with migration strategies that may be robustly encoded by simple sub-networks) takes place first. With increasing "evolutionary" time, more complex organisms,

represented by migration strategies with higher fitness on one hand, but a higher degree of organization within the network on the other hand, starts to dominate. We find simple and complex strategies coexist. Both aspects also occur in natural systems. The frequency with which different strategies are represented within a population depends on the fitness of the strategies, the degree or organization and robustness of the strategy-encoding networks, the fraction of networks that encode a certain strategy and on the mutation rate of the evolutionary process. Within our model, a higher degree of network organization was always accompanied with a smaller robustness against mutations, since mutations more easily affect functional elements or relationships in highly organized, than in simple networks in which the degree of organization is low. The classification of species turned out to be a non-trivial issue. The Hamming distance between the binary sequences that encode the boolean rule tables is no appropriate measure to classify "similar" behavior. The state of some elements, encoded by some regions of the sequence, are crucial to a particular strategy, others may be completely irrelevant hence mutation of the latter is neutral and does not affect the cell phenotype. We classified the strategies by analyzing the network elements and wiring that are known to be characteristic for a certain strategy for each species, for example by analysis of the mutual information. In addition, we performed knockouts in order to assess the importance of elements for a successful phenotype (e.g. if a cell is capable to detect a signal source in space and remain on it). The analysis of the network rules by the mutual information alone does not always lead to a correct assessment of the cell phenotype (the search and migration strategies). The ultimate test for whether a network element is indispensable or not for a given, or an alternative successful phenotype is by a knockout experiment, in which the respective element is knocked out.

We find most species adopt the phenotype with has the highest fitness (i.e. it uses the optimum migration strategy) but many species also adopt a sub-optimal phenotype (which performs mixed migration strategies). The evolved regulatory networks that control the phenotype (the migration strategy) show features that are known from gene regulatory networks such as robustness and redundancy which are closely related. We find that in general the cell phenotype is more robust to perturbations of internal elements than of input (or output) elements but even input elements can be replaced if the information can be sensed in an alternative way (here: direct detection of a signal source vs. detection of a gradient; hence the informations 'signal source' and 'no gradient' is redundant). However, knockouts can lead to an ambiguous input information for example, if a cell receives two contradictory input informations on the state of its environment as a consequence of a dysfunction of the sensing machinery; in this case the cell may not be able to distinguish which of the two competing input signals is informative and which may be a consequence of a dysfunction, and thus has no "fit" phenotype.

If the objective is to assess whether a certain (reduced) network is potentially able to encode a successful phenotype, an evolutionary optimization should be performed after a knockout, too. In should be noted, however, that the knockout strategy

Element	state 0	state 1	type
A	signal source not found	signal source found	input
B,C	encoding of gradient		input
D	deterministic movement	stochastic movement	output
E	$x \to x-1$	$x \to x+1$	output
F	stop	move	output
G,H,I	memory elements		internal

Table 1. Elements of the Boolean which allows the cells to establish a (1) a straight, (2) a chemotactic, (3) a return point, (4) a random movement, and (5) mixed strategies to find the signal source. Elements $A-C$ are input, $G-I$ are internal elements, $D-F$ are output elements. Element F is only evaluated if element $D=0$, element E is only evaluated if elements $D=0$ and $F=1$. $(B,C)=(0,0)$ corresponds to no morphogen gradient, $(B=1,C=0)$ to $\partial c/\partial x<0$ ($c(x,t)$: concentration of morphogen at position x at time t), $(B=0,C=1)$ to $\partial c/\partial x>0$. $(B,C)=(1,1)$ does not occur (note, that B,C are input elements). (We have also assessed species with internal elements in $d=2$ for selected runs but did not find significant differences to those without internal elements. However, the presence of internal elements increases the search space such that the emergence of successful phenotypes becomes very improbable and and their analysis almost impossible.)

does not necessarily allow to specify whether the knocked-out element has been used prior to the knockout; in case the same information is (or can be) encoded in another way, the phenotype may not have changed although the information flux within the regulatory network that controls the phenotype has changed. In this case the mutual information is more useful.

There are many directions into which the work may be extended. For example one could allow the regulatory network to grow or shrink during the evolutionary process, i.e, increase or decrease the number of elements during the evolutionary process. Since this is naturally expected to result in continuously growing networks, this increase must be accompanied by an increased need for nutrition, the latter constituting the "penalty term" in order to avoid "over-fitting". However, since we found internal elements are those which were the most dispensable ones we do not expect that this would significantly modify our results in case the selection pressure is as in our simulations. Our concept may in a further step be generalized to other representations of gene regulation networks [7]. Furthermore our concept may be applied to very specific networks for which the precise function is known to predict which network changes would be expected in case selection on another function would happen. Thereby it may be a starting point from which also the emergence of organization in networks may be studied, which now becomes also accessible to direct experimentation [4].

strategy	max. fitness		avg. fitness		min. fitness	
no realizations	1	20	1	20	1	20
straight movement	0.647	0.647	0.647	0.647	0.647	0.647
random walk	1.76	1.755	0.775	1.6870	0.317	0.367
return point	1.953	1.951	1.731	1.811	0.317	0.350
chemotaxis	1.960	1.960	1.960	1.960	1.917	1.919
mixed	1.960	1.955	1.674	1.746	0.023	0.062

Table 2. Fitness values for different strategies for the parameter choice of Fig. 3a. For "no realizations" = 1 a cell in the evaluation process started from each initial position exactly once (cf. Fig. 1). For "no realizations" = 20 we started the cell from each initial position 20 times and investigated the fitness from the average over the 20 realizations. This reduces random fluctuations in the fitness for those strategies that contain proportions of a random movement. As reflected by the average fitness for random walks the effect of averaging over many realizations is particularly important to properly consider the effect of F_2, i.e., whether a cell finally finds the signal source after a fixed time ($F_2 = 1$) or not ($F_2 = 0$).

x_0	straight movement	random walk	return point	chemotaxis	mixed
7	0.000557	0.017842	0.001620	0.000037	0.979977
4	0.000682	0.018516	0,002144	0.000011	0.978646

Table 3. Fraction of networks that is classified to a particular strategy for $T = 100$, $\Theta = 0.9$ for random sampling of 12×10^7 networks for $x_0 = 7$ and $x_0 = 4$.

