Topology-based modeling of cellular regulatory networks

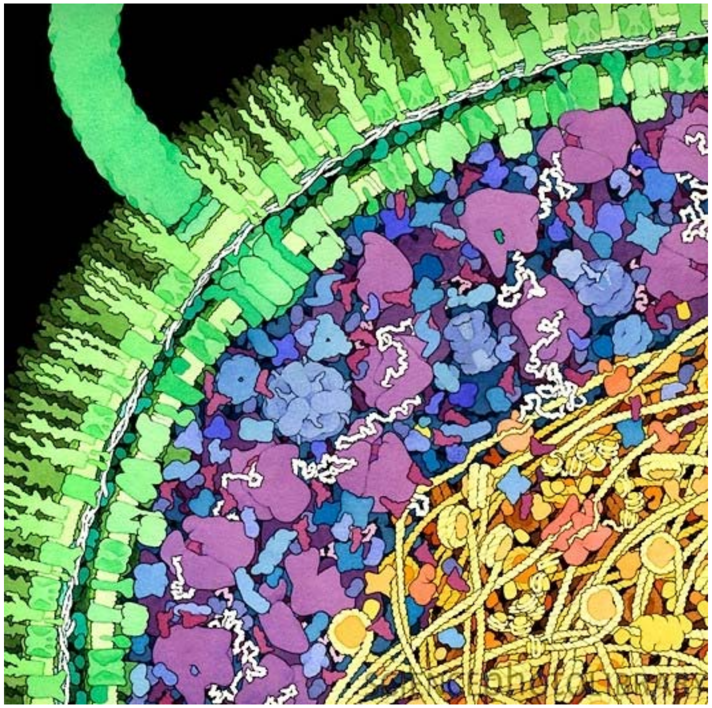
Réka Albert
Departments of Physics and Biology,
Huck Institutes for the Life Sciences
Pennsylvania State University

Life at the cellular level

- Gene ⟶ protein
- Proteins
  - provide structure to cells and tissues
  - work as molecular motors
  - sense chemicals in the environment
  - drive chemical reactions
  - regulate gene expression

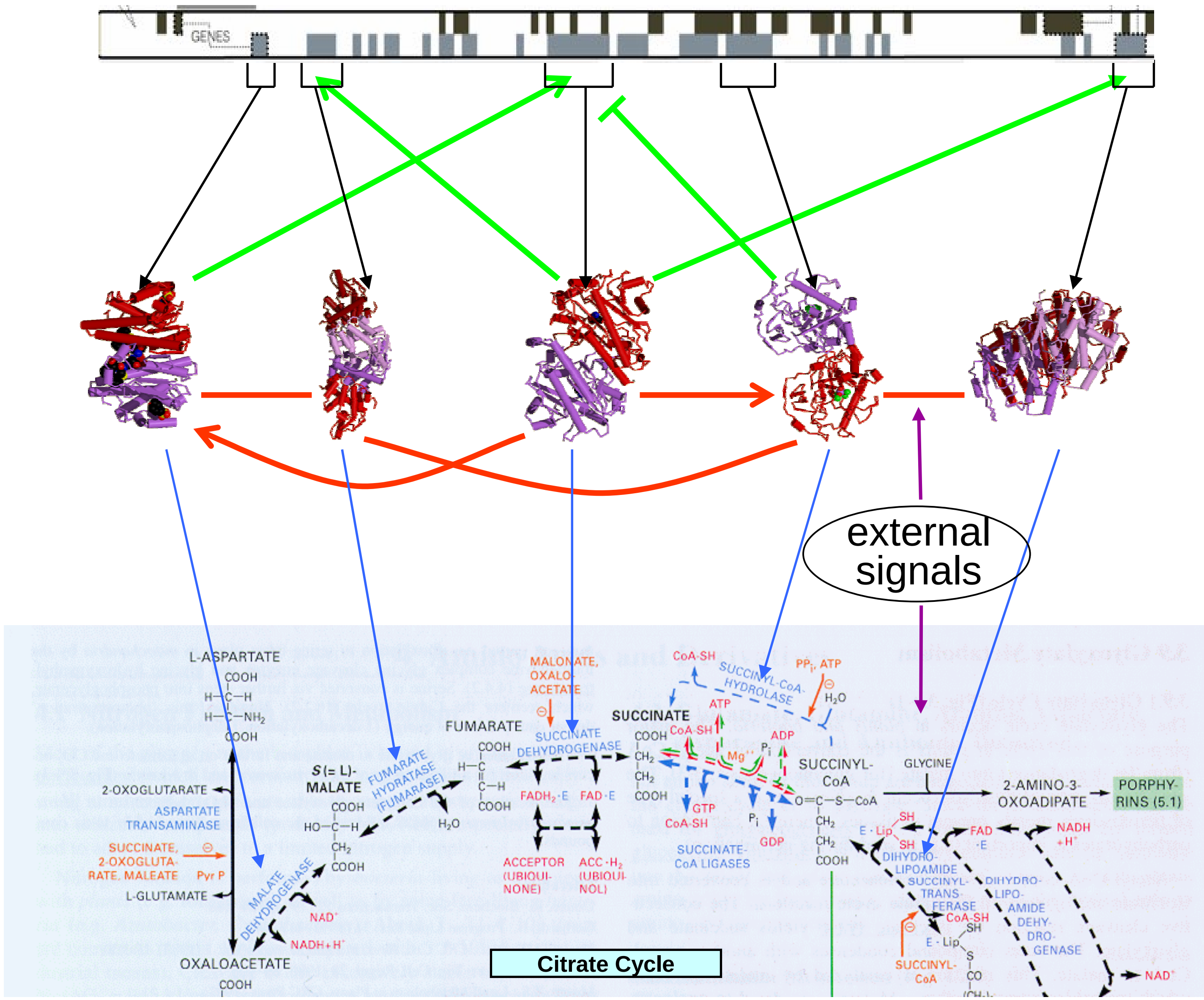
- Cellular functions rely on the coordinated action of gene products.

- Interconnections between components are the essence of a living process.



David Goodsell/ Science Photo Library

Many non-identical elements connected by diverse interactions



GENOME

gene regulation

PROTEOME

protein-protein interactions

signal transduction

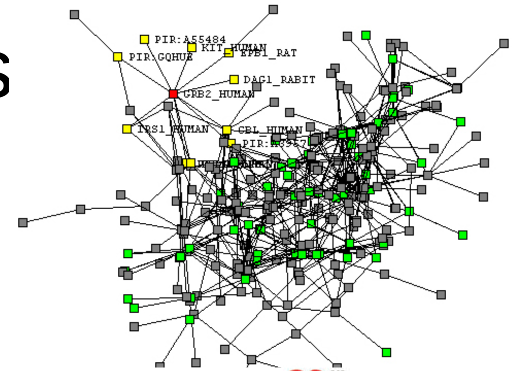
METABOLISM

Bio-chemical reactions

Definition of cellular networks

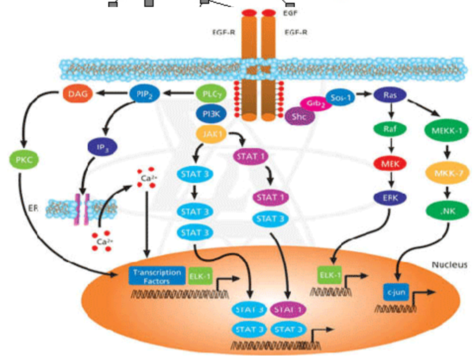
1. Protein interaction networks

Nodes: proteins
Edges: protein-protein interactions (binding)



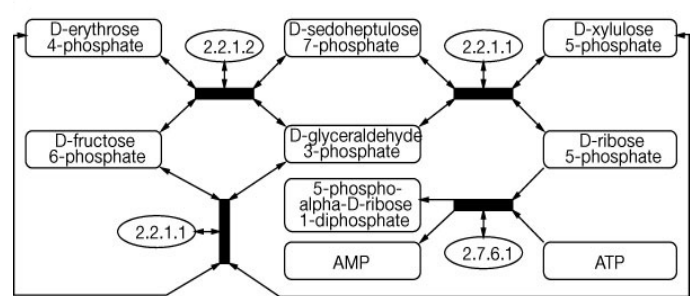
2. Signal transduction networks

Nodes: proteins, molecules
Edges: reactions and processes reflecting information transfer



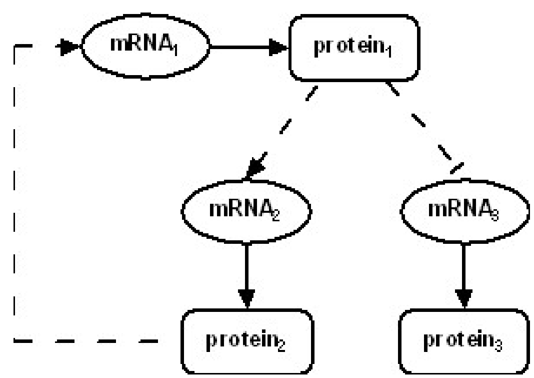
3. Metabolic networks

Nodes: metabolites, enzymes
Two types of edges: mass flow or catalysis.



