Connections matter: Boolean modeling of gene regulatory networks

Réka Albert
University of Minnesota

Acknowledgements: Hans G. Othmer, NIH Grant GM 29123

Complex systems and biology

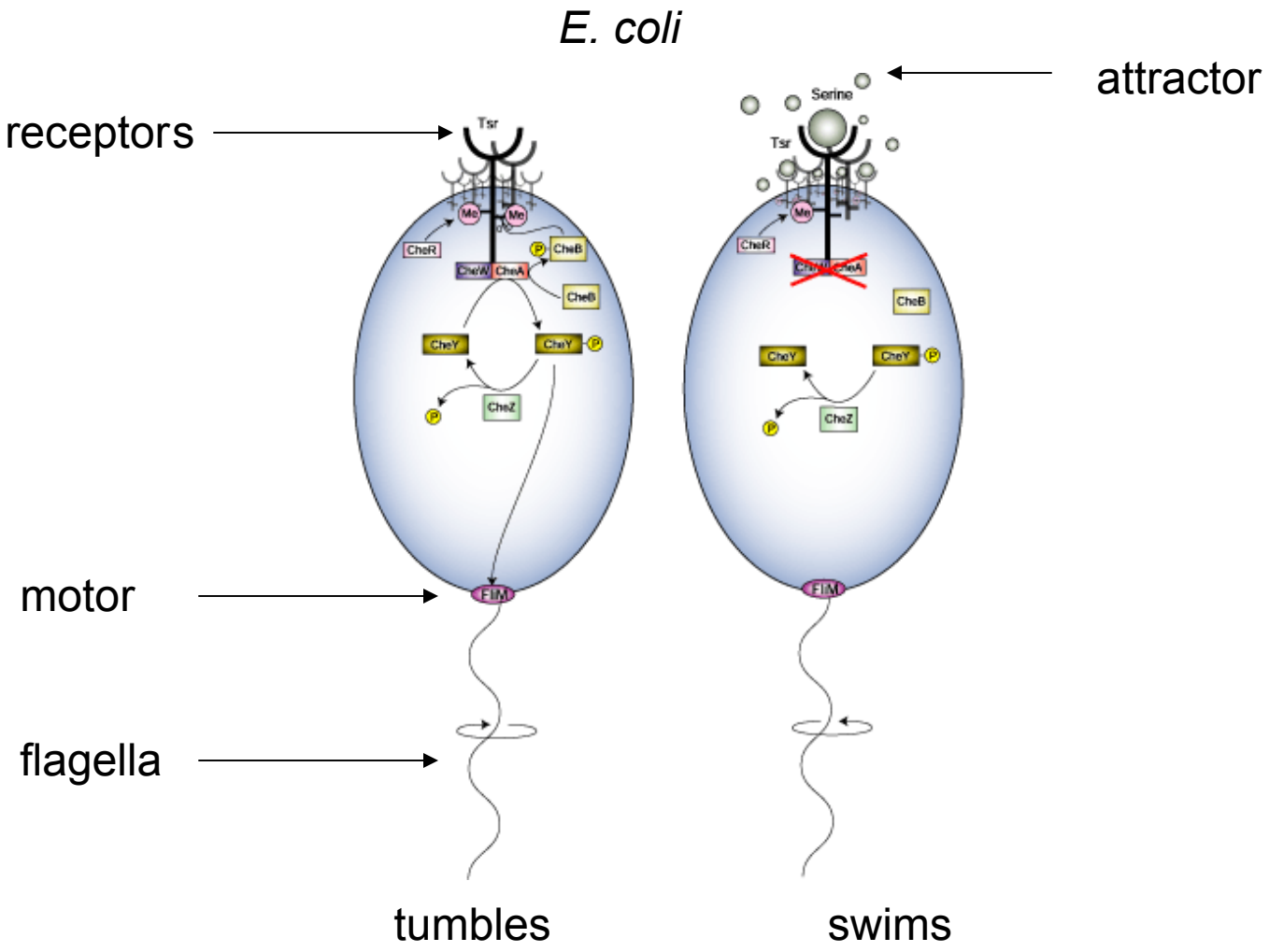
Complex Systems - cannot be described in reductionist terms

- numerous interacting components
- nonlinear interactions
- adaptive, self-organized behavior

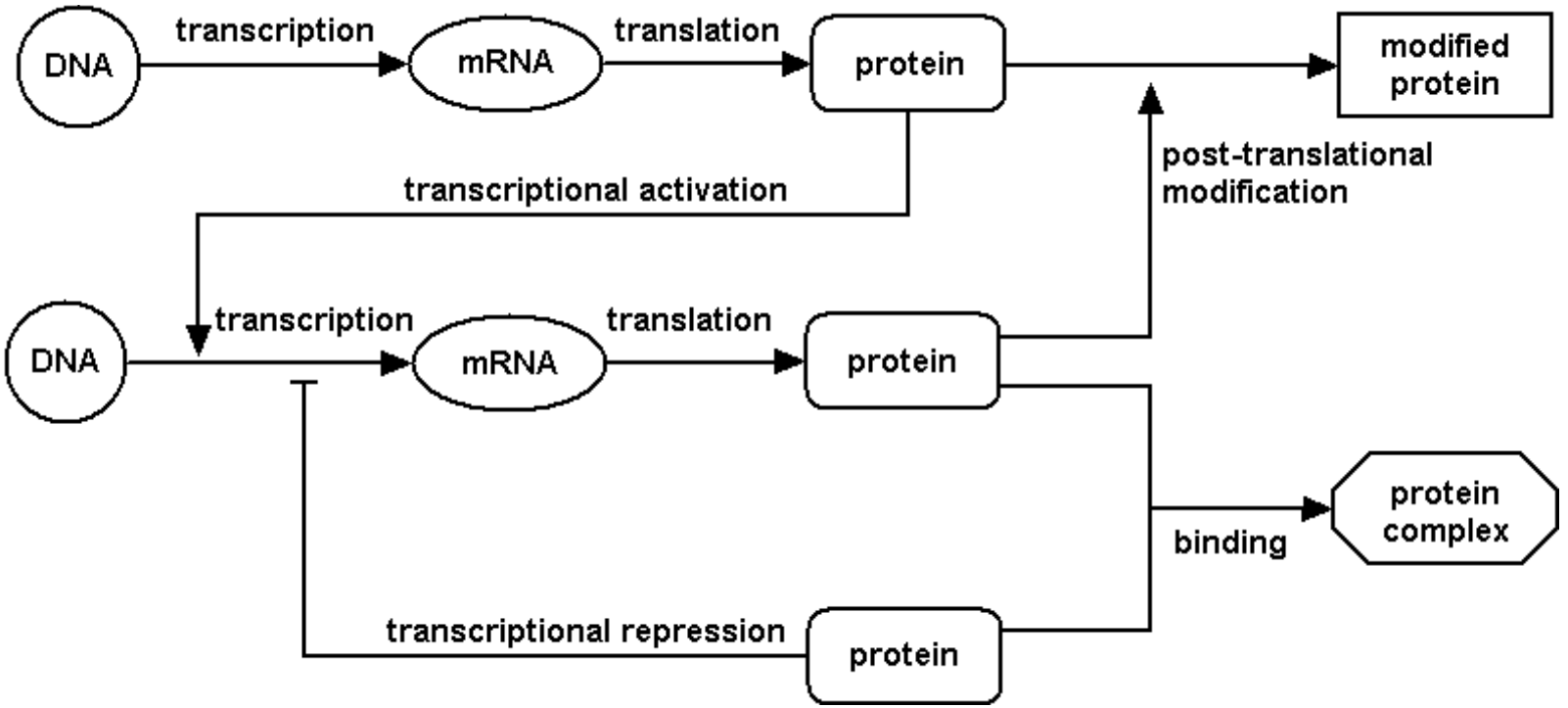
Biological systems

- functionally diverse elements,
- nonlinear interactions that form a network,
- have a function that needs to be performed.

Example: signal transduction networks



Example: gene regulatory networks



Understanding biological systems

Systems-level understanding of biology is a two-step process:

1. Discovering the components and their functions : genome sequences, protein structure.

2. Mapping the network of interactions between components.

Major lines of research:

- Knowledge discovery - extract patterns from huge quantities of experimental data
- Model-based analysis – formulate models based on experimental knowledge, provide predictions to be tested *in vivo.*

Modeling biological systems

Input: components.

Hypotheses: network of interactions; kinetics.

Validation: capture known behavior.

Output: predictions and insights into the function of the system.

My study: The segment polarity genes of the fruit fly.

Input: segment polarity genes

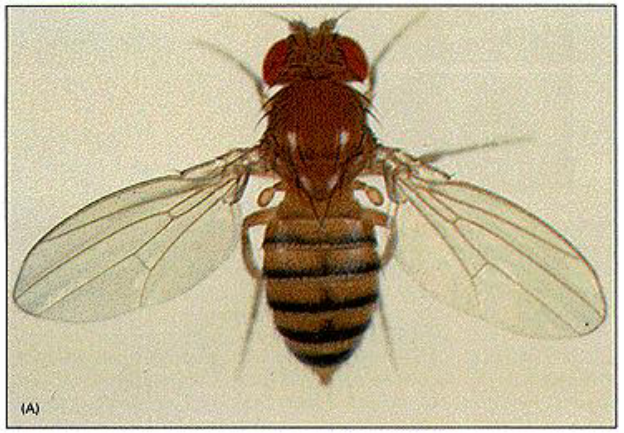
Hypotheses: boolean interactions

Validation: reproduces all known gene expression patterns.

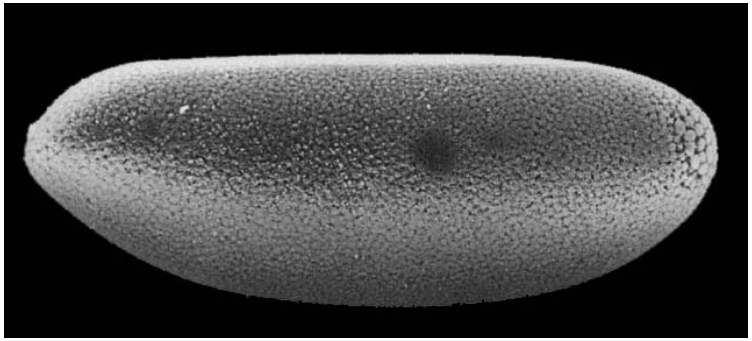
Insight: topology is a main source of robustness.