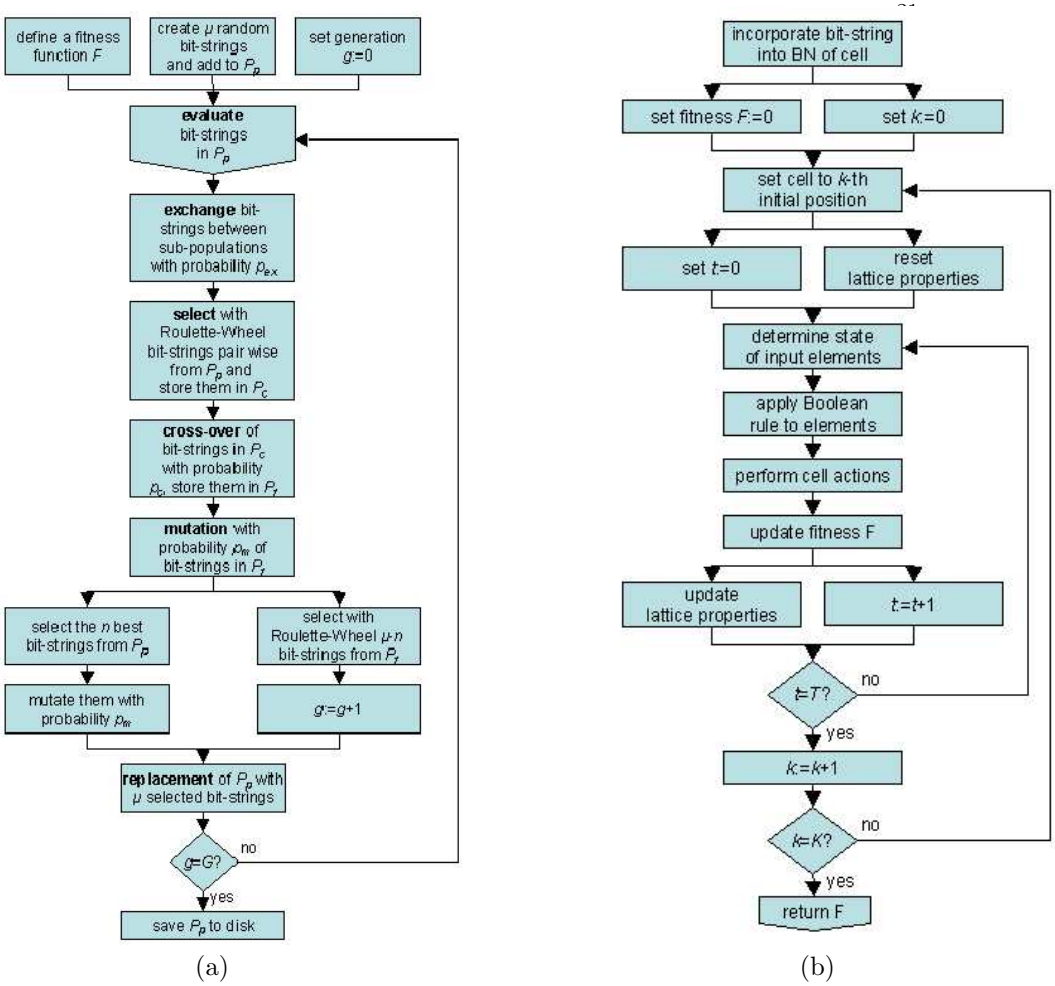

Fig. 1. Schematic illustration of the simulation steps carried out during (a) the evolution process, (b) the evaluating process.

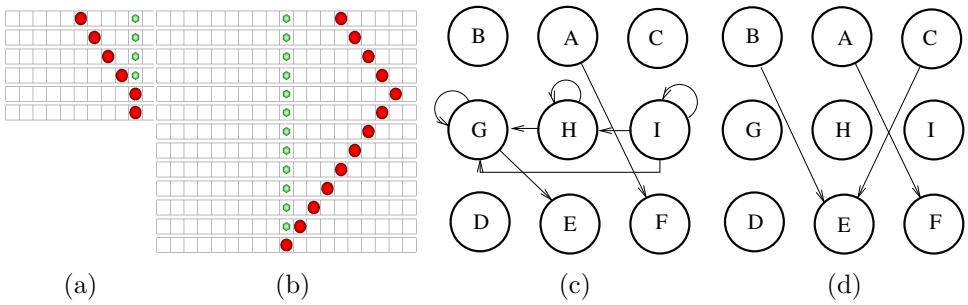

Fig. 2. Finding of a signal source based on deterministic return point (RP) strategy. In (a) the cell (large point) which starts left of the signal source (small circle) directly moves to the signal source and stops. In (b) the cell starts to the right of the signal source, moves $2^e - 1$ steps to the right (e is the number of internal elements) and then returns to the left until it stops on the signal source. (c) shows the effective wiring of the network used for the pure RP strategy. (d) shows the network for pure chemotaxis, where no internal element is used. Neither in (c), nor in (d), element D, that allows a switch into the stochastic mode, needs to be used. In the initial state of the simulation the wiring allowed that each input element was connected to each internal and each output element, and each internal element was connected to each internal element (including itself) and each output element (cf. table 1).

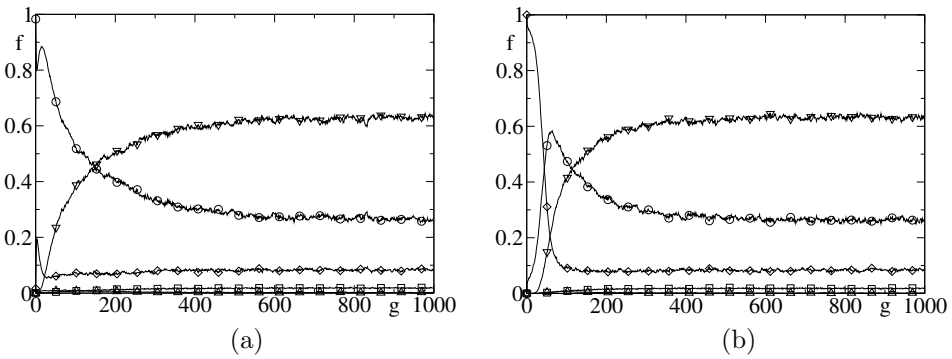

Fig. 3. Fraction of species $f(g)$ in a population with straight movement (squares), return point strategy (RP; triangles up), random walk (RW; diamonds), chemotactic movement (CHT; triangles down), or mixed strategies (circles) as a function of the generation g for the reference parameter set with $x_0 = 7$, $p = 0.01$, $T = 100$, $\Theta = 0.9$. In (a) the initial bit-strings of all cells were chosen randomly and independently, in (b) all $M\mu$ cells initially contain identical copies of a bit string that encodes a random walk strategy. In (a) the subpopulation performing a random strategy is small (maximum at $F_{RW}(g) \approx 0.2$). For both initial choices of the bit-string the dominant strategy at large g is chemotaxis which is the strategy with the highest fitness, and the second frequent strategy are mixed strategies while random motion is the third frequent strategy. Note, that the asymptotic distribution of strategies does not depend on the initial distribution.

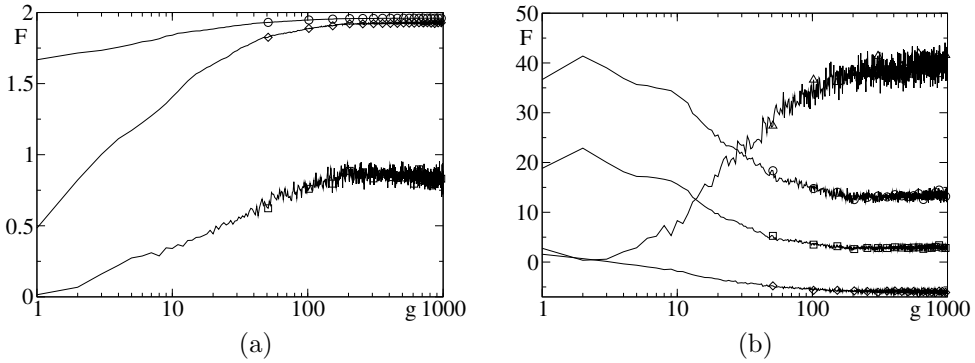

Fig. 4. (a) Maximum (circles, full lines), average (diamonds, dashed-dotted lines), and minimum (squares, dashed lines) fitness in the population with the parameters used in Fig. 3(a). The fitness value increases accompanied by a change of the dominating cell phenotype to a chemotaxis movement. The fitness values include successful strategies (where the cell eventually stops on the signal source ($F_2 = 1$)) and those which were not successful. (b) shows the standard deviation (circles, full lines), variance (squares dashed lines), scaled skewness (diamonds, dashed-dotted, lowest curve), and scaled kurtosis (triangles, dotted, uppermost curve) of the same population. The standard deviation and the variance have been multiplied by a factor of 80. Both decrease by a factor of 2 since all species arrange themselves close to the maximum fitness of the chemotaxis phenotype at the end of the evolutionary process. The skewness changes sign, and kurtosis significantly increases in agreement with the changes of the fitness distribution (compare Fig. 5).