4. Gene regulatory networks

Two types of nodes: mRNA , protein
Two types of edges: mass flow or regulation (activating or inhibiting).



Mapping of cellular interaction networks

Experimental advances allow the construction of genome-wide cellular interaction networks

- Protein networks

  Uetz et al. 2000, Ito et al., 2001, Krogan et al. 2006 – *S. cerevisiae,*
  Giot et al. 2003 – *Drosophila melanogaster* , Li et al. 2004 – *C. elegans,* Rual et al 2005 - Human interactome
- Transcriptional regulatory networks

  Shen-Orr et al. 2002 – *E. coli*,
  Guelzim et al 2002, Lee et al. 2002 - *S. cerevisiae,*
  Davidson et al. 2002 – sea urchin
- Signal transduction networks

  Ma'ayan et al. 2005 – mammalian hippocampal neuron

Graph analysis uncovered common architectural features of cellular networks: Connected, short path length, heterogeneous (scale-free), overexpressed interaction motifs

Importance of a dynamical understanding

Only subsets of the genome-wide interaction networks are active in a given external condition

Han et al. 2004 – dynamical modularity of protein interaction networks – date hubs and party hubs
Luscombe et al. 2004 – endogeneus and exogeneus transcriptional subnetworks

Network topology needs to be complemented by a description of network dynamics – states of the nodes and changes in the state

Quantitative dynamic description is only feasible on smaller networks (modules):
Signal transduction in bacterial chemotaxis, the yeast cell cycle, the mammalian circadian clock

Dynamic modeling

- Network = backbone of process
- Node states + transfer functions → outcome

Ingredients: components of the system; interactions, states of components

Hypotheses: transfer functions, kinetics, parameters.

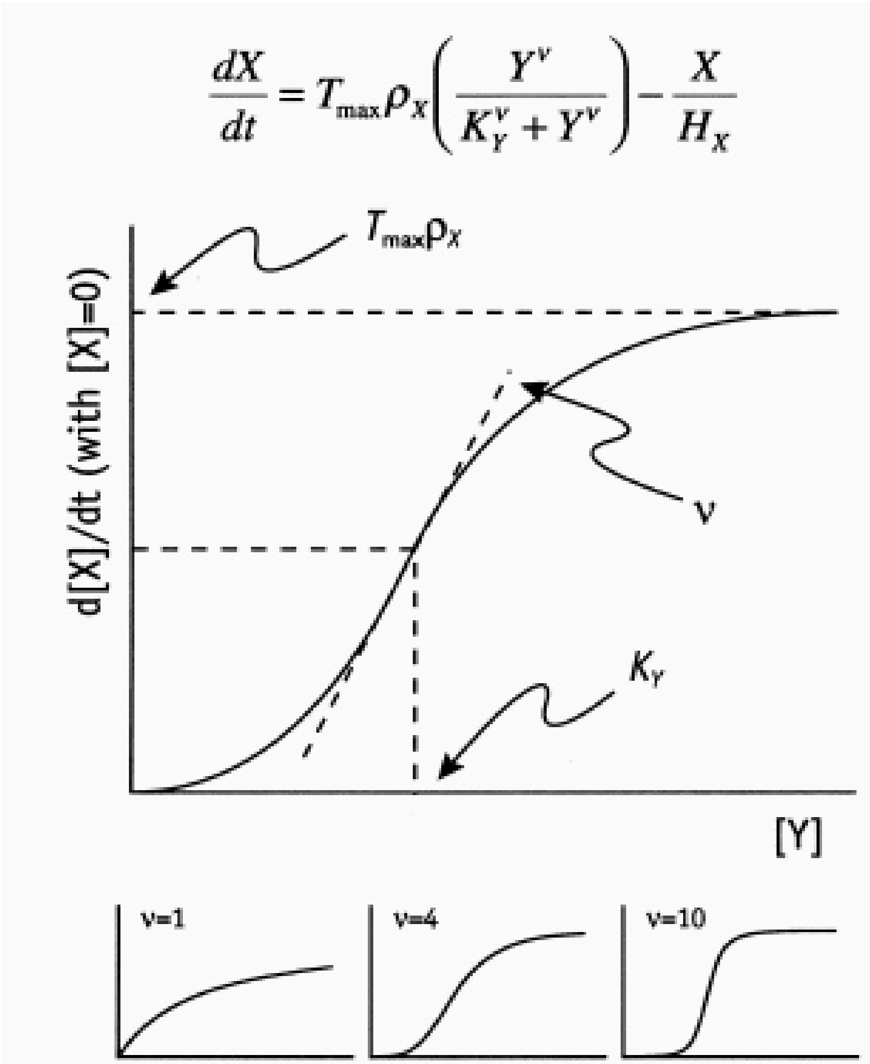
Validation: capture known behavior.

Explore: study cases that are not accessible experimentally
change parameters, change assumptions
gain insight into why complexity is necessary

Qualitative models: the regulatory network is more important than the kinetic details of the individual interactions.

Essential information: inhibitors, conditional activation, independent activation, decay, relative timing (when known)

Dose-response curves for regulated processes



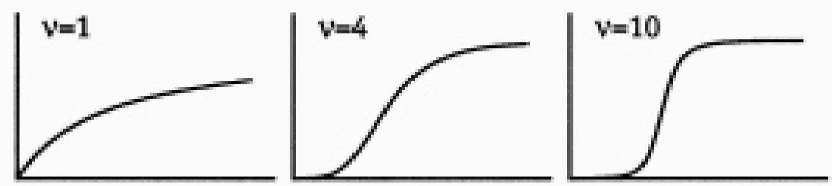
- Y – regulator (e.g. transcriptional activator)
- X – target (e.g. mRNA)
- Synthesis is a nonlinear function of activator
- Decay is un-catalyzed
- Parameters:
  - maximum rate $T_{max}\rho_X$
  - Half-maximal activation K_Y
  - Hill coefficient ν
  - Half- life H_X

Combinatorial regulation of synthesis is approximated with similar sigmoidal curves.

$$\frac{dX_j}{dt} = T_j^{max}\, g\left(\sum_i R_{ij} X_i\right) - \frac{X_j}{H_j}$$

From dose-response curves to switches

$$\frac{dX}{dt} = T_{\max}\rho_X\left(\frac{Y^\nu}{K_Y^\nu + Y^\nu}\right) - \frac{X}{H_X}$$



X – mRNA
Y – transcriptional activator

If ν is large, the dose-response curve becomes a switch

If $Y>K_Y$ $dX/dt>0$
If $Y<K_Y$ $dX/dt<0$
The activation threshold is K_Y
If activation is weak, mRNA can decay.

Boolean simplification:
$Y>K_Y \longrightarrow Y=ON$
$Y<K_Y \longrightarrow Y=OFF$

X* = Y
Activation:
If Y=ON X produced
X*=ON
Decay:
If Y= OFF X decays
X*=OFF

Refinement of Boolean model

- Discrete decay times and threshold durations

$$X^* = Y \text{ and not } X^{t-\tau_X} \qquad X^* = \bigcup_{i=0}^{i_{max}} \text{and } Y^{t-i}$$

- Asynchronous update

- Hybrid model (Glass & Kauffmann 1973): each node is characterized by both a continuous and a Boolean variable.

$$\frac{d\hat{X}}{dt} = Y - \hat{X}$$

- X_i is defined by the threshold rule

$$X = \begin{cases} 0, & \text{if } \hat{X} < 0.5 \\ 1, & \text{if } \hat{X} > 0.5 \end{cases}$$

Modeling the segment polarity gene network

Input: segment polarity genes

Hypotheses:

continuous model: transcription factors act as enzymes

Boolean model: mRNA and protein activity is switch-like

Validation: reproduces known gene expression patterns.

Explored: changes in kinetic parameters

knock-out mutations

changes in initial conditions

Insight: topology is a main source of robustness.

