PLoS one

A Software Tool to Model Genetic Regulatory Networks. Applications to the Modeling of Threshold Phenomena and of Spatial Patterning in *Drosophila*

Rui Dilão*, Daniele Muraro

Nonlinear Dynamics Group, Instituto Superior Técnico, Lisbon, Portugal

Abstract

We present a general methodology in order to build mathematical models of genetic regulatory networks. This approach is based on the mass action law and on the Jacob and Monod operon model. The mathematical models are built symbolically by the *Mathematica* software package *GeneticNetworks*. This package accepts as input the interaction graphs of the transcriptional activators and repressors of a biological process and, as output, gives the mathematical model in the form of a system of ordinary differential equations. All the relevant biological parameters are chosen automatically by the software. Within this framework, we show that concentration dependent threshold effects in biology emerge from the catalytic properties of genes and its associated conservation laws. We apply this methodology to the segment patterning in *Drosophila* early development and we calibrate the genetic transcriptional network responsible for the patterning of the gap gene proteins Hunchback and Knirps, along the antero-posterior axis of the *Drosophila* embryo. In this approach, the zygotically produced proteins Hunchback and Knirps do not diffuse along the antero-posterior axis of the embryo of *Drosophila*, developing a spatial pattern due to concentration dependent thresholds. This shows that patterning at the gap genes stage can be explained by the concentration gradients along the embryo of the transcriptional regulators.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: rui@sd.ist.utl.pt

Introduction

A genetic regulatory network is an ensemble of interactions in a biological process involving proteins, genes and mRNAs. The interactions between different proteins and genes can be done by transcriptional activation and repression at the level of the genes, by protein-protein interactions, and by protein-mRNA interactions.

A genetic regulatory networks is described by a graph where vertices represent genes, proteins, enzymes or other chemical substances. The edges represent transformations, *e. g.*, phosphorylation and dephosphorylation, or activation and inhibitory actions through transcription regulators.

More precisely, a genetic regulatory networks is described by a double graph $G = (V, E_a, E_r)$, where V is the set of vertices or nodes of the graph and E_a and E_r are two sets of ordered pairs of vertices of the double graph. Each ordered pair of vertices defines the activation or the repression mechanism of the first node over the second. In classical graph theory, $G = (V, E_a)$ and $G = (V, E_r)$ are two graphs with a common set of vertices. For example, in Figure 1, we show the graph of a genetic network associated with the production of the proteins Bicoid (BCD), Hunchback (HB), Knirps (KNI) and Tailless (TLL) in *Drosophila* early development, [1,2]. In this example, we have $V = \{bcd, hb, BCD, HB, KNI, TLL\}$, $E_a = \{(bcd, BCD), (hb, HB), (BCD, HB), (BCD, KNI)\}$ and $E_r = \{(HB, KNI), (KNI, HB), (TLL, KNI)\}$.

The graph of Figure 1 has a clear biological meaning. It expresses the fact that BCD is a transcriptional activator of both HB and KNI, HB and KNI proteins both repress each other, and TLL is a repressor of KNI. The vertices of the graph of Figure 1 can represent mRNAs, as in the case of hb and bcd, or proteins, as in the case of BCD and TLL, or genes and proteins simultaneously, as in the case of HB and KNI.

Here we propose a set of rules in order to construct the model equations associated with a genetic regulatory networks described by a double graph $G = (V, E_a, E_r)$. This paper is an attempt to delineate a methodological approach for the construction of mathematical models of gene expression regulation from the principles of chemical kinetics and chemical bound. In the literature, it is often found examples of mathematical models of biological systems described by different sets of equations and characterized by different sets of parameters that are difficult to interpret and to measure experimentally. Making qualitative predictions with these different models has a limited predictable value. For a review on the different approaches see, for example, [3,4].

In the construction of models for generic regulatory networks, we assume that models can be built with rate equations reflecting a mean field view of the stochastic random motion occurring at the molecular scale. This mean field approach, also called mass action law, is derived from the probabilistic collision laws occurring at the molecular scale. The models originated from this view are
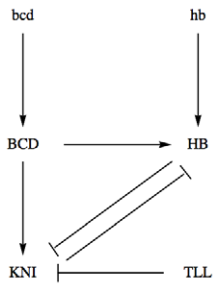
Figure 1. Graph describing the genetic network associated with the production of proteins Bicoid (BCD), Hunchback (HB), Knirps (KNI) and Tailless (TLL) in *Drosophila* early development. mRNAs *hunchback* and *bicoid* are represented by hb and bcd respectively. Arrows represent activations and are listed in the set of ordered pairs E_a. Lines with perpendicular endings represent repressions and are listed in the set E_r.
doi:10.1371/journal.pone.0010743.g001

described by ordinary differential equations with polynomial vector fields, [5–7].

One of the advantages of the mass action law approach is that the mean field rate equations have a direct microscopic interpretation, being associated with the collision mechanism that are in the origin of every reactive process. For model refinement, fluctuations can also be studied through the corresponding master equation. From the experimental point of view, microbiology techniques are strongly anchored in the mass action law or mean field approach, [5,8].

For genetic regulatory networks described by graphs with a large number of vertices, and a complex structure of edges, the rate equations describing the evolution in time of concentrations are in general difficult to build, and are critically dependent on the assumptions done about the biological and the chemical interactions involved. During the development of these complex models, it is often necessary to test different graph configurations, and to change parameters and initial conditions. Writing by hand all this information is both time-consuming and error-prone.

In order to perform these tasks automatically, we have developed two *Mathematica* software packages, *Kinetics* and *GeneticNetworks*, that execute the symbolic computations associated with the construction of the model equations for a genetic regulatory network. The result of the analysis is in symbolic form, and can be used in *Mathematica*, C or any other simulation software for further numerical integration and graphical analysis. The software packages *GeneticNetworks* and *Kinetics* are freely distributed, [9].

The *Kinetics* package implements the mass action law in its polynomial exact form, computing symbolically the associated rate equations and conservation laws. The parameters generated within *Kinetics* are chemical rate constants.

The package *GeneticNetworks* implements a particular model for protein-gene regulation. This model for the protein-gene regulation is based on the operon model of Jacob and Monod [10] in prokaryotes, and its basic properties have been previously introduced in [11]. The tools in the *GeneticNetworks* software package implement a simplified view of the Molecular Biology Dogma, [12], for protein encoding, translation and transcription, and is consistent with the mass action law. For eukaryotes, transcriptional regulation in is a much more complex issue, involving many redundant binding sites dispersed along genomic sequences. Therefore, in this case, the modeling approach proposed here should be understood as a descriptive approximation to the not well understood eukaryotic regulation mechanisms.

In order to obtain a dimensional reduction on the number of variables in the equations obtained by *Kinetics* and *GeneticNetworks*, it is possible to construct, by a steady state approximation, Hill's function models, [4,8,13].