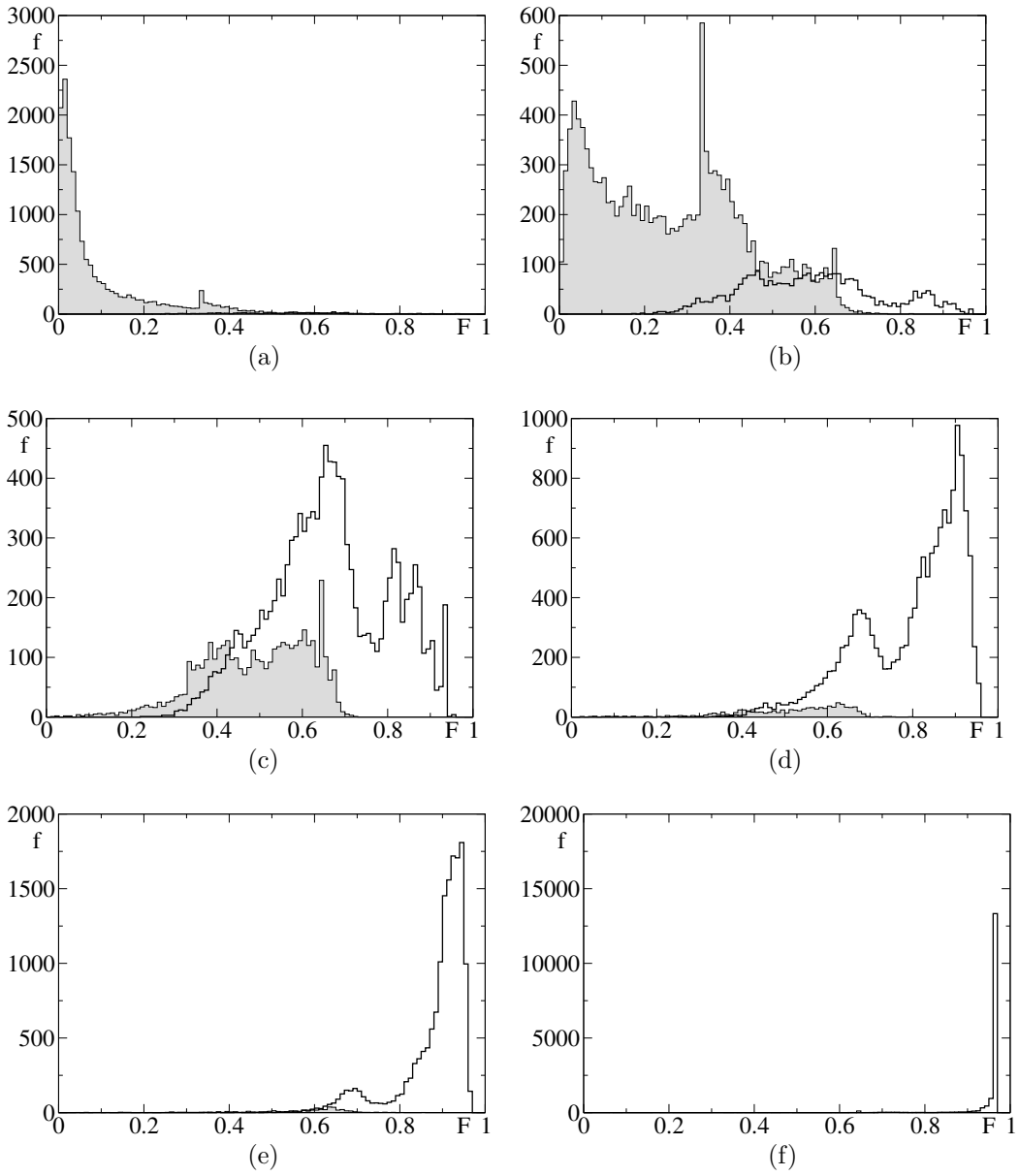

Fig. 5. Frequency f of fitness values during evolution for different generations g for the parameters of Fig. 3. The fat lines denote individuals that finally reach the signal source and stop on it from all initial positions. For presentation purposes we subtracted the benefit of "$F_2 = 1$" by drawing $f(F - F_2)$ instead of $f(F)$ for the fat curves which represent the species that stop on the signal source once they have detected it (in which case $F_2 = 1$ in the fitness function, cf. eqn. (1)). (a) $g = 0$ (initial generation with random generated binary strings), (b) $g = 1$, (c) $g = 5$, (d) $g = 20$, (e) $g = 50$, (f) $g = 1000$ (end of evolution). Initially none of the species is able to stop on the signal source (a) while eventually all species have acquired a network that permits them to stop on the signal source (f). Intermediately the initial fitness distribution smears out indicating that the species, although not able to stop at the signal source, adopt more complex strategies with a higher fitness (b), then integrate the element necessary to stop on the signal source (c), and finally increase their fitness towards the maximum possible value (d-e). Note the scale changes. The fitness values in Fig. 4 use the cumulative fitness distributions without the shift of the fat curve by $-F_2$.

