Boolean Modeling of Genetic Regulatory Networks

Réka Albert

Department of Physics, Pennsylvania State University, University Park, PA 16802, USA

Abstract. Biological systems form complex networks of interaction on several scales, ranging from the molecular to the ecosystem level. On the subcellular scale, interaction between genes and gene products (mRNAs, proteins) forms the basis of essential processes like signal transduction, cell metabolism or embryonic development. Recent experimental advances helped uncover the qualitative structure of many gene control networks, creating a surge of interest in the quantitative description of gene regulation. We give a brief description of the main frameworks and methods used in modeling gene regulatory networks, then focus on a recent model of the segment polarity genes of the fruit fly Drosophila melanogaster.

The basis of this model is the known interactions between the products of the segment polarity genes, and the network topology these interactions form. The interactions between mRNAs and proteins are described as logical (Boolean) functions. The success in reproducing both wild type and mutant gene expression patterns suggests that the kinetic details of the interactions are not essential as long as the network of interactions is unperturbed. The model predicts the gene patterns for cases that were not yet studied experimentally, and implies a remarkable robustness toward changes in internal parameters, initial conditions and even some mutations.

The success of this approach also suggests a wide applicability of real-topology-based Boolean modeling for gene regulatory networks. In cases when the information about the system is incomplete, Boolean modeling can verify the sufficiency of interactions and can propose ways to complete the network. After a coherent picture is obtained, more realistic kinetic models can be used to gain additional insights into the functioning of the system.

1 Introduction

1.1 Complex Networks in Genetic Regulation

Recent remarkable progress in molecular biology has led to a complete map of the genomes of many organisms, and the identification and classification of the proteins is well under way. The next major challenge is to determine all the interactions between genes, proteins and other cellular components and to integrate this knowledge into a system-level understanding. It is now widely recognized that the networks of interaction and regulation between cellular entities are highly complex, and their understanding needs a concerted effort between experiment, modeling and theory.

R. Albert, Boolean Modeling of Genetic Regulatory Networks, Lect. Notes Phys. **650**, 459–481 (2004)
http://www.springerlink.com/

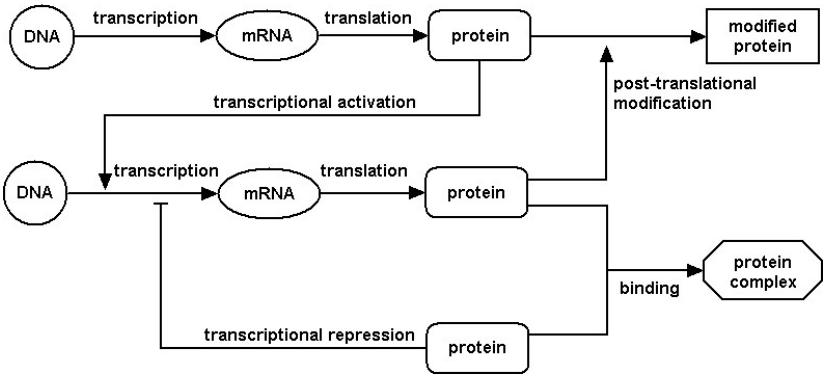

Fig. 1. Genes regulate each other's activity through regulatory networks. Gene transcription into mRNA is influenced by transcription factors, themselves products of other genes. In addition, post-translational modifications lead to proteins with modified properties.

Genes and gene products interact and form networks on several levels [1]. On the genomic level, a class of proteins called transcription factors can activate or inhibit the transcription of genes into mRNAs. Since these transcription factors are themselves products of genes, the ultimate effect is genes regulating each other's expression by forming so called gene regulatory networks. Proteins can participate in diverse chemical reactions that lead to modified proteins with different functions than the originals. Several proteins can also bind to form protein complexes with new roles. Proteins that are members of a class called enzymes catalyze the biochemical reactions forming the base of cellular processes (e.g. metabolism). In many cases the different-level interactions are integrated, for example in signal transduction networks the presence of an external signal (a chemical attractor or a hormone) triggers a cascade of interactions that can involve both biochemical reactions and regulation.

The focus of this article is gene regulatory networks whose components are genes, mRNAs and proteins, and the interactions include transcription, translation, transcriptional regulation and posttranslational reactions (see Fig. 1). We can realize from this description that gene regulatory networks cannot be completely described by a standard graph of nodes and edges. Specifically, the nodes have distinct identities as they correspond to diverse cellular components, and the edges can have two different signatures (signs) corresponding to activation and inhibition.

Gene regulatory networks play a crucial role during development, the process in which a unicellular egg gives rise to an adult [2]. Each cell in a developing embryo has the same DNA, but at no time in their life cycle are all of their genes expressed, i.e. transcribed into mRNA and synthesizing protein. The basis of cell differentiation is differential gene expression, and this is accomplished by interactions between genes, i.e. gene regulatory networks [3].

1.2 Modeling Gene Regulatory Networks

When trying to understand the role and functioning of a gene regulatory network, the first step is to assemble the components of the network and the interactions between them. This structural information needs to be complemented with information or hypotheses regarding the kinetics of the interactions. Since development is a dynamic process in which the expression of genes can constantly change, gene network models need to have a dynamical aspect, i.e. they need to define a state variable for each component, and study how this state changes by the interactions in the network. This state variable can correspond to the concentration of mRNAs and proteins, or it can be a binary value corresponding to the qualitative statement that a gene is expressed or not.

A complete gene regulatory network model incorporates experimental knowledge about the components and their interactions as well as the initial state of these components, and leads to the known final state or dynamical behavior of the network. Validated models then are able to investigate cases that cannot be explored experimentally, for example changes in the initial state, in the components or in the interactions, and they can lead to predictions and insights into the functioning of the system.

1.3 Pioneering Work in Modeling Gene Regulatory Networks

The experimental advances in the mapping of gene regulatory networks are fairly recent, but modeling general aspects of gene regulatory networks dates back to the end of 1960s thanks to the pioneering work of Stuart Kauffman and René Thomas.

In the absence of experimental results, Stuart Kauffman considered an idealized representation of a typical (random) gene network [4,5]. He assumed that genes are equivalent, and their interactions form a directed graph in which each gene receives inputs from a fixed number K of randomly selected neighbors. The state of genes is described by binary (ON/OFF) variables, and the dynamic behavior of each variable, that is, whether it will be ON or OFF at next moment, is governed by a Boolean function. In general, a Boolean or logical function is written as a statement acting on the inputs using the logical operators "and", "or" and "not" and its output is 1(0) if the statement is true (false). In a Random Boolean Network (RBN) the functions governing the state of each node are randomly selected from the 2^{2^K} possible K-input Boolean functions, and kept fixed afterward. Kauffman studied the dynamics of these RBNs, focusing on the attractors (usually cycles) in the state space of the whole network. He discovered the existence of a phase transition in an RBN of size N depending on the value of the parameter K. For $K > 2$ there are around N/e possible cycles whose length scales exponentially with N, however, for $K = 2$ both the number and length of the limit cycles is only $\sqrt{N}$. Kauffman proposed to identify the number of attractors of a gene regulatory network with the number of possible cell types, and noted that the number of cell types seems to increase approximately with the square root of the number of genes per cell, suggesting that gene regulatory

networks are in the ordered regime, or on the edge between order and chaos. The RBN models spawned a lot of research in the physics literature, see e.g. [6–9].