The advantages of using the *Mathematica* computing environment are (i) the possibility of obtaining an exact form for the model equations; (ii) to perform, if necessary, further symbolic simplifications on the models; (iii) to modify the initial theoretical assumptions of the model without having to re-introduce or choose new parameter values; (iv) to make the numerical and graphical analysis of models within the same computing environment; (v) to use of a natural language without a deep knowledge of programming; (vi) to use an easy interface for other programming environments.

To build a model of given genetic regulatory network, the only necessary input to *GeneticNetworks* is the activation and inhibition relationship between genes, mRNAs and proteins. This input is given in the form of the two order pairs of vertices E_a and E_r of the network graph. Then, the generation of model equations is done with the *GeneticNetworks* commands. The model equations can be analysed within the *Mathematica* environment or introduced in other programming environments as COPASI, [14], and Potters-Whell, [15], for simulation and parameter estimation. These programs are powerful general propose tools in order to numerically simulate solutions of ordinary differential equation and to simulate stochastic models for system biology. At the time of writing this paper, in the site of the Systems Biology Markup Language, http://sbml.org, there were more than 180 registered systems biology simulation programs.

This paper is organized as follows. In the Methods subsection, we briefly review the mass action law of chemical kinetics and we introduce the collision graphs associated with the mass action law. We derive the basic mass action rate equations. A special emphasis is done on mass action conservation laws, an important feature that is in the very foundations of threshold effects in biology. In other approaches, threshold do not result as emergent phenomena, but must be imposed through *ad hoc* regulatory functions (see for example [3] or [4, pp. 237]). We describe the mechanism of genetic regulation based on the Jacob and Monod operon model, [10], and we introduce the modeling assumptions for the construction of the mathematical models of genetic networks described by double graphs. Finally, we give an overview of the *GeneticNetworks* software package.

In the Results section, we show three different applications of the quantitative approach developed here. In the first application, we show, with a very simple example of auto-regulation, that the conservation law constant is a bifurcation parameter for the regulation model, inducing a concentration dependent threshold effect in the model for the production of proteins. This solves the problem of the introduction of *ad hoc* threshold effects in biological simulations, [16]. In the second application, we give a genetic regulatory network inducing a localized spiky pattern along a spatial domain. In this case, the spatial spiky patterns appears without the necessity of other transport mechanisms, as diffusion or advection, but is a consequence of the concentration dependent threshold effect. In the third example, we analyze the experimental data associated with the KNI and HB inhibitory cross regulation in *Drosophila* early development, described by the double graph of Figure 1, and we calibrate this model with the experimental data, without the need of a diffusion hypothesis for the zygotically produced proteins HB and KNI. In the final section, we summarize and discuss the main biological conclusions of the paper.

Methods

The mass action law framework of chemical kinetics

In general, an ensemble of chemical reactions is represented by the following collision diagram,

$$v_{i1}A_1 + \cdots + v_{im}A_m \overset{r_i}{\to} \mu_{i1}A_1 + \cdots + \mu_{im}A_m \qquad (1)$$
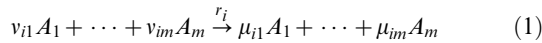
where $i = 1, \ldots, n$. The A_j, for $j = 1, \ldots, m$, represent chemical substances, as for example, $A_j = H_2O$. The constants v_{ij} and μ_{ij} are the stoichiometric coefficients, in general, non-negative integers, and the constants r_i are the rate constants. If $v_{ij} = \mu_{ij} > 0$, the corresponding substance A_j is a catalyst and, if $\mu_{ij} > v_{ij} > 0$, A_j is an autocatalyst. In the diagram (1), there are m chemical substances and n rate constants or chemical reactions.

Under the hypothesis of homogeneity of the solution where reactions occur, the mass action law asserts that the time evolution of the concentrations of the chemical substances is described by the system of ordinary differential equations,

$$\frac{dA_j}{dt} = \sum_{i=1}^{n} r_i(\mu_{ij} - v_{ij})A_1^{v_{i1}} \cdots A_m^{v_{im}} \qquad (2)$$

where $j = 1, \ldots, m$, and we use the same symbol to represent both the chemical substance and its concentration. The rate equations (2) are derived under the following assumptions: (i) chemical reactions, when they occur, are due to elastic collisions between the reactants, (ii) homogeneity of the reacting substances in the solution, and (iii) thermal equilibrium of the solution. All the kinetics aspects related with the dependence of the velocity of the reactions on the temperature or pressure are contained in the rate constants r_i. For details see [5].

At the atomic and molecular scale, chemical reactions between molecules can occur only if molecules collide or approach each other to small distances where bounding forces become meaningful. These chemical bounding forces are of electrical or quantum origin, and at distances larger than the mean free path they become less important when compared with the kinetics associated with the molecular motion. As chemical reactions only occur if the chemical substances involved collide, the vector fields associated with the right hand side of (2) are in general quadratic, representing binary collisions. Higher order polynomial vector fields are possible but, at the microscopic level, they are associated to triple or higher order collisions, a situations that occurs with a very low probability. Therefore, we will restrict our examples to models with two-body collisions.

The equations (2) can also be written in the matrix form,

$$\frac{dA}{dt} = \Gamma \omega(A) \qquad (3)$$

where Γ is a $n \times m$ matrix, $A^T = (A_1, \ldots, A_m)$, and,

$$\omega(A) = \begin{pmatrix} r_1(\mu_{1j} - v_{1j})A_1^{v_{11}} \cdots A_m^{v_{1m}} \\ \vdots \\ r_n(\mu_{nj} - v_{nj})A_1^{v_{n1}} \cdots A_m^{v_{nm}} \end{pmatrix}$$

In general, $n \neq m$, and the equations in system (2) are not all independent. Let us denote by r the rank of the matrix Γ. The dimension of the null space of Γ relates with its rank by,

$\dim(Null(\Gamma)) + r = m$ (number of rows of Γ). Let $v_1, \ldots, v_{m-r}$ be a basis of the null space of Γ, then, $\Gamma v_k = 0$, for $k = 1, \ldots m - r$. So, by (3), we have,

$$\frac{dA}{dt}.v_k = \frac{d}{dt}(A.v_k) = (\Gamma \omega).v_k = \omega.(\Gamma v_k) = 0 \qquad (4)$$

Hence, associated with the differential equations (2), we have the conservation laws,
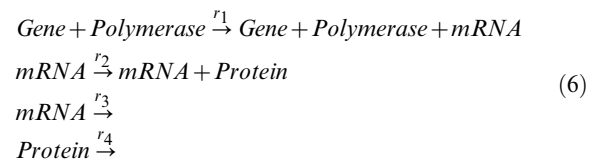
$$A.v_k = cons \qquad (5)$$

where, $k = 1, \ldots m - r$.

The *Mathematica* software package *Kinetics* calculates the rate equations (2) describing the time evolution of the concentrations of the substances involved in the reactions described by the collision diagram (1). The package calculates also the corresponding conservations laws (5).

The input of the package is the ensemble of chemical reactions, and the output of the package is the set of differential equations derived by the mass action law. Then, the output can be later analyzed and studied by the analytical and numerical tools in the software package *Mathematica*. In order to avoid long development times, the names of the rate constants are chosen automatically by the program.