R. Albert, H. G. Othmer, Journal of Theoretical Biology, to appear.

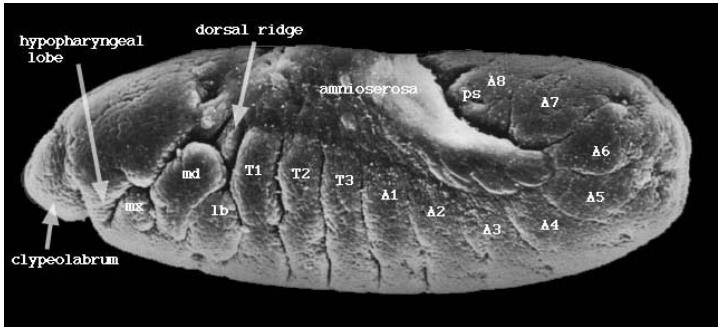
Segmentation of the fruit fly



Drosophila melanogaster

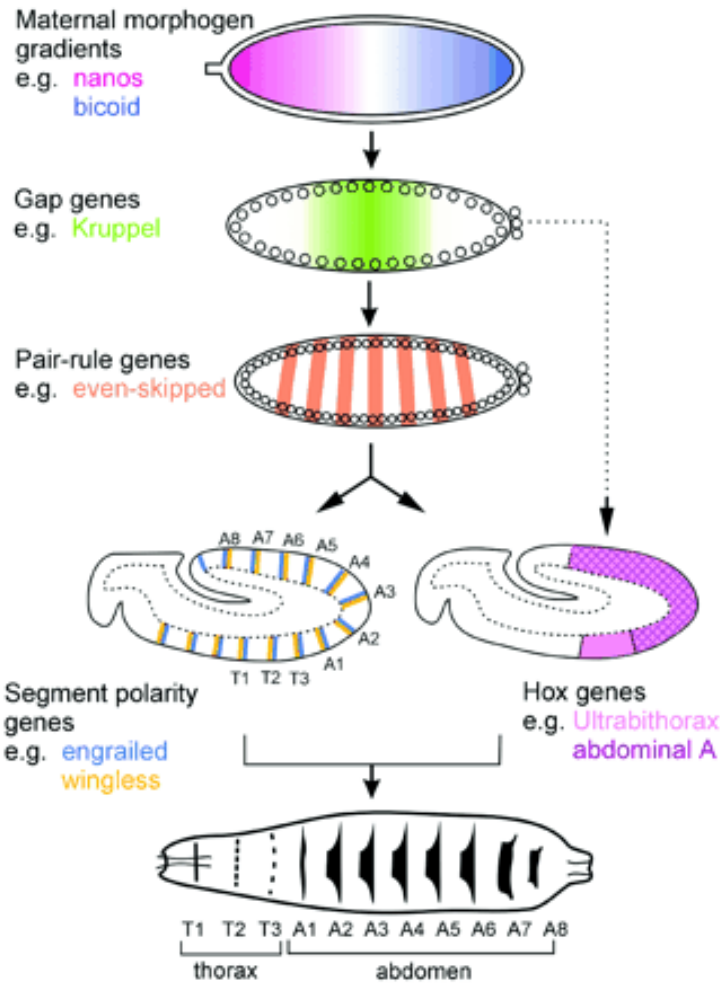


Syncytial blastoderm, 1h



End of gastrulation, 7h

Segmentation is governed by a cascade of genes



Transient gene products, initiate the next step then disappear.

The role of the segment polarity genes

- The segment polarity genes are initiated by the pair-rule genes
- Several segment polarity genes are expressed (active) in stripes that are repeated in every fourth cell.
- These genes interact via a complex regulatory network.
- The expression pattern of the segment polarity genes is maintained for 3 hours.
- The parasegment borders appear between the cells expressing the two most important segment polarity genes, *engrailed* and *wingless*.

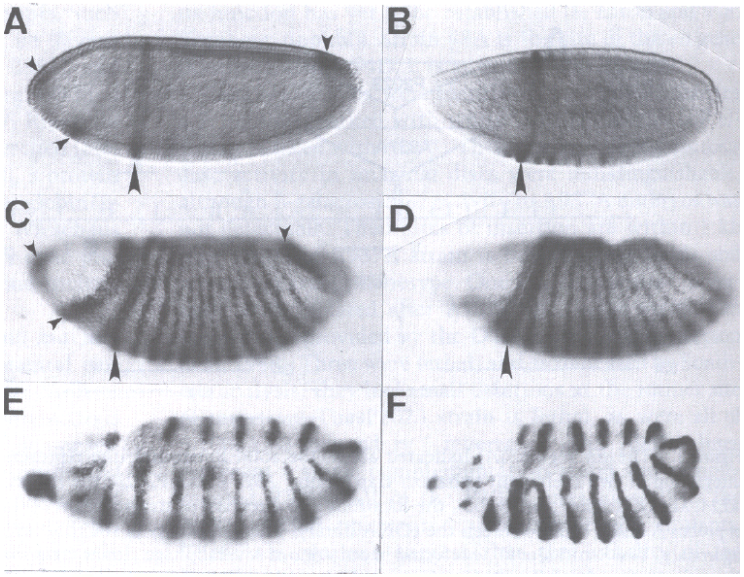
Segment polarity genes

Genes		Proteins
•*wingless (wg)*	→	Wingless protein (WG) - secreted
•*hedgehog (hh)*	→	Hedgehog protein (HH) - secreted
•*engrailed (en)*	→	Engrailed protein (EN) - transcription factor
•*patched (ptc)*	→	Patched protein (PTC) - receptor
•*smoothened (smo)*	→	Smoothened protein (SMO) - receptor
•*sloppy paired (slp)*	→	Sloppy paired protein (SLP) - transcription factor
•*cubitus interruptus (ci)*	→	Cubitus interruptus protein (CI) Cubitus activator (CIA) - transcription factor Cubitus repressor (CIR) - transcription factor

Gene products form a network that maintains a gene expression pattern initiated in an earlier stage.

Evolution of gene expression patterns

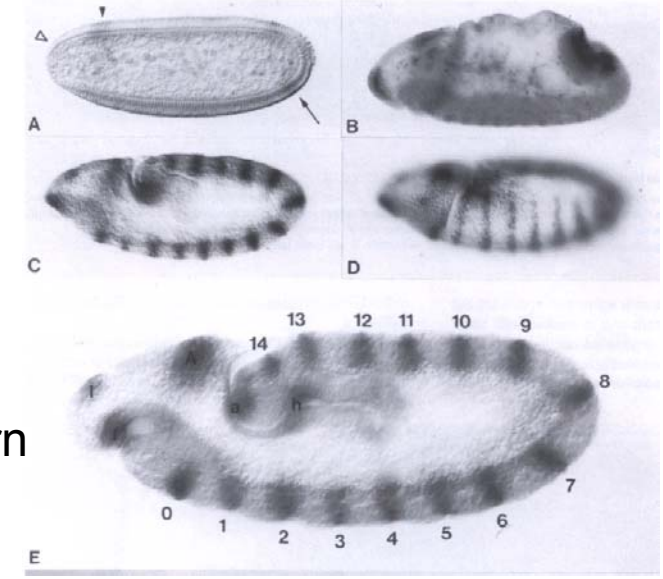
en *hh* *wg*

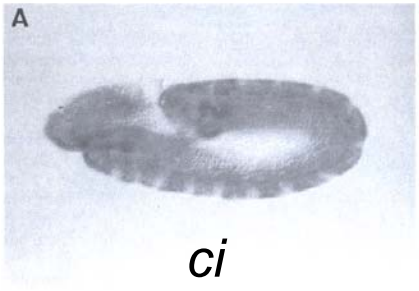


early stages
2:50 h

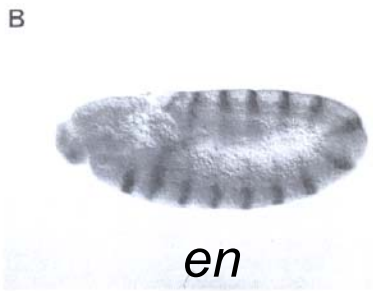
pre-pattern
3:00-3:30 h

stable pattern
4:20-7:20 h





ci

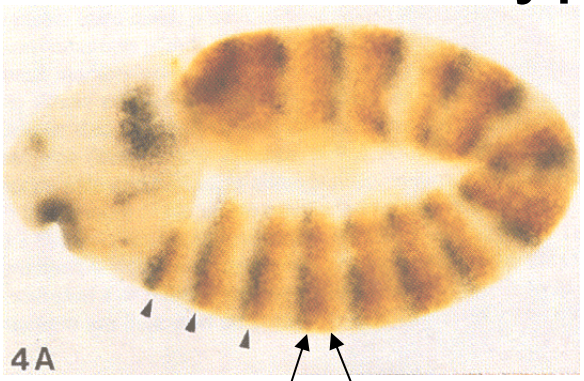


en

ptc

3:30 h

Wild type, stable gene patterns

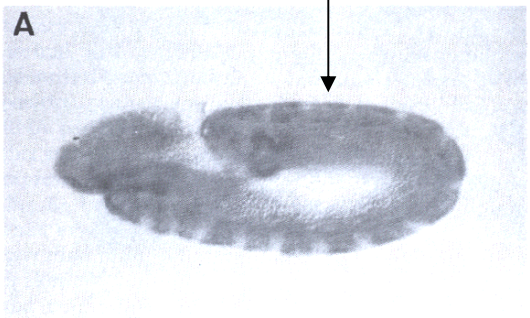


wg en

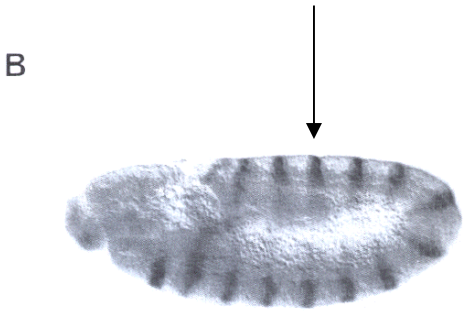
- *en* is expressed in the anterior part of the parasegment.
- *wg* is expressed in the posterior part of the parasegment.
- parasegmental grooves form between the *wg* and *en* stripes.

