Morphogenesis of *Drosophila* early development: the even-skipped protein

Beatriz Albergaria*
Nonlinear Dynamics Group, Departamento de Física, Instituto Superior Técnico – IST, Universidade de Lisboa – UL, Avenida Rovisco Pais 1, 1049, Lisboa, Portugal
(Dated: October 2, 2020)

Drosophila produces, in its early development, a series of proteins which regulate themselves in a cascade manner, such that the first have an ill defined pattern and the last have a high frequency pattern. The first two groups of these proteins have already been modeled considering reaction and diffusion mechanisms. Nonetheless, the pair-rule class, with a seven stripes profile and which includes the *Even-skipped* protein, is still missing a model to predict its steady-state and time evolution. In this thesis, we show that the local reaction mechanisms of the preceding proteins are insufficient to regulate such profile. Moving to reaction-diffusion mechanisms, and using the Brusselator as a toy model for Turing patterns formation, we are able to qualitatively predict the pair-rule pattern evolution and the setting of the final seven stripe profile. This same approach is undertaken for the last class of embryonic proteins, the segment-polarity, and its fourteen stripe pattern is obtained was well.

I. Introduction

A. *Drosophila* early development

In insects, the definition of periodic band structures along the antero-posterior axis of the body occurs early in embryo development. This periodic structure determines the segment organisation of the embryo, conditioning the following morphogenic processes. These processes are going to set the differentiation of adult functional areas that follow the insect development. For this periodic structure to emerge, just like all morphological structures, several complex networks of gene regulatory pathways need to be developed at the cellular level. *Drosophila melanogaster*, also known as the fruit fly, is an example of one of these insects.

Drosophila's shape is firstly set before fertilisation, when the mother places mRNA of *bicoid (bcd)* and *caudal (cad)* at the poles of the oocyte. After fertilisation, the maternal genes are translated into proteins, and these are going to determine the larger body parts (*Bcd* defines the anterior axis while *Cad* defines the posterior).

The first 13 nuclear divisions (that are not followed by cytoplasm division) give rise to a *syncytial blastoderm*, in which all the nuclei share the same cytoplasm and have migrated along the cell's surface. It is, therefore, very easy for morphogenes and nutrients to spread between nuclei. During this stage, the *zygotic genes* are expressed, and these are going to be responsible for the progressive differentiation of the fly's body pattern. The zygotic genes are transcribed in certain regions of the embryo's syncytial blastoderm, and the resulting proteins will act as transcription factors that regulate genes that will be transcribed afterwards. Accordingly, we observe a hierarchy, or cascade, of genes: the genes to be expressed earlier regulate the activity of those to be expressed later, where the former control large domains - entire body regions - and the later control minor domains, that result in segmentation patterns [1] [2].

The first zygotic genes to be expressed the are the *gap genes*, that affect entire regions of the body, and these are regulated by maternal-effect and gap proteins. These gap genes include giant *(gt)*, kruppel *(kr)*, hunchback *(hb)* and knirps *(kni)*. After that, the expression of *pair-rule genes* occurs, which are the first proteins to show a periodic structure, creating a seven band structure perpendicularly to the antero-posterior axis; these include fushi-tarazu *(ftz)*, even-skipped (*eve*), runt *(run)*, sloppy-paired *(slp)*, odd-paired *(opa)*, odd-skipped *(odd)*, hairy (*h*) and paired *(prd)* - see figure 1 for the final concentration profile of the first two, together with maternal *bcd* and *cad*.

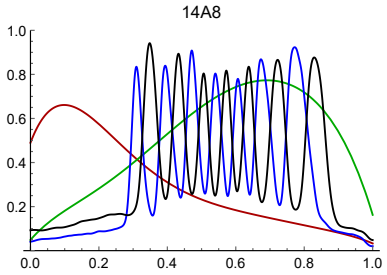

Figure 1. Normalized concentration through normalized embryo length of maternal proteins *Bcd* (red), *Cad* (green), *Eve* (blue) and *Ftz* (black) in relation to *Eve* concentration. These plots were made with fluorescence images from the FlyEx database, and the concentration profiles were normalized to *Eve* maximum at cycle 14A8. We used the central part of the embryo and convolved the data with a gaussian filter.

B. Stripe regulation

The majority of experiments made in order to understand pair-rule's stripe formation have a biological approach, and are based on the *principle stripe mechanism*. According to this view, the pair-rule gene expression is controlled in transcription, such that transcript levels are enhanced within stripe domains and diminished between them, that is, the stripe region itself is activated whereas the interstripe region is repressed. Taking into account that the *Drosophila* genes are expressed in a cascade manner, the pair-rule genes are thought to respond directly to gap gene positional cues via extensive upstream promoters with independent regulatory elements for individual stripes. In other words, the stripes are thought to be formed individually - each of them is activated and repressed

* beatriz.albergaria@tecnico.ulisboa.pt

in its boundaries by a specific set of genes [3].

Accordingly, several experiments have been made where a control pattern of *eve* with some stripes is observed, together with an embryo where selected genes have been mutated. The experiments typically concern the genes *bcd*, *kr*, *kni*, *gt* and *hb*. For mutations in *bcd*, it is observed an enormous reduction of stripe 7 intensity and an anterior expansion and shift of stripe 3 [4].

Moving to the gap genes, in mutants lacking *kr* function it is reported that there is a more pronounced expression of stripe 2, and stripes 4, 5, 6 and 7 are transformed into two new different stripes [5]. In what concerns gap gene *kni*, [4] reports that *eve* pattern shows significant alterations in the posterior part of its mutants: although stripes 1 and 2 remain normal, there is a broad staining that encompasses stripes 3 through 7. In [4], stripes 3 and 7 of *eve* in *hb* mutants are presented, where we observe that the anterior border of stripe 3 is expanded, as well as the posterior border of stripe 7. Concerning *gt* mutants, [5] reports that, without the expression of this gap gene, the anterior border of *eve*'s stripe 2 becomes substantially extended to the anterior part. Furthermore, in these mutants stripes 5 and 7 are merged into a broad staining.

Moving to the pair-rule regulation, it is reported that *prd*, *odd*, *opa* and *slp* do not appear to affect the establishment or the maintenance of the *eve* pattern. Yet, *h* and *run* are essential for the normal maintenance of the *eve* expression during gastrulation: in *run* gastrulating embryos, *eve* pattern is overexpressed and *ftz* products disappear prematurely; in contrast, in *h*- the *eve* expression fades prematurely while *ftz* is overexpressed [6]. In what respects *eve* itself, genetic studies also suggest that this gene contains auto-regulatory elements [7]. In what concerns *eve* and *ftz* mRNAs, it is reported in [8] that, although *ftz* pattern initiation and evolution does not require *eve* function, its maintenance and refinement does. Regarding *ftz* regulation, in [9] its pattern was observed as mutations in the primary pair-rule *eve*, *run* and *h* were separately made, and the conclusion was barely the same for the three of them: none of these genes is key for *ftz* pattern formation, as the stripes' initial elaboration is not affected by the mutations, nonetheless, the pattern fades faster when the mentioned pair-rule are not present. The *ftz* gene has also been studied on the genetic level, and its regulation is reported to be controlled by two elements: the *zebra element*, which confers the striped pattern by mediating the effects of other genes, and the *upstream element*, an enhancer element requiring *ftz* activity for its autoregulatory action [10]. In summary, there are evidences that the pair-rules *ftz* and *eve* are regulated by other genes as well as by themselves, but no pair-rule has shown to be fundamental for their development.