While Kauffman proposed a dynamic view on randomly connected gene regulatory networks, René Thomas developed a detailed logical description of the mechanisms governing transcriptional regulation, including the effects of DNA domains such as promoters, initiators, terminators, and the concepts of genetic dominance and recessivity [10]. This formalism was later refined to include multilevel variables and used to study feedback loops, i.e. circular chains of interaction. These loops can be classified into two categories based on the number of negative (inhibitory) interactions in the loop: if this number is even, the loop is positive, and if the number of negative interactions is odd, the loop is a negative feedback loop. Thomas found that a positive feedback loop is a necessary condition for the existence of multiple steady states, while a negative feedback loop with two or more elements is a necessary condition for stable limit cycles [11]. Biologically this means that cell differentiation is based on positive feedback loops, and homeostasis (stability to small perturbations) is based on negative feedback loops. The logical framework introduced by René Thomas was successfully applied to various gene regulatory networks playing a role in the flower morphogenesis of the wall cress *Arabidopsis thaliana* [12] and the development of the fruit fly *Drosophila melanogaster* [13,14].

1.4 Current Models

Broadly speaking, the modeling approaches to gene regulatory networks can be divided into two main groups. In the 'discrete-state' approach each network node (mRNA or protein) is assumed to have a small number of discrete states and the regulatory interactions between nodes are described by logical functions similar to those used in programming. Typically time is also quantized, and the network model that describes how gene products interact to determine the state at the next time gives rise to a discrete dynamical system [12–18].

A more detailed level of description is used in the 'continuous-state' approach, in which the levels of mRNAs and proteins are assumed to be continuous functions of time, and their evolution is modeled by differential equations with mass-action kinetics or other rate laws for the production and decay of all components [19–21]. In order to compare with usually ON/OFF type experimental gene expression profiles, the continuous concentrations are transformed into binary variables using thresholds.

In this paper I shall focus on a recent model of the segment polarity genes developed in collaboration with Hans Othmer [18]. In this model we concentrate on the products (mRNAs, proteins) of segment polarity genes. We reconstruct the network of interaction between these components from experimental data, and assume that these interactions can be expressed as Boolean functions. We find that the dynamic behavior of this model always leads to steady states, and these steady states are in very good agreement with the experimental data on the gene expression pattern of wild type and mutant *Drosophila* embryos. In

addition, the model leads to insights into the functioning of this network, the most important being that the network topology is a main source of robustness.

2 The Segment Polarity Gene Network

The genes involved in embryonic pattern formation in the fruit fly *Drosophila melanogaster*, as well as the majority of the interactions between them, are known (for recent reviews see [22–24]). As in other arthropods, the body of the fruit fly is composed of segments, and determination of the adult cell types in these segments is controlled by about 40 genes organized in a hierarchical cascade of gene families [25]. These gene families are expressed in consecutive stages of embryonic development and have a spatial expression pattern that is successively more precisely-defined (see Fig. 2). The genes at one step initiate or

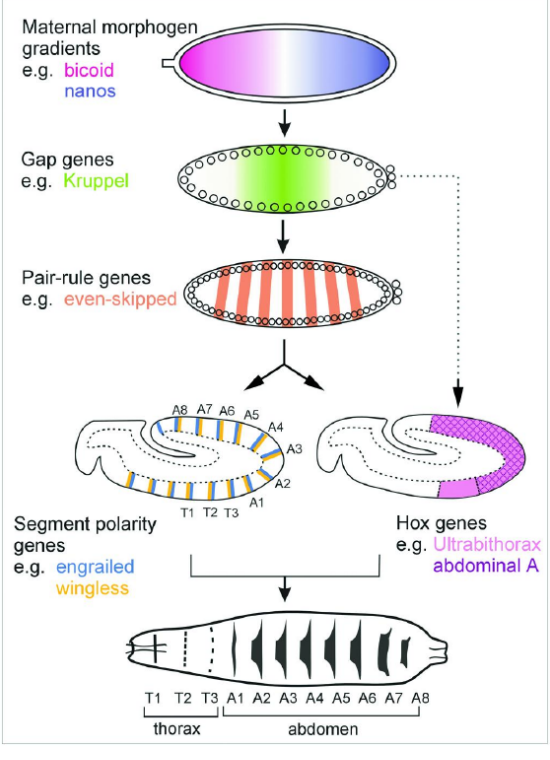

Fig. 2. The segmentation of the fruit fly embryo is governed by a hierarchy of gene families, starting from maternal genes. Each of these genes encode for transcription factors, and are responsible for the initiation of the genes in the next family. While the genes in the first three steps are transient, the segment polarity genes maintain a stable pattern for three hours. Reproduced with permission from [23].

modulate the expression of those involved in the next step of the cascade. While most of these genes act only transiently, the segment polarity genes are expressed throughout the life of the fly, and their periodic spatial pattern is maintained for at least 3 hours of embryonic development.

The best characterized segment polarity genes include *engrailed* (*en*), *wingless* (*wg*), *hedgehog* (*hh*), *patched* (*ptc*), *cubitus interruptus* (*ci*), *smoothened* (*smo*) and *sloppy paired* (*slp*)[1]. The segment polarity genes encode for diverse proteins including the transcription factors Engrailed (EN), Sloppy Paired (SLP), and Cubitus Interruptus (CI), the secreted proteins Wingless (WG) and Hedgehog (HH), and the transmembrane proteins Patched (PTC) and Smoothened (SMO)[2].

2.1 Wild Type Patterns of the Segment Polarity Genes

The segment polarity genes are activated by the pair-rule genes at about 3 hours after fertilization. The initial state of the segment polarity genes includes two-cell-wide SLP stripes followed by two-cell-wide stripes not expressing SLP [26], single-cell-wide *wg*, *en* and *hh* stripes followed by three cells not expressing them, and three-cell-wide stripes for *ci* and *ptc* [2]. This pattern is maintained almost unmodified for three hours, during which time the initially homogeneous-looking embryo is divided into 14 parasegments (the embryonic counterparts of the adult segments) by regularly - distributed furrows. The position of these furrows coincides with the space between the *wg* and *en* -expressing cells, thus the periodicity of the gene expression drives the future external appearance of the embryo [25]. The cells in a parasegment are counted from anterior (toward the head) to posterior (toward the tail). According to this notation, *wg* is expressed in the most posterior cell of each parasegment, and *en* in the most anterior cell.

The segment polarity genes refine and maintain their expression through the network of intra- and intercellular regulatory interactions shown in Fig. 3. The stable expression pattern of these genes (specifically the expression of *wingless* and *engrailed*) defines and maintains the borders between different parasegments and contributes to subsequent developmental processes, including the formation of denticle patterns and of appendage primordia [2,25]. Homologs of the segment polarity genes have been identified in vertebrates, including humans, which suggests strong evolutionary conservation of these genes.

The pair-rule gene product SLP activates *wg* transcription and represses *en* transcription. The WG protein is secreted from the cells that synthesize it [25, 27] and initiates a signaling cascade leading to the transcription of *en* [28]. EN promotes the transcription of the *hh* gene [29] and represses the transcription of *ci* [30] and possibly *ptc* [31,32]. The HH protein is also secreted, and binds to the HH receptor PTC on a neighboring cell [22]. The intracellular domain

[1] Many of these genes were named for the phenotypic changes their mutations cause, e.g. a *wingless* mutant fruit fly does not have wings.

[2] These notations follow the convention that names of genes and mRNAs are italicized, while names of proteins are capitalized.