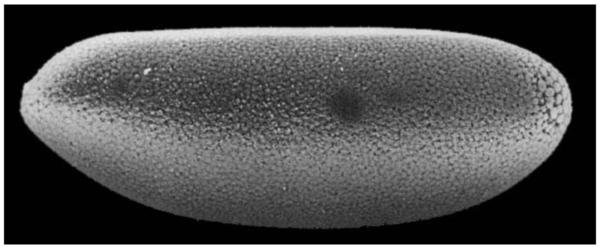
G. von Dassow et al., Nature 406, 188 (2000)

R. Albert, H. G. Othmer, Journ. Theor. Biol. 223, 1 (2003)

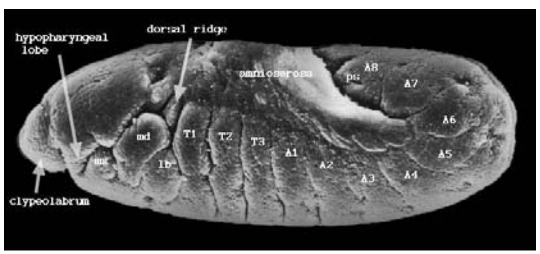
M. Chaves, R. Albert, E. Sontag Journ. Theor. Bio. 235, 431 (2005).

M. Chaves, E. Sontag, R. Albert, IEE Proc. Systems Biology 153, 154 (2006).

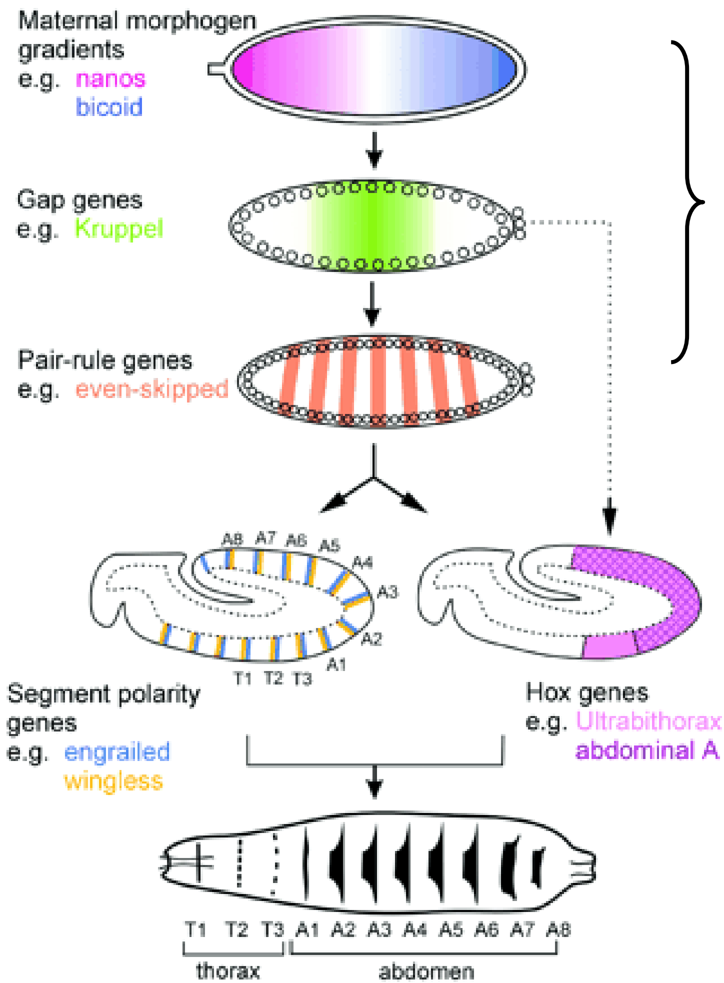
Segmentation of the fruit fly embryo



1h after fertilization



7h after fertilization



Transient gene products, initiate the next step then disappear.