The package *Kinetics* has the usual help commands, and we provide the *Mathematica* notebook *KineticsTest.nb* with several self-explanatory examples and computations, [9].

For example, let us describe now a simple protein production model with *Kinetics*. The Molecular Biology Dogma asserts that genes are the templates for protein production, and the standard mechanism for protein production can be represented by the collision diagrams,

$$\begin{aligned} Gene + Polymerase &\overset{r_1}{\to} Gene + Polymerase + mRNA \\ mRNA &\overset{r_2}{\to} mRNA + Protein \\ mRNA &\overset{r_3}{\to} \\ Protein &\overset{r_4}{\to} \end{aligned} \qquad (6)$$

Using the symbols G, Pol, R and P to represent gene, polymerase, mRNA and protein concentrations, respectively, the collision equations (6) are the input for *Kinetics*, with the syntax,

$$\text{input} = \{G + Pol \to G + Pol + mRNA, mRNA \to$$
$$mRNA + P, mRNA \to W1, P \to W2\}$$

where $W1$ and $W2$ are waist products.

For the collision mechanism (6), the rate equations for the protein concentration, and calculated by the package *Kinetics*, are,

$$\begin{aligned} G' &= 0 \\ Pol' &= 0 \\ R' &= r_1 G.Pol - r_3 R \\ P' &= r_2 R - r_4 P \end{aligned} \qquad (7)$$

and the rate constants have been chosen automatically by the software package. In this model, genes, polymerase and mRNAs are catalysts, and these equations have the exact solutions,

$$
\begin{aligned}
G(t) &= G(0) \\
Pol(t) &= Pol(0) \\
R(t) &= R(0)e^{-r_3 t} + \frac{r_1 G(0) Pol(0)}{r_3}(1 - e^{-r_3 t}) \\
P(t) &= G(0)Pol(0)\frac{r_1 r_2}{r_3 r_4} \\
&+ \left(P(0) + \frac{R(0)r_2 r_4 - r_1 r_2 G(0)Pol(0)}{r_4(r_3 - r_4)}\right)e^{-r_4 t} \\
&+ \frac{R(0)r_2 r_3 - r_1 r_2 G(0)Pol(0)}{r_3(r_4 - r_3)}e^{-r_3 t}
\end{aligned}
\tag{8}
$$

To simplify the model equations of protein production and maintaining the catalytic properties of genes, in the following, instead of the collision mechanism (6), we use the simplified or reduced kinetic mechanism,

$$
\begin{aligned}
Gene &\xrightarrow{s_1} Gene + Protein \\
Protein &\xrightarrow{s_2}
\end{aligned}
\tag{9}
$$
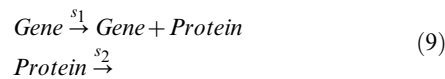
To the reactions (9) correspond the rate equations,

$$
\begin{aligned}
G' &= 0 \\
P' &= s_1 G - s_2 P
\end{aligned}
\tag{10}
$$

The rate equations (10) have the solutions,

$$
\begin{aligned}
G(t) &= G(0) \\
P(t) &= G(0)\frac{s_1}{s_2} + (P(0) - G(0)\frac{s_1}{s_2})e^{-s_2 t}
\end{aligned}
\tag{11}
$$

Comparing the protein solutions in (11) and (8), we conclude that the steady state of the protein in both models is unique and is proportional to the gene concentration. The proportionality constant is different for both models, depending on the rates of the reactions involved. The steady state $G(0)\frac{s_1}{s_2}$ of model (9) has a direct biological meaning: s_1 is the rate of protein production and s_2 is the rate of protein degradation, and $G(0)$ is the initial gene concentration.

In the following, and in order to simplify the description of the transcriptional regulation of genes, we will adopt the mechanism (9) to describe the associated protein production.

In both rate equations (10) and (7), the concentration of gene is constant along time, and therefore the gene concentration is a conservation law. In the following, we will show that the linear conservation laws of the form (5), will have an important role in the determination of steady states and in bifurcations associated with threshold effects.

A mass action framework to describe genetic regulatory networks based on the protein-gene interaction

In the previous section, we have described a basic model model for the production of proteins, in the framework of the mass action law. Based on this framework, we generalize this view in order to include the case of transcriptional regulation of genes by proteins.

In order to keep general the approach presented here and to maintain the biological reality of model parameters, we make the following basic modeling assumptions:

1) In order to describe quantitatively the protein production (concentration) within the Molecular Biology Dogma, we only consider genes and proteins. Intermediate substances in the regulatory mechanism like catalysts, polymerases and mRNAs are not considered. A model of protein production has been presented and analyzed at the end of the previous section.

2) The regulation of protein production by the template gene is based on the Jacob and Monod operon model, [10]. Namely, every gene has associated a certain number of binding sites where transcription factors can bind — activators or repressors, Figure 2. The regulation of activations and repressions occurs only through the binding sites. For a given double graph of interactions, the number of binding sites of a gene is determined by the number of edges that end up in the corresponding graph node.

3) Transcription factors are the proteins associated with the vertices that activate or inhibit the production of other proteins. The vertex of a graph represent a transcription factor only if it is the initial point of a edge of activation or inhibition. If a vertex has incoming and outgoing edges of any type, this vertex represents symbolically a protein and a gene with several binding states.

4) Each transcription factor has its own binding site in the gene strand, or each gene has only one binding site for all the regulators. Both cases are treated separately in the model. We assume that when at least one activator is bound to a gene, the transcription is activated with a particular production rate for each combinatorial possibility of all the remaining binding sites.

For example, in the double graph of the biological mechanism of Figure 1, we have the following chemical substances,

$$
\begin{array}{ll}
\text{bcd, hb} & \text{mRNAs} \\
\text{BCD, HB, KNI, TLL} & \text{proteins} \\
HB_0^0,\ HB_0^{BCD},\ HB_{KNI}^0,\ HB_{KNI}^{BCD} & \text{operons} \\
KNI_{0,0}^0,\ KNI_{0,0}^{BCD},\ KNI_{HB,0}^0,\ KNI_{HB,0}^{BCD} & \text{operons} \\
KNI_{0,TLL}^0,\ KNI_{0,TLL}^{BCD},\ KNI_{HB,TLL}^0,\ KNI_{HB,TLL}^{BCD} & \text{operons}
\end{array}
\tag{12}
$$

The description of the time evolution of protein concentrations of the mechanisms of Figure 1 involves one rate equation for each substance in (12), except eventually for bcd and hb. As proteins are produced from a gene template, the symbol associated to each vertex of the graph represents a protein. The operon states are represented by the same symbol with superscripts and lowerscripts. The superscripts positions indicate the binding or unbinding of transcriptional activators. The lowerscripts positions indicate the binding or unbinding of transcriptional repressors. In the *GeneticNetworks* software package, bcd, hb, BCD, HB, KNI and TLL are the names of the vertices of the regulation graphs, but the operon variables in (12) are generated by the software.