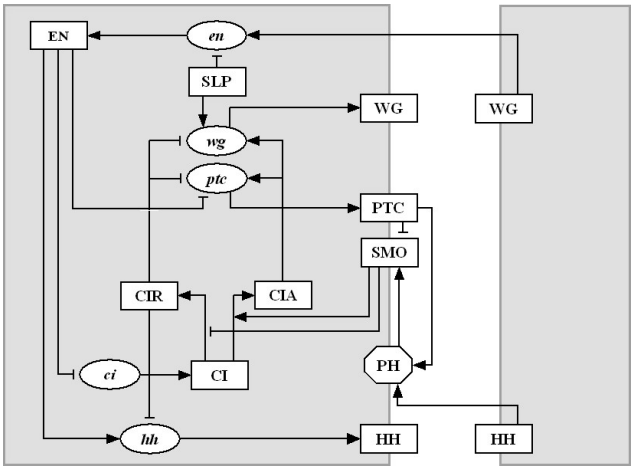

Fig. 3. The network of interactions between the segment polarity genes. The shape of the nodes indicates whether the corresponding substances are mRNAs (ellipses), proteins (rectangles) or protein complexes (octagons). The edges of the network signify either biochemical reactions (e.g. translation) or regulatory interactions (e.g. transcriptional activation). The edges are distinguished by their signatures, i.e. whether they are activating ($\rightarrow$) or inhibiting ($\dashv$). Terminating arrows ($\rightarrow$) indicate translation, post-translational modifications (in the case of CI), transcriptional activation or the promotion of a post-translational modification reaction (e.g., SMO determining the activation of CI). Terminating segments ($\dashv$) indicate transcriptional inhibition or in the case of SMO, the inhibition of the post-translational modification reaction CI$\rightarrow$CIR.

of PTC forms a complex with SMO [33] in which SMO is inactivated by a post-translational conformation change [34]. Binding of HH to PTC removes the inhibition of SMO, and activates a pathway that results in the modification of CI [34]. The CI protein can be converted into one of two transcription factors, depending on the activity of SMO. When SMO is inactive, CI is cleaved to form CIR, a transcriptional repressor that represses *wg*, *ptc* [35] and *hh* transcription [36,37]. When SMO is active, CI is converted to a transcriptional activator, CIA, that promotes the transcription of *wg* and *ptc* [35,37–39].

3 Description of the Model

In the model, each mRNA or protein is represented by a node of a network, and the interactions between them are encoded as directed edges (see Fig. 3). The state of each node is 1 or 0, according as the corresponding substance is present or not. The states of the nodes can change in time. We choose a time interval that is larger or equal to the duration of all transcription and translation processes, and we use this interval as the length of a unit timestep. The next state of node i is determined by a Boolean function of its state and the states of those nodes that that have edges incident on it.

3.1 Updating Rules

The functions determining the state of each node are constructed from the interactions between nodes displayed in Fig. 3 according to the following rules (see also Fig. 4)

(i) mRNAs/proteins are synthesized in one timestep if their transcriptional activators/mRNAs are present;

(ii) the effect of transcriptional activators and inhibitors is never additive, but rather, inhibitors are dominant;

(iii) mRNAs decay in one timestep if not transcribed;

(iv) transcription factors and proteins undergoing post-translational modification decay in one timestep if their mRNA is not present.

For example, EN is translated from en, and therefore the state of EN at time $t+1$, $EN^{t+1} = 1$ if $en^t = 1$. Since EN is a transcription factor, it is assumed that its expression will decay sufficiently rapidly that if $en^t = 0$, then $EN^{t+1} = 0$. These two assumptions mean that

$$EN^{t+1} = en^t. \tag{1}$$

Table 1 gives an overview of the Boolean functions for each node. In each case, subscripts signify spatial position (i.e. cell number) and superscripts signify time[3].

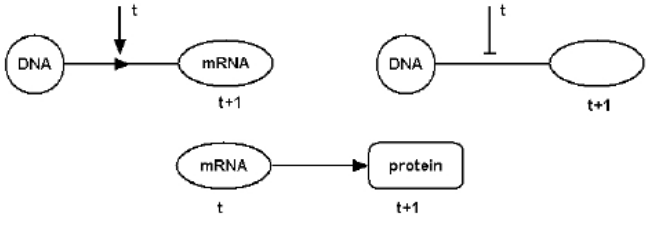

Fig. 4. Assumptions for the kinetics of the interactions. We assume that the timescale for turning ON or OFF is the same. Transcription requires the presence of activators and the absence of inhibitors; translation requires the presence of the mRNA.

3.2 Representing the State of the System

Expression of the segment polarity gene occurs in stripes that encircle the embryo, and therefore we treated the two-dimensional pattern as one-dimensional. We considered a line of 12 cells corresponding to three parasegment primordia (i.e. the spatial regions that will become the parasegments), and imposed

[3] In coding these rules we have used an equivalent description of a function giving the output of every possible combination of inputs.

Table 1. The Boolean functions used in the model. The functions are based on the known interactions between mRNAs and proteins shown in Fig. 3, and on the temporal assumptions listed above. In general the updating rule gives the expression of a node at time $t+1$ as a function of the expression of its effector nodes at time t. However, there are three exceptions: we assume that the expression of SLP does not change, and that the activation of SMO and the binding of PTC to HH are instantaneous.

Node	Boolean updating function
SLP_i	$SLP_i^{t+1} = SLP_i^t = \begin{cases} 0 \text{ if } i\%4 = 1 \text{ or } i\%4 = 2 \\ 1 \text{ if } i\%4 = 3 \text{ or } i\%4 = 0 \end{cases}$
wg_i	$wg_i^{t+1} = (CIA_i^t \text{ and } SLP_i^t \text{ and not } CIR_i^t)$ or $[wg_i^t \text{ and } (CIA_i^t \text{ or } SLP_i^t) \text{ and not } CIR_i^t]$
WG_i	$WG_i^{t+1} = wg_i^t$
en_i	$en_i^{t+1} = (WG_{i-1}^t \text{ or } WG_{i+1}^t) \text{ and not } SLP_i^t$
EN_i	$EN_i^{t+1} = en_i^t$
hh_i	$hh_i^{t+1} = EN_i^t \text{ and not } CIR_i^t$
HH_i	$HH_i^{t+1} = hh_i^t$
ptc_i	$ptc_i^{t+1} = CIA_i^t \text{ and not } EN_i^t \text{ and not } CIR_i^t$
PTC_i	$PTC_i^{t+1} = ptc_i^t \text{ or } (PTC_i^t \text{ and not } HH_{i-1}^t \text{ and not } HH_{i+1}^t)$
PH_i	$PH_i^t = PTC_i^t \text{ and } (HH_{i-1}^t \text{ or } HH_{i+1}^t)$
SMO_i	$SMO_i^t = \text{not } PTC_i^t \text{ or } HH_{i-1}^t \text{ or } HH_{i+1}^t$
ci_i	$ci_i^{t+1} = \text{not } EN_i^t$
CI_i	$CI_i^{t+1} = ci_i^t$
CIA_i	$CIA_i^{t+1} = CI_i^t \text{ and } (SMO_i^t \text{ or } hh_{i-1}^t \text{ or } hh_{i+1}^t)$
CIR_i	$CIR_i^{t+1} = CI_i^t \text{ and not } SMO_i^t \text{ and not } hh_{i\pm1}^t$

periodic boundary conditions on the ends. We used four cells per parasegment primordium because when expression of the segment polarity genes begins, a given gene is expressed in every fourth cell. The state of the system includes a 12-cell wide 1-dimensional periodic pattern for each node in the network which we represented as a series of black/gray boxes corresponding to cells in which the given node is ON/OFF (see Fig. 5). To make the periodicity of the pattern clear, we separated the patterns corresponding to distinct parasegments by short white spaces.

4 Modeling the Wild Type Segment Polarity Genes

The first step in validating the model is testing whether it captures the wild type behavior of the system. Therefore we started from the known initial pattern of the segment polarity genes and updated their states according to the rules presented on Table 1, checking whether they become stationary.