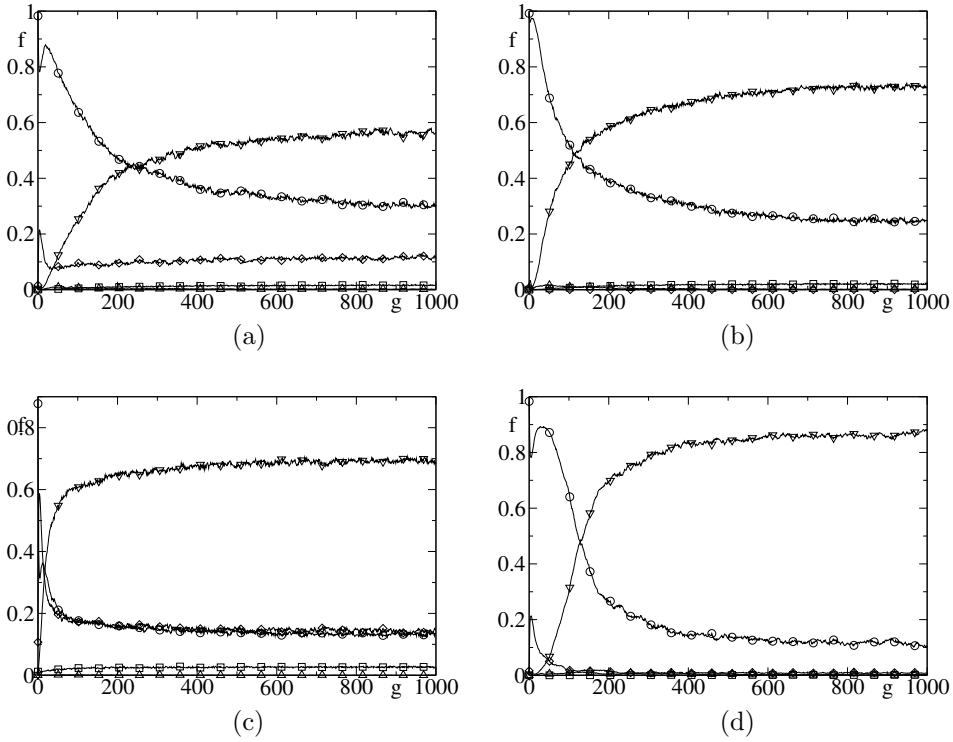

Fig. 6. (a) Fraction $f(g)$ of species in a population with straight movement (squares), return point strategy (RP; triangles up), random walk (RW; diamonds), chemotactic movement (CHT; triangles down), or mixed strategies (circles) as a function of the generation g. Different from the reference parameter set in Fig. 3(a), $x_0 = 7$, $p = 0.01$, $T = 100$, $\Theta = 0.9$, here in (a) $x_0 = 4$, (b): $T = 30$, (c): $\Theta = 0.5$, (d): $p = 0.001$. The intermittent random walk peak is the higher, the larger is T (increases fitness of RW), and the smaller are Θ (most mixed strategies have a large RW proportion) and x_0 (since x_0 affects fitness of RW-strategy non-linearly, see text). Asymptotically the dominant strategy is always chemotaxis which has the largest fitness. However, even at large g many different strategies coexist since complex strategies as chemotaxis require a well orchestrated interplay of the network elements and consequently can easily be destroyed under mutations. Hence the smaller the mutation rate is the larger is the fraction of cells that perform chemotaxis.

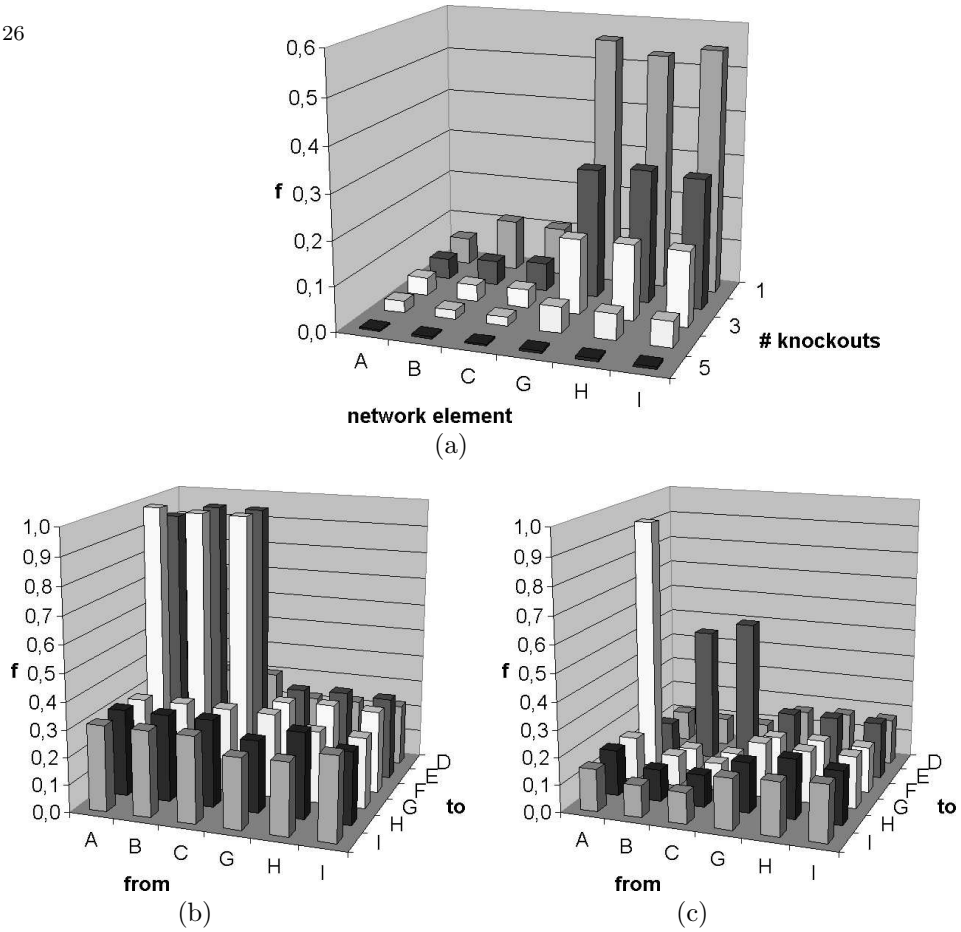

Fig. 7. (a) Fraction of cells that are still able to detect the signal source after knockout. Details see text. A is the most sensitive element. It encodes whether a cell is able to stop when it is located on the signal source (stop is encoded by $F = 0$). The internal elements are less important for the phenotype. Adding further internal elements is thus expected to have almost no effect on the phenotype. Note that even in case of 5 knockouts still a fraction of cells are able to find the signal source and stop on it reflecting a remarkable robustness. (b) Result of m-analysis. Shown is the fraction of individuals in which an influence from elements A, B, C, G, H, I to elements I, H, G, F, E, D was identified. In almost 100% of the individuals the element A has an influence on element F. Furthermore in $\sim 50\%$ of the individuals an influence from element B on E as well as from element C to E exists. B, C encode the gradient information, E encodes the movement direction. In principle the stop information can be encoded by B, C in the presence of a morphogen, since at the location of the signal source the gradient vanishes. Note that although the information of whether the gradient is larger or smaller than zero can be encoded by either element B or element C, a suppression of the expression of element B (or C) does not yield a functional phenotype since the cell cannot distinguish whether the missing expression is informative ($B = 0$ encodes $\partial c/\partial x < 0$) or the consequence of a perturbation that represents a dys-function. Hence the cell phenotype is sensitive to an element knockout that leads to an ambiguity due to the conflict with another information source. Beside the direct dependencies between input and output elements there exists on a lower level dependencies between input and output elements via internal elements. In general the phenotype is much more stable against knock-outs of internal elements. However, a comparison of (a) and (b) shows that the mutual information analysis overestimates the influence between the elements which is, since it cannot distinguish between wirings that are relevant for the phenotype and wirings that are not. (c) A strict application of REVEAL underestimates the influence between network elements since not all network transitions can be observed, so the knowledge on the network remains incomplete. The ultimate test on whether an element is important is therefore the knockout experiment.