II. Motivation and structure

The maternal and the gap genes have been analytically modeled and calibrated, considering their genetic regulatory network [11] [12]. For the *eve* gene, several proposals have been made for the complete regulatory network that underlines *eve* development, taking into account the mutation experiments described above, and according to the principle stripe mechanism. Via statistical relationships between the stripes positions and the concentration profiles of gap and maternal genes, each of *Eve* stripe's expression has been successfully predicted, and a group of repressors and activators for each has been identified [13] [14]. Enhancers and promoters for stripes have also been identified [7] [15], but it is not clear that each stripe has an individual set of promoters.

The objective of this thesis is, therefore, to model both the pair-rule steady-state profile and its evolution in time, without recurring to the principle stripe mechanism. In section IV, we propose a model which considers activation and repression of *Eve* by the earlier stage proteins, since this is the approach considered in the majority of the literature. In section VI, we move to test a reaction-diffusion model, where *Eve* pattern is produced by a Turing mechanism. Finally, in section VII, we propose a dynamical system identification algorithm to accurately identify the model equations.

III. Protein regulation

The central dogma of molecular biology is a well-established theory for the transmission of information in living organisms: DNA is copied to other DNA molecules in cell division in *replication*, and it is also copied to an mRNA molecule via *transcription* in order to form a template for protein synthesis through *translation* [16]. We now focus on how activation and repression mechanisms may happen in gene expression. In the original DNA molecule there is a *binding site* for a specific gene, to which a *transcription factor* (TF), which is a protein, may bind. There are DNA sequences, the *enhancers*, which are binding sites for regulatory proteins that affect RNA polymerase activity. If the TF is a *promoter*, the RNA polymerase is free to bind to the binding site and initiate the transcription process; if the TF is a *repressor*, the RNA polymerase is unable to connect and gene expression does not occur. There is also the possibility that the protein which is translated is the promotor or the repressor of its own transcription, and in these situation we are dealing with self-activation or auto-catalysis and self-repression, respectively [17].

In order to study these processes in a quantitative manner, we need to describe the kinetics of the reactions involved: given the individual reactions that are likely to occur between all possible chemical species in the system, if we want to model the rates at which the system's species concentrations evolve with time, and how these rates depend upon those concentrations, we then need to write the *reaction rate equations*. The rate at which these reactions occur is k, the *reaction rate constant*: normally, these rates are independent of concentration but are not independent of temperature (and we will not analyze the latter case). Regarding the formalism of a chemical reaction, the system we consider has a total number of m chemical substances and n chemical reactions. Representing the species labeled j by A_j, the reaction rate of the i^{th} reaction by k_i and the stoichiometric coefficients of the reactants and products by ν_{ij} and μ_{ij} respectively, the reactions occurring in the media can be represented by n collision diagrams [18]:

$$\nu_{i1}A_1 + \ldots + \nu_{im}A_m \rightarrow^{k_i} \mu_{i1}A_1 + \ldots + \mu_{im}A_m \,(i = 1, ..., n). \tag{1}$$

Furthermore, if $\nu_{ij} = \mu_{ij} > 0$, the corresponding substance A_j

is a *catalyst* and, if $\mu_{ij} > \nu_{ij} > 0$, A_j is an *autocatalyst* [19]. Under these conditions, the time evolution of the concentration of all the chemical substances is described by the set of m ordinary differential equations, the *law of mass-action* [18]:

$$\frac{dA_j}{dt} = \sum_{i=1}^{n} k_i(\mu_{ij} - \nu_{ij}) A_1^{\nu_{i1}} \dots A_m^{\nu_{im}}. \tag{2}$$

If there is mass conservation, we can also write the *conservation equations* $\sum_{j=1}^{n} \alpha_{jk} A_j = constant_k$, where $k = 1, \dots, s$, for some constants α_{jk} and an integer s.

A. A mathematica model for protein gradients

In order to write the differential equations for the mathematical model that will translate a genetic network into a gradient of protein concentrations, we need to have the kinetic diagrams (1) that represent the interactions of the system, that will afterwards return the differential equations according to the mass-action law (2) and according to the conservations of the system. When these equations are solved, with a numerically method in the large majority of the cases, the steady-state solutions will describe the experimental steady-state of protein concentration.

We start by defining a model for protein production from mRNA, and mRNA production from DNA. The complex binding site plus gene will be called an *operon* - see figure 2 .

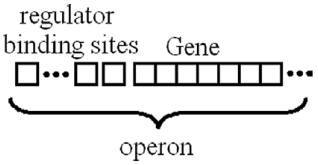

Figure 2. Jacob and Monod operon model where transcriptional regulators bind to the *regulator binging sites*, repressing or activating the translation of the *gene* into mRNA. Image taken from [17].

We start by considering the case where a transcription activator binds to an operon site, following the initiation of transcription and translation - the case of positive regulation. If we represent the activator concentration by A, the concentration of the gene to be transcribed by G and the binded DNA-activator complex by G_A, the mechanism for protein production is:

$$\begin{aligned} &A + G \underset{k_2}{\overset{k_1}{\rightleftharpoons}} G_A \\ &G_A \xrightarrow{a} G_A + P \\ &P \xrightarrow{d} \\ &A \xrightarrow{d_A} \end{aligned} \tag{3}$$

where the last reaction may be omitted if we do not consider that the activator degrades.

The case for negative regulation will be analogous, but instead of considering an activator A, we consider a repressor R, and instead of considering the complex G_A, we consider G_R:

$$\begin{aligned} &R + G \underset{k_4}{\overset{k_3}{\rightleftharpoons}} G_R \\ &G \xrightarrow{a} G + P \\ &P \xrightarrow{d} \\ &R \xrightarrow{d_R} \end{aligned} \tag{4}$$

IV. Reaction only model

In [12], an analytical model based on the mass action law (local reaction type) is shown to be sufficient to reproduce gap and maternal proteins regulatory network and their expression. [13] is also able to reproduce maternal, gap and pair-rule expression using a thermodynamic description for the genes regulation by transcription factors and, assuming independent contribution from multiple enhancers, it is able to very accurately calibrate experimental data and predict unknown parts of *eve*'s regulatory network. Nevertheless, the assumption that each stripe border is independently regulated by a given group of maternal and gap proteins [13] does not have a clear biological ground. It is possible that this is an *ad hoc* theory to explain the sudden increase of spatial frequency that we observe from gap to pair-rule proteins. Therefore, we keep the theory that *Eve* is the result of a set of local activations and repressions, but consider that these mechanisms occur through all embryo length, which seems more biological plausible than having a specific controlling proteins for each stripe.