The model associated with a given double graph contains the rate equations for the proteins and the operons in its different
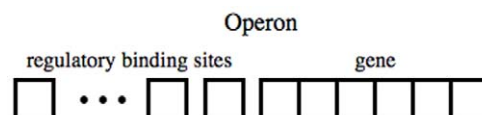
Figure 2. Jacob and Monod operon model for the regulation of protein production. The transcription is regulated by the activators and the repressors binding to the binding sites of the gene.
doi:10.1371/journal.pone.0010743.g002

states. We also assume by default that proteins always degrade and genes are autocatalytic substances that never degrade. The first assumption implies that protein concentrations remain bounded in time, and the second assumption implies the existence of a conservation law for the concentration of the operon states. Using the symbolic tools of *Mathematica*, other assumptions can be introduced at this stage of model construction.

The *GeneticNetworks* software package

GeneticNetworks is a software package that generates the rate equations for the concentrations of genes and proteins in a regulatory network, [9]. The starting point is the double graph of activations and repressions, $G = (V, E_a, E_r)$. The inputs for *GeneticNetworks* are two strings, *activation* and *repression*, that describe the transcriptional activations and repressions of proteins on genes. In the graph, the same symbol is used to denote both a gene and the corresponding produced protein. As we have seen in (12), the set of variables for the regulation model is constructed with the vertex symbols.

For example, using as input for *GeneticNetworks* the interaction strings,

$$
\begin{aligned}
activation &= \{A \rightarrow B\} \\
repression &= \{R \rightarrow B\}
\end{aligned} \tag{13}
$$

the double graph of the genetic network (13) is shown in Figure 3. In this case, the double graph $G = (V, E_a, E_r)$ is characterized by the sets, $V = \{A, B, R\}$, $E_a = \{(A, B)\}$ and $E_r = \{(R, B)\}$.

In the interaction mechanism (13), protein A activates gene B, and protein R represses gene B. Therefore, the variables of the mechanism (13) are,

$$
\begin{aligned}
&A, B, R &&\text{proteins} \\
&B_0^0, B_0^A, B_R^0, B_R^A &&\text{operons}
\end{aligned} \tag{14}
$$

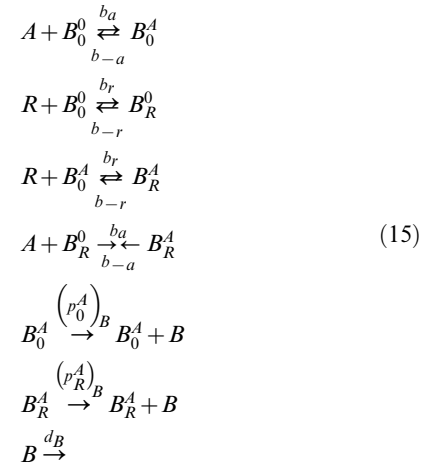The following functions are defined in the *GeneticNetworks* package:

- NetworkGraph, ManipulateGraph
- Reactions, ReactionsOneSite, ReactionGraph
- SubstanceNames, SubstanceVariables, SubstanceInitialConditions
- ParameterNames, ParameterInput
- Equations
- ConservationLaws

With these functions, we calculate the model equations associated with the input strings (13), calculate automatically the number of variables of the model, define all the relevant
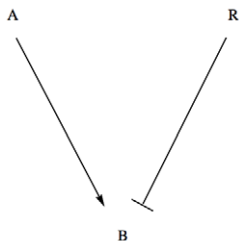
Figure 3. Double graph associated with the input strings (13) for the *GeneticNetworks* software package.
doi:10.1371/journal.pone.0010743.g003

parameters and calculate the rate equations. For example, to the genetic regulatory network (13), the *GeneticNetworks* package builds the mass action law type collision diagrams,

$$
\begin{aligned}
A + B_0^0 &\underset{b_{-a}}{\overset{b_a}{\rightleftarrows}} B_0^A \\
R + B_0^0 &\underset{b_{-r}}{\overset{b_r}{\rightleftarrows}} B_R^0 \\
R + B_0^A &\underset{b_{-r}}{\overset{b_r}{\rightleftarrows}} B_R^A \\
A + B_R^0 &\underset{b_{-a}}{\overset{b_a}{\rightarrow \leftarrow}} B_R^A \\
B_0^A &\overset{\left(p_0^A\right)_B}{\rightarrow} B_0^A + B \\
B_R^A &\overset{\left(p_R^A\right)_B}{\rightarrow} B_R^A + B \\
B &\overset{d_B}{\rightarrow}
\end{aligned} \tag{15}
$$

To these collision diagrams, we have the mass action law rate equations,

$$
\begin{aligned}
A' &= -b_a A . B_R^0 + b_{-a} B_R^A - b_a A . B_0^0 + b_{-a} B_0^A \\
B' &= \left(p_R^A\right)_B B_R^A + \left(p_0^A\right)_B B_0^A - d_B B \\
R' &= -b_r R . B_0^A + b_{-r} B_R^A - b_r R . B_0^0 + b_{-r} B_R^0 \\
\left(B_0^0\right)' &= -b_a A . B_0^0 + b_{-a} B_0^A - b_r R B_0^0 + b_{-r} B_R^0 \\
\left(B_0^A\right)' &= b_a A . B_0^0 - b_{-a} B_0^A - b_r R . B_0^A + b_{-r} B_R^A \\
\left(B_R^0\right)' &= -b_a A . B_R^0 + b_{-a} B_R^A + b_r R . B_0^0 - b_{-r} B_R^0 \\
\left(B_R^A\right)' &= b_a A . B_R^0 - b_{-a} B_R^A + b_r R . B_0^A - b_{-r} B_R^A
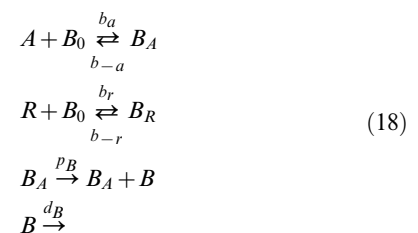\end{aligned} \tag{16}
$$

and the conservation law,

$$
B_0^0(t) + B_0^A(t) + B_R^0(t) + B_R^A(t) = B_0^0(0) + B_0^A(0) + B_R^0(0) + B_R^A(0) \tag{17}
$$

From the conservation law (17), we can eliminate one of the equations in (16). In this genetic network, we have assumed that the protein concentrations of A and R are constant along time.

The rate equations (16) define a mass action law based model for the genetic regulatory network of Figure 3.