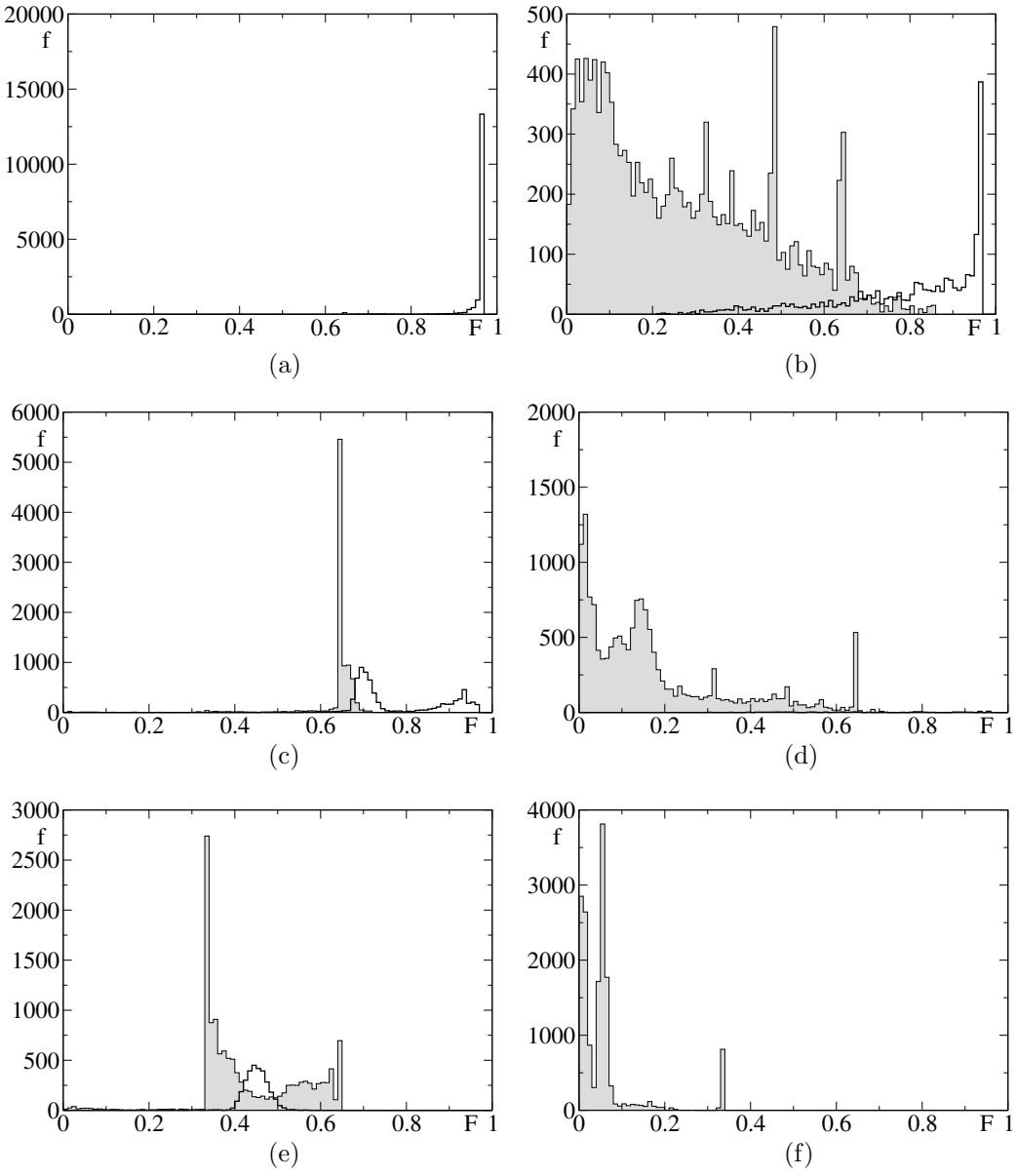

Fig. 8. Frequency f of fitness values in case of knockouts after evolution. The cells have been evolved with a completely functional network, without knockouts (a). For the cells evolved in (a), knockouts have been performed and the fitness distribution has been re-calculated (b-f). By knockout we understand a permanent down-regulation of an element, i.e., the state element is permanently set to zero. In (b) the cell is unable to sense signal source ($A = 0$), in (c) cell senses gradient only partially ($B = 0$), in (d) the cell is unable to sense signal source and unable to determine the gradient from all directions ($A = 0 \wedge B = 0$), in (e) cell is unable to detect any gradient information ($B = 0 \wedge C = 0$), in (f) the cell unable to sense signal source or any gradient information ($A = 0 \wedge B = 0 \wedge C = 0$). If A (b), B (or C) (c), and if B and C (e) are both knocked out, some of the cells are still able to detect the signal source and stop on it. Note the change of the y-scale. In (b) the cells use the gradient information to stop (no gradient $\rightarrow$ stop), in (c) it performs a random walk from one side, and a directed walk using the gradient information from the other side of the signal source, while in (e) it uses a random walk from each initial position to reach the signal source. So even this very simple networks show a remarkable robustness and redundancy. A knockout of internal elements has usually less dramatic effects such that a major fraction remains able to still adopt a successful strategy.

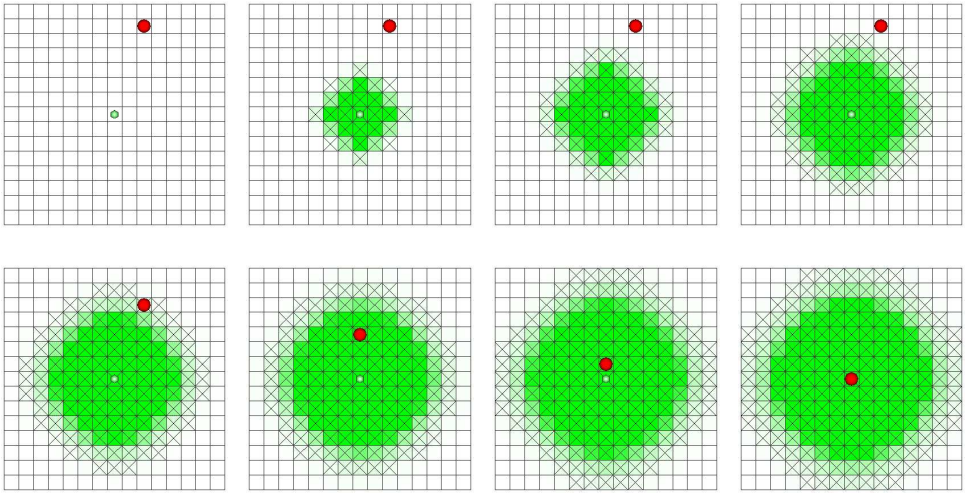

Fig. 9. A typical scenario of a cell migrating to the signal source by chemotaxis in $d = 2$. The cell does not move until it is able to sense the gradient of a diffusing morphogen secreted by a source (upper sequence of pictures). As soon as the cell is able to sense the morphogen gradient it moves straight forward to the signal source. An alternative scenario is shown in Fig. 10.

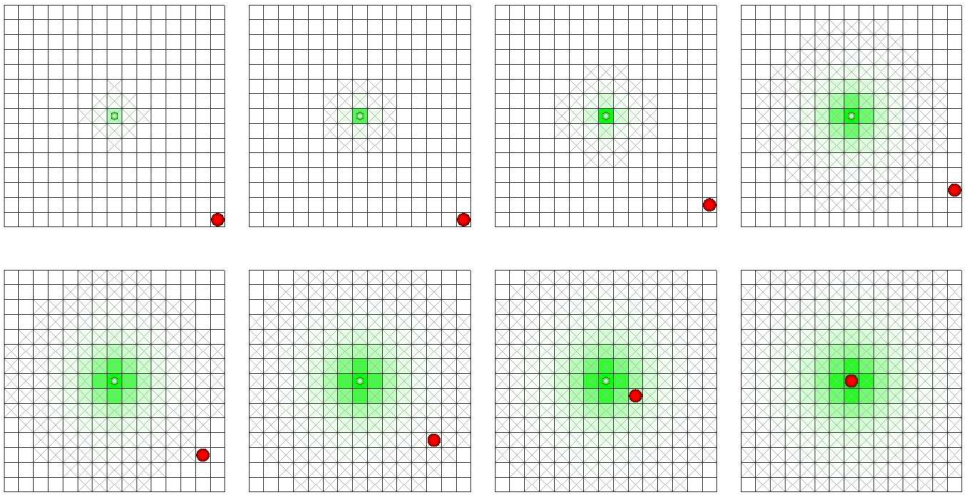

Fig. 10. As an alternative, equally frequent scenario to that in Fig. 9 we found that a cell performs a random walk as long as it cannot sense a morphogen gradient (upper sequence of pictures) and switches to chemotaxis when the morphogen gradient becomes detectable.