A. Implementation

Following the mathematical model for protein production in III, we consider that each pair-rule gene is positively regulated - see mechanism (3) - by a set of n gap or maternal genes, $A_1(x), \dots, A_i(x), \dots, A_n(x)$, that have too a distribution over the embryo length x; moreover, it is also negatively regulated - see mechanism (4) - by a set of m gap or maternal genes, $R_1, \dots, R_m$. The set of activators can be replaced by an effective activator, $A(x) = A_1(x) + \dots + A_i(x) + \dots + A_n(x)$, and the set of activators can be replaced by an effective repressor, $R(x) = R_1(x) + \dots + R_i(x) + \dots + R_n(x)$. As *Eve* and *Ftz* proteins disapear later in development, we consider their degradation; on the contrary, since gap and maternal proteins are approximately constant in time, we do not consider the degradation neither of the effective activator nor of the effective repressor. This way, our global mechanism for the production of a pair-rule protein is going to be:

$$\begin{aligned} &G + A \underset{l_2}{\overset{l_1}{\rightleftharpoons}} G_A \\ &G_A \xrightarrow{k_p} G_A + P \\ &G + R \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} G_R \\ &P \xrightarrow{\delta} \end{aligned} \tag{5}$$

According to the mass-action law (2), the time evolution equations are going to be:

$$
\begin{aligned}
A'(t) &= l_{-1}G_A(t) - l_1 A(t)G(t)\\
R'(t) &= k_{-1}G_R(t) - k_1 G(t)R(t)\\
P'(t) &= k_p G_A(t) - \delta P(t)\\
G'_A(t) &= -l_{-1}G'_A(t) + l_1 A(t)G(t)\\
G'_R(t) &= -k_{-1}G_R(t) + k_1 G(t)R(t)\\
G'(t) &= k_{-1}G_R(t) + l_{-1}G_A(t) - A(t)G(t)l_1 - G(t)R(t)k_1,
\end{aligned}
\tag{6}
$$

with the conservation equations:

$$
\begin{aligned}
A(t) - G(t) + R(t) &= A_0 - G_0 + R_0\\
-A(t) + G(t) + G_R(t) &= -A_0 + G_0\\
A(t) + G_A(t) &= A_0,
\end{aligned}
\tag{7}
$$

where $R_0 = R(t=0)$, $A_0 = A(t=0)$ and $G_0 = G(t=0)$. We also consider that the transcription starts at $t = 0$, so that $G_R(t=0) = G_A(t=0) = P(t=0) = 0$. For the positive and negative regulation, we should have $l_1 \gg l_{-1}$ and $k_1 \gg k_{-1}$. We start by considering the evolution in time only, that is, the activator, the repressor and the gene initial concentration have a uniform distribution is space. The numerical solution of the equations for $A(t)$, $R(t)$ and $P(t)$ until steady-state ($t \approx 50$) is reached is depicted in figure 3.

We start by considering the evolution in time only, that is, the activator, the repressor and the gene initial concentration have a uniform distribution is space. The numerical solution of the equations for $A(t)$, $R(t)$ and $P(t)$ until steady-state ($t \approx 50$) is reached is depicted in figure 3.

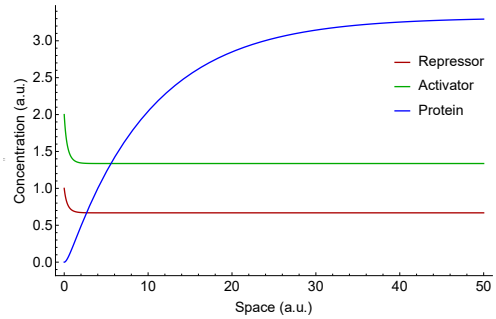

Figure 3. Evolution of the concentrations of the activator, repressor and protein, for $R_0 = 2.0$, $A_0 = 1.0$ and $G_0 = 1.0$, according to (6) and (7).

We then investigate how a pattern with a spatial distribution is altered by the action of an activator and a repressor. We propose an activator and a repressor with the initial spatial distribution of $A(x,0) = 0.5 + 0.1\exp\left(-\frac{(x-0.8)^2}{0.001}\right)$ and $R(x,0) = 0.5 + 0.3\exp\left(-\frac{(x-0.5)^2}{0.001}\right)$ respectively, which are depicted in figure 4, as well as the steady-state distribution in space of the protein, at $t = 50$.

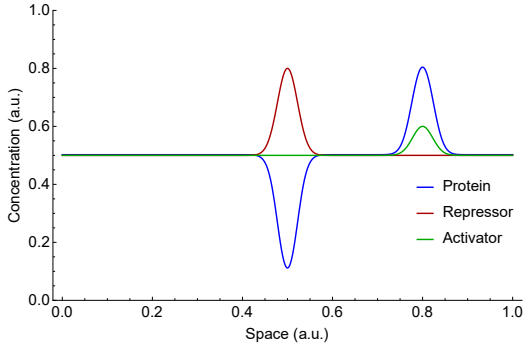

Figure 4. Solution of the steady state ($t = 50$) distribution of the protein for $G_0 = 1.0$, in the presence of an activator and a repressor with the initial distribution seen in figure 4, according to (6) and (7). We subtracted -1.74 to the protein profile for better visualization.

The results are what we intuitively predict: the protein has an increasing of concentration at the position of the activator, and a decreasing of concentration at the position of the repressor.

B. Results

We use this model to fit the normalized *Eve* pattern at cycle 14A8, with the maternal proteins *Bcd* and *Cad*, plus the gap proteins *Kr*, *Kni*, *Gt*, *Hb* and *Tll* as possible activators or repressors, with their profiles at cycle 14A1. We looped possible values for the rate constants between 0 and 10, and thresholded these until a given protein was either an activator or a repressor, and repeated the procedure until the best χ^2 was obtained.The results are shown in figure 5.

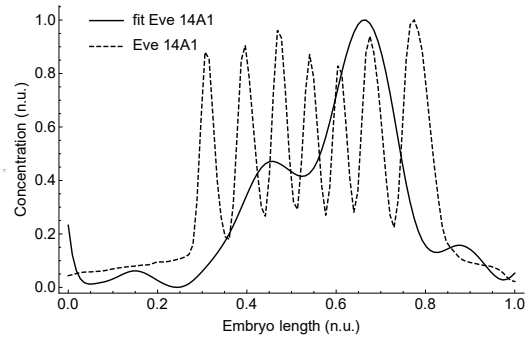

Figure 5. The best fit was obtained with $\chi^2 = 18.2$ after 8000 loops (5 rounds) using 112 points, with the rate constants of table I.

	Bcd	Cad	Kr	Kni	Gt	Hb	Tll
Activation		2.6		4.4		3.2	
Repression	8.9		1.2		6.8		2.1

Table I. Activating and repressing proteins for the best fit, depicted in figure 5, and their respective rate constants ratios.