The initial state of each parasegment primordium includes SLP present in the last (posterior) two cells, *wg* present in the last cell, *en* and *hh* expressed in the first (anterior) cell, and *ci* and *ptc* expressed in the posterior three cells [25,26,29,31,32]. Since the proteins are translated after the mRNAs are transcribed, we assumed that the proteins are not expressed in the initial state. The

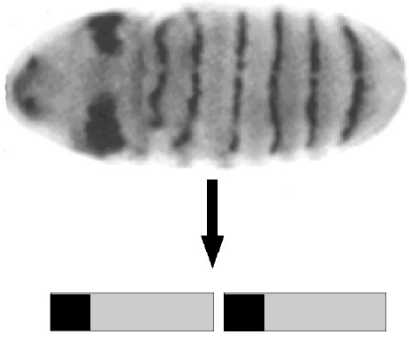

Fig. 5. Because the expression pattern of the segment polarity genes is symmetrical, we represent it with a one-dimensional pattern corresponding to the anterior-posterior axis. We identify the state of each node in the network with a periodic succession of black/ gray squares corresponding to cells that express/do not express the given node. Each segment of four cells corresponds to a parasegment primordium.

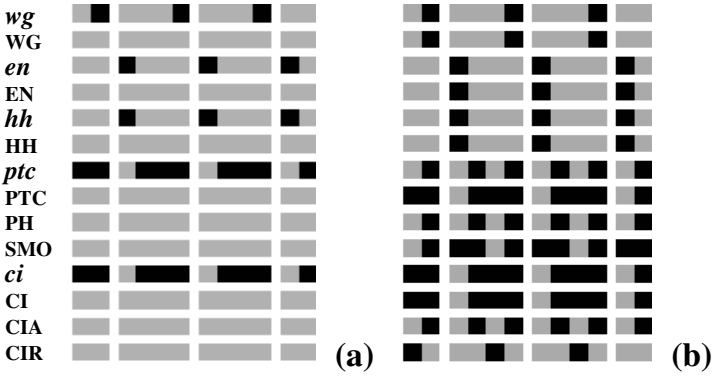

Fig. 6. Wild-type expression patterns of the segment polarity genes. Here and hereafter left corresponds to anterior and right to posterior in each parasegment. Horizontal rows correspond to the pattern of individual nodes - specified at the left side of the row - over two full and two partial parasegments. Each parasegment is assumed to be four cells wide. A black (gray) box denotes a node that is ON (OFF). (a) The experimentally-observed initial state. *en, wg* and *hh* are expressed in every fourth cell, while the broad *ptc* and *ci* stripes are complementary to *en*. (b) The steady state given by the model when initialized with the pattern in (a). This pattern is in excellent agreement with the observed gene expression patterns. After [18].

one-dimensional representation of the mRNA and protein patterns is shown in Fig. 6a.

We iterated the dynamical system defined by the rules in Table 1 starting from the initial state described above. We found that after only 6 time steps, the expression pattern stabilizes in a time-invariant spatial pattern (see Fig. 6b) that coincides with the experimentally observed stable expression of the segment

polarity genes. Indeed, *wg* and WG are expressed in the most posterior cell of each parasegment [40], while *en*, EN, *hh* and HH are expressed in the most anterior cell of each parasegment [29,40], *ptc* is expressed in two cells, one on each side of the *en*-expressing cells [25,31]. SMO is present in a broad region ranging from the *wg*-expressing to the *en*-expressing cells [41]. *ci* is expressed almost ubiquitously, with the exception of the cells expressing *en* [30,38]. CIA is expressed in the neighbors of the HH-expressing cells, while CIR is expressed far from the HH-expressing cells [35].

Thus the model [18] demonstrates that the interaction between the segment polarity genes is able to maintain their expression after initialization. The success in reproducing the stable expression pattern of these genes is a strong indication that the kinetic details of the interactions do not matter, just their signature and the regulatory network they form. This conclusion is in agreement with the results of the continuous-state model of von Dassow *et al.* [20].

5 The Functional Topology of the Segment Polarity Network

The success of our model demonstrates that the topology of the regulatory network has a determining role in its dynamics. Nevertheless, knowledge of the topology alone is not enough to determine what will happen in the network. Moreover, the presence of dual interaction signatures precludes us from using standard graph theoretical tools to analyze this network.

To obtain a better insight into the connection between topology and dynamics, we proposed the construction of an expanded graph that reflects the function of the network. The first step of this expansion is adding complementary pseudo-nodes corresponding to every node whose negated state enters the Boolean rules on Table 1. The second step is to introduce composite pseudo-nodes for nodes whose states are terms of a conjunction in these rules (see Table 2). Consider the transcription of the *hh* gene. Figure 3 shows that *hh* has two incoming edges, one from EN and one from CIR, and Table 1 shows that transcription of the *hh* gene requires both the presence of the EN protein and the absence of the CIR protein. We introduce complementary pseudo-node, $\overline{\text{CIR}}$, that is expressed whenever CIR is not, and connect it to CIR with a symmetrical edge. Then we add the composite pseudo-node $\text{E}\overline{\text{CR}}$, and we draw two directed edges starting from EN and $\overline{\text{CIR}}$ and ending in $\text{E}\overline{\text{CR}}$, to represent the dependence of $\text{E}\overline{\text{CR}}$ on the expression of EN and $\overline{\text{CIR}}$ (see Fig. 7). Now *hh* receives inputs only from $\text{E}\overline{\text{CR}}$.

Figure 8 shows the nodes and edges corresponding to the mRNAs and proteins in the second cell of the parasegment together with the pseudo-nodes these mRNAs and proteins interact with, both cell-autonomously, and in the neighboring cells. Although the introduction of the pseudo-nodes increases the number of nodes in the network, it eliminates the distinction between edges based on their signatures; all directed edges in Fig. 8 now signify activation. However, there are differences in the way multiple activating edges are taken into account: multiple

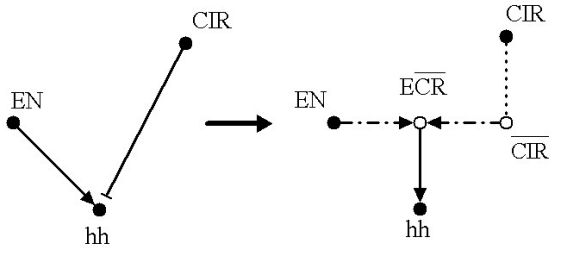

Fig. 7. Illustration of the network expansion process used to construct the functional topology. To express the logical rule governing the transcription of *hh* graphically, we introduce the complementary node $\overline{\text{CIR}}$ and the composite node $\text{E}\overline{\text{CR}}$. The expanded network contains real nodes (filled circles) and pseudo-nodes (open circles), an interdependence relation between CIR and $\overline{\text{CIR}}$ (dotted line), edges corresponding to the activation of $\text{E}\overline{\text{CR}}$ (dash-dotted lines) and a single edge expressing the activation of *hh* transcription. After [18].