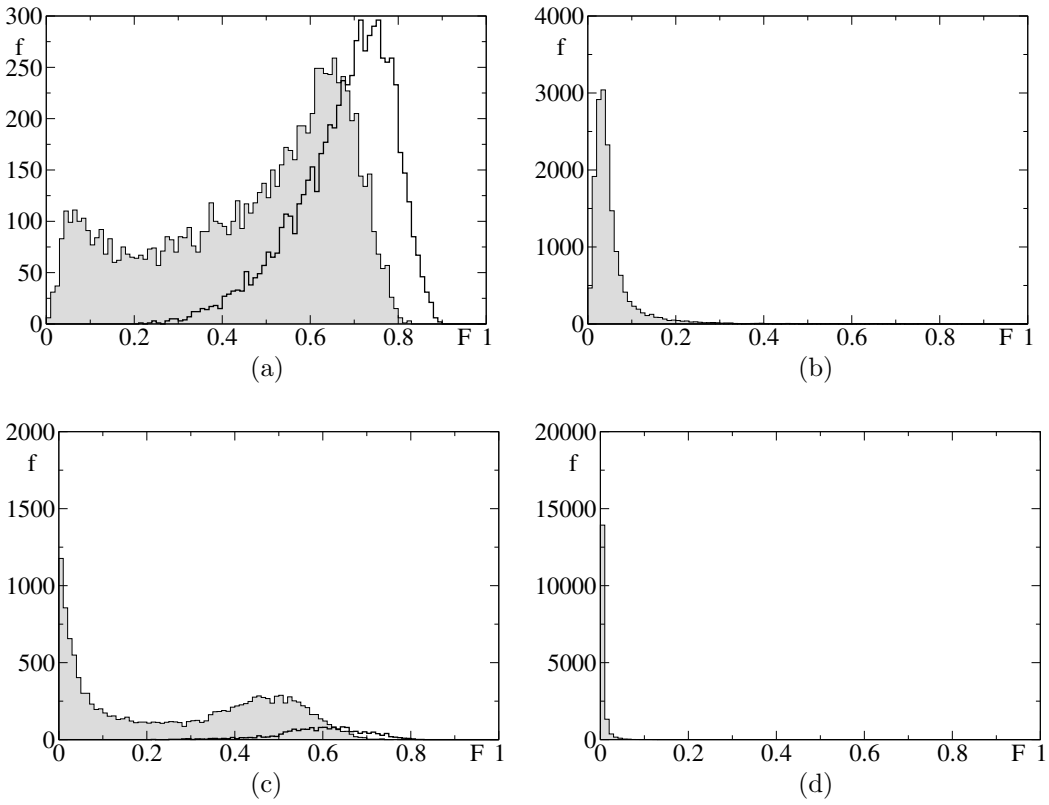

Fig. 11. Frequency of fitness values (a) without, and (b-d) with knockouts in $d = 2$. The knockouts have been performed in the networks of the cells in (a). In (b), A has been knocked out, in (c), one of the four elements (in $d = 2$) that encode the gradient information has been knocked out, in (d) element F that in $d = 2$ compares the morphogen concentrations at two successive points of time (chemokinesis) has been knocked out. Only if element B (or C, D, E) have been knocked out, the cell can adopt an alternative strategy which enables it to detect the signal source and stop on it. $F = 1$ if $c(t + 1) > c(t)$, i.e., the if a cell walks up a gradient, $F = 1$. The selection thus occurs only on the network rules at fixed $F = 1$, while for $F = 0$ the network is never evolutionarily optimized.

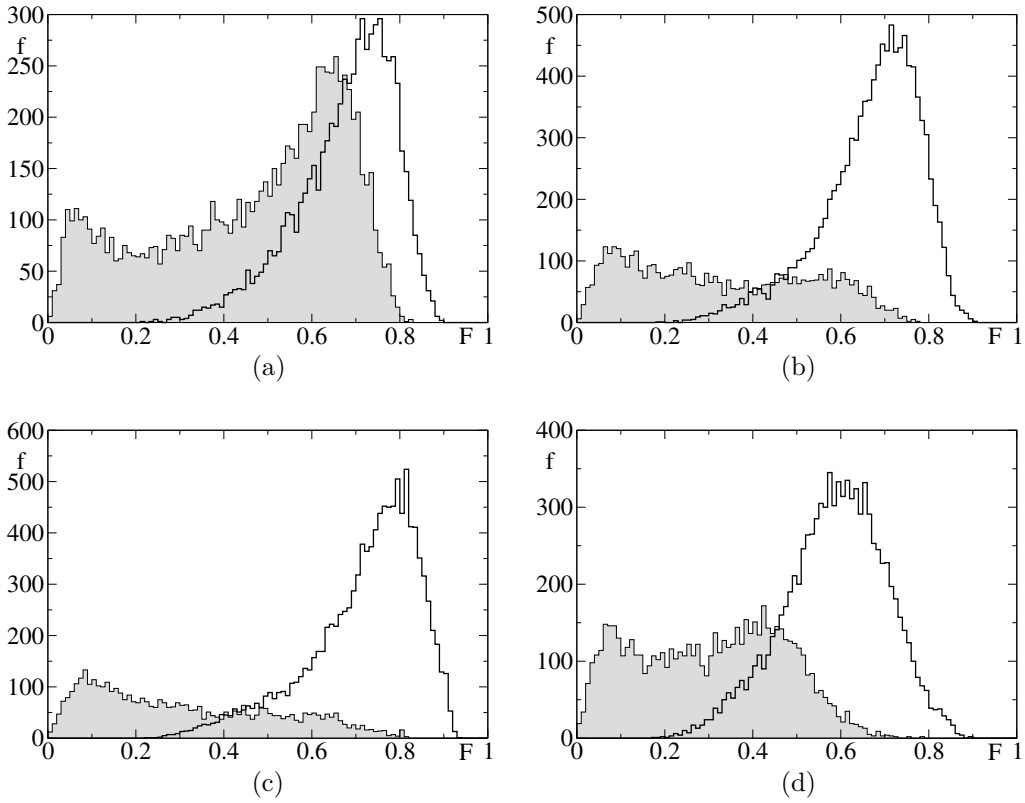

Fig. 12. Same as in Fig. 11 but here the evolution was carried out after the knockout of (b) element A, (c) element B, and (d) element F. Different from Fig. 11 the cell is now able to adopt successful strategies for all knockouts with a similar fitness distribution as without knockout (a).