This way, we conclude that, although we perceive some regularities between the gap and the maternal proteins pattern, as

well as correspondences between possible regions for activation and repression of the pair-rule stripes, these proteins' profiles are not sufficient to produce such a high frequency spatial pattern.

C. Model for pattern formation

V. Reaction-diffusion mechanisms

Alan Turing proposed in his seminal work that morphogenesis, which is the ensemble of processes that determines form, shape and patterns in organisms, is the result of a dynamical system that considers not only reaction but also diffusion [20].

We begin with an homogeneous system with the presence of irregularities which are going to break this homogeneity in the presence of the appropriate kind of instability. This is going to trigger reaction and diffusion mechanisms, that may result in a steady-state patterns, contrary to what happens in the majority of systems with diffusion, which tend to homogeneity. Without the presence of these triggering instabilities, no pattern is formed. Moreover, the pattern that is formed in the steady state may depend on form of the irregularities that prompted it. In what concerns diffusion, this *reaction-diffusion* system typically produces stable patterns when the two species diffuse at very different rates, nonetheless, systems with species diffusing at the same rate have been found too [21].

A. Model for pattern formation

We shall then consider a system of n species, each jth species A_j with a diffusion coefficient D_j: the evolution of a species concentration is a result of a source term due to the reactions with all the other substances plus a term concerning the diffusion, according to the Fick's equation:

$$\frac{dA_j}{dt} = f_j(A_1,\ldots,A_n) + D_j \Delta A_j, \tag{8}$$

where $\Delta = \left(\frac{\partial^2}{\partial x_1^2} + \ldots + \frac{\partial^2}{\partial x_k^2}\right)$ is the k-dimensional Laplace operator. From now on, we will consider a 1-dimensional system, *i.e.* $k = 1$. If the jth substance diffuses (that is, $D_j > 0$) we say that this substance is a *morphogene*, *form producer* or *evocator*. If $D_j = 0$, we identify this species as a non-diffusing substance [18].

B. The Brusselator model

The Brusselator is a simple model able to produce Turing patterns with two species. It mimics an autocatalytic process and has the following kinetic mechanism:

$$\begin{aligned} &\mathrm{A} \xrightarrow{\mathrm{k_1}} \mathrm{X} \\ &\mathrm{B} + \mathrm{X} \xrightarrow{\mathrm{k_2}} \mathrm{Y} + \mathrm{D} \\ &2\,\mathrm{X} + \mathrm{Y} \xrightarrow{\mathrm{k_3}} 3\,\mathrm{X} \\ &\mathrm{X} \xrightarrow{\mathrm{k_4}} \mathrm{E}, \end{aligned} \tag{9}$$

where X is the autocatalytic species. If we apply the law of mass action (2), we obtain the differential equations which will determine the reaction part of the reaction-diffusion system:

$$\begin{aligned} \frac{dX}{dt} &= k_1 A - k_2 BX + k_3 X^2 Y - k_4 X \\ \frac{dY}{dt} &= k_2 BX - k_3 X^2 Y \\ \frac{dA}{dt} &= -k_1 A \qquad \frac{dB}{dt} = -k_2 BX \\ \frac{dE}{dt} &= k_4 X \qquad \frac{dD}{dt} = k_2 BX. \end{aligned} \tag{10}$$

This system obeys to the conservation laws:

$$\begin{aligned} &B(t) + D(t) = B(0) + D(0) \\ &X(t) + Y(t) + A(t) + E(t) = X(0) + Y(0) + A(0) + E(0). \end{aligned} \tag{11}$$

If we now assume that A and B are constants, we obtain the following one-dimensional reaction-diffusion system for species X and Y, taking the first two equations from (10) as the reaction terms:

$$\begin{aligned} \frac{\partial X(x,t)}{\partial t} =& k_1 A - (k_2 B + k_4) X(x,t) + k_3 X^2(x,t) Y(x,t) \\ & - k_4 X(x,t) + D_X \Delta X(x,t) \\ \frac{\partial Y(x,t)}{\partial t} =& k_2 BX(x,t) - k_3 X(x,t)^2 Y(x,t) + D_Y \Delta Y(x,t), \end{aligned} \tag{12}$$

where D_X and D_Y are the diffusion coefficients for species X and Y respectively. For the numerical integration of the system, we use Euler's method for time evolution and finite differences for space evolution - for the latter, we consider Neumann boundary conditions (zero flux). We used $dt = 0.001$ for Euler's method step and $dx = \sqrt{\frac{max(D_X, D_Y)}{\gamma} dt}$ for the finite differences, with $\gamma = \frac{1}{6}$, since this relationship is reported to lead to optimal convergence to the solution of the system [22]. In figure 6, we reproduce the Turing pattern with the parameters $A = 2.0$, $B = 15.0$, $D_X = 0.1$ and $D_Y = 1.0$, $k_1 = k_2 = k_3 = k_4 = 1$. We chose to simulate the system with $M = 250$ lattice sites, thus the length of the spatial domain is $M\sqrt{\frac{\Delta t}{\gamma} max(D_X, D_Y)} \approx 19.52$; the simulation was run until the concentrations of both species in all lattice sites were less that 0.001% different that its previous value (time), and this was our criterion to detect the steady-state while taking into account small numerical deviations. Moreover, for the instabilities that are going to trigger the reaction-diffusion mechanisms, we chose an initial profile with the values at the fixed point plus a random profile $\delta(x)$ with values between 0.0 and 0.5, $X(0,x) = \frac{k_1 A}{k_4} + \delta(x)$ and $Y(0,x) = \frac{k2 k4 B}{A k1 k3} + \delta(x)$[18].

VI. Reaction-diffusion model

A. Development of the initial pattern

Since it was not possible to form a pair-rule's pattern via local activation or repression by the gap and maternal genes, we now propose that these early stage proteins are only responsible for the setting of the initial pattern, which then develops stripes via a Turing mechanism. This hypothesis has

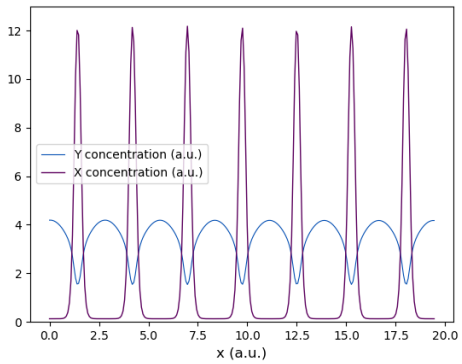

Figure 6. Turing pattern of the Brusselator reaction-diffusion system (12).

already been proposed in [23] and [24]. Moreover, [3] also suggests that these two pair-rule genes are subject to positive auto-regulation control, and [25] and [26] suggest that the stripe-like activation of genes is a result of an autocatalytic feedback, which is experimentally supported as we mentioned in I B.

For the setting of the initial pattern, and following the method described in IV B, we fitted *Eve* pattern at cycle 14A1: the result is shown in figure 7 and the reaction constants in table II.