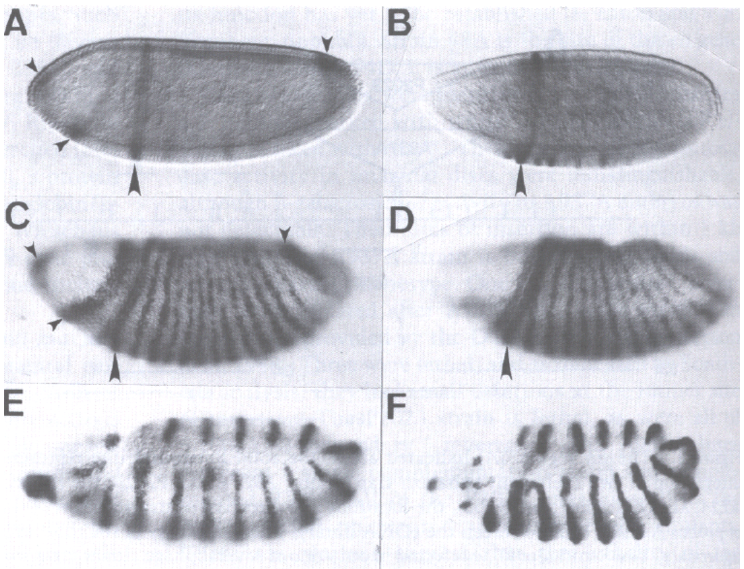
Segment polarity genes

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

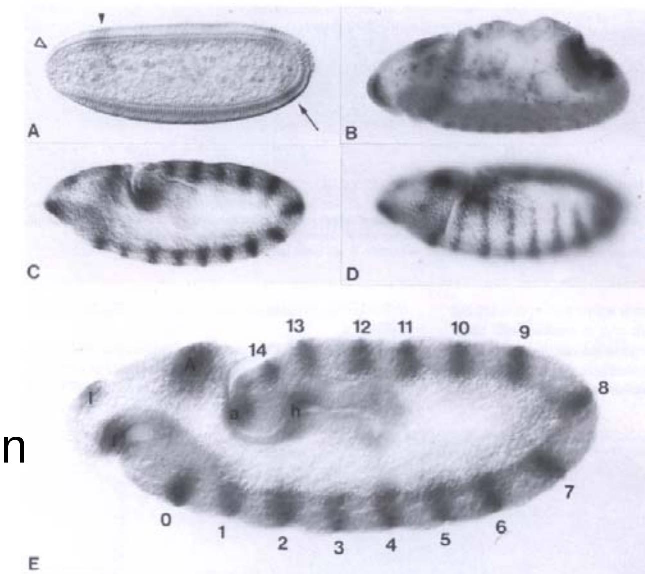
en *hh* *wg*



early stages
2:50 h

pre-pattern
3:00-3:30 h

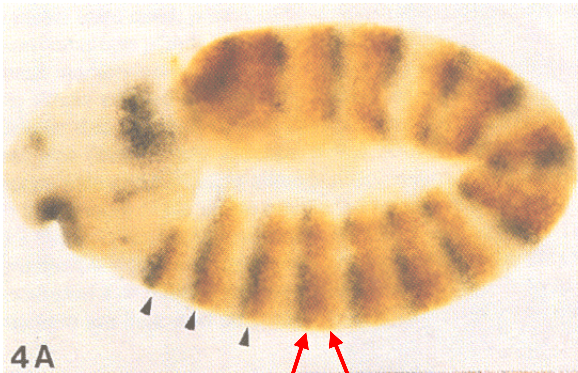
stable pattern
4:20-7:20 h

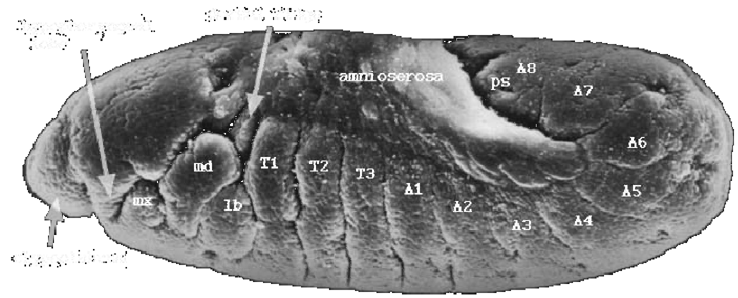


3:30 h

ci *en* *ptc*

Wild type, stable gene patterns





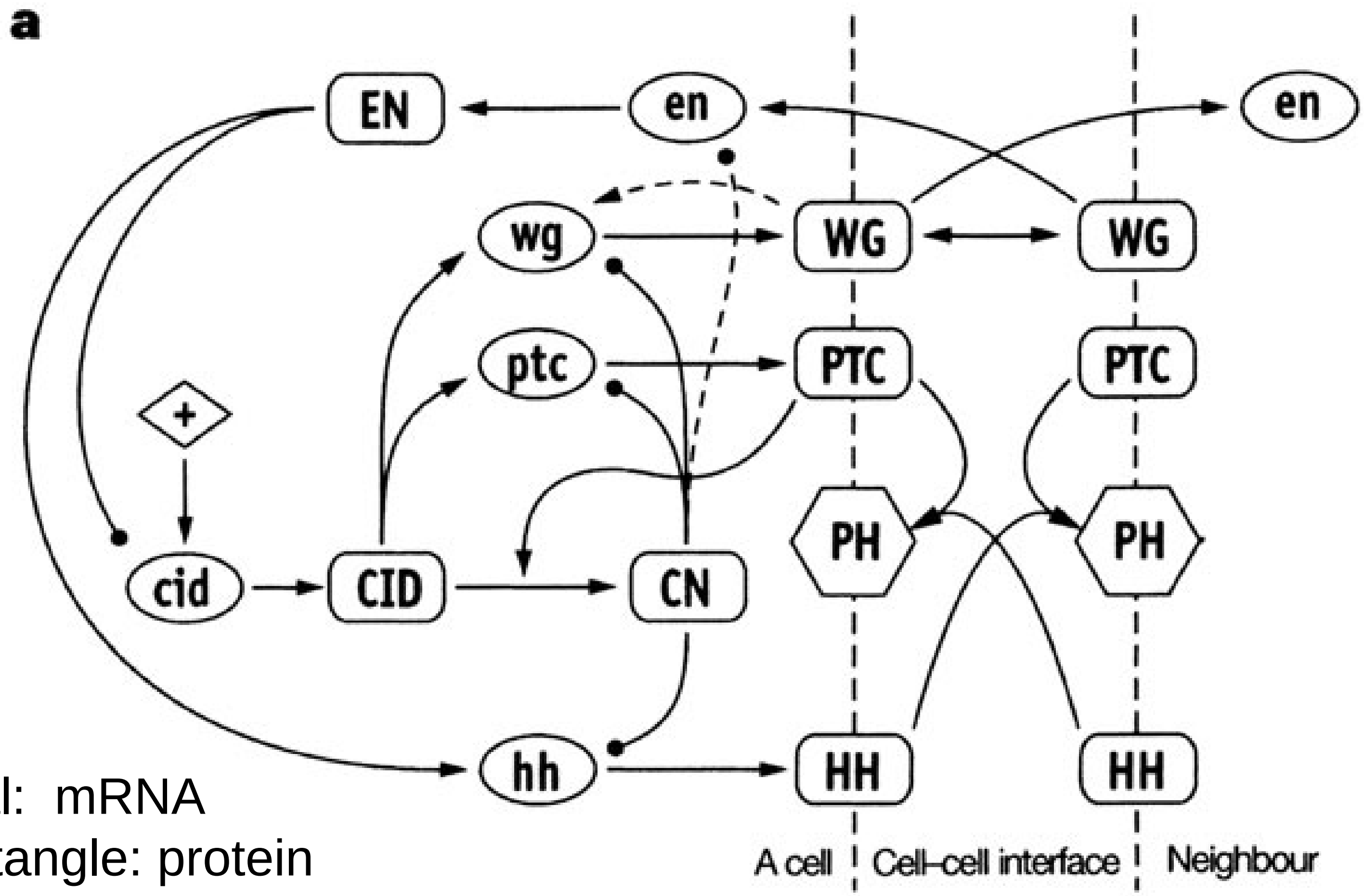
wg en

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.

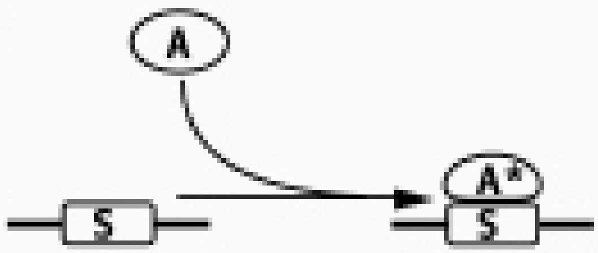
- *two ptc* stripes in each parasegment.
- *ci* pattern is complementary to that of *en.*

Gene interaction network in the von Dassow model



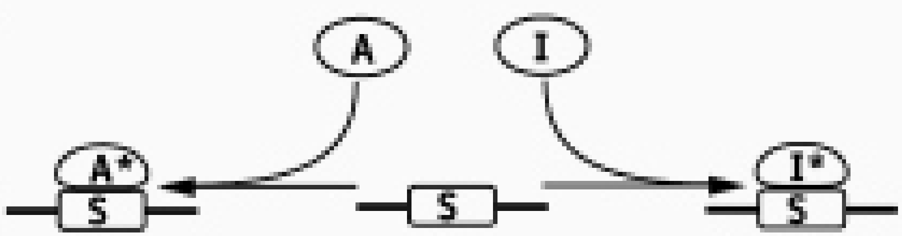
oval: mRNA
rectangle: protein
hexagon: protein complex
inhibition
transport

Transcriptional activation



$$\Phi([A],K_A,\nu_A) = \frac{\left(\frac{[A]}{K_A}\right)^{\nu_A}}{1+\left(\frac{[A]}{K_A}\right)^{\nu_A}} = \left(\frac{[A]^{\nu_A}}{K_A^{\nu_A}+[A]^{\nu_A}}\right)$$

Competition between a transcriptional activator and a transcriptional repressor

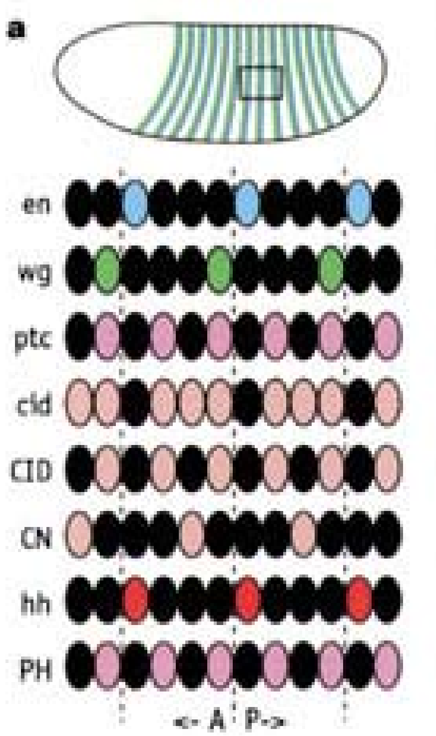


$$[AS] = \frac{\Phi(A)(1-\Phi(I))}{1-\Phi(A)\Phi(I)}$$

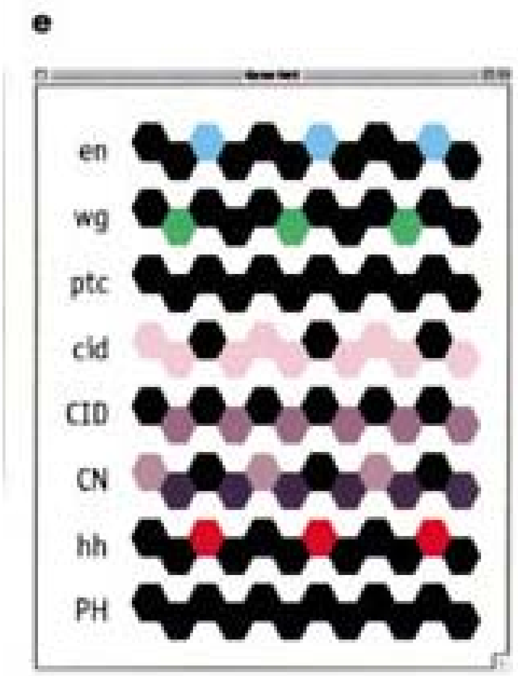
$$= \left(\frac{[A]^{\nu_A}\left(1-\frac{[I]^{\nu_I}}{K_I^{\nu_I}+[I]^{\nu_I}}\right)}{K_A^{\nu_A}+[A]^{\nu_A}\left(1-\frac{[I]^{\nu_I}}{K_I^{\nu_I}+[I]^{\nu_I}}\right)}\right)$$

Gene expression patterns

The 2D pattern is reduced to 1D, assume cells are hexagonal
The gene expression is essentially binary (ON in some cells, OFF in others)



wild type



model solution

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

Robustness to parameter changes

- Start from the wild type initial condition for *en* and *wg*

- Generate a set of kinetic parameters from the biologically relevant range (48 unknown parameters)