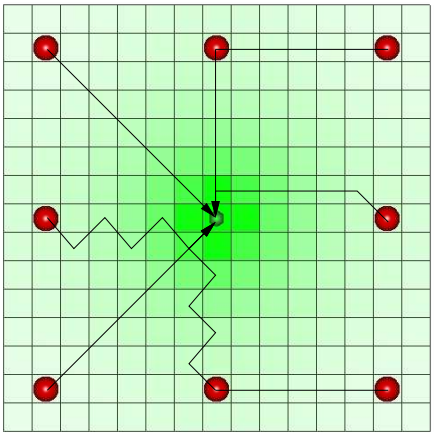

Fig. 13. Examples for a successful cell migration strategy if one of the input elements (B) that participate in determining the morphogen gradient is knocked out ($B = 0$). A cell where $B = 0$, cannot distinguish between situations in which ($\frac{\partial c}{\partial x} = 0, \frac{\partial c}{\partial y} = 0$) and ($\frac{\partial c}{\partial x} = 0, \frac{\partial c}{\partial y} < 0$), and between ($\frac{\partial c}{\partial x} = 0, \frac{\partial c}{\partial y} > 0$) and ($\frac{\partial c}{\partial x} < 0, \frac{\partial c}{\partial y} = 0$), and between ($\frac{\partial c}{\partial x} < 0, \frac{\partial c}{\partial y} < 0$) and ($\frac{\partial c}{\partial x} < 0, \frac{\partial c}{\partial y} = 0$), and between ($\frac{\partial c}{\partial x} > 0, \frac{\partial c}{\partial y} = 0$) and ($\frac{\partial c}{\partial x} > 0, \frac{\partial c}{\partial y} > 0$). The phenotype shown circumvents each of this difficulties.

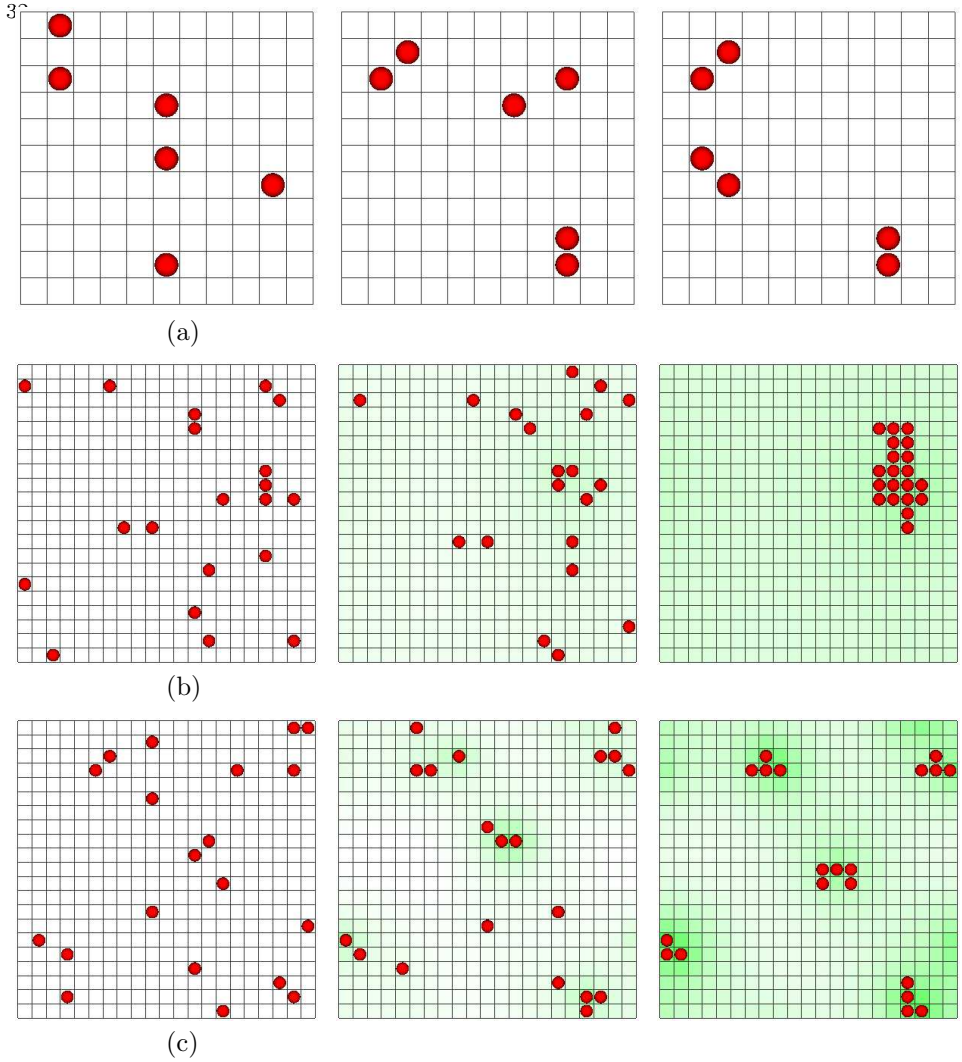

Fig. 14. Typical time-series during cell-cell aggregation, where (a) cells can only perform a random walk strategy, (b,c) cells can perform chemotaxis strategy. For (b) large morphogen diffusion coefficient, large aggregates of cells form while for (c) small morphogen diffusion coefficients, small cell aggregates form.

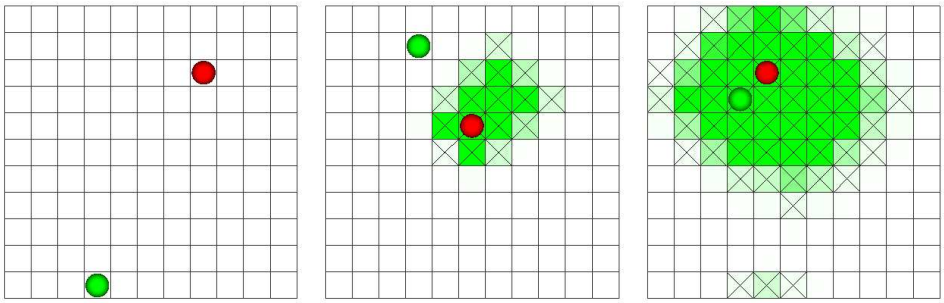

Fig. 15. Migration of two cells with different cell types. Cells of one type release a chemoattractant which can be measured by cells of a second type.

Element	state 0	state 1	type
A	signal source not found	signal source found	input
B,C,D,E	encoding of gradient		input
F	$c(t) < c(t-1)$	$c(t) \geq c(t-1)$	input
G	deterministic movement	stochastic movement	output
H	$x \rightarrow x$	$x \rightarrow x \pm 1$	internal
I	$y \rightarrow y$	$y \rightarrow y \pm 1$	internal
J	$x \rightarrow x - 1$	$x \rightarrow x + 1$	internal
K	$y \rightarrow y - 1$	$y \rightarrow y + 1$	internal

Table 4. Elements of the Boolean which allows the cells to establish a (1) a straight, (2) a chemotactic, (3) a chemokinetic, (4) a random movement, and (5) mixed strategies to find the signal source in two dimensions. Elements $A - F$ are input, G is output, $H - K$ are output elements. Element H and I are only evaluated if element $G = 0$, element J is only evaluated if elements $G = 0$ and $H = 1$, element K is only evaluated if elements $G = 0$ and $I = 1$.

parameter	description
x_0	initial distance between cell and signal source
T	maximum simulation time of cell movement
k	number of initial positions
ψ	morphogen segregation rate
δ	morphogen diffusion rate
γ	morphogen decay rate
M	number of sub-populations
μ	size of each parental sub-population P_p
λ	size of each filial sub-population P_f
n	number of elitists in each sub-population
p_e	probability of exchange of individuals between sub-populations
p_c	probability of cross-overs per individual
p_m	probability of bit-flips per loci
G	maximum evolution time

Table 5. Parameter descriptions in our simulations.

parameter	typical value
x_0	$= 7$
T	$= 30$ (signal source finding) and $= 300$ (cell-cell-aggregation)
k	$= 3$ $(1d)$ and $= 9$ $(2d)$
ψ	$= 1$
δ	$= 0.2$ (signal source finding) and $= 0.05 \ldots 0.2$ (cell-cell-aggregation)
γ	$= 0.01$
M	$= 4$
μ	$= 80$
λ	$= 200$
n	$= 4$
p_e	$= 0.1$
p_c	$= 0.8$ (default) and $= 0.0$ (quantitative measurements)
p_m	$= 0.01$
G	$= 500$

Table 6. Typical choices in our simulations.