- *two ptc* stripes in each parasegment.
- *ci* pattern is complementary to that of *en.*

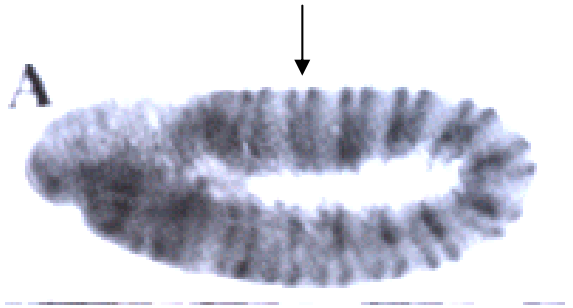
ci



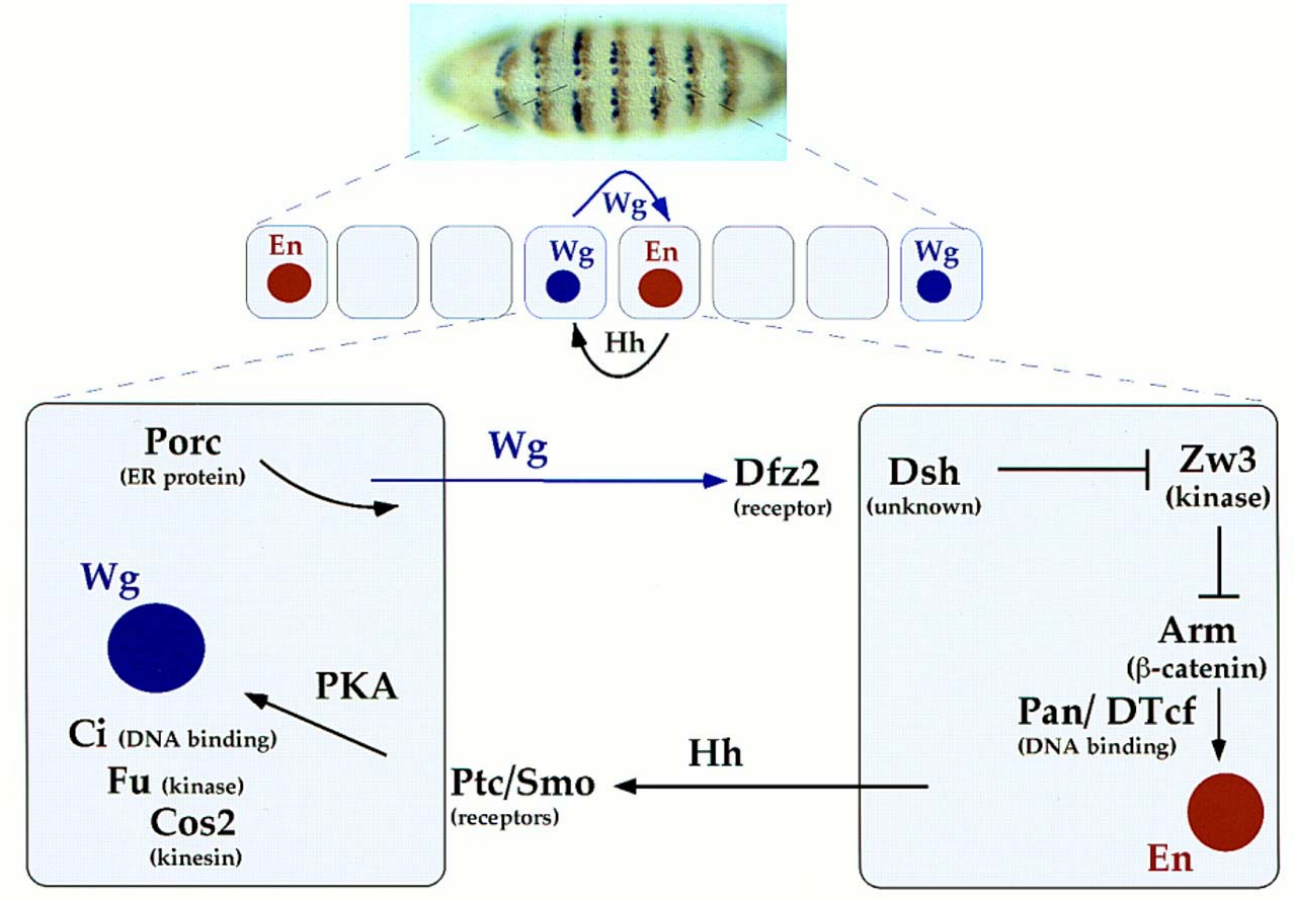
en



ptc

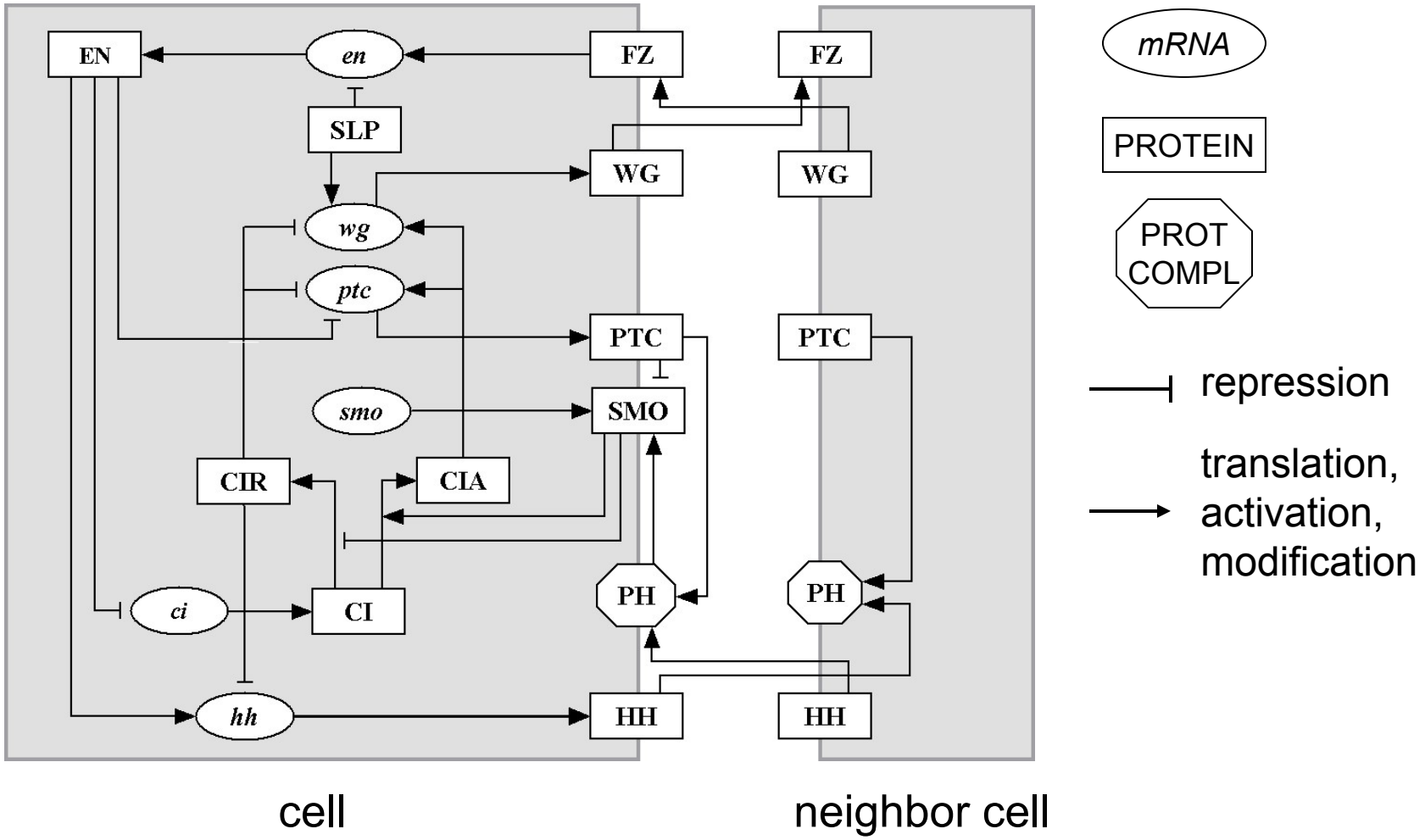


The segment polarity genes interact intercellularly



Cadigan, Nusse, Genes & Development 11, 3286 (1997)

Gene interaction network



Previous work assumes that gene regulation is similar to biochemical reactions

von Dassow et al., Nature 406, 188 (2000)

Transcription and translation processes follow dose-response curves.

48 unknown parameters.

$$\frac{d[prot]}{dt} = T_{\max} \frac{[mRNA]^{\nu}}{K^{\nu} + [mRNA]^{\nu}} - \frac{[prot]}{\tau}$$

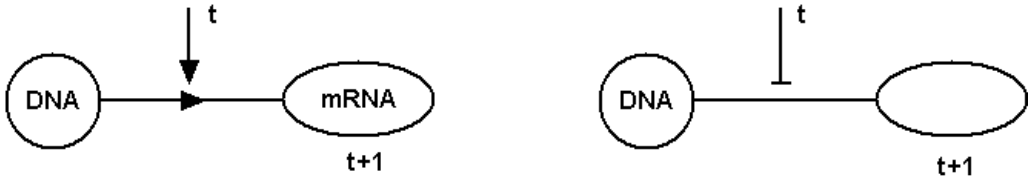
Systematic search shows that 1 in 46 parameter combinations lead to wild type final patterns.