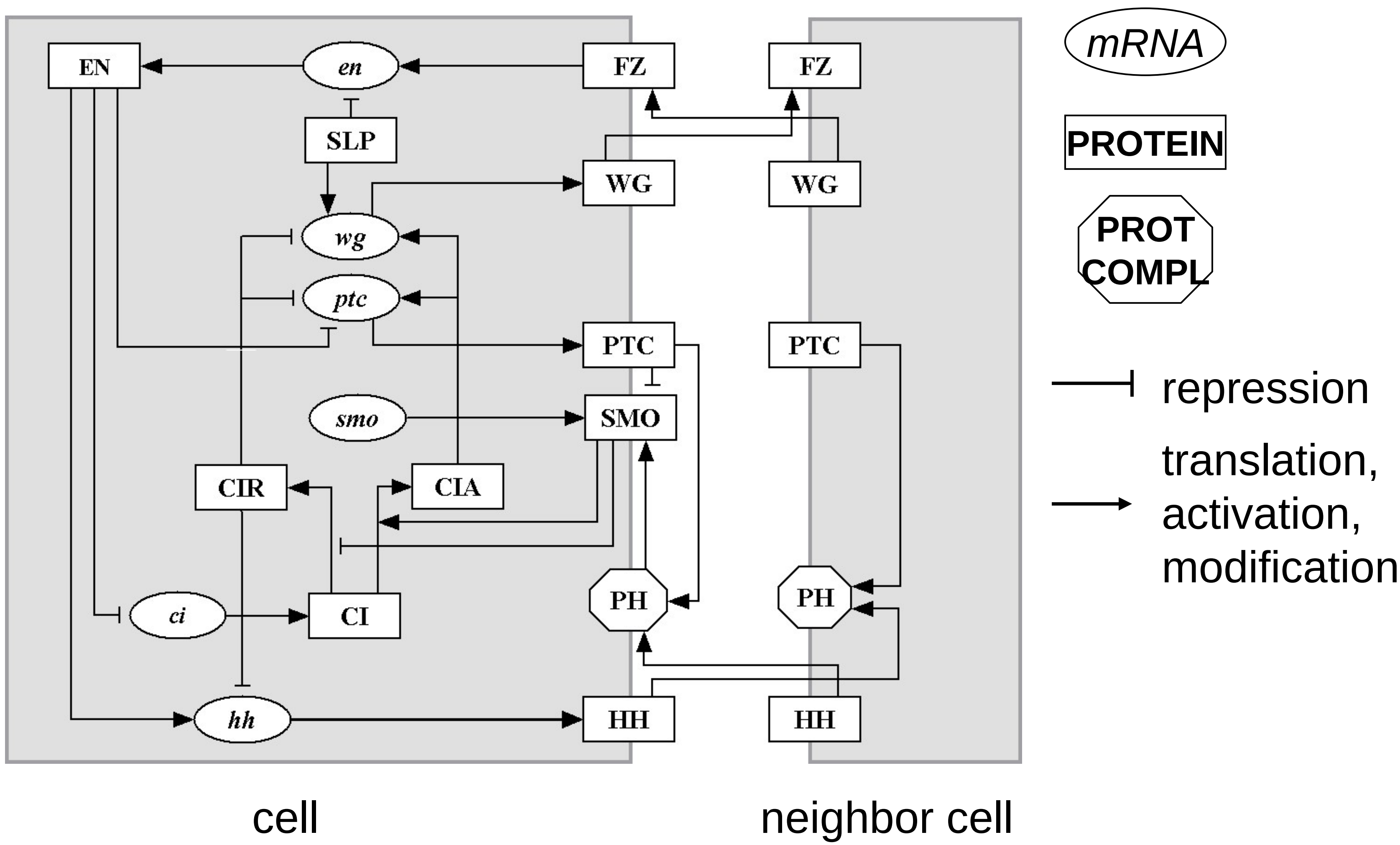
- Run the simulation until steady state is reached.
- Use threshold (>6% of maximal concentration) to decide whether node is ON or OFF.
- Compare with wild type pattern, if the same accept as a solution.

- 1 in every 46 parameter combinations lead to wild type final patterns.

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

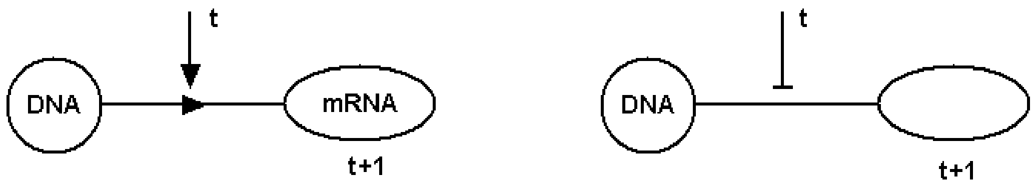
It is not the fine-tuning of the kinetic rates but the overall network topology what matters.

Second reconstruction of the segment polarity gene interaction network

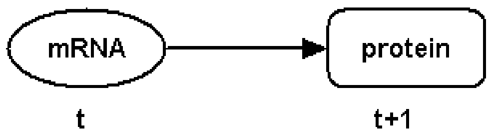


Qualitative (Boolean) model

- Transcripts and proteins are either ON (1) or OFF(0).
- Transcription depends on transcription factors; inhibitors are dominant.



- Translation depends on the presence of the transcript.



- Transcripts and most proteins decay if not produced.
- Synchronous update: transcription, translation, mRNA/protein decay on the same timescale, protein binding faster

R. Albert, H. G. Othmer, Journ. Theor. Bio. 223, 1 (2003).

- Asynchronous update & hybrid model: post-translational processes faster than pre-translational