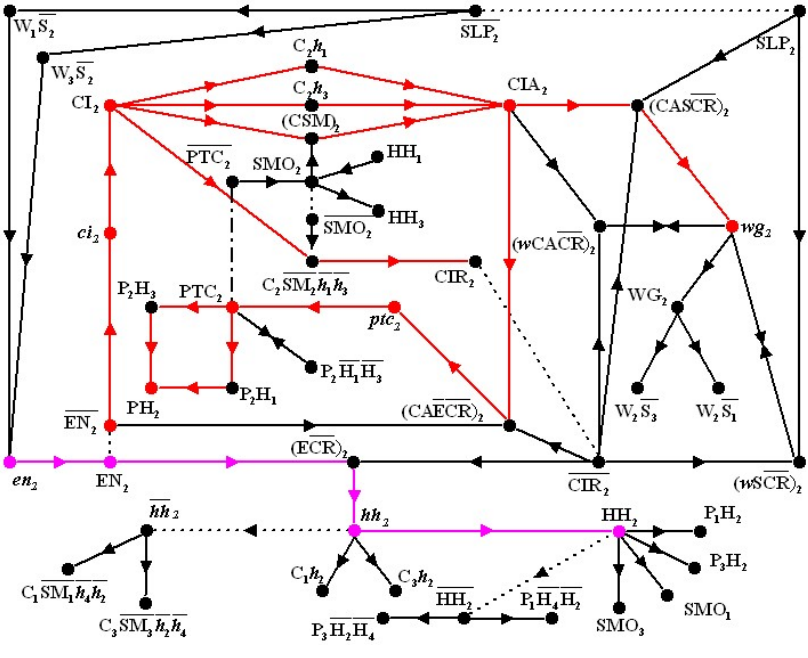

Fig. 8. Functional topology of the network affecting the second cell of the parasegment. Pseudo-nodes with multiple indexes correspond to intercellular interactions and either receive some of their inputs from the neighboring cells, or contribute to the expression of the nodes in the neighboring cells (not shown). Symmetrical edges between nodes and their complementaries are drawn with dotted lines, double arrows denote a pair of oppositely directed edges. The colored edges illustrate two antagonistic activating clusters starting from *en* (purple) and $\overline{\text{EN}}$ (red).

Table 2. Definition of the symbols for pseudo-nodes used in Fig. 8. The state of each composite node is determined from the logical function giving its relation to the state of its "parent" nodes.

Symbol of pseudo-node	Relation to parent node(s)
Complementary nodes	
$\overline{\mathrm{EN}}$	not EN
$\overline{hh}$	not *hh*
$\overline{\mathrm{HH}}$	not HH
$\overline{\mathrm{PTC}}$	not PTC
$\overline{\mathrm{SLP}}$	not SLP
$\overline{\mathrm{SMO}}$	not SMO
Composite nodes corresponding to a single cell	
$(\mathrm{CA}\overline{\mathrm{ECR}})_2$	CIA_2 and $\overline{\mathrm{EN}_2}$ and $\overline{\mathrm{CIR}_2}$
$(\mathrm{CAS}\overline{\mathrm{CR}})_2$	CIA_2 and SLP_2 and $\overline{\mathrm{CIR}_2}$
$(\mathrm{CSM})_2$	CI_2 and SMO_2
$(\mathrm{E}\overline{\mathrm{CR}})_2$	EN_2 and $\overline{\mathrm{CIR}_2}$
$(w\mathrm{CA}\overline{\mathrm{CR}})_2$	wg_2 and CIA_2 and $\overline{\mathrm{CIR}_2}$
$(w\mathrm{S}\overline{\mathrm{CR}})_2$	wg_2 and SLP_2 and $\overline{\mathrm{CIR}_2}$
Composite nodes corresponding to intercellular interactions	
$\mathrm{C}_i h_j$	CI_i and hh_j
$\mathrm{C}_i\overline{\mathrm{SM}_i h_j h_k}$	CI_i and $\overline{\mathrm{SMO}_i}$ and $\overline{hh_j}$ and $\overline{hh_k}$
$\mathrm{P}_i\overline{\mathrm{H}_j\mathrm{H}_\mathrm{k}}$	PTC_i and $\overline{\mathrm{HH}_j}$ and $\overline{\mathrm{HH}_k}$
$\mathrm{P}_i\mathrm{H}_j$	PTC_i and HH_j
$\mathrm{W}_i\mathrm{S}_j$	WG_i and SLP_j

edges ending in composite pseudo-nodes are added by the operator "and", while multiple edges ending in real nodes are cumulated by the operator "or".

Figure 8 illustrates the heterogeneous functional topology of the segment polarity network. The majority of nodes have few edges, but there are key nodes with a large number of incoming or outgoing edges. For example, $\overline{\mathrm{CIR}}$ has 5 outgoing edges, while HH has 4. The important role of HH in the network is reflected in the fact that it affects the future expression of 4 other proteins (CIA, CIR, PTC and PH) in the neighboring cells, for a total of 8 nodes. Other nodes such as CIA have several incoming edges, indicating that they can be activated in many ways. A single node, SLP, has only outgoing edges because it is constitutively present; all others have both incoming and outgoing edges (the apparent exceptions interact with nodes in the neighboring cells).

The functional network of Fig. 8 gives insight into the time-evolution of the expression of the segment polarity genes. For example, we can determine the cluster of nodes that can be activated by the expression of a given node (see colored nodes in Fig. 8). The absence of EN (or conversely the presence of $\overline{\mathrm{EN}}$) gives the largest activated cluster, containing *ci*, CI, CIA, *ptc*, PTC, PH, *wg*, and WG. A separate activated cluster starts with the presence of *en*, and contains EN, *hh* and HH. These activating clusters indicate that the cells expressing *en* and *hh* never express *wg*, *ptc* or *ci*. This separation into anterior and posterior

compartments expressing different genes is well-known, in fact, it is the basis for calling these genes "segment polarity genes" [2].

While the majority of the activating effects propagate outside the cell, there are three cases in which an activation can return to its source. In other words, three short positive feedback loops exist in the network of Fig. 8. The first two loops connect wg_2 with $(w\mathrm{CA}\overline{\mathrm{CR}})_2$ or $(w\mathrm{S}\overline{\mathrm{CR}})_2$ and the third connects PTC_2 with $\mathrm{P}_2\overline{\mathrm{H}_1}\overline{\mathrm{H}_3}$. These loops ensure the maintenance of *wg* and PTC if all the conditions for the expression of the pseudo-node in the cycle are met. The successful activation of the *wg* cycle can induce the stable expression of *en* and *hh* in those neighboring cells where neither SLP nor CIR is expressed, and stable expression of PTC leads to stable CIR expression two cells removed from *en* expression.

6 Gene Mutations

An important method for inferring gene interactions experimentally is to silence selected genes by mutations. These null mutant genes are not able to synthesize protein, and if that protein is a transcription factor, the effects of the mutation propagate through the system (see Fig. 9). Our model is able to simulate the effect of null mutations by setting the state of the transcript to OFF and not updating it during the evolution of the system.

Our results indicate that if any of *en*, *wg* or *hh* are blocked, the steady state is a pattern with no *en*, *wg*, *ptc* or *hh*, as in Fig. 10a. We can see from Fig. 8 that each of these mutations disrupts intercellular signaling, causing ubiquitous expression of CIR, which in turn leads to ubiquitous repression of transcription. This result is in excellent agreement with all experimental observations regarding *en*, *wg* and *hh* mutant embryos [29,31,42–44].

If the *ptc* gene is blocked, we obtain a pattern with broad *wg*, *en* and *hh* stripes (see Fig. 10b). Indeed, the network in Fig. 8 shows that if *ptc* is deactivated, $\overline{PTC}$ will be ubiquitous, causing all CI to be transformed into CIA, which leads to two-cells-wide *wg* and *en*/*hh* expression. This pattern agrees with the experimental results on *ptc* mutants [29,43–45]. Moreover, our results are in agreement with all experimental observations of double mutants as well [29,40, 42,46].

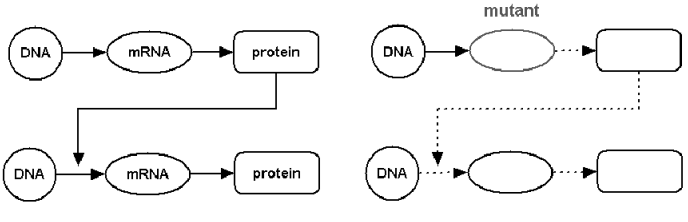

Fig. 9. Gene mutations that disrupt transcription factors propagate through the system, affecting multiple nodes.