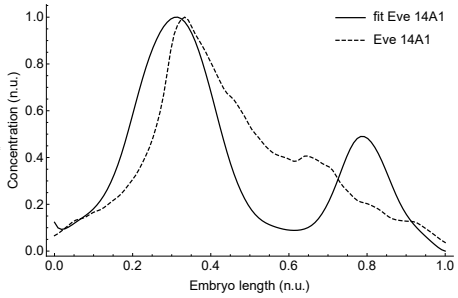

Figure 7. The best fit was obtained with $\chi^2 = 20.7$ after 8000 rounds, using 112 points, with the rate constants of table II. The experimental profile is normalized.

	Bcd	Cad	Kr	Kni	Gt	Hb	Tll
Activation					0.9	0.3	
Repression	9.4	4.3	0.4	0.8			6.4

Table II. Activating and repressing proteins for the best fit, depicted in figure 7, and their respective rate constant ratios.

We have obtained the anterior protein *Bcd* and the posterior proteins *Cad* and *Tll* as major repressors, whereas the other gap genes have a residual influence on this pattern formation, and we also observe an odd concavity at $x \approx 0.6$, caused by *Tll* repression. The fact that *Bcd* and the majoriity of gap genes are repressors meets very well the experiments with mutations described in chapter I B, where the mutations of these caused and enlargement of *Eve* stripes. The exceptions are the gap proteins *Gt* and *Hb*, here identified as activators instead of repressors. We got a very good fit for the anterior part of the profile, until $x \approx 0.3$, where the concentration has its maximum, as the intersept and the second derivatives (positive at first and negative at last) very closely match the experimental data. Thereafter, we fitted *Ftz* pattern at cycle 14A1: the result is shown in figure 8 and the reaction constants in table III.

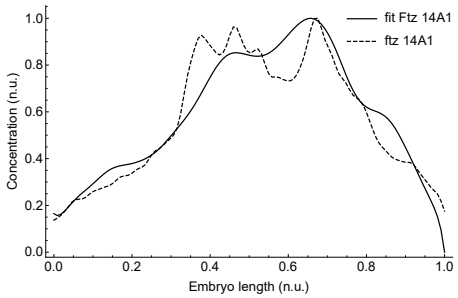

Figure 8. The best fit was obtained with $\chi^2 = 1.7$ after 16000 rounds, with the rate constants of table III.The experimental profile is normalized.

	Bcd	Cad	Kr	Kni	Gt	Hb	Tll
Activation		9.1	0.5	3.5		6.7	
Repression	9.2				6.0		5.9

Table III. Activating and repressing proteins for the best fit, depicted in figure 8, and their respective rate constant ratios.

In what concerns *Ftz*, our fit very closely matches the experimental data, and we once again obtained *Bcd* as a major repressor, and *Cad* is now an activator instead. Moreover, the gap proteins *Kni*, *Gt*, *Hb* and *Tll* no longer have a residual influence on the profile initiation. Nonetheless, the fit does not reproduce the small irregularities in central part: this suggests that, even at this early stage, the proteins *Eve* and *Ftz* have already started interacting.

In summary, these eary stage fits improved very much when compared to the last stage fit at figure 5, which means that our reaction-diffusion hypothesis may have a closer match to reality.

B. Development of stripes

As the initial pattern of *Eve* and *Ftz* are already established, we propose that these two proteins start interacting with each other, using the Brusselator (12) as a template for this reaction, where the two species that feed the system, A and B, are going to be a combination of maternal and gap genes. Since this combination alone is going to set the initial pattern for *Eve* and *Ftz*, we are the going to use $A(x) = Eve14A1$ and $B(x) = Ftz14A1$. We tested all the combinations of 0.01, 0.1 and 1.0 for the diffusion coefficients of species X (D_X) and Y (D_Y), and we obtained three different patterns.

We only obtain a Turing pattern when D_Y is 10 or 100 bigger than D_X, which was expected since in the majority of Tur-

ing patterns one diffusion coefficient is way larger than the other. If we fix $D_Y = 1.0$, there will be Turing patterns for (at least) $D_X \in [0.01, 0.1]$, with 12 stripes for $D_X = 0.01$ and 6 stripes for $D_X = 0.1$; more specifically, we find a 7 stripe pattern for $D_X \in [0.04D_Y, 0.06D_Y]$, and the pattern for the values in this interval is the same, and it is depicted in figure 9.

In order to explore if this is a strict or a reasonable range for the diffusion coefficient, further experiments to measure this constant are necessary.

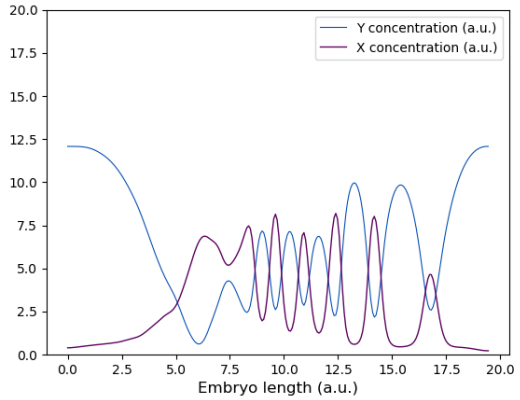

Figure 9. Steady-state pattern of (12) using $A(x)$=Eve14A1, $B(x)$=Ftz14A1 and D_Y=1.00, $D_X \in [0.04, 0.06]$. Y pattern has been enlarged $\times 8$ for better comparison.

For the identification with the time evolution of figure 1, in figures 10 and 11 we show eight frames of the integration of the system for which the steady-state is depicted in figure 9, separated in equal time intervals. The first four stages match quite closely the experimental data: the maximum concentration increases from 1 to 2 and is constant from that time onwards, and the first stripe projections appear in stage 2 in both experimental data and simulations. In the pattern which derives from the Brusselator model, these projections begin to elongate, forming stripes, like what happens in the *Eve* evolution. This way, we conclude that a model that considers reaction-diffusion, with auto-catalysis and interactions for the two species, suits the biological formation of *Eve* stripes.

We also supposed that the species Y with which X interacts and forms a Turing patterns would be the pair-rule complementary *Ftz*. Nonetheless, the Y concentration pattern at cycle 14A8 in figure 9 does not exactly match the steady-state pattern of *Ftz* (see figure 1): although both patterns have the same maximums and minimums (that is, both are complementary with *Eve* or X), their second derivatives have opposite signs in the striped region and, in the side regions without the pattern (for embryo length $\sim< 5.0$ and $\sim> 17.5$), Y concentration decreases towards the center, while *Ftz* increases towards the center. This means that the interaction term in the Brusselator model $\pm X^2Y$ does not describe the pair-rule interaction, and a correct identification of the dynamical system is required. Furthermore, this interaction is essential to form both species' stripes, and it was observed experimentally that *Eve* and *Ftz* can form their patterns without each other, which means that these two proteins may interact with other pair-rules.