In the implementation of *GeneticNetworks*, we have two possible modeling choices. In one choice, each different regulator has its own binding site, and the model diagrams (15) have been constructed with this assumption. For the second choice, we consider that there is only one binding site in the operon where all the regulators bind. In this case, the collision diagrams associated with the genetic network (13) and calculated in the *GeneticNetworks* package are,

$$
\begin{aligned}
A + B_0 &\underset{b_{-a}}{\overset{b_a}{\rightleftarrows}} B_A \\
R + B_0 &\underset{b_{-r}}{\overset{b_r}{\rightleftarrows}} B_R \\
B_A &\overset{p_B}{\rightarrow} B_A + B \\
B &\overset{d_B}{\rightarrow}
\end{aligned} \tag{18}
$$

By the mass action law, the ReactionsOneSite command leads to the rate equations,

$$
\begin{aligned}
A' &= b_{-a}B_A - b_a A.B_0 \\
B' &= p_B B_A - d_B B \\
R' &= b_{-r}B_R - b_r B_0.R \\
(B_0)' &= -b_a A.B_0 + b_{-a}B_A - b_r B_0.R + b_{-r}B_R \\
(B_A)' &= b_a A.B_0 - b_{-a}B_A \\
(B_R)' &= b_r B_0.R - b_{-r}B_R
\end{aligned}
\tag{19}
$$

The equations (19) have the conservation law,

$$
B_0(t) + B_A(t) + B_R(t) = B_0(0) + B_A(0) + B_R(0)
\tag{20}
$$

The two models (15) and (18) for the genetic network (13) are different and these two choices are implemented in the *GeneticNetworks* package. For the dynamical analysis of a particular case of the distinction between the two models (15) and (18), see [11]. Below, we will show with a specific example that these two different regulation choices lead to qualitatively and quantitatively similar results.
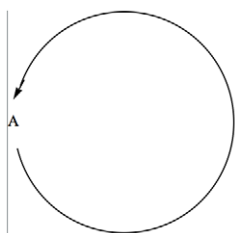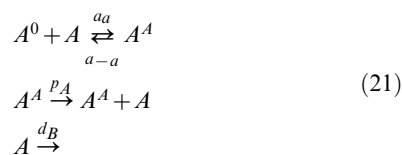
Results

An emerging concentration threshold in the dynamics of a self-activating protein

As an application of the rules describing a genetic regulatory network just introduced, we discuss now the basic role of the conservation laws in the occurrence of threshold effects in regulation mechanisms. We study the case of a self-activating protein, where the produced protein activates its own production, Figure 4.

The simplest self-activating genetic network is described by the input tables,

$$
\begin{aligned}
activation &= \{A \rightarrow A\} \\
repression &= \{\}.
\end{aligned}
$$

The reactions and the parameters involved in this activation can be obtained by the *GeneticNetworks* command Reactions, followed by the command ReactionGraph,

$$
\begin{aligned}
A^0 + A &\underset{a_{-a}}{\overset{a_a}{\rightleftharpoons}} A^A \\
A^A &\overset{p_A}{\rightarrow} A^A + A \\
A &\overset{d_B}{\rightarrow}
\end{aligned}
\tag{21}
$$

Figure 4. Regulation graph describing a self-activating protein.
doi:10.1371/journal.pone.0010743.g004

where A^0 and A^A are the operon states and A is the corresponding protein.

With the command Equations, we get the rate equations,

$$
\begin{aligned}
A' &= a_a A.A^0 + a_{-a}A^A + p_A A^A - d_A A \\
(A^0)' &= -a_a A.A^0 + a_{-a}A^A \\
(A^A)' &= a_a A.A^0 - a_{-a}A^A
\end{aligned}
\tag{22}
$$

Finally, with the command ConservationLaws, we find,

$$
A^0(t) + A^A(t) = A^0(0) + A^A(0) = c
\tag{23}
$$

where c is a constant.

Introducing (23) into (22), the independent set of rate equations describing the process (21) is,

$$
\begin{aligned}
A' &= a_a A.A^0 + a_{-a}(c - A^0) + p_A(c - A^0) - d_A A \\
(A^0)' &= -a_a A.A^0 + a_{-a}(c - A^0)
\end{aligned}
\tag{24}
$$

We analyze now the steady state and the phase space structure of the solutions of equations (24). Equations (24) have two steady states with coordinates,

$$
(A^0_{(1)}, A_{(1)}) = (c, 0)
$$

and,

$$
(A^0_{(2)}, A_{(2)}) = \left(\frac{d_A a_{-a}}{p_A a_a}, \frac{c p_A a_a - d_A a_{-a}}{d_A a_a} \right)
$$

As the coordinate of the two steady sates are dependent of c, by (23), the steady state coordinates are dependent of the initial concentrations of the operon.

Let $J_{(i)}$, $i=1,2$ be the Jacobian of equation (24) evaluated at the fixed points. As,

$$
\det J_{(1)} = -\det J_{(2)}
$$

and,

$$
\det J_{(1)} < 0 \Leftrightarrow a_{-a}d_A < c a_a p_A
$$

then, we have,

$$
\text{if } a_{-a}d_A < c a_a p_A
\begin{cases}
(A^0_{(1)}, A_{(1)}) \text{ is of saddle type} \\
(A^0_{(2)}, A_{(2)}) \text{ is a stable node}
\end{cases}
\tag{25}
$$

$$
\text{if } a_{-a}d_A > c a_a p_A
\begin{cases}
(A^0_{(1)}, A_{(1)}) \text{ is a stable node} \\
(A^0_{(2)}, A_{(2)}) \text{ is of saddle type}
\end{cases}
\tag{26}
$$

As, for $c < a_{-a}d_A/a_a p_A$, $A_{(2)}$ is negative, the protein concentration at the steady state of the rate equations (24) is zero ($A_{(1)}=0$). For $c > a_{-a}d_A/a_a p_A$, the protein concentration at equilibrium is $A_{(2)} = \frac{c p_A a_a - d_A a_{-a}}{d_A a_a}$. Therefore, the conservation law (23) tune a bifurcation for $c = a_{-a}d_A/a_a p_A$ (transcritical bifurcation), implying the existence of a threshold effect tuned by the conservation law parameter c.
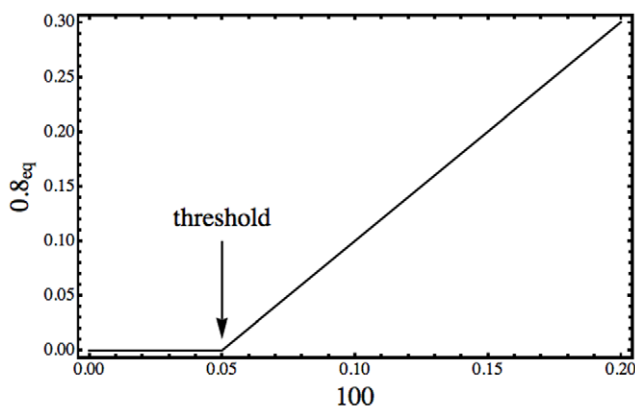
In Figure 5, we show the dependence of the protein steady state on the total concentration of the gene. In this simple regulation model, where both protein and gene concentrations are modeled, the steady state of the protein depends on the initial concentration of the corresponding operon states. On the other hand, the initial concentration of the operon states induce a bifurcation from a quiescent state to a non zero steady state. This is a threshold effect that emerges from the dynamics (24). As the steady state protein concentration depend on c, the "after threshold" concentration values depends continuously on the operon initial concentration c. In the following, we will see that in networks with more than one node, the steady state depends also on the concentration of other transcriptional regulators, and these concentration dependent thresholds can be in the origin of spatial patterning.