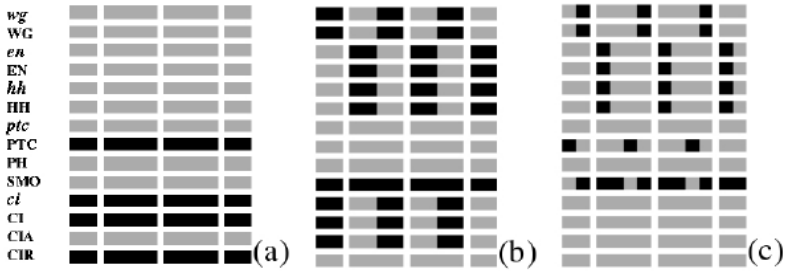

Fig. 10. Segment polarity gene expression patterns predicted for gene mutations. (a) Pattern with no segmentation. This pattern arises if any of *wg*, *en* or *hh* is kept OFF in the model and is independent of the initial state of the other genes. (b) Broad expression pattern. The stripes of *en*, *wg* and *hh* are two-cells-wide, while the *ci* stripe narrows and CIR is not expressed. This state arises when *ptc* is kept OFF, regardless of the initial state of other genes. (c) Almost normal pattern obtained for *ci* mutants and wild type initial conditions.

If all the other genes are initiated normally, we find that the effect of a *ci* deletion does not affect the *en*, *wg* and *hh* patterns (see Fig. 10c). Indeed, Fig. 8 shows that the deactivation of *ci* leads to the disappearance of CIA and CIR, but wild-type *wg* can still be maintained by SLP[39,44]. In conclusion, the model is in agreement with every observed gene pattern in mutants (see Fig. 11) and provides predictions for genes whose expression was not studied experimentally.

7 Determination of the Steady States and Their Domains of Attraction

The fact that the model reproduces the results of numerous experiments remarkably well suggests that the structure of the model is essentially correct, and warrants exploration of problems that have not been studied experimentally. For example, we can determine the complete set of stable steady state patterns of segment polarity gene expression, and estimate the domain of attraction of these states. The former can be done analytically by noting that these are fixed points of the discrete dynamical system, and so $x_i^{t+1} = x_i^t$, where x corresponds to any node in the network. Thus a steady state is the solution of the system of equations obtained from Table 1 by simply removing the time indices (see [18]). We obtain 10 solutions that correspond to four distinct patterns (see Fig. 12) and their slight variations.

The first steady state is the pattern with no segmentation first presented in Fig. 10a. The second corresponds to the wild-type pattern first shown in Fig. 6b. The third steady state has two-cells-wide *en* and *wg* stripes like the *ptc* mutant (see Fig. 12c). In the fourth distinct steady state *wg* is expressed in the anterior neighbor of its wild-type position, while the *en*/*hh* stripe is displaced posteriorly (see Fig. 12d). This expression pattern corresponds to an embryo

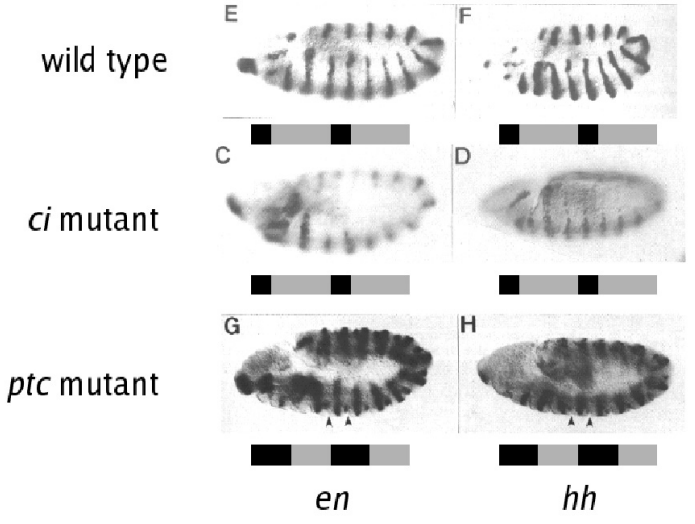

Fig. 11. Comparison between the experimental results (embryo pictures showing the expression pattern of *en* and *hh*, after [29]) and predictions of the model (black and gray patterns) for two gene mutations. The model indicates that *ci* mutation can preserve the pattern of *en* and *hh*, while *ptc* mutation doubles their expression, in agreement with experiments.

with no parasegmental grooves, since the end of the *wg* stripe does not meet the beginning of the *en* stripe.

While each of the steady states can be obtained starting from suitable nearby states, the number of initial conditions leading to a chosen stable pattern, i.e., its domain of attraction, can be very different. Consider first the number of initial states that lead to the wild-type steady state. If we fix all nodes but one in their wild-type pattern, there are $2^4 = 16$ distinct initial patterns corresponding to the four cells of the parasegment. We do this for each of the 14 variable nodes in turn (we do not change the expression of SLP) and find that the number of initial patterns leading to the wild-type steady state is 3 for *wg* or WG variation, 4 for *en*, EN, *hh* or HH variation, 8 for *ptc*, PTC, CI or CIA variation, and 16 for PH, SMO, *ci* or CIR variation. When the initial pattern of all 14 nodes can vary, there are $3^2 \cdot 4^4 \cdot 8^4 \cdot 16^4 \sim 6 \times 10^{11}$ prepatterns that lead to the wild type steady state, which is a fraction of 8×10^{-6} of the total number of initial states $N_{st} = 16^{14}$.

We find that the network is very robust with respect to missing initial expression of nodes. We have determined that the minimal prepatterning that leads to wild-type stable expression is as follows.

- *wg* **is wild type,**
- *en* **and** *hh* **are not expressed,**
- *ptc* **is expressed in the third cell of the parasegment primordium,**
- *ci* **and the proteins are not expressed.**

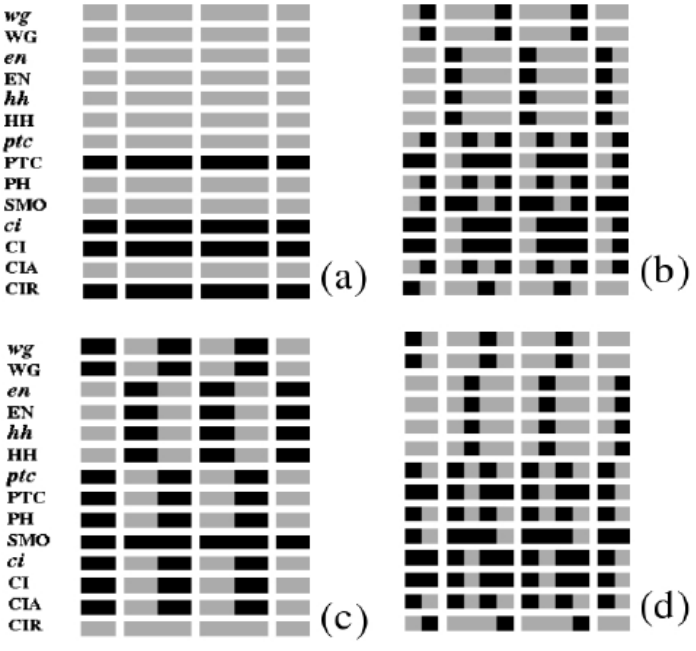

Fig. 12. Various stable patterns of the segment polarity genes. (a) Steady state with no segmentation. (b) Wild-type type expression pattern. (c) Steady state with broad *en*, *wg* and *hh* domains. (d) Ectopic pattern with displaced *wg*, *en* and *hh* stripes.

In summary, it is enough to initiate the expression of two genes in two cells per parasegment primordium, and the interactions between the segment polarity genes will initiate the others. This result suggests a remarkable error-correcting ability for the segment polarity gene network.