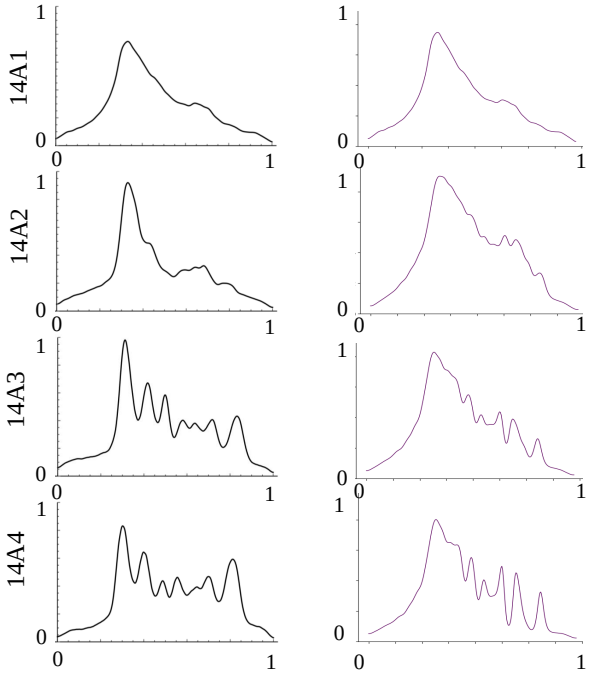

Figure 10. Normalized concentration through normalized embryo length. *Eve* experimental patterns for cycles 14A1, 14A2, 14A3 and 14A4 (left figures) and comparison with four initial frames from the numerical integration of the Brusselator model (12) (right figures), using $A(x)$=Eve14A1, $B(x)$=Ftz14A1 and $(D_X, D_Y) = (0.05, 1.00)$.

C. Segment-polarity stripes

We also propose a similar mechanism for the formation of the segment-polarity stripes. These are the last genes to be expressed in the *Drosophila* embryo, and their establishment occurs during the late cellular blastoderm stage, such that their pattern should determined by the pair-rule proteins, and they have twice the number of stripes (fourteen) as these (seven). Studies have shown that one of these segment polarity, *en*, is activated by both *Eve* and *Ftz*, whereas another segment-polarity protein that develops at its side, *wg*, is repressed by the same pair-rule [27]. These two proteins are reported to react with each other and diffuse [28], what is also consistent with the Turing pattern formation mechanism. Moreover, in *Eve* and *Ftz* embryo mutants, *en* and *wg* have few and broader stripes than in wild type, which tells us that these pair-rule have an important role in these segment-polarity setting [27].

Analogously to what we did with the development of the pair-rule, we propose that these initialize the segment polarity pattern, which then evolves into stripes via a reaction-diffusion mechanism. With the same model for transcription and translation we used for the initiation of *Eve* in section IV A, we present in figure 12 a protein (at green) which was activated by *Eve* and *Ftz*, in equal proportions, and another proteins (at red) which was repressed by the same pair-rule.

Accordingly, these may mimic the initial pattern of *wg* and *en* respectively.

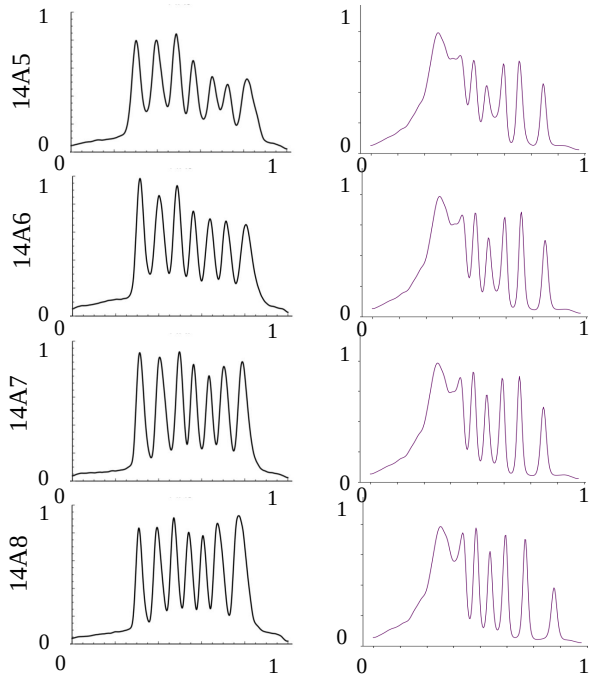

Figure 11. Normalized concentration through normalized embryo length - same as figure 10 but for 14A5, 14A6, 14A7 and 14A8.

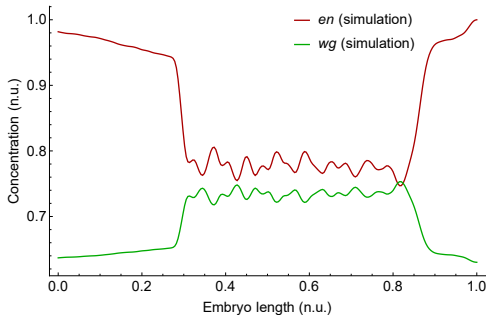

Figure 12. Possible initial pattern for *en* in red (repressed by *Eve* and *Ftz*) and for *wg* in green (activated by same proteins).

For the reaction-diffusion part which is going to refine the stripes, we use the Brusselator as template once again, plus $A(x) = wg$ and $B(x) = en$. We tested all combinations of 0.01, 0.1 and 1.0 for the diffusion coefficients D_X and D_Y, and we obtained three different patterns.

As it happened before with the pair-rule, we obtained a Turing pattern when D_Y is 10 or 100 bigger than D_X. Hence, we looked for Turing patterns once more fixing $D_Y = 1.0$ and searched in the range $D_X \in [0.01, 0.1]$, and we could only find 14 stripes for $D_X = 0.03$, for which steady-state profile is depicted in figure 13. We once more suggest that this protein's diffusion coefficient should be measured in order to evaluate this result.

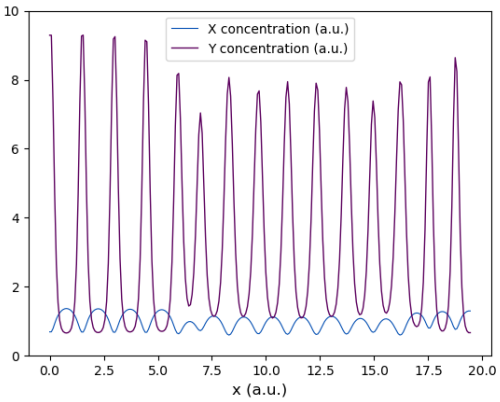

Figure 13. Steady-state pattern of (12) using $A(x) = wg$(simulation) and $B(x) = en$(simulation). This profile was obtained for $(D_X, D_Y) = (0.03, 1.0)$.