Spatial distribution and steady states

We have focused on genetic regulatory models without specifying a spatial localization. In genetic networks describing some biological process, the initial concentration of proteins can significantly vary across tissues. For example, in some developmental processes, proteins show a non-uniform concentration along embryos, with very sharp slopes, playing a basic role in the establishment of body plans of organisms. A well known case is the *Drosophila* segmentation, where variations on protein concentrations across the embryo induces protein patterning, [17–19]. One of such genetic regulatory networks is the one represented in Figure 1, [2].

To show that patterning can be explained by the non homogeneity of initial conditions of regulators across tissues, we analyze a genetic regulatory network for the production of a protein B, regulated by one activator A and two repressor proteins R and S, Figure 6. To simplify our analysis, we take the competitive case, where the activator and the repressors bind to the same binding site of the operon of protein B.

To simplify further, we assume that the spatial distribution of the proteins A, R and S are constant in time. Under these conditions and with the regulation model developed in the package *GeneticNetworks*, we obtain the system of linear rate equations,

$$
\begin{aligned}
B' &= p_B B_A - d_B B \\
(B_A)' &= b_a A B_0 - b_{-a} B_A \\
(B_R)' &= b_r R B_0 - b_{-r} B_R \\
(B_S)' &= b_s S B_0 - b_{-s} B_S
\end{aligned}
\tag{27}
$$

where the derivative is in order to time, $B = B(x,t)$, $B_0 = B_0(x,t)$, $B_A = B_A(x,t)$, $B_R = B_R(x,t)$, $B_S = B_S(x,t)$, $A = A(x)$, $R = R(x)$ and $S = S(x)$. These concentration variables are defined in a spatial one-dimensional bounded region of the real line ($x \in I \subset \mathbb{R}$). The following conservation law holds,

$$
B_0(x,t) + B_A(x,t) + B_R(x,t) + B_S(x,t) = c(x)
$$

where $c(x)$ is a constant, depending eventually of the spatial independent coordinate x. The system of rate equations (27) has one steady state with coordinates,

$$
\bar{B} = \frac{p_B \bar{B}_A}{d_B} \quad , \quad \bar{B}_A = \frac{cAb_a b_{-r} b_{-s}}{D}
$$
$$
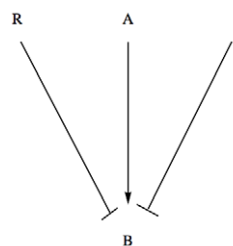\bar{B}_R = \frac{cRb_{-a} b_r b_{-s}}{D} \quad , \quad \bar{B}_S = \frac{cSb_{-a} b_{-r} b_s}{D}
$$

where,

$$
D = b_{-a} b_{-r} b_{-s} + Ab_a b_{-r} b_{-s} + Rb_{-a} b_r b_{-s} + Sb_{-a} b_{-r} b_s
$$

Choosing $b_a = b_r = b_s$, and $b_{-a} = b_{-r} = b_{-s}$, the steady state concentration of the protein B, is,

$$
\bar{B}(x) = \frac{c(x)A(x)p_B}{(1 + A(x) + R(x) + S(x))d_B}
\tag{28}
$$

In Figure 7, we show the steady state concentration (28) of protein B, as a function of a spatial coordinate, $x \in I = [0,1]$. We have considered the initial distributions $A(x) = 0.8$, $R(x) = 83e^{-7x}$, $S(x) = 83e^{-7(1-x)}$, $c(x) = 100$, and the parameter value $p_B/d_B = 1$. In this case, due to the inhibitory regulation of the repressor proteins R and S, the steady distribution of protein B is spiky. We have analyzed the same genetic network of Figure 6 with a model with different binding sites for each regulator. The final result is similar with the one shown in Figure 7. This shows that the two model approaches in *GeneticNetworks*, with one binding site and with several binding sites in the operon, give similar qualitative results. When, the calibration and validation of models is not a problem, we can use the simplest one binding site regulation model in order to describe a given genetic regulatory

Figure 5. Dependence of the protein steady state on the total concentration of the gene. Below the bifurcation or threshold $c = a_{-a} d_A / a_a p_A$, the equilibrium value of protein concentration A_{eq} is zero, while above it takes the value $A_{eq} = (p_A c a_a - d_A a_{-a})/(d_A a_a)$. The parameters are: $a_a = 1.0$, $a_{-a} = 0.1$, $p_A = 0.2$ and $d_A = 0.1$.
doi:10.1371/journal.pone.0010743.g005

Figure 6. Genetic regulatory network for the production of protein B with one activator A and two repressors R and S.
doi:10.1371/journal.pone.0010743.g006
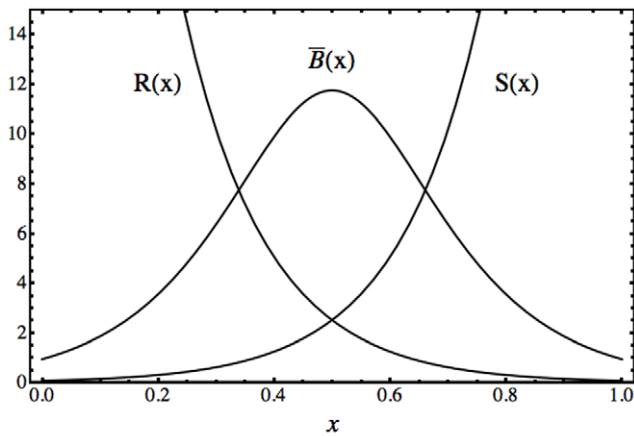
Figure 7. Steady states of proteins B, R **and** S **for the genetic regulatory network of Figure 6.** The steady state of protein B shows a spiky profile, resulting from the inhibitory action of proteins R and S. In this model, we have considered that the concentrations of R and S are constant in time and non-homogeneous in space. The activator protein A has been considered constant along the spatial region.
doi:10.1371/journal.pone.0010743.g007

Figure 8. Concentration of protein Hunchback (HB) at the end of cleavage cycle 13, and of Bicoid (BCD) and Tailless (TLL) proteins at the cleavage cycle 14, along the antero-posterior axis of the embryo of *Drosophila***.** The embryo length has been scaled from 0 to 100, and the units in the vertical axis are proportional to protein concentration. The data has been taken from the FlyEx database. The continuous curves represent the fitted mean distribution of the concentration of proteins calculated from the data of 954 embryos. These curves are the initial conditions for a model obtained with the software package *GeneticNetworks* for the production of the proteins KNI and zygotically produced HB, during cleavage cycle 14.
doi:10.1371/journal.pone.0010743.g008

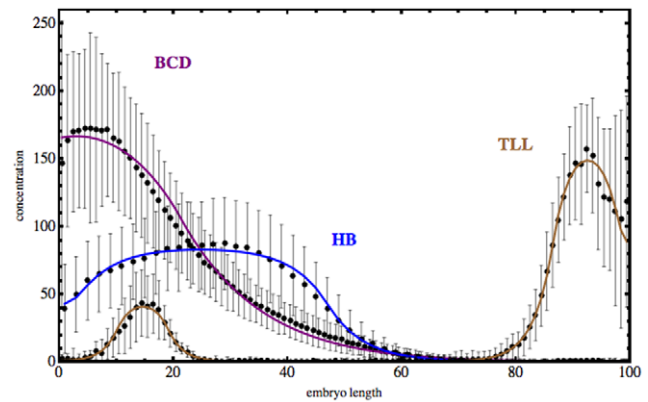network. The one-regulator site framework leads to simpler mathematical models.