References

[1] R. Albert and H. Othmer. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in drosophila melanogaster. *J. Theor. Biol.*, 223:1–18, 2003.

[2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *The Cell*. Garland Science Publ., New York, 2002.

[3] E. Ben-Jacob, I. Cohen, and H. Levine. Cooperative self-organization of microorganisms. *Adv. in Phys.*, 49(4):395–554, 2000.

[4] T. Bulter, S. Lee, W. Wong, E. Fung, M. Connor, and J. Liao. Design of artificial cell-cell communication using gene and metabolic networks. *Proc. Natl. Acad. Sci. (USA)*, 101(8), 2004.

[5] J. Dallon and H. Othmer. A continuum analysis of the chemotactic signal seen by dictyostelium discoideum. *J. theor. Biol.*, 194:461–483, 1998.

[6] L. Davidson, M. Koehl, R. Keller, and G. Oster. How do sea urchins invaginate? using bio-mechanics to distinguish between mechanisms of primary invagination. *Development*, 121:2005–2018, 1995.

[7] H. DeJong. Modeling and simulation of genetic regulatory systems: A literature review. *J. Comp. Biol.*, 9(1):67 – 103, 2002.

[8] S. Dormann and A. Deutsch. Modeling of self-organized avascular tumor growth with a hybrid cellular automaton. *In Silico Biology*, 2:0035, 2002.

[9] D. Drasdo and G. Forgacs. Modelling the interplay of generic and genetic mechanisms in cleavage, blastulation and gastrulation. *Dev. Dyn.*, 219:182–191, 2000.

[10] D. Drasdo, R. Kree, and J. McCaskill. Monte-carlo approach to tissue-cell populations. *Phys. Rev. E*, 52(6):6635–6657, 1995.

[11] N. Dunn, S. Lockery, J. Pierce-Shimomura, and J. Conery. A neural network model of chemotaxis predicts functions of synaptic connections in the nematode caenorhabditis elegans. *J. Comp. Neuros.*, 17(2):137 – 147, 2004.

[12] M. Eigen. Selforganization of matter and evolution of biological macromolecules. *Naturwissenschaften*, 58:465 – 523, 1971.

[13] M. Eigen, J. McCaskill, and P. Schuster. Molecular quasi-species. *J. Phys. Chem.*, 92:6881 – 6891, 1988.

[14] M. Eigen, J. McCaskill, and P. Schuster. The molecular quasi-species. *Adv. Chem. Phys.*, 75:149 – 263, 1989.

[15] P. Friedl. Prespecification and plasticity: shifting mechanisms of cell migration. *Curr. Opin. Cell. Biol.*, 16(1):14 – 23, 2004.

[16] J. Galle, M. Loeffler, and D. Drasdo. An individual-cell based model of the growth regulation of the epithelial cell populations in vitro. *German Conference on Bioinformatics*, pages 87 – 91, 2003.

[17] J. Galle, M. Loeffler, and D. Drasdo. Modelling the effect of deregulated proliferation and apoptosis on the growth dynamics of epithelial cell populations in vitro. *Biophys. J.*, 88:62–75, 2005.

[18] S. Gilbert. *Develoment*. Sinauer Associates Inc., New York, 1997.

[19] P. Hogeweg. Evolving mechanisms of morphogenesis: on the interplay between differential adhesion and cell differentiation. *J. Theor. Biol.*, 203:317 – 333, 2000.

[20] P. Hogeweg. Shapes in the shadow: Evolutionary dynamics of morphogenesis. *Artificial Life*, 6:85 – 101, 2000.

[21] S. Kauffman. *The Origins of Order: Self-organization and Selection in Evolution*. Oxford University Press, Oxford, 1993.

[22] A. Levchenko and P. Iglesias. Models of eukaryotic gradient sensing: Application to chemotaxis of amoebae and neutrophilis. *Biophys. J.*, 82:50–63, 2002.

[23] S. Liang, S. Fuhrman, and R. Somogyi. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac Symp Biocomput.*, 1:18–29, 1998.

[24] A. Maree, A. Panfilov, and P. Hogeweg. Migration and thermotaxis of dictyostelium discoideum slugs, a model study. *J. theor. Biol.*, 199:297–309, 1999.

[25] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chkloviskii, and U. Alon. Network motifs: Simple building blocks of complex network. *Science*, 298:824 – 827, 2002.

[26] K. Missal, M. Cross, and D. Drasdo. Inference of gene regulatory networks for incomplete expression data. *submitted*, 2005.

[27] J. Moreira and A. Deutsch. Cellular automata models of tumour development - a critical reviwe. *Adv. in Complex Systems*, 5(1):247 – 267, 2002.

[28] G. Odell, G. Oster, P. Alberch, and B.Burnside. The mechanical basis of morphogenesis. *Dev. Biol*, 85:446–462, 1981.

[29] E. Palsson and H. Othmer. A model for individual and collective cell movement in dictyostelium discoideum. *Proc. Natl. Acad. Sci. USA*, 12(18):10448–10453, 2000.

[30] F. Pasemann, U. Steinmetz, M. Hülse, and B. Lara. Robot control and the evolution of modular neurodynamics. *Theory in Biosciences*, 20:311–326, 2001.

[31] P. Schuster, W. Fontana, P. Stadler, and I. Hofacker. From sequences to shapes and back: A case study in RNA secondary structures. *Proc. Roy. Soc. Lond. B*, 255:279–284, 1994.

[32] C. Siu. Cell-cell adhesion molecules in dictyostelium. *BioEssays*, 12:357 – 362, 1990.

[33] R. Somogyi and S. Fuhrman. Distributivity, a general information theoretic network measure, or while the whole is more than the sum of its parts. *Proc. international workshop on information processing in cells and tissues (IPCAT)*, 1997.

[34] A. Stevens. The derivation of chemotaxis equations as limit dynamics of moderately interacting stochastic many-particle systems. *SIAM J. APPL. MATH.*, 61(1):183–212, 2000.

[35] A. Stevens. A stochastic cellular automaton modeling of gliding and aggregation of myxobacteria. *SIAM J. APPL. MATH.*, 61(1):172–182, 2000.

[36] P. Van Haastert. Transduction of the chemotactic camp signal across the plasma membrane. In *Dictyostelium – a Model system for Cell and Developmental Biology*. Universal Academy Press/Yamada Science Foundation, 1997.

[37] E. Vaughan, E. A. Di Paolo, and I. Harvey. The evolution of control and adaptation in a 3d powered passive dynamic walker. In J. Pollack, M. Bedau, P. Husbands, T. Ikegami, and R. Watson, editors, *Artificial Life IX: Proceedings of the Ninth Interational Conference on the Simulation and Synthesis of Life*, pages 139–145. MIT Press, 2004.

[38] M. Zajac, G. Jones, and J. Glazier. Model of convergent extension in animal morphogenesis. *Phys. Rev. Lett.*, 85(9):2022 – 2025, 2000.