The parameter combinations leading to wild type steady states are distributed homogenously in the biologically relevant parameter space.

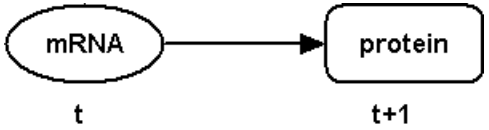
No need for kinetic parameters?

We propose a model without any kinetic parameters and a single timescale

- Transcripts and proteins are either ON (1) or OFF(0).
- The expression of a node at timestep *t* is given by a logical rule of the expression of its effectors at time *t-1*.
- Transcription depends on transcription factors; repressors are dominant.



- Translation depends on the presence of the transcript.



- Transcripts and proteins decay in one step if not produced.

Updating rules

$$hh_i^{t+1} = EN_i^t \text{ and not } CIR_i^t$$

$$en_i^{t+1} = (WG_{i-1}^t \text{ or } WG_{i+1}^t) \text{ and not } SLP_i^t$$

$$ptc_i^{t+1} = CIA_i^t \text{ and not } EN_i^t \text{ and not } CIR_i^t$$

transcription

$$ci_i^{t+1} = \text{not } EN_i^t$$

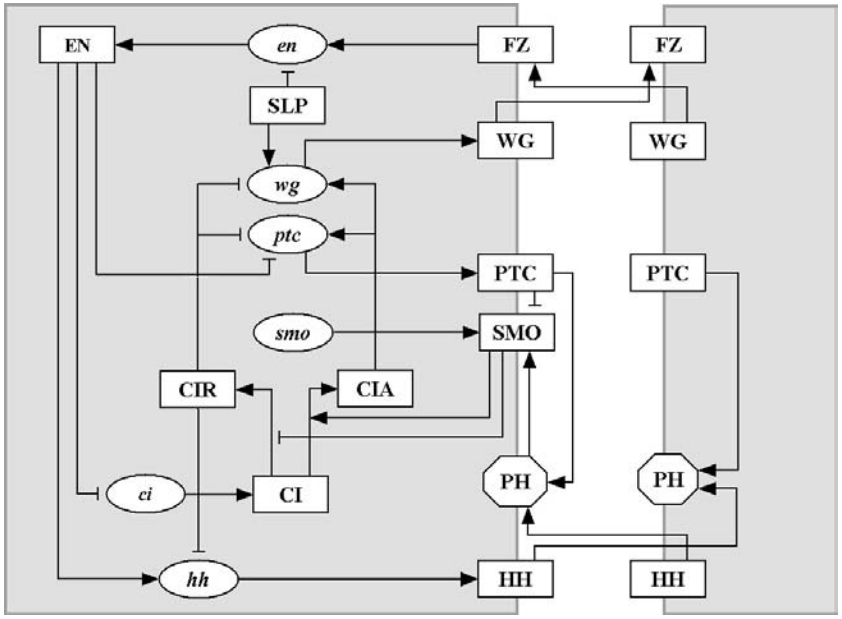
$$EN_i^{t+1} = en_i^t$$

$$WG_i^{t+1} = wg_i^t$$

$$CI_i^{t+1} = ci_i^t$$

$$HH_i^{t+1} = hh_i^t$$

translation

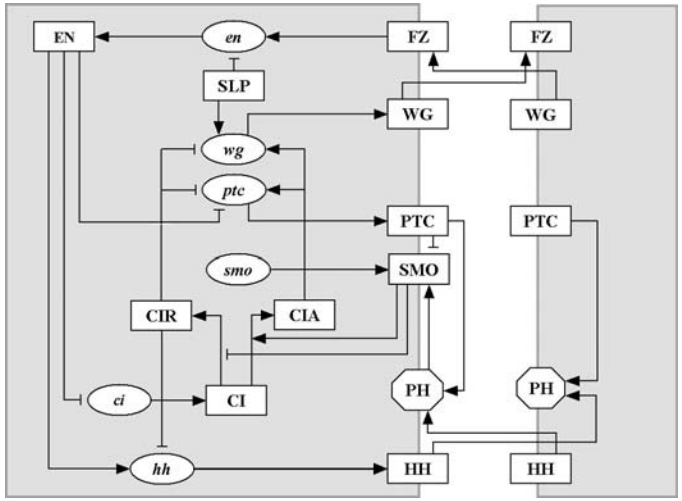


Rules for post-translational processes

$$PH_i^t = PTC_i^t \text{ and } (HH_{i-1}^t \text{ or } HH_{i+1}^t)$$

instantaneous

$$SMO_i^t = \text{not } PTC_i^t \text{ or } HH_{i-1}^t \text{ or } HH_{i+1}^t$$



$$\overbrace{SMO_i^{t+\varepsilon}}$$

$$CIA_i^{t+1} = CI_i^t \text{ and } (\overbrace{SMO_i^t} \text{ or } hh_{i-1}^t \text{or } hh_{i+1}^t)$$

$$CIR_i^{t+1} = CI_i^t \text{ and not } SMO_i^t \text{ and not } hh_{i-1}^t \text{and not } hh_{i+1}^t$$

wg, PTC and SLP are more stable than other proteins

$$wg_i^{t+1} = (CIA_i^t \text{ and } SLP_i^t \text{ and not } CIR_i^t) \text{ or}$$
$$[wg_i^t \text{ and } (CIA_i^t \text{ or } SLP_i^t) \text{ and not } CIR_i^t]$$

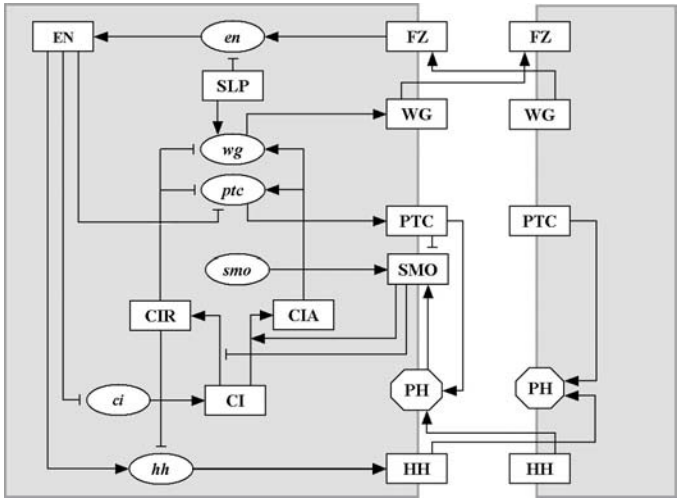
Either of the activators can counter mRNA decay.

$$PTC_i^{t+1} = ptc_i^t \text{ or } (PTC_i^t \text{ and not } HH_{i-1}^t \text{ and not } HH_{i+1}^t)$$

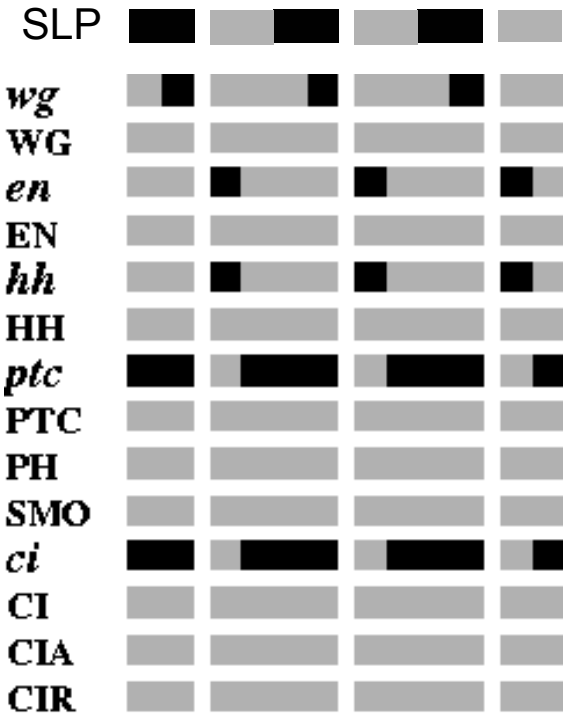
Free PTC does not decay .

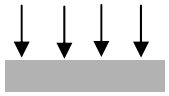
$$SLP_i^{t+1} = SLP_i^t$$

SLP is a source in the segment polarity network.



Start the model from an initial state giving the prepattern of all nodes





Parasegment with four cells

Two cells of a previous/ next parasegment

cell with node ON

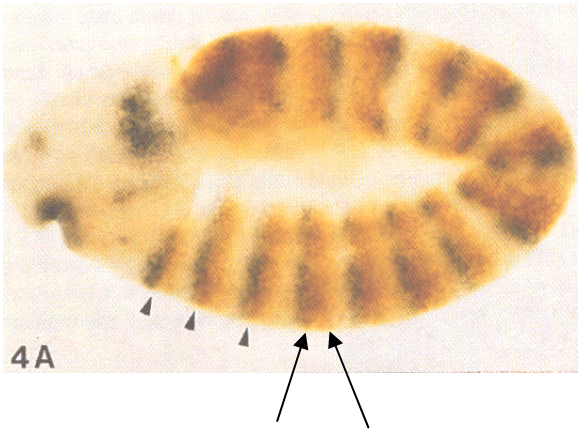
cell with node OFF

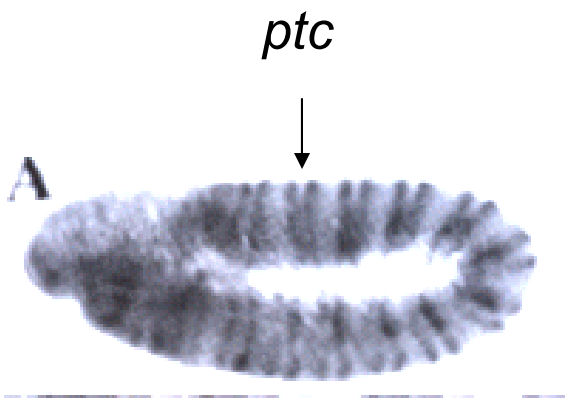
The patterns are assumed to be identical in each parasegment.