M. Chaves, R. Albert, E. Sontag Journ. Theor. Bio. 235, 431 (2005).

M. Chaves, E. Sontag R. Albert, IEE Proc. Syst. Bio. 153, 154 (2006).

Updating rules

$$hh_i^* = EN_i \text{ and not } CIR_i$$

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

$$ptc_i^* = CIA_i \text{ and not } EN_i \text{ and not } CIR_i$$

transcription

$$ci_i^* = \text{not } EN_i$$

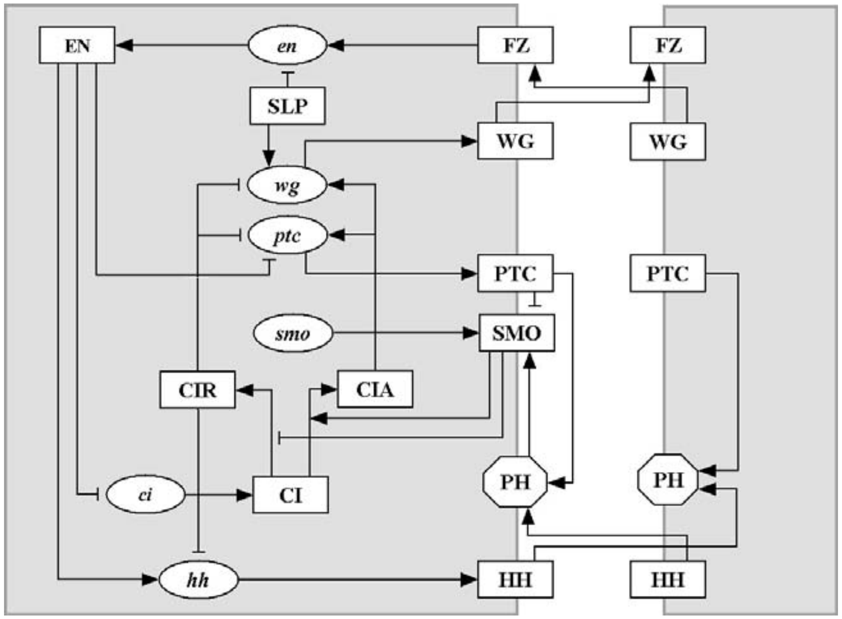
$$EN_i^* = en_i$$

$$WG_i^* = wg_i$$

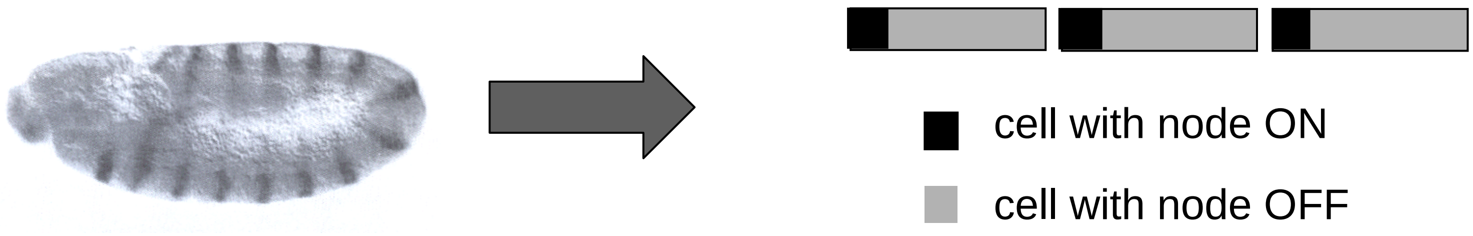
$$CI_i^* = ci_i$$

$$HH_i^* = hh_i$$

translation



State: 14 mRNAs and proteins x 4 cells



Initial state - updating rules– steady state

$$hh_i^* = hh_i(T_{hh}^k) = EN_i \text{ and not } CIR_i$$

Synchronous: $T_{hh}^k = k$

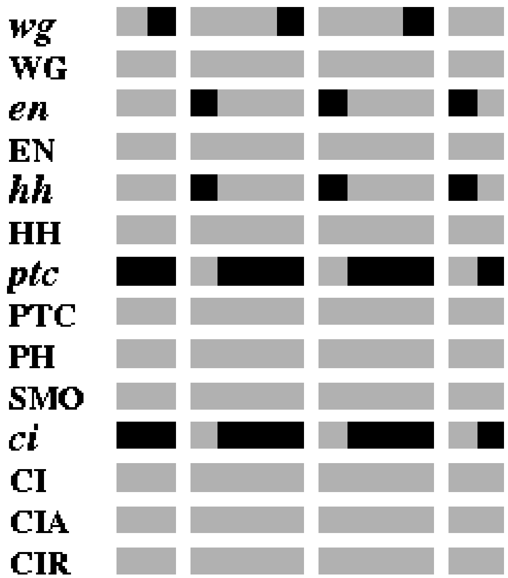
Asynchronous: $T_{hh}^k = k\gamma_{hh}$

Hybrid: $$\frac{d\,\hat{hh}_i}{dt} = \alpha_{hh}\left(EN_i \text{ and not } CIR_i - \hat{hh}_i \right)$$

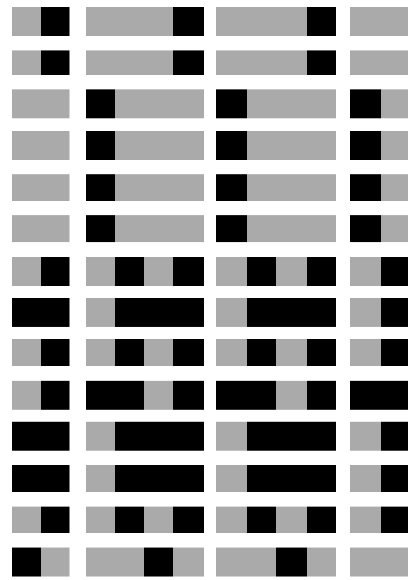
The steady state repertoire is independent of durations.

Start with the synchronous model, then explore whether conclusions change by asynchronicity.

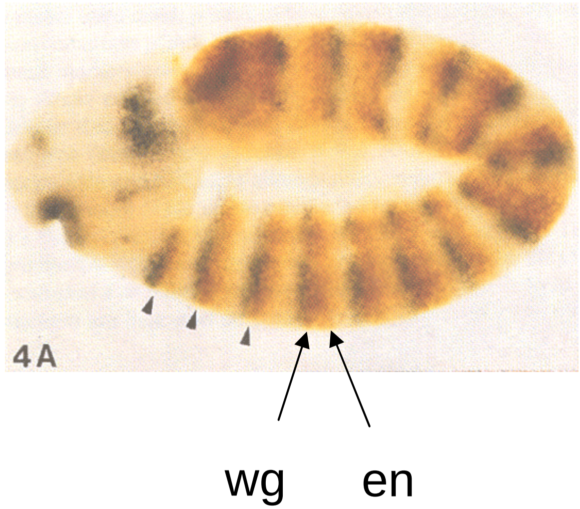
The model reproduces the wild type steady state

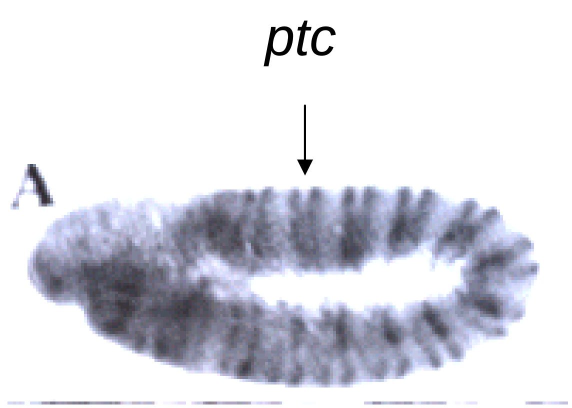


wt initial state



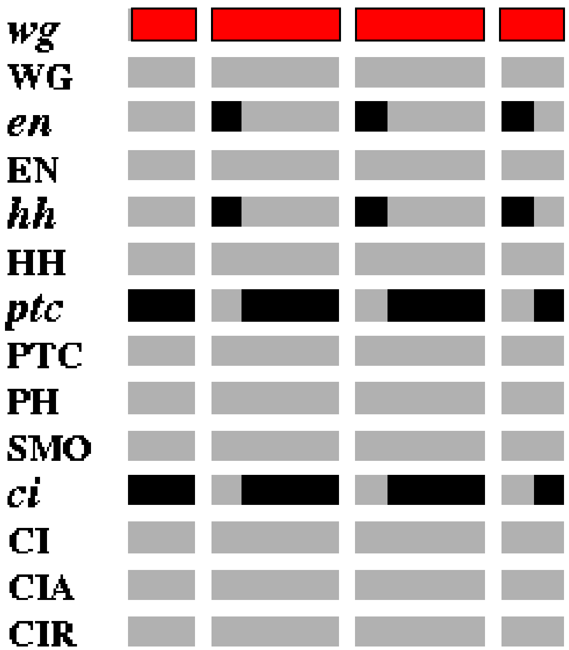
steady state



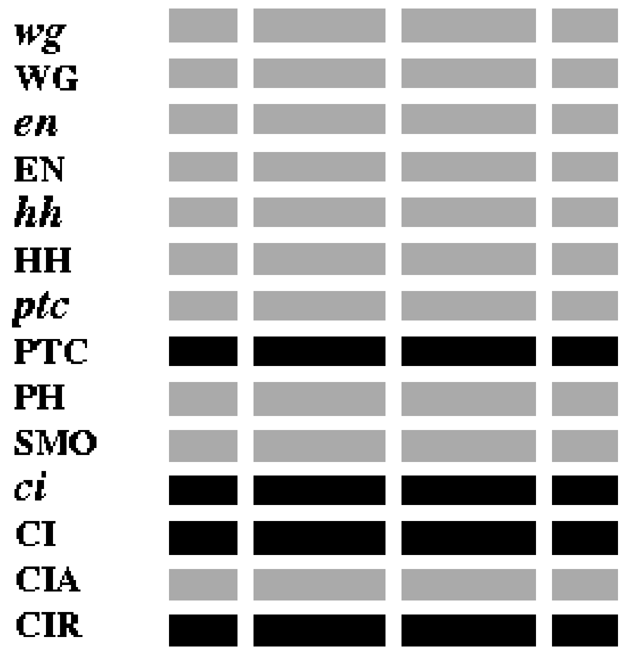


The interaction network and the net effect of the interactions (with reasonable assumptions on timing) is enough to capture the functioning of the network.

wg, en or *hh* mutations are lethal



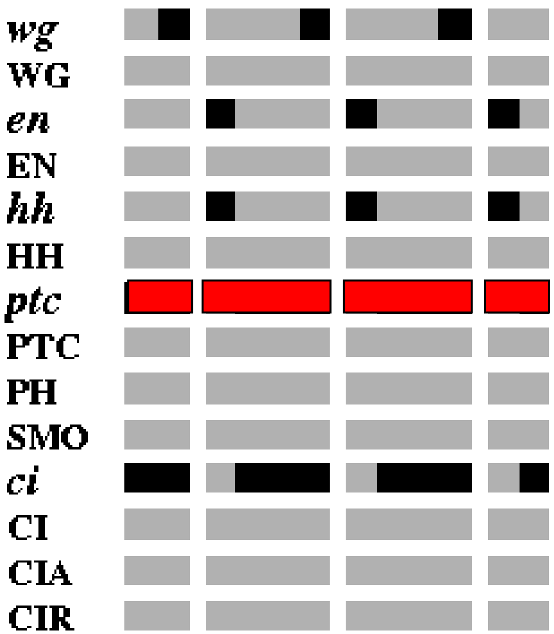
initial state



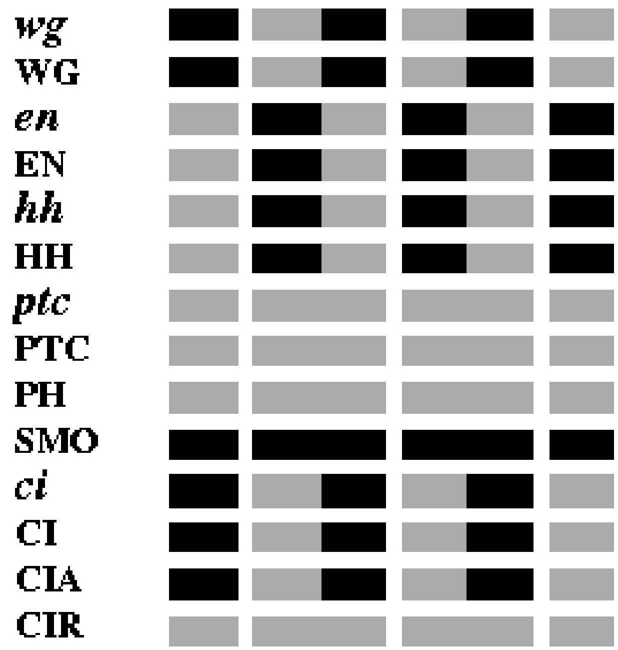
final state

No *wg*, *en* and *hh* stripes, no segmentation, regardless of initial state or interaction durations.