Note that the minimal prepattern contains the wild-type expression of *wg*. If *wg* is not expressed initially we find that the final pattern is like Fig. 12a, regardless of the initial pattern of the other nodes. Consequently, a fraction of at least $1/16^{th}$ of the initial states leads to the pattern of Fig. 12a. This finding suggests that *wg* has a special role in the functioning of the segment polarity network, and has to be activated at a specific time and specific cells in order to obtain wild-type gene expression.

In the other limit, broader than wild-type initial expression of any node except PH, SMO, *ci* and CIR leads to the pattern with broad stripes as in Fig. 12c. This pattern is obtained in the vast majority of prepatterns, comprising about 90% of the total number of initial state and its features were frequently observed in overexpression experiments [43,44,47].

The minimal prepattern needed for the ectopic pattern with displaced *wg* and *en* stripes (Fig. 10c) is *wg* expression in the third cell of the parasegment primordium (the same as its steady pattern), and *ptc* expression in the last cell of the parasegment primordium, where the wild-type stripe of *wg* would normally

be. Note that this minimal initial condition is simply a shifted version of the minimal condition for the wild-type steady state. In practice the simultaneous ectopic initiation of several nodes is very improbable, and indeed, this steady state has never been observed.

In addition to steady state analysis we have performed a systematic analysis of the dynamics of the network when the initial expression of genes differs from the wild-type initial condition. In principle the attractor for some initial conditions could be periodic in time, but we have found that the only stable attractors are steady states. Since the purpose of the segment polarity network is to stabilize and maintain the parasegment borders, this result is biologically realistic.

8 Possible Changes in the Assumptions

Our goal in constructing this model was to base it on the topology of the regulatory interactions and have as few additional parameters as possible. However, there remain a few assumptions that might not be necessary or, on the contrary, could reveal essential constraints on the network.

8.1 Equal Timescale for Synthesis and Decay

We assumed that the expression of mRNAs/proteins decays in one time step if their transcriptional activators/mRNAs are switched off. This conjecture is probably too severe, as the decay time of proteins is usually longer than the time their synthesis takes. Therefore we studied a variant of the model in which the expression of a protein is maintained for at least two steps [18]. We find that this variant leads to exactly the same steady states as the original model, and these states have approximately the same basins of attraction[4]. The two-step model reaches the wild-type steady state shown in Fig. 6b if it is started from the initial pattern of Fig. 6a, and leads to the same states for gene mutations. The only change is in the intermediate states visited en route to the final state: both the wild-type and the broad type pattern stabilizes on average 30% faster using the two-step assumption. On the other hand, the pattern with no segmentation is reached at a slightly lower rate than in the original model. In conclusion, the two-step assumption provides a more realistic modeling of the decay of the proteins without changing the conclusions of the model.

8.2 Assumptions for WG and PTC

The model contains two exceptions to the one-step decay rule through the assumptions of persistence of existent *wg* and PTC expression. The stability of these nodes has a major role in stabilizing the expression of the segment polarity genes, reflected in the existence of the cycles in the functional topology

[4] Note that in this case limit cycles are possible.

of the network (Fig. 8) and in the fact that the steady states are completely determined by the pattern of *wg* and PTC. It is therefore important to check what happens if *wg* and PTC decay in one step as other mRNAs and proteins.

If we assume that

$$wg_i^{t+1} = CIA_i^t \text{ and } SLP_i^t \text{ and not } CIR_i^t, \tag{2}$$

it is still possible to arrive at the wild-type steady state, but only for much more restricted initial states. Furthermore, the resilience of the network to mutations in *ci* is destroyed, and all initial states lead to the steady state with no segmentation. Since it is observed experimentally that *ci* null mutants still display almost normal segmentation [44], we can conclude that the stability of *wg* is required for the functioning of the segment polarity genes. This suggests a special role for SLP as the main activator of *wg*, and underlies the need for its stability.

If we do not assume the maintenance of initial PTC expression, the pattern of PTC will follow that of its transcript and split into two stripes. This will cause the complete disappearance of CIR and the only steady state will be the pattern with broad stripes as in Fig. 10b. Thus the persistence of PTC is a major requirement for the function of the segment polarity network, and suggests that the protein has special structural properties.

8.3 Four-Cell-Wide Parasegments

During the three hours of stable segment polarity gene patterning the parasegment is enlarged due to two rounds of divisions [2]. While the *wg* stripe remains a single cell wide, the *en* stripe widens to three cells. The maintaining of this *en* requires WG transport, and, indeed, wingless protein is seen to diffuse over a distance of 2-3 cell diameters [48]. In order to determine if our model is able to describe the segment polarity gene patterns in later stages, we applied it to the transition between a four- and eight-cell-wide parasegment. We started with the wild-type pattern of Fig. 6a and assumed that each cell divides into two identical cells, with the same genes expressed in each of the two. We also assumed that WG and HH can be transported through the nearest neighbors of the cells expressing their mRNAs. The model leads to the steady state represented in Fig. 13, with a single cell wide *wg* stripe, three cell wide *en* and *hh* stripes, and two *ptc* stripes flanking the *en* domain. This steady state agrees perfectly with the wild-type pattern observed in 8hr old embryos [25,30,31,41,38].

8.4 Stable SLP

Throughout our analysis we assumed that the expression of SLP does not change. To test whether this assumption is necessary, we have studied the effects of inactivated and overexpressed SLP. We obtain seven final states for inactivated SLP, but none of them corresponds to the wild-type pattern. The closest state, obtained when we start from wild-type initial conditions, has a wild-type *wg* and *ptc* pattern, but *en* and *hh* are expressed on both sides of the *wg* stripe and CIR

Fig. 13. Stable expression pattern of the segment polarity genes after a round of cell division, as obtained from our model. We assume that at this stage WG and HH can be transported through the neighboring cells. This pattern is in good agreement with experimental observations of 8hr old embryos. After [18].

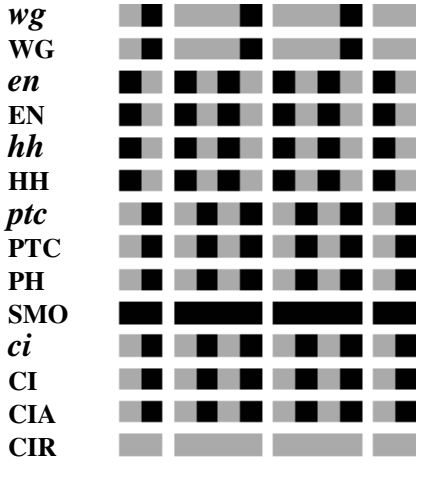

Fig. 14. The pattern obtained from our model when we start from wild-type initial conditions, but SLP is not functional. Note that *en* is expressed on both sides of the *wg* stripe. After [18].

is not absent (Fig. 14). At this point this state is a theoretical prediction that can be verified by conditional SLP mutants (i.e., mutants that have normal pair-rule activity, but no segment polarity activity). We also find that ubiquitously expressed SLP leads to the state with no segmentation presented in Fig. 10b. This finding is in agreement with experimental results [26]. Based on these results,

and the important role of SLP in maintaining *wg* transcription, we conclude that the SLP protein plays a vital role in this network.