Wild type initial state: *wg* in the last cell of the parasegment, *en/hh* in the first cell of the parasegment, *ptc* and *ci* complementary to *en,* no proteins.

Within six steps the model reproduces the wild type steady state

initial state steady state





The net effect of the interactions is enough to capture the functioning of the network.

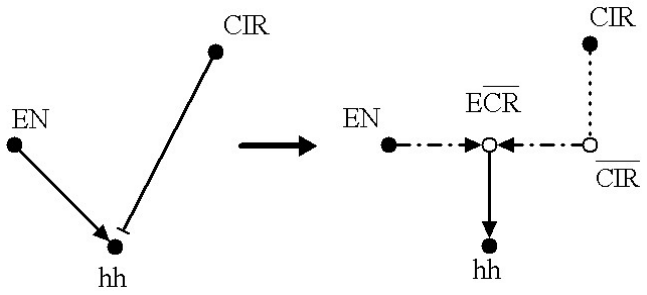
The kinetic details of the interactions can vary as long as their overall effect is maintained – robustness.

Building a "functional topology"

We propose a novel approach: integrate the Boolean rules into the network to express function through topology.

- The future expression of a node depends on a combination of the expression of other nodes.

$$hh_i^{t+1} = EN_i^t \text{ and not } CIR_i^t$$



- Introduce "complementary" nodes.

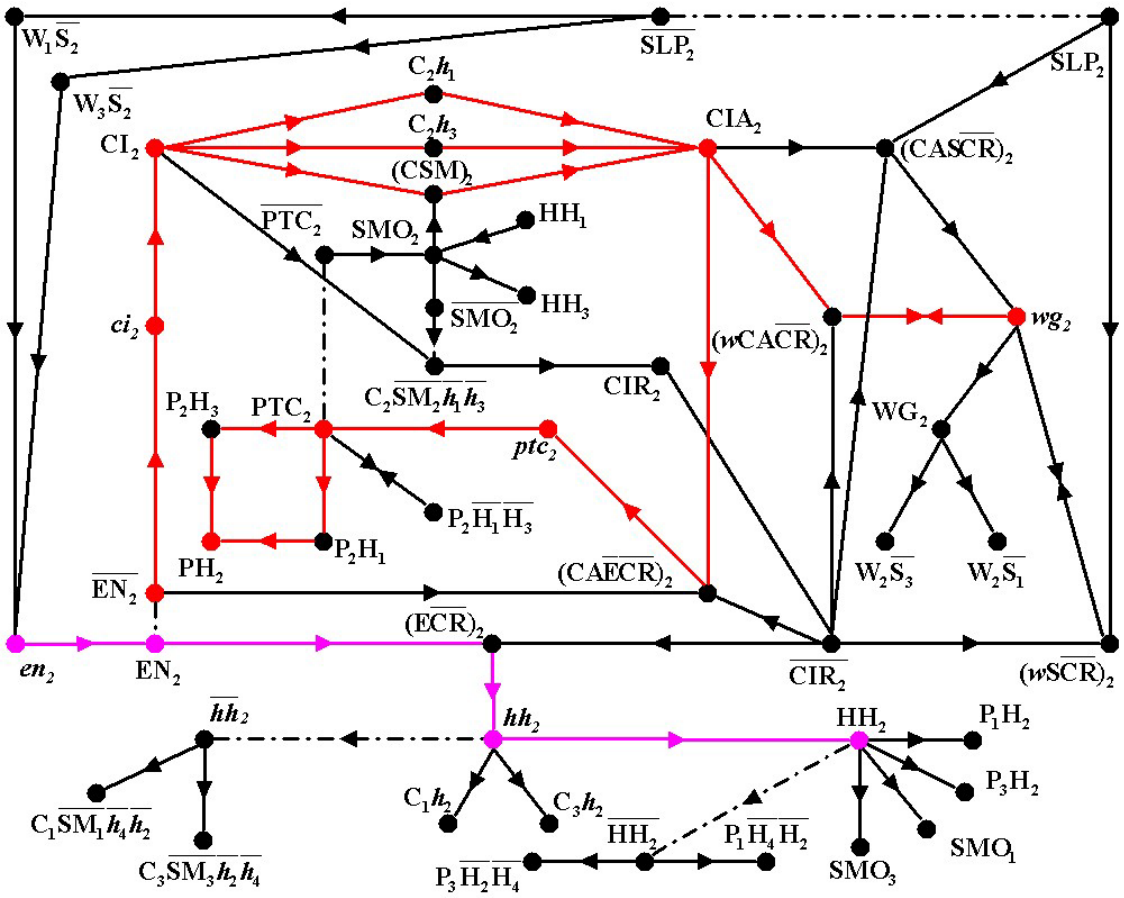
$$\overline{CIR}_i \equiv \text{not } CIR_i$$

- Associate pseudo-nodes to node combinations.

$$(E\overline{CR})_i \equiv EN_i \text{ and } \overline{CIR}_i$$

- The future expression of nodes depends on the expression of pseudo-nodes. $hh_i^{t+1} = (E\overline{CR})_i^t$

The functional graph reveals activating paths



— · — complement

Cycles:

$wg_2 \leftrightarrow (wCA\overline{CR})_2$

$wg_2 \leftrightarrow (wS\overline{CR})_2$

$\mathbf{PTC_2 \leftrightarrow P_2\overline{H_1}\,\overline{H_3}}$

Activating paths determine segment polarity

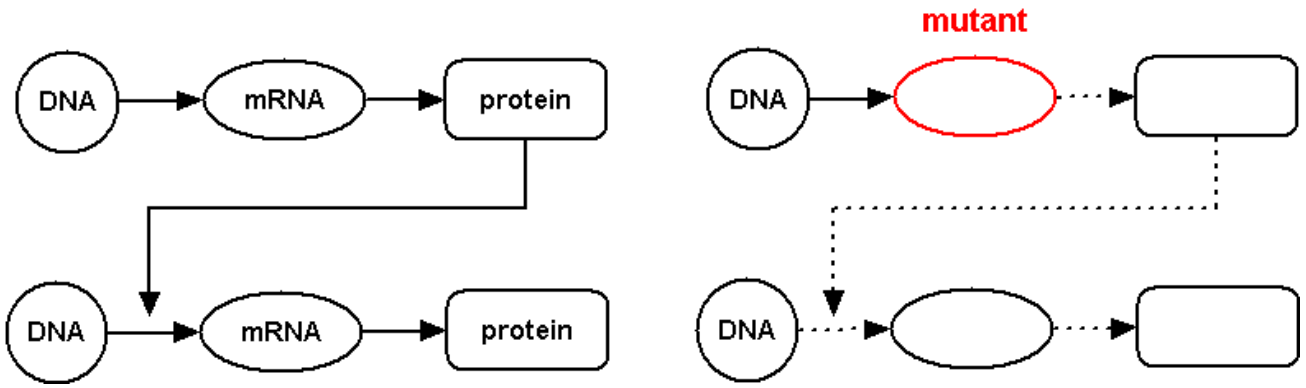
- *en*, EN, *hh*, HH
- *ci*, CI, CIA, *wg*, WG, *ptc*, PTC

What happens if the components are perturbed?

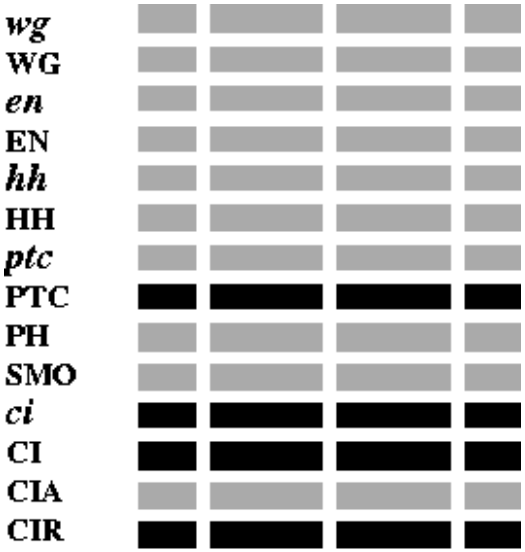
The most severe perturbation is caused by gene mutations.

To model a null mutation, we assume that the mRNA is kept OFF, thus the protein cannot be translated.

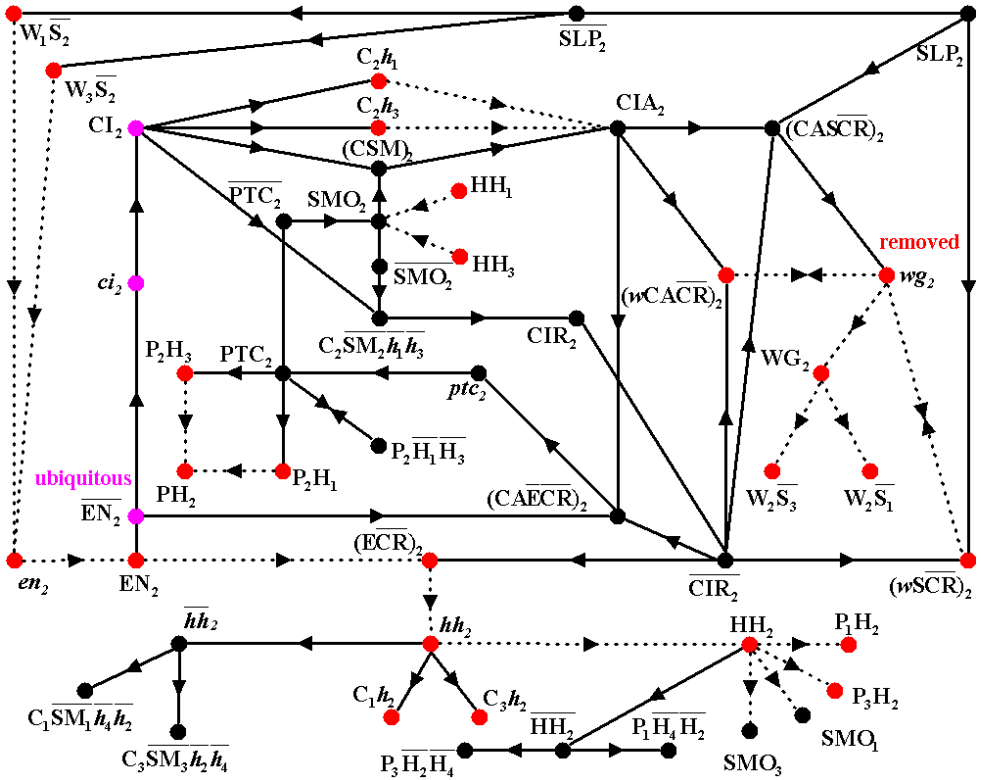
The effects of the mutation propagate throughout the network.