The solution (28) shows that steady states can depend on the initial conditions of the regulators ($A(x)$, $R(x)$ and $S(x)$) and on the initial conditions of the catalytic agents ($c(x)$), showing that spatial patterning can be a direct consequence of the non homogeneity of initial conditions.

In this model, we have considered that the concentrations of A, R and S are constants, implying that the model equations (27) describe a system where A, R and S have a fast recovery time. This situation only occurs in thermodynamically open systems, as is the case of biological systems.

Cross-regulation in *Drosophila*

In *Drosophila* early development, some proteins as Bicoid (BCD) and Hunchback (HB) are translated from mRNA of maternal origin. Early in the first developmental stages of *Drosophila*, at cleavage stage 13, BCD and HB proteins form a stable concentration gradient along the antero-posterior axis of the *Drosophila* embryo. In Figure 8, we show the data for these protein gradients, taken from the FlyEx database, [20,21, http://flyex.ams.sunysb.edu/flyex/]. These stable gradients have are established by diffusion processes occurring in the syncytial blastoderm of the embryo, [17–19,22]. At a latter stage, in the 14th cleavage stage, other proteins as Knirps (KNI) show segments characterized by spiky concentration patterns along the antero-posterior axis of the embryo, [1,2,23–26].

We show now that the patterning of HB and KNI proteins as observed at late cleavage stage 14 of the embryo of *Drosophila* is due to the concentration gradients of proteins at an early developmental stage. This results follows from the concentration dependent threshold effect, just described previously, without assuming diffusion for KNI and for zygotically produced HB. For that, we have taken the genetic regulatory network of Figure 1, describing the genetic regulation of HB and KNI, and we have used the package *GeneticNetworks* to describe the production of proteins Hunchback and Knirps during the cleavage cycles 14 of the developing embryo of *Drosophila*.

Hunchback and Knirps proteins are both activated by the maternally produced protein Bicoid, Figure 1, and they mutually repress each other. The protein Knirps is also repressed by the protein Tailless. Therefore, the genetic regulatory network model obtained with the package *GeneticNetworks* should lead to the experimental profiles of HB and KNI, as observed at cleavage cycle 14. As the model obtained with the *GeneticNetworks* package is a system of ordinary differential equations that depend on initial data, we have assumed that, at the end of cleavage cycle 13 and beginning of the 14th, the proteins BCD, TLL and HB of maternal origin have a non homogeneous distribution along the embryo, as shown in Figure 8.

In Figure 9, we show the experimental profiles of HB and KNI proteins at cleavage cycle 14, as well as a fit of the experimental data obtained with the model built with the software package *GeneticNetworks*. The model equations is a system of 14 ordinary differential equations for the proteins and corresponding operon states, and have 23 free parameters. In the ordinary differential equations of the model, we have considered that time is also a free parameter. The protein profiles shown in Figure 9 are out of equilibrium patterns, obtained with the integration time $t = 7.80$ s. These equations and the fitted parameter values are presented in Text S1.

To fit the experimental data with the mathematical model, we have assumed that the initial protein concentrations of BCD and TLL are constant over time, and we have also assumed that each regulator has its own binding site. To find the numerical values of the model parameters in order to calibrate the model, we have used an optimization technique based on a genetic evolutionary algorithm, minimizing a sum of chi square functions, [23,27].

The fitted curves in Figure 9 show a very good agreement with the experimental data. HB fits well in the embryo length range [5%,80%], and KNI fits well in the embryo length range [20%,100%]. The quality of the fits has been measured by the penalized chi square test. For the two fits in Figure 9, we have obtained the reduced chi square values $\chi^2_{HB} = 0.57$ and $\chi^2_{KNI} = 0.68$. This calibration of the genetic regulatory network
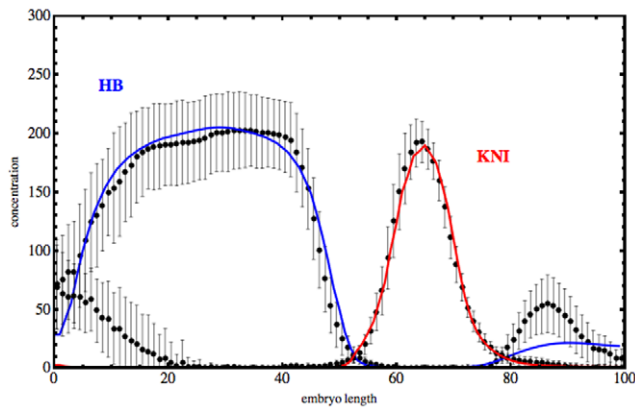
Figure 9. The dots with error bars are the mean experimental profiles of the proteins HB and KNI at cleavage cycle 14, taken along the antero-posterior axis of the embryo, for several *Drosophila* **embryos.** The mean distribution of the concentration of proteins has been calculated from the data of 954 embryos, taken from the FlyEx database. The embryo length has been scaled from 0 to 100, and the units in the vertical axis are proportional to protein concentration. The continuous curves show the predictions of the model of production of proteins HB and KNI. The model equations have been obtained with the software package *GeneticNetworks*.
doi:10.1371/journal.pone.0010743.g009

of Figure 1 suggests that the anterior and posterior regions of the embryo are under the control of additional regulators. This same conclusion has been obtained in [25], but within a different regulation model. On the other hand, this mass action law model without diffusion at the level of gap genes is simpler than other full diffusion models, and is consistent with the reverse engineering methodology described in [26].

Discussion

We have presented a software tool to build mathematical models of genetic regulatory networks. The input of the package is the graph containing the list of transcriptional activators and repressors of the network. The software implements an approach based on the mass action law and on the operon regulation model in prokaryotes. We have also assumed in general that genes are catalytic substances presented in any genetically controlled biological process. For eukaryotic organisms, the modeling approach proposed here should be understood as a descriptive approximation to the not well understood eukaryotic regulation mechanisms.

Within this approach, the usual threshold concept in biology emerges as a bifurcation phenomenon of the model equations. These bifurcations are tuned by the conservation law constants of the equations, resulting from the catalytic role of genes. This

corroborates the view that threshold effects should be anchored on bifurcation phenomena, [16].