ptc mutation broadens the stripes



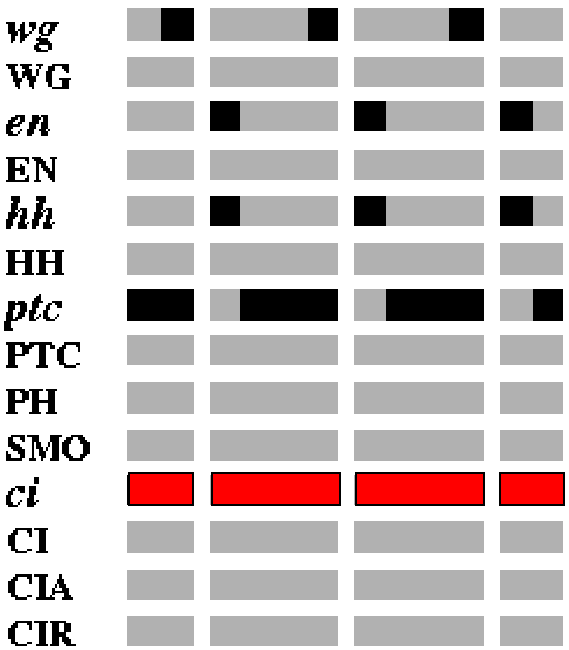
initial state



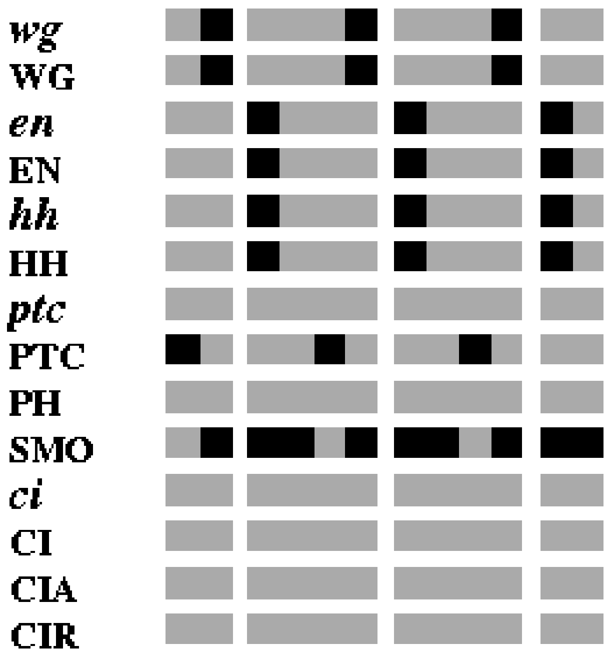
final state

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

ci mutation can preserve the prepattern



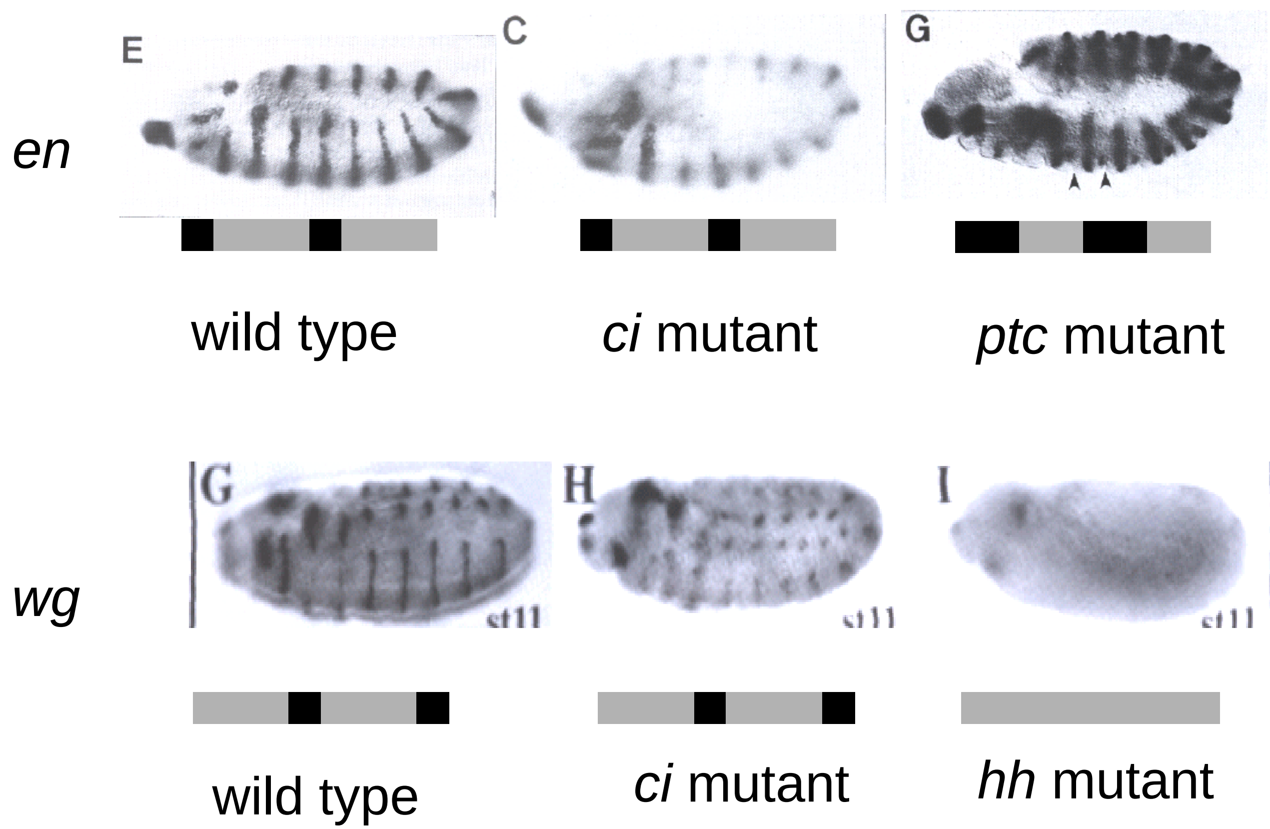
initial state



final state

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

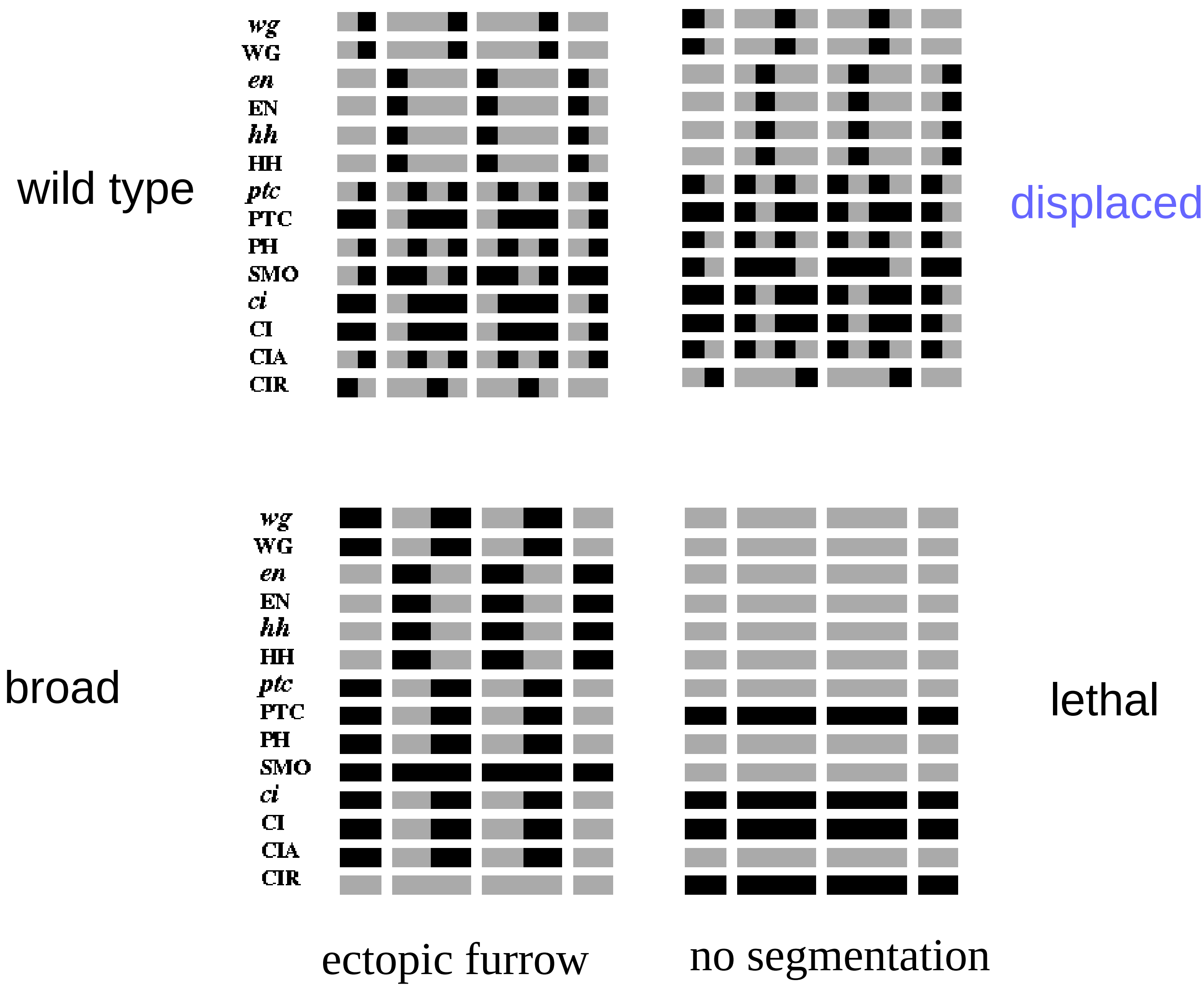
Model correctly reproduces experimental results on knock-out mutants



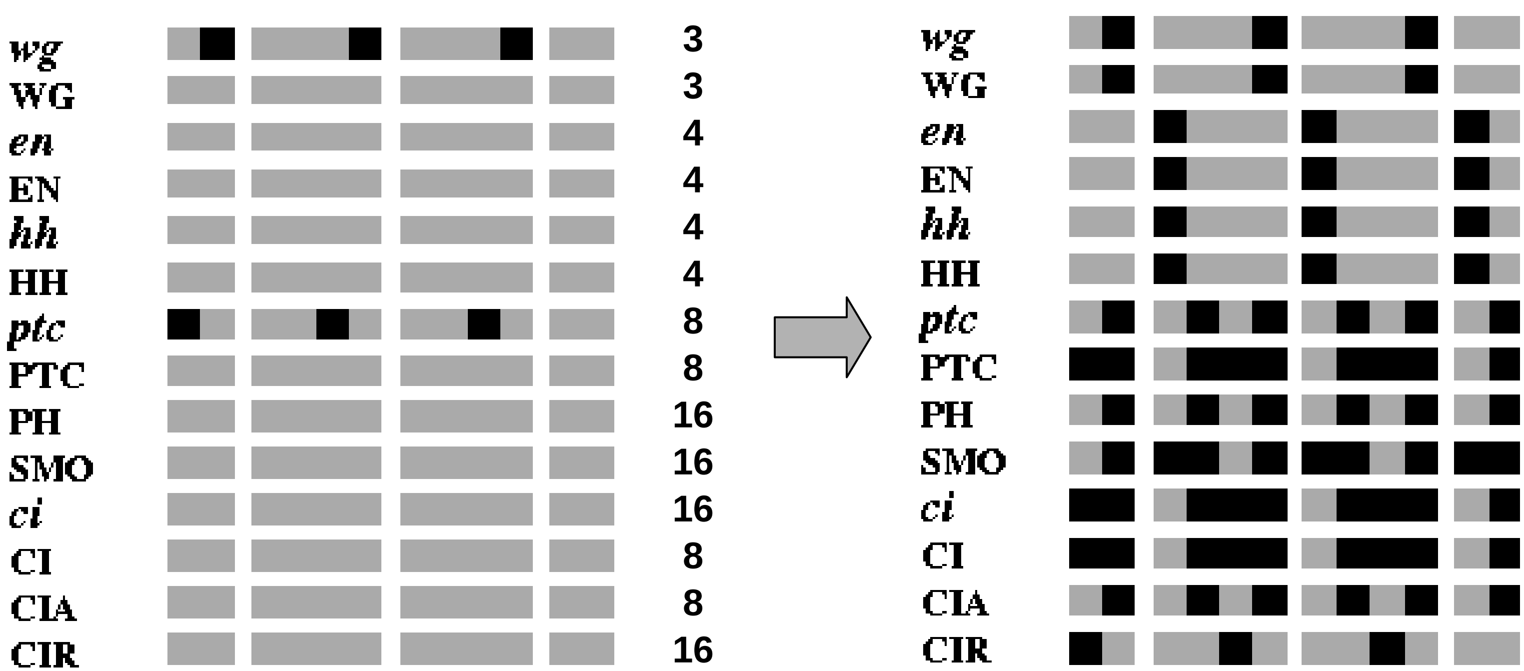
Tabata, Eaton, Kornberg, Genes & Development 6, 2635 (1992)

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