Modeling *wg, en* or *hh* mutations

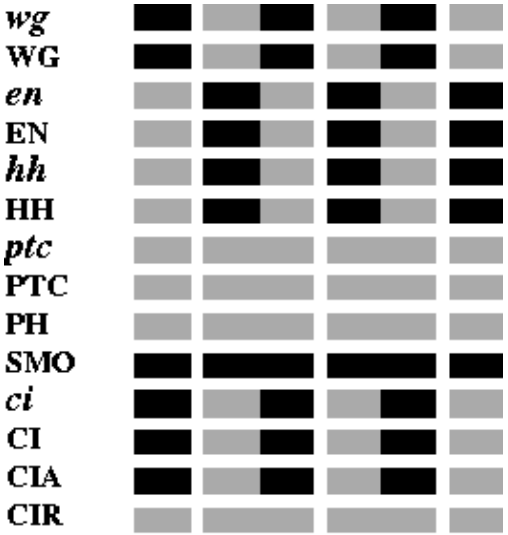


final state

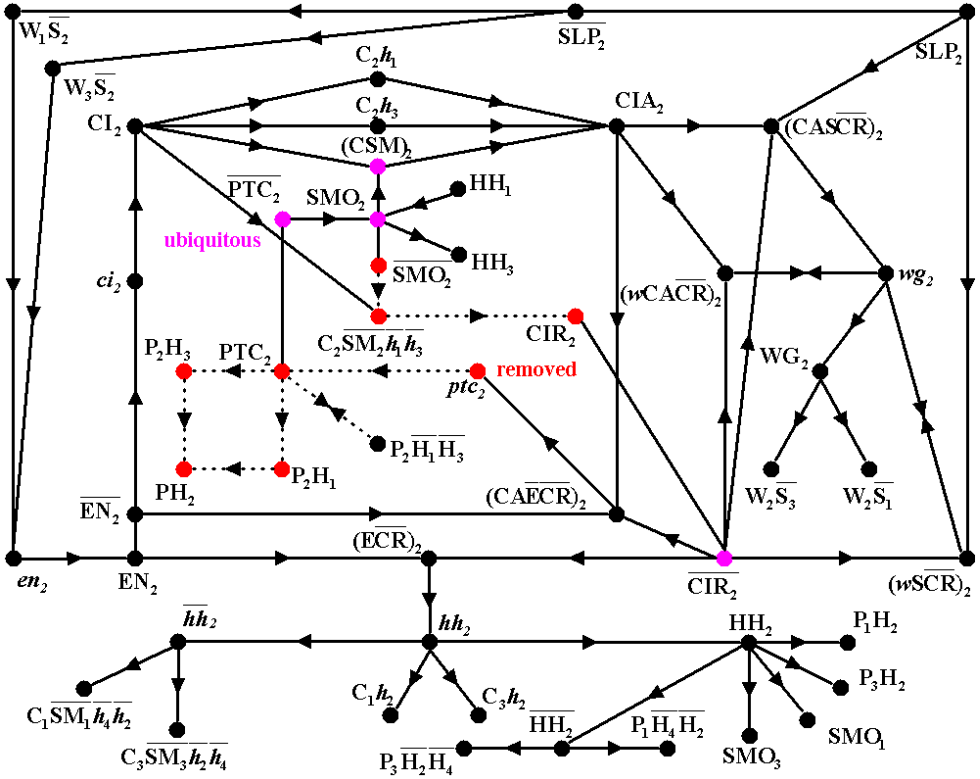


No segmentation, regardless of initial state: lethal mutation.

ptc mutation broadens the stripes

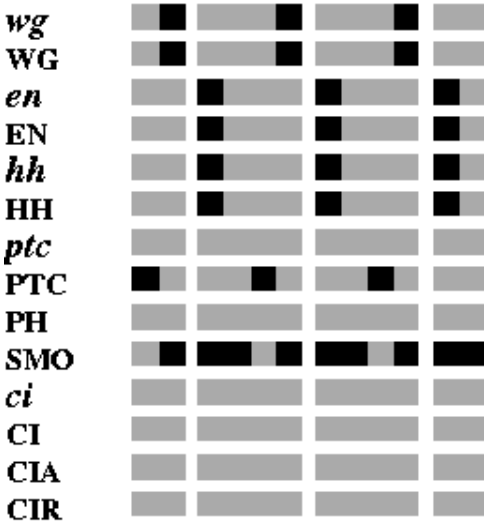


final state

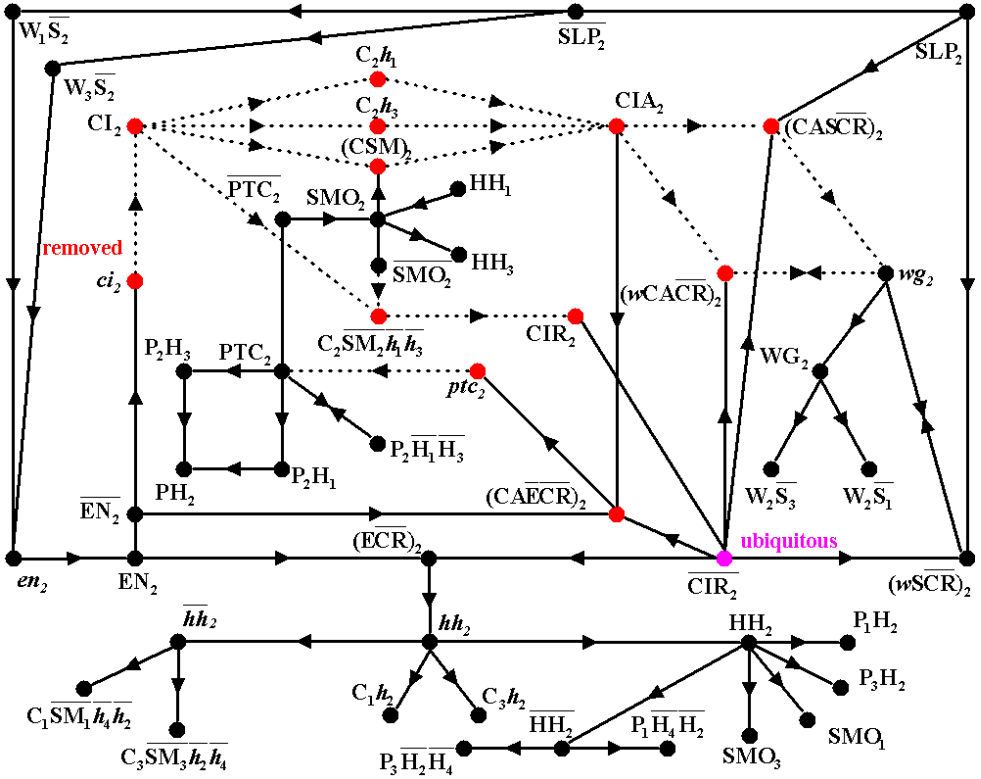


The *wg*, *en* and *hh* stripes broaden, regardless of initial state.

ci mutation can preserve the prepattern



final state for wild type prepattern

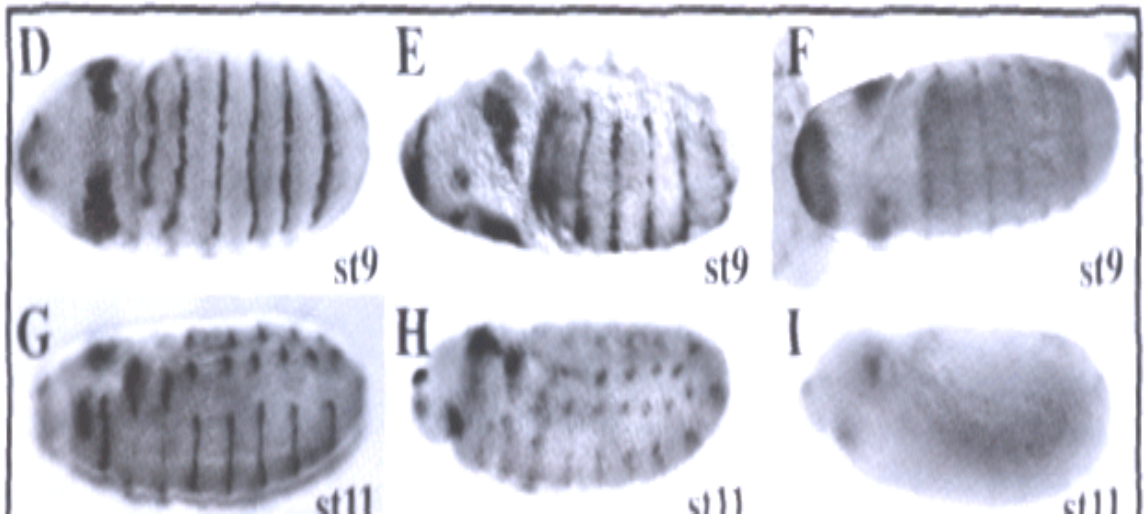


The effect of *ci* mutation depends on the initial state.
For wild type prepattern, the *wg, en, hh* stripes remain.