We note that, in *en*- embryos, *wg* shows a broader ectopic transcription (vice-versa for *en* and *wg* mutants), but the same results are obtained for mutant embryos of *nkd* and *ptc*, which means that the refinement and maintenance processes does not happen only between these two, just like we proposed for the *Eve* and *Ftz* development [29].

VII. The SINDy algorithm

We now aim to identify the system which regulates the pair-rule formation by taking advantage of an algorithm that gives a sparse identification of nonlinear dynamical systems, the SINDy algorithm, proposed by S. L. Brunton and colleagues [30]. In their work, the authors combine sparsity-promoting techniques with nonlinear dynamical systems to discover governing equations from input data, where the only assumption made on the system is that its equations are sparse in the number of terms, which is reasonable for the great majority of physical systems. For this method, we give as input data the n-dimension system values at m different time instances, $\mathbf{X} \in \mathbb{R}^{m \times n}$. With this, we create a matrix with p candidate functions, $\Theta(\mathbf{X}) \in \mathbb{R}^{m \times p}$. Since only a few candidate functions are going to be selected to describe the dynamical system, we are left with determining the sparse matrix $\Xi \in \mathbb{R}^{p \times n}$. This matrix is successively thresholded by a parameter λ, until a solution which provides both sparsity and fit quality is obtained. The authors propose to determine this parameter as a compromise in these two quantities, so that it is in their Pareto front.

VIII. Identification of the Brusselator steady-state Turing pattern

The SINDy model is reported to be efficient to determine the dynamical system of the damped oscillator, of the Lorenz system and of the Navier-Stokes equations. As benchmarking, we test if this algorithm is able to identify the equations for the steady-state Turing pattern of the Brusselator model, depicted in figure 6, using combinations of X and Y until 4^{th} order as candidate functions. Since this pattern is achieved at a steady-state, the solution verifies a one-dimension system. We then imported the data of the steady-state pattern into the SINDy

algorithm in order to see how it detects the linear and nonlinear functions of this system. By running the algorithm for several values and orders of magnitude of the cut-off parameter λ, we found that we could only fit one species at a time: each time the algorithm converged for one of them, the system identification was rather poor for the other. We obtained a correct prediction of the X species system when $\lambda \in [1.2, 6.7]$ and for Y species when $\lambda \in [0.1, 0.6]$. Besides this, we did not find the correct solutions as a compromise between sparsity and fit quality, as it was proposed by the authors of the algorithm. This means that we need a more profound study on how to choose the sparsification parameter that correctly identifies the dynamical system.

After that, we test the Brusselator system but with forced terms, using the same parameters as before (reaction and rate constants), but making $A = Cad(x)$ and $B = Bcd(x)$ instead. Using polynomials until 3^{rd} order of $X(x)$, $Y(x)$, $Bcd(x)$ and $Cad(x)$ as candidate functions there is an interval of $\lambda \in [0.1, 0.9]$ where both species' equations are correctly identified, unlike what happened in the identification of the simple Brusselator system. Hence, we conclude that SINDy is able to identify the equations of a forced system.

In order to identify the interaction between the pair-rule, we now fit the central part of the pattern only, which for simplicity we approximate with $X(x) = 1 + \sin(x)$, $Y(x) = 1 + \sin(x - \pi)$, and this regulation was identified by the algorithm as $\frac{d^2X(x,t)}{dx^2} = 0.125(Y^2 - X^2)(X + Y)$, $\frac{d^2Y(x)}{dx^2} = 0.125(X^2 - Y^2)(X + Y)$ ($\lambda \in [0.00, 0.15]$, with $\chi_X = \chi_Y = 0$). For bigger values of λ, all terms cancel, which means that in this particular case we do not need a criterion to balance fit quality and sparsity. Although the system's terms are clearly identified, the reaction-diffusion equations diverged with all combinations of 1.00, 0.1 and 0.01 for the diffusion coefficients, using constant profiles with noise as initial conditions as before.

This means that, at least for some systems, the convergence may depend on more specific initial profiles. Moreover, for further identification of the complete pair-rule pattern, other solutions that we lose when we make the system reduction to one-dimensional should also be studied, which can be done by using the transient states.

IX. Conclusion

The genetic network of *Drosophila* is regulated in a cascade manner and the first two classes of proteins, maternal and gap, have already been analytically modeled [12], considering activation and repression interactions between the genes and proteins of the regulatory network, as well as the diffusion of these proteins. Statistical relationships between the pattern of the gap genes and the pattern of the pair-rule have also been studied, and a group of repressors and activators for each stripe has been found. Nevertheless, it has not been proven that each stripe is regulated individually, and it is physically more plausible that each protein is globally regulated by a group of repressors and a group of activators. In this sense, a model with this assumption and that could predict the pair-rules evolution and steady-state still misses.

Firstly, we reviewed the model presented in [17] which translates a genetic network of activation and repression mechanisms into protein gradients, and used it to fit the steady-state of *Eve* what would, in principle, enable us to find which proteins would be activators and which proteins would be repressors. This fit was rather poor, and the resulting pattern was nowhere near to produce a seven stripe pattern. For this reason, we concluded that local type reactions were not sufficient to produce such regularities, and a more robust model would be necessary.

Accordingly, we then tested if the gap and maternal proteins can regulate the 14A1 pattern of the pair-rule proteins *Eve* and *Ftz*, and this earlier stage fit was in a closer agreement with the experimental data. We proposed that these initial patterns, as they are the combination of repressing and activating proteins, are feeding the evolution of the pair-rule; moreover, we consider diffusion and an interaction between them - which means that we proposed that the seven stripes are a Turing pattern, developed via a reaction-diffusion mechanism. Based on the Brusselator, a reaction-diffusion system with Turing patterns, and using the pair-rules initial patterns for the intermediary species of the Brusselator, we were able to obtain steady-state and transient profiles which were in a very close agreement with the experimental data. Different number of stripes were obtained for different values of the diffusion coefficients, and seven stripes patterns were achieved when $D_X \in [0.04D_Y, 0.06D_Y]$.

We also extended this approach for the segment-polarity proteins. The biological experiments indicate that there is one segment-polarity, *en*, which is activated by both *Eve* and *Ftz*, and another segment-polarity, *wg*, which is, in contrary, repressed by the same two pair-rule. Once more, beginning with the reaction model for these regulations, we were able to outline the initial patterns for these proteins. From these early patterns on, and considering the Brusselator system for the reaction-diffusion mechanism, we were able to obtain the fourteen stripe pattern when $D_X = 0.03D_Y$.

Finally, we aimed to identify the dynamical system which regulates the interaction between the pair-rule or between the segment-polarity. To do this, we used the regression algorithm SINDy which, from steady-state profiles, is able to identify the nonlinear functions that compose the system, balancing fit quality and the system sparsity. This algorithm was able to correctly identify the Brusselator system as well as a modification with space-dependent functions, and the interaction terms for the pair-rules and segment-polarity were then successfully identified. Nonetheless, this system diverged when integrated, for all combinations of 1.0, 0.1 and 0.01 for the diffusion coefficients.