9 Conclusions

The model demonstrates that the topology of the regulatory network plays a determining role in the dynamics and the stability of the segment polarity genes. The success in reproducing wild type and mutant gene patterns indicates that the kinetic details of the interactions do not matter, as long as their net effect is preserved. Our simulations also suggest a remarkable robustness and error correcting ability of the segment polarity gene network. We found that a large fraction of initiation delays can be rescued, and the network can compensate even for some gene mutations . The model also gives numerous predictions that can be tested experimentally. First, we concluded that the *wingless* gene plays a key role in the system, and it is imperative that it be initiated at the right time in the right pattern. However, non-initiation of *engrailed* and *hedgehog* can be rescued by the activity of the network. Experiments with conditional mutants defective in initiation could verify these predictions. Second, we found that the state of the segment polarity genes can evolve into a pattern with displaced stripes if initiated in a certain way. While this ectopic initiation is difficult, it should be possible. Finally, we concluded that the stable expression of SLP is a crucial requirement; this could be tested by the isolation of the segment polarity- and pair-rule roles of SLP.

The two-step model represents a step toward modeling the transition from the initial state to the final steady state of the segment polarity network. A more realistic model would assume different time intervals (expressed in number of steps) for the decay for mRNAs and proteins. While this extension would involve unknown parameters, the condition of reaching the same steady states as the original model would provide constraints on the variability of the decay rates. Another direction where the model could be extended is to consider a two-dimensional array of cells. It is known experimentally that the stripes of segment polarity genes are not initiated as straight lines, but have jagged borders [2]. During the functioning of the segment polarity network these stripes straighten, and the parasegment borders become sharp. A two-dimensional simulation of our model could lead to important insights into this process.

The model presented here is part of a larger family of models using a logical approach to gene regulatory networks [13–17,50]. As illustrated by these models, this approach enables the integration of qualitative observations on gene interactions into a coherent picture, while adding a minimum of additional kinetic hypotheses. The analysis of a Boolean model is more tractable than that for a model based on differential equations, which inevitably has numerous unknown parameters, and a Boolean model facilitates a more systematic study of the possible steady states and their basins of attraction[5]. We envision realistic

[5] Note that Boolean logic can be extended to so-called polynomial logic applicable to multi-level variables, see [49].

topology-based Boolean modeling as an important first step in understanding the interplay between the topology and dynamics of gene regulatory networks and testing the completeness of available topological information. While the segment polarity gene network was successfully modeled by a simple synchronous binary Boolean model, other networks require more detailed models incorporating asynchronous updating and/or multi-level variables [13,14]. Of course there are undoubtedly systems, such as metabolic networks, for which a Boolean approach might not be an appropriate first level of analysis.

Acknowledgements

The original work presented here was done in collaboration with Hans G. Othmer and supported in part by NIH Grant #GM 29123.

References

1. B. Alberts *et al.*, *Molecular Biology of the Cell*, 4th edn. (2003).
2. L. Wolpert, R. Beddington, J. Brockes, T. Jessell, P. Lawrence and E. Meyerowitz, *Principles of Development*, (Current Biology Ltd., London 1998).
3. E.H. Davidson *et al.*, Science **295**, 1669 (2002).
4. S. A. Kauffman, J. Theor. Biol. **22**, 437 (1969).
5. S. A. Kauffman, *The origins of Order*, (Oxford University Press, New York, 1993).
6. B. Derrida and Y. Pomeau, Europhys. Lett. **1**, 45 (1986).
7. G. Weisbuch and D. Stauffer, J. Phys. (Paris) **48** 11 (1987).
8. B. Luque and R. V. Solé, Phys. Rev. E **55**, 257 (1997).
9. R. Albert and A.-L. Barabási, Phys. Rev. Lett. **84**, 5660 (2000).
10. R. Thomas, J. Theor. Biol. **42**, 563 (1973).
11. R. Thomas and R. D'Ari, *Biological Feedback* (CRC Press, Boca Raton, Ann Arbor, Boston, 1990).
12. L. Mendoza, D. Thieffry and E. R. Alvarez-Buylla, Bioinformatics **15**, 593 (1999).
13. L. Sánchez and D. Thieffry, J. Theor. Biol. **211**, 115 (2001).
14. A. Ghysen and R. Thomas, BioEssays **25**, 802 (2003).
15. J. W. Bodnar, J. Theor. Biol. **188**, 391 (1997).
16. J. W. Bodnar and M. K. Bradley, Cell Biochem. and Biophys. **34**, 153 (2001).
17. C.-H. Yuh, H. Bolouri and E. H. Davidson, Development **128**, 617 (2001).
18. R. Albert and H. G. Othmer, J. Theor. Biol **223**, 1 (2003).
19. J. Reinitz and D. H. Sharp, Mechanisms of Development **49**, 133 (1995).
20. G. von Dassow, E. Meir., E. M. Munro and G. M. Odell, Nature **406**, 188 (2000).
21. V. V. Gursky, J. Reinitz and A. M. Samsonov, Chaos **11**, 132 (2001).
22. P. W. Ingham and A. P. McMahon, Genes Dev. **15**, 3059 (2001).
23. B. Sanson, EMBO Reports **2**, 1083 (2001).
24. V. Hatini and S. DiNardo, Trends in Genetics **17**, 574 (2001).
25. J. E. Hooper and M. P. Scott, The Molecular Genetic Basis of Positional Information in Insect Segments. In: *Early Embryonic Development of Animals*, ed. by W. Hennig (Springer, Berlin 1992) pp. 1-49.
26. K. M. Cadigan, U. Grossniklaus and W. J. Gehring, Genes Dev. **8**, 899 (1994).
27. S. Pfeiffer and J.-P. Vincent, Cell & Dev. Biol. **10**, 303 (1999).

28. K. M. Cadigan and R. Nusse, Genes Dev. **11**, 3286 (1997).
29. T. Tabata, S. Eaton and T. B. Kornberg, Genes Dev. **6**, 2635 (1992).
30. S. Eaton and T. B. Kornberg, Genes. Dev. **4**, 1068 (1990).
31. A. Hidalgo and P. Ingham, Development **110**, 291-301 (1990).
32. A. M. Taylor, Y. Nakano, J. Mohler and P. W. Ingham, Mechanisms of Development **42**, 89 (1993).
33. M. van den Heuvel and P. W. Ingham, Nature **382**, 547 (1996).
34. P. W. Ingham, EMBO J. **17**, 3505 (1998).
35. P. Aza-Blanc and T. B. Kornberg, Trends in Genetics **15**, 458 (1999).
36. J. T. Ohlmeyer and D. Kalderon, Nature **396**, 749 (1998).
37. N. Méthot and K. Basler, Cell **96**, 819 (1999).
38. C. Alexandre, A. Jacinto and P. W. Ingham, Genes Dev. **10**, 2003 (1996).
39. T. von Ohlen and J. E. Hooper, Mechanisms of Development **68**, 149 (1997).
40. P. W. Ingham, A. M. Taylor and Y. Nakano, Nature **353**, 184 (1991).
41. J. Alcedo, Y. Zou and M. Noll, Molecular Cell **6**, 457 (2000).
42. S. DiNardo, E. Sher, J. Heemskerk-Jongens, J. A. Kassis and P. H. O'Farrell, Nature **332**, 45 (1988).
43. C. Schwartz, J. Locke, C. Nishida and T. B. Kornberg, Development **121**, 1625 (1995).
44. A. Gallet, C. Angelats, S. Kerridge and P. P. Thérond, Development **127**, 5509 (2000).
45. A. Martinez-Arias, N. Baker and P. W. Ingham, Development **103**, 157 (1988).
46. A. Bejsovec and E. Wieschaus, Development **119** 501 (1993).
47. J. Heemskerk, S. DiNardo, R. Kostriken and P. H. O'Farrell, Nature **352**, 404 (1991).
48. A. Bejsovec and A. Martinez-Arias, Development **113**, 471 (1991).
49. R. Laubenbacher and B. Stigler, Polynomial models for biochemical networks (preprint, 2003).
50. I. Shmulevich, E. Dougherty and W. Zhang, Proc. IEEE **90**, 1778 (2002).