Model matches experimental results on mutants

wg initial

wg final



prediction for *wg*

wild type

ci mutant

hh mutant

Gallet et al., Development 127, 5509 (2000)

Comparison between *ci* and *ptc* mutants

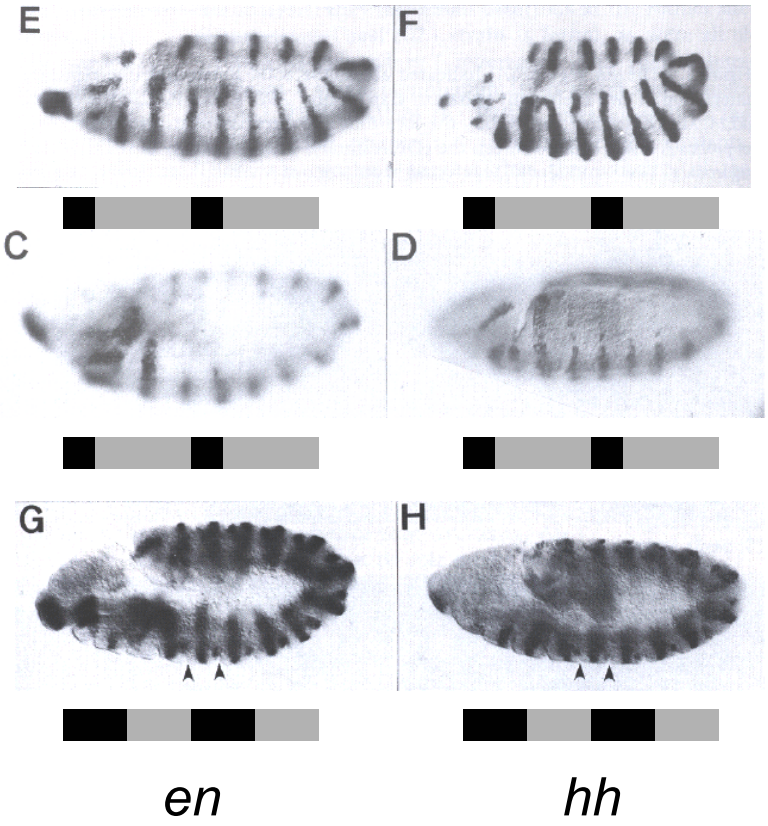
wild type

ci mutant

prediction, *ci* mutant

ptc mutant

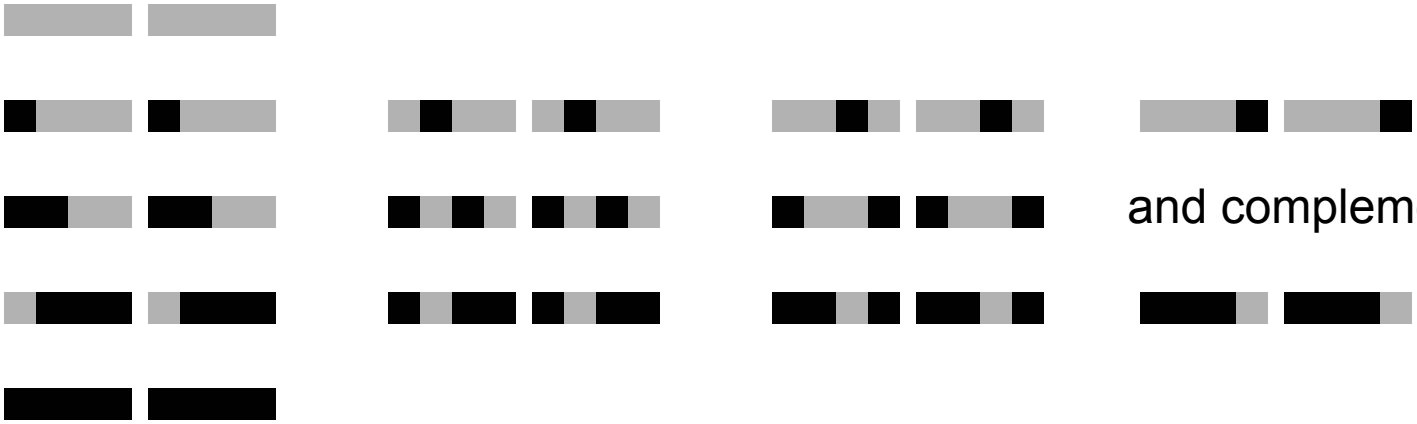
prediction, *ptc* mutant



Tabata, Eaton, Kornberg, Genes & Development 6, 2635 (1992)

Sensitivity to initial states

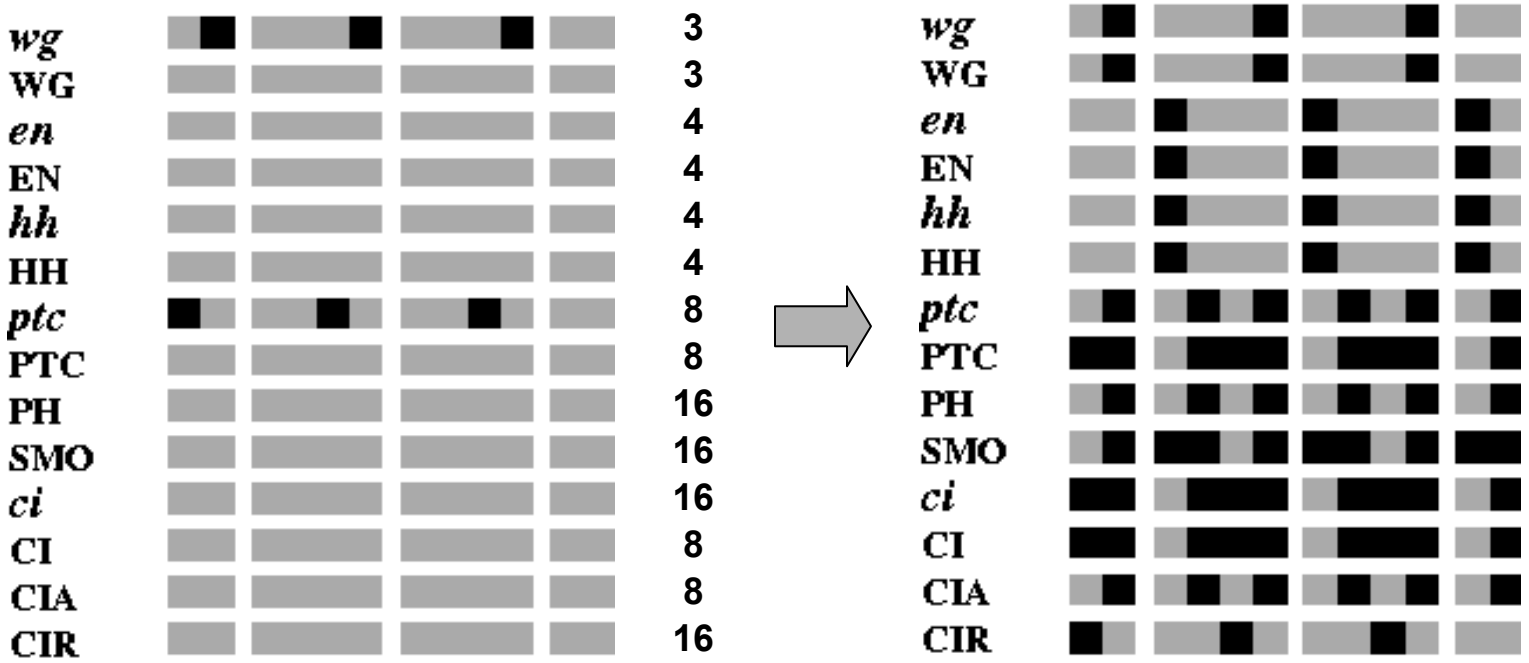
Possible number of prepatterns for a single node: 16



Total number of network-wide prepatterns: $N_i = 16^{15}$

All initial states lead to steady states within 10-15 steps! - robustness

How many initial states lead to the experimentally observed wild type steady state?



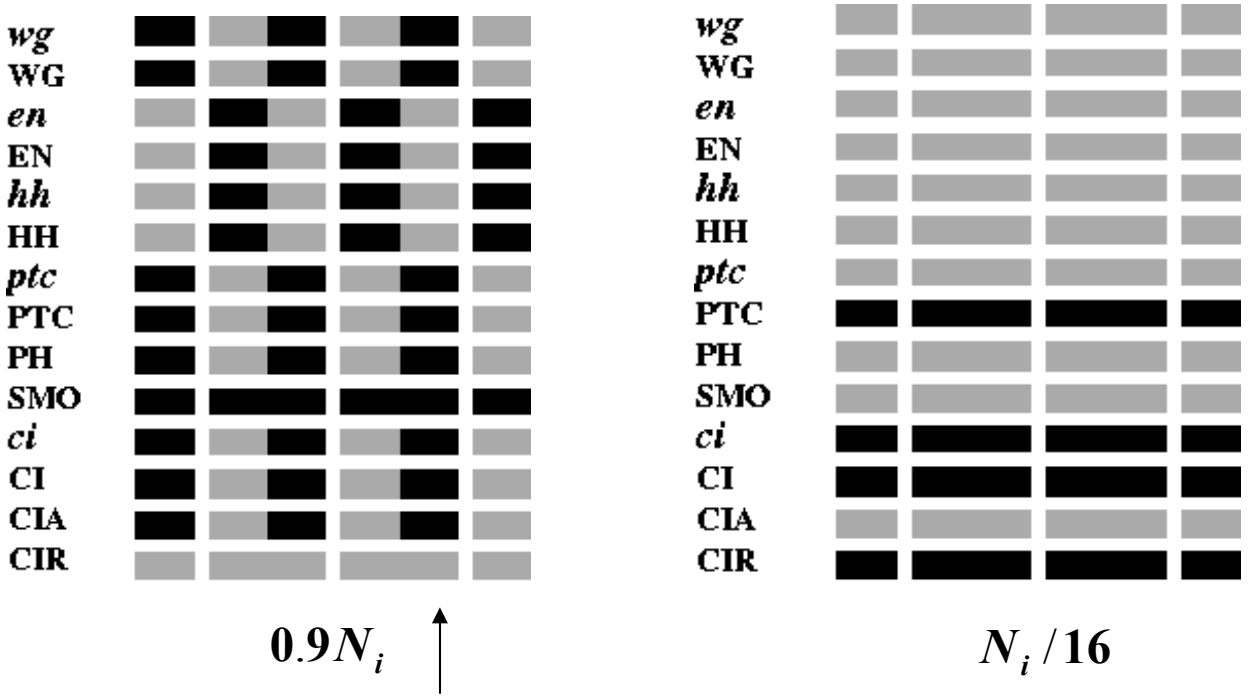
minimal prepattern

steady state

Total number of wild-type inducing prepatterns: $6\times10^{11} = 8\times10^{-6} N_i$

There are two other frequently occurring steady states

Broader initiation of *wg, en* or *hh* leads to broad stripes.
Absence of *wg* leads to a state with no segmentation.