In summary, we showed that, despite what is suggested by most literature, the local regulation via activation and repression by the earlier proteins, gap and maternal, is not sufficient to produce a high frequency pattern as the seven stripes pattern is. Moving to the reaction-diffusion hypothesis, we simulated the development and setting of the pair-rule pattern, as well as the segment-polarity steady-state pattern, using the Brusselator system as template model. Our simulations were in a very good agreement with the experimental data, as well as with the regulatory mechanisms proposed in the literature. Using

a regression algorithm for nonlinear systems, the proper interaction terms which generate the central part of the pattern were also identified, but the system was not integrated. For this reason, we suggest a more complete characterization of the solution space of the algorithm in the future, as well as their dependence on the initial profile.

Also as an extension of this work, the investigation of a criterion to determine whether a reaction-diffusion system is able or not to produce Turing patterns would be usefull, because it would enable us to immediately test the SINDy solutions, and no numerical integration would be necessary. Additionally, the observation of *Ftz* pattern with mutations in the gap and maternal proteins would enable us to determine if we correctly identified its repressors and activators. Still concerning the biological part, according to our reaction-diffusion hypothesis the stripes are only formed when the two pair-rule interact with each other. Nonetheless, it is observed experimentally that a mutation in *Ftz* does not affect substantially the pattern of *Eve* and vice-versa. We then suggest to mutate several pair-rules simultaneously, to observe if the interaction mechanisms happen not between two species but between more, and this suggestion applies to the segment-polarity as well. Since we could not compare the evolution of our simulation with the transient profiles of the segment polarity, we suggest to analyze this data.

[1] C. Nüsslein-Volhard, *Coming to life - how genes drive development*. Yale University Press, 2006.

[2] F. Alves, *A mathematical model for segment formation in Drosophila melanogaster*. PhD thesis, Instituto Superior Técnico, 2006.

[3] S. Parkhurst and D. Ish-Horowicz, "Mis-regulating segmentation gene expression in *Drosophila*," *Development*, vol. 111, pp. 1121–1135, 1991.

[4] S. Small, A. Blair, and M. Levine, "Regulation of two pair-rule stripes by a single enhancer in the drosophila embryo," *Developmental Biology*, vol. 175, pp. 314–324, 1996.

[5] D. Stanojevic, S. Small, and M. Levine, "Regulation of a segmentation stripe by overlapping activators and repressors in the *Drosophila* embryo," *Science*, vol. 254, pp. 1385–7, 1991.

[6] M. Frasch and M. Levine, "Complementary pattern of *even-skipped* and *fushi-tarazu* expression involve their differential regulation by a common set of segmentation genes in *Drosophila*," *Genes & Development*, vol. 1, pp. 981–995, 1987.

[7] K. Harding, T. Hoey, R. Warrior, and M. Levine, "Autoregulatory and gap gene response elements of the even[U+2010]skipped promoter of drosophila," *The EMBO journal*, vol. 8, pp. 1205–12, 1989.

[8] K. Harding, C. Rushlow, H. J. Doyle, T. Hoey, and M.Levine, "Cross-regulatory interactions among pair-rule genes in *Drosophila*," *Science*, vol. 233, pp. 953–959, 1986.

[9] Y. Yu and L. Pick, "Non-periodic cues generate seven *ftz* stripes in the *Drosophila* embryo," *Mechanisms of Development*, vol. 50, pp. 163–175, 1995.

[10] Y. Hiromi and W. J. Gehring, "Regulation and function of the drosophila segmentation gene *fushi tarazu*," *Cell*, vol. 50, pp. 963–934, 1987.

[11] R. Dilão and D. Muraro, "Calibration and validation of a genetic regulatory network model describing the production of the protein hunchback in drosophila early development," *Comptes Rendus Biologies*, vol. 333, pp. 779–788, 2010.

[12] F. Alves and R. Dilão, "Modeling segmental patterning in *Drosophila*: Maternal and gap genes," *Journal of Theoretical Biology*, vol. 241, pp. 342–359, 2006.

[13] M. Samee and S. Sinha, "Quantitative modeling of a gene's expression from its intergenic sequence," *PLoS Computational Biology*, vol. 10, 2014.

[14] G. R. Ilsley, J. Fisher, R. Apweiler, A. H. DePace, and N. M. Luscombe, "Cellular resolution models for even skipped regulation in the entire drosophila embryo," *PLoS Computational Biology*, vol. 2, 2013.

[15] D. Stanojevic, T. Hoey, and M. Levine, "Sequence-specific dna-binding activities of the gap proteins encoded by *hunchback* and *Kruppel* in *Drosophila*," *Nature*, vol. 341, pp. 331–335, 1989.

[16] F. Crick, "Central dogma of molecular biology," *Nature*, vol. 227, pp. 561–563, 1970.

[17] F. Alves and R. Dilão, "A simple framework to describe the regulation of gene expression in prokaryotes," *Comptes Rendus Biologies*, vol. 328, pp. 429–44, 2005.

[18] R. Dilão, "The reaction-diffusion approach to morphogenesis," *Proceedings of 4th Brazilian Symposium on Mathematical and Computational Biology*, vol. 1, 2004.

[19] R. Dilão and D. Muraro, "A software tool to model genetic regulatory networks," *PLoS ONE*, vol. 5, pp. 1–10, 2010.

[20] A. M. Turing, "The chemical basis of morphogenesis," *Philosophical Transactions of the Royal Society of London*, vol. 237, pp. 5–72, 1952.

[21] D. E. Strier and S. P. Dawson, "Turing patterns inside cells," *PLoS ONE*, vol. 2, 2007.

[22] R. Dilão and J. Sainhas, "Validation and calibration of models for reaction-diffusion systems," *International Journal of Bifurcations and Chaos in Applied Sciences and Engineering*, vol. 8, pp. 1163–1182, 1998.

[23] K. Howard, "The generation of periodic patterns during early *Drosophila* embryogenesis," *The Company of Biologists Limited*, vol. 104, pp. 35–50, 1988.

[24] S. Kondo and T. Miura, "Reaction-diffusion model as a framework for understanding biological pattern formation," *Science*, vol. 329, pp. 1616–1620, 2012.

[25] H. Meinhardt, "Hierarchical inductions of cell states: a model for segmentation in *Drosophila*," *Journal of Cell Science*, vol. 4, pp. 357–381, 1986.

[26] J. M. W. Slack, *Essential Developmental Biology*. Wiley-Blackwell, 2001.

[27] P. W. Ingham, "The molecular genetics of embryonic pattern formation in *Drosophila*," *Nature*, vol. 335, p. 744, 1988.

[28] S. F. Gilbert, *Developmental Biology*. Sinauer Associates, 2000.

[29] A. M. Arias, N. E. Baker, and P. W. Ingham, "Role of segment polarity genes in the definition and maintenance of cell states in the drosophila embryo," *The Company of Biologists*, vol. 103, pp. 157–170, 1988.

[30] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *PNAS*, vol. 113, pp. 3932–3937, 2016.