Another consequence of the modeling approach presented here is that positional information in developmental processes can be described by the non-homogeneity in the spatial distribution of the concentration of regulators, and is not necessarily associated with other physical processes of transport or diffusion. Other models for *Drosophila* development include a balance between protein diffusion and degradation, [2,28–30]. Recently, some criticism to the diffusion-degradation hypothesis for proteins, [31], suggest that it is important to search for other mechanisms of pattern formation, [22,32,33]. The results presented here show that other mechanisms of spatial patterning are possible.

In conclusion, we have calibrated a genetic regulatory network for the production of Hunchback and KNI during the 14th cleavage stage of the embryo of *Drosophila*, without assuming the hypothesis of protein diffusion for KNI and zygotically produced HB, and we have presented evidence that gap gene protein segments are out of equilibrium patterns. The genetic regulatory network of Figure 1 describes well the gap gene protein concentration of HB in the embryo length region [5%,80%], as well as the gap gene protein concentration of KNI in the embryo length region [20%,100%]. The out of equilibrium pattern hypothesis has been suggested in [28] in the framework of a model assuming that the HB and KNI proteins diffuse along the antero-posterior axis of the embryo of *Drosophila*. In [30], patterning at the gap gene stage is associated with the existence of attractors in an high dimensional phase-space, implying that gap gene patterns are obtained as steady state patterns. The necessity of concentration dependent thresholds in the gap-gene *Drosophila* patterning has been discussed in [24], and modeled through a Hill type response function with diffusion. Here, with mass action law approach, gap-gene patterns emerge from the concentration dependent thresholds that result from the catalytic role of genes in organisms.

Supporting Information

Text S1 Ordinary differential equation model describing the genetic regulatory network of Figure 1.
Found at: doi:10.1371/journal.pone.0010743.s001 (0.06 MB PDF)

References

1. Sánchez L, Thieffry D (2001) A logical analysis of the Drosophila gap-gene system. J Theo Bio 211: 115–141.
2. Alves F, Dilão R (2006) Modeling segmental patterning in Drosophila: Maternal and gap genes. J Theo Bio 241: 342–359.
3. Jong HD (2002) Modelling and simulations of genetic regulatory systems: a literature review. J Comput Biol 9: 67–103.
4. Klipp E, Herwig R, Kowald A, Wierling C, Lehrach H (2005) Systems Biology in Practice. Weinheim: Wiley-VCH. pp 282–286.
5. van Kampen NG (1992) Stochastic Processes in Physics and Chemistry. Amsterdam: North-Holland. pp 166–173.
6. Volpert AI, Volpert VA, Volpert VA (1994) Traveling Wave Solutions of Parabolic Systems. ProvidenceRhode Island: American Mathematical Society. pp 299–300.
7. Horn F, Jackson R (1972) General Mass Action Kinetics. Arc Rat Mec Anal 47: 187–194.
8. Nicolis G, Progogine I (1977) Self-Organization in nonequilibrium systems. New York: Wiley & Sons. pp 395–396.
9. Dilão R, Muraro D (2009) Software packages can be downloaded from the site https://sd.ist.utl.pt/Download/download.html.
10. Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3: 318–356.
11. Alves F, Dilão R (2005) A simple framework to describe the regulation of gene expression in prokaryotes. Com Rend Biologies 328: 429–444.
12. Crick F (2006) Central Dogma of Molecular Biology. Nature 227: 561–563.
13. Ellner SP, Guckenheimer J (2006) Dynamic Models in Biology. Princeton: Princeton Uni. Press. pp 8–12.

14. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, et al. (2006) COPASI Ñ a COmplex PAthway SImulator. Bioinformatics 22: 3067–74.

15. Maiwald T, Timmer J (2008) Dynamical modeling and multi-experiment fitting with PottersWheel. Bioinformatics 245: 2037–2043.

16. Tyson JJ (2007) Bringing cartoons to life. Nature 445: 823.

17. Driever W, Nüsslein-Volhard C (1988) A gradient of bicoid protein in Drosophila embryos. Cell 54: 83–93.

18. Gilbert SF (1997) Developmental Biology. Sunderland: Fifth Edition, Sinauer. pp 547–548.

19. Nüsslein-Volhard C (1992) Gradients that organize embryo development. Scientific American 275(2): 54–61.

20. Pisarev A, Poustelnikova E, Samsonova M, Reinitz J (2009) FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. Nucleic Acids Research 37: D560–D566.

21. Poustelnikova E, Pisarev A, Blagov M, Samsonova M, Reinitz J (2004) A database for management of gene expression data in situ. Bioinformatics 20: 2212–2221.

22. Dilão R, Muraro D (2010) mRNA diffusion explains protein gradients in Drosophila early development. J Theo Bio doi:10.1016/j.jtbi.2010.03.012.

23. Dilão R, Muraro D (2010) Calibration of a genetic network model describing the production of gap gene proteins in Drosophila. Pre-print: arXiv:0912.4391v1.

24. Jaeger J, Surkova S, Blagov M, Janssens H, Kosman D, et al. (2004) Dynamic control of positional information in the early Drosophila blastoderm. Nature 430: 368–371.

25. Jaeger J, Sharp DH, Reinitz J (2007) Known maternal gradients are not sufficient for the establish-ment of gap domains in Drosophila melanogaster. Mech Dev 124: 108–28.

26. Perkins TJ, Jaeger J, Reinitz J, Glass L (2006) Reverse Engineering the Gap Gene Network of Drosophila melanogaster. PLoS Comp Biol 2: e51, 417–428.

27. Dilão R, Muraro D, Nicolau M, Schoenauer M (2009) Validation of a morphogenesis model of Drosophila early development by a multi-objective evolutionary optimization algorithm. In: Pizzuti C, Ritchie MD, Giacobini M, eds. EvoBIO 2009, Lecture Notes in Computer Science 5483. pp 176–190.

28. Fomekong-Nanfack Y, Kaandorp JA, Blom JG (2007) Efficient parameter estimation for spatio-temporal models of pattern formation: Case study of Drosophila melanogaster. Bioinformatics 23: 3356–3363.

29. Houchmandzadeh B, Wieschaus E, Leibler S (2005) Precise domain specifica-tion in the developing Drosophila embryo. Phys Rev E 72: 061920.

30. Manu, Surkova S, Spirov A, Gursky V, Janssens H, et al. (2009) Canalization of Gene Expression and Domain Shifts in the Drosophila Blastoderm by Dynamical Attractors. PLoS Comp Biol 5: e1000303.

31. Kerszberg M, Wolpert L (2007) Specifying Positional information in the embryo: Looking Beyond Morphogens. Cell 130: 205–209.

32. Reinitz J (2007) A ten per cent solution. Nature 448: 420–421.

33. Coopey M, Berezhokovskii AM, Kim Y, Boettinger AN, Shvartsman SY (2007) Modeling the bicoid gradient: Diffusion and reversible nuclear trapping of a stable protein. Dev Biol 312: 623–630.