This state has been observed in heat-shock experiments.
Gallet et al., Development 127, 5509 (2000)

The steady states can be determined analytically from the Boolean rules

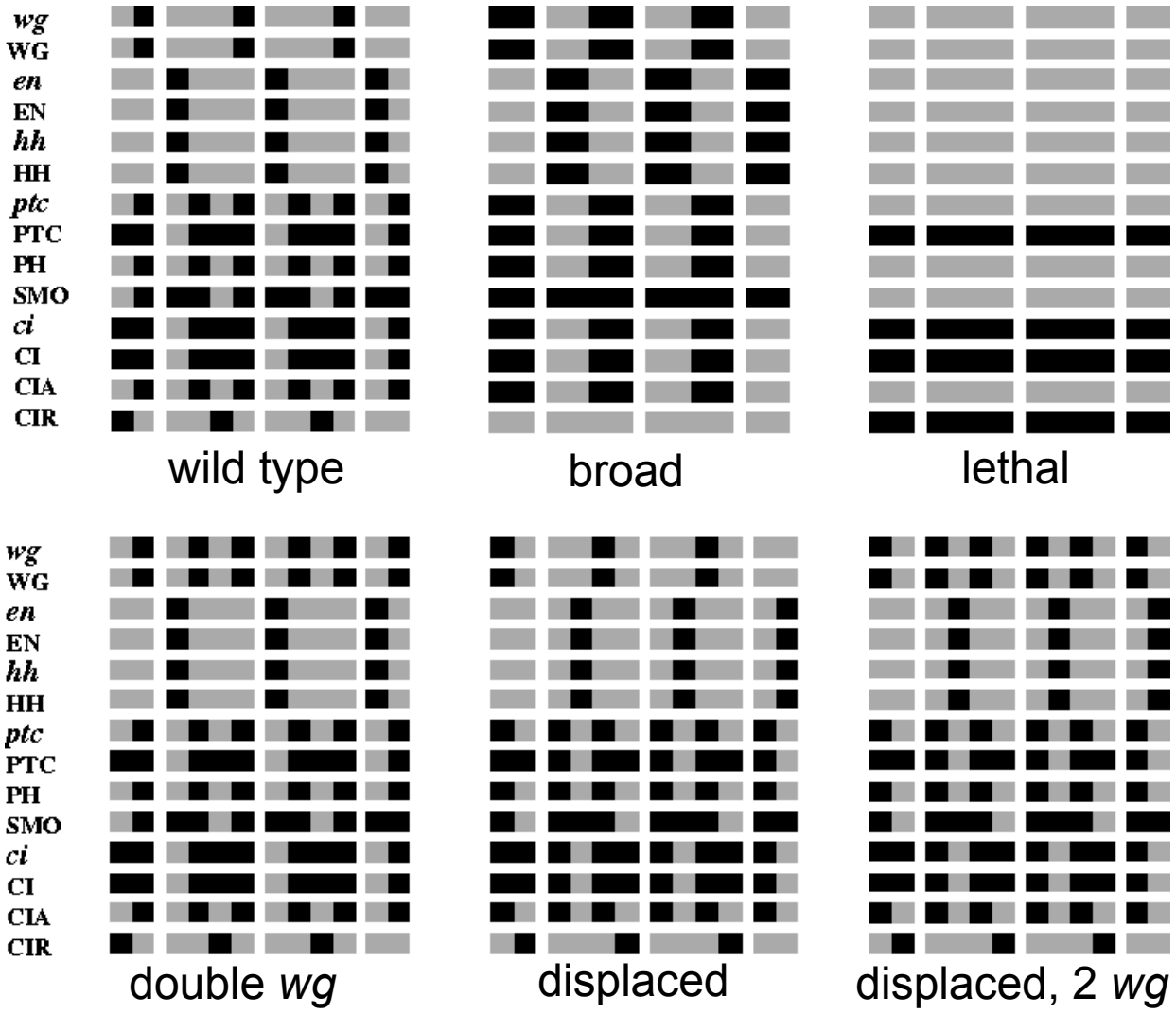
- In the stable state $x_i^{t+1} = x_i^t$
- Use the fact that $SLP_1 = SLP_2 = 0$ and $SLP_3 = SLP_4 = 1$
- The set of equations reduces to:

$$\begin{cases} wg_1 = wg_1 \text{ and not } wg_2 \text{ and not } wg_4 \\ wg_2 = wg_2 \text{ and not } wg_1 \text{ and not } wg_3 \\ wg_3 = wg_1 \text{ or } wg_3 \\ wg_4 = wg_2 \text{ or } wg_4 \end{cases}$$

$$\begin{cases} PTC_1 = (\text{not } wg_2 \text{ and not } wg_4) \text{ or } (PTC_1 \text{ and not } wg_1 \text{ and not } wg_3) \\ PTC_2 = (\text{not } wg_1 \text{ and not } wg_3) \text{ or } (PTC_2 \text{ and not } wg_2 \text{ and not } wg_4) \\ PTC_3 = PTC_4 = 1 \end{cases}$$

Reflects the assumption of stability of *wg* and PTC.

Model's prediction for the only possible stable patterns



The first three have been observed experimentally.
The latter states have very small probability.

How robust is the model to changes in the basic assumptions?

1. Node expression decays in one step if not renewed.

   Change: proteins decay in two steps; $EN_i^{t+1} = en_i^t \text{ or } en_i^{t-1}$

   Same steady states, only the path leading to them changes.

   More overall stability.

2. *wg* and PTC expression is easier to maintain than others.

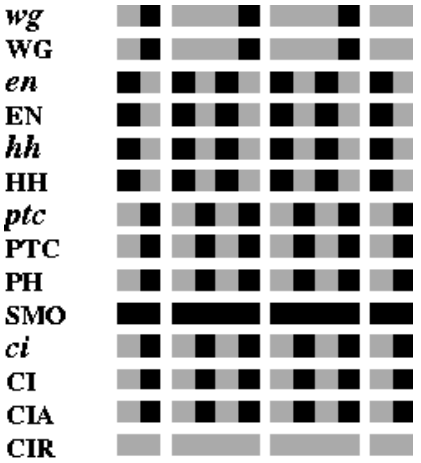
   Change: *wg* and PTC decay if not renewed.

   ci mutants have no segmentation if *wg* decays;

   no wild type steady state if PTC decays.

The model can be modified to include different timescales for transcription, translation, post-transcriptional modifications, mRNA and protein decay.

Can we relax assumptions of time-independence?

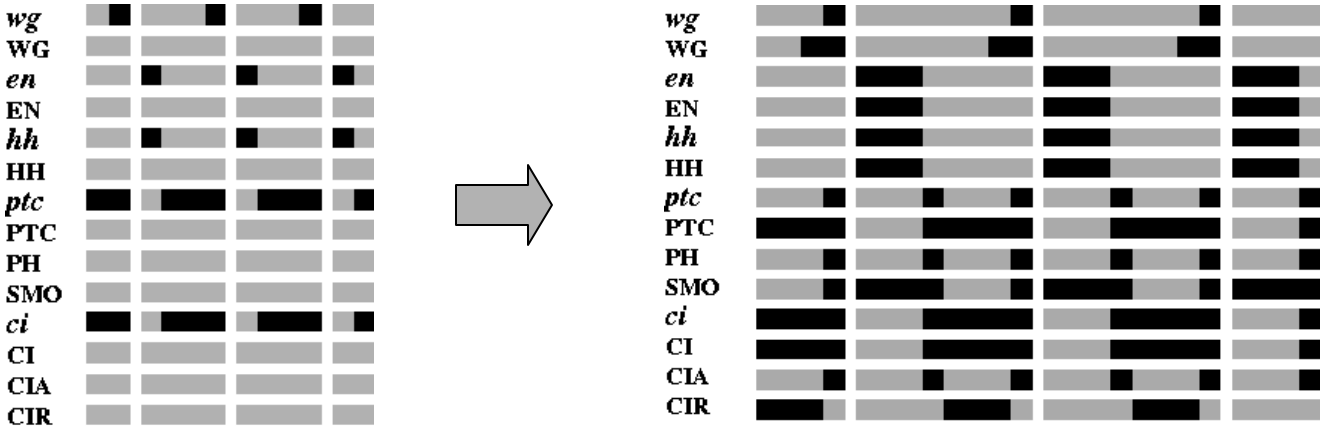


3. SLP expression is a necessary input.

   No wild type steady state in a SLP mutant.

4. The parasegments remain 4 cells wide.

   Cell division can be incorporated into the model.



Conclusions

Boolean modeling successfully reproduced all known patterns of the segment polarity genes, both wild type and mutant.

Functional topology reveals the robustness of the segment polarity network

- The kinetic details of the interactions do not matter.
- The wild type steady state is robust to perturbations.
- The function is maintained even for some gene mutations.

The model predicts the existence of three additional steady states, and illustrates the crucial role played by the *wingless* and *sloppy paired* genes.