Dynamic repertoire: four steady states



How many initial states lead to the wild type steady state (in the synchronous model) ?



minimal prepattern steady state

Total number of wild-type inducing prepatterns: $\mathbf{6 \times 10^{11} = 8 \times 10^{-6}} N_i$

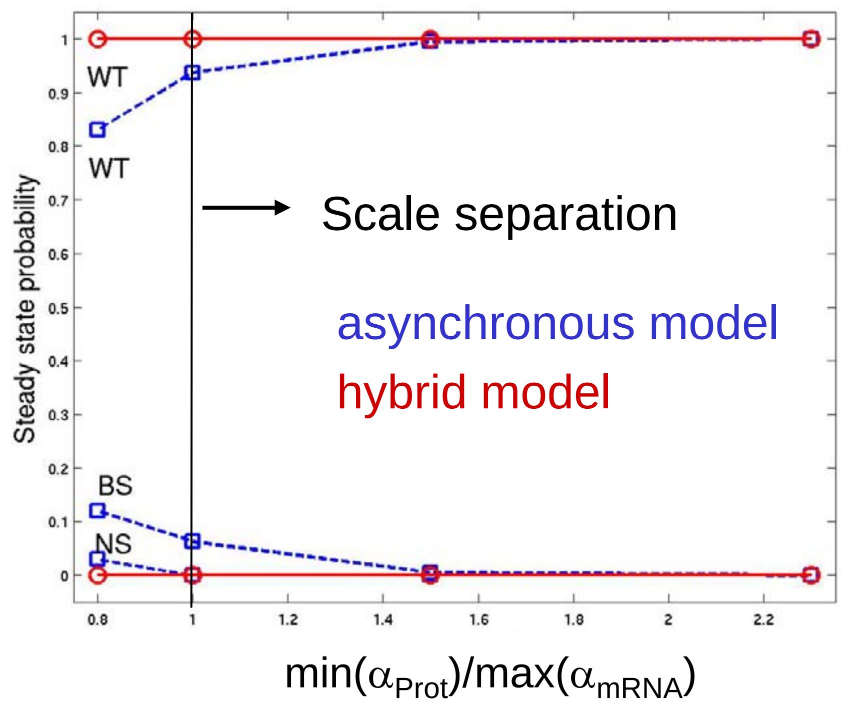
Divergence from wild-type development

Asynchronous model: each node updated at multiples of γ_i , γ_i chosen from a uniform distribution on (0, 2) or a subset thereof.

Hybrid model: Individual activation threshold $\theta_i \in (0,1)$,
individual synthesis/ decay rate α_i, $\alpha^{-1} = \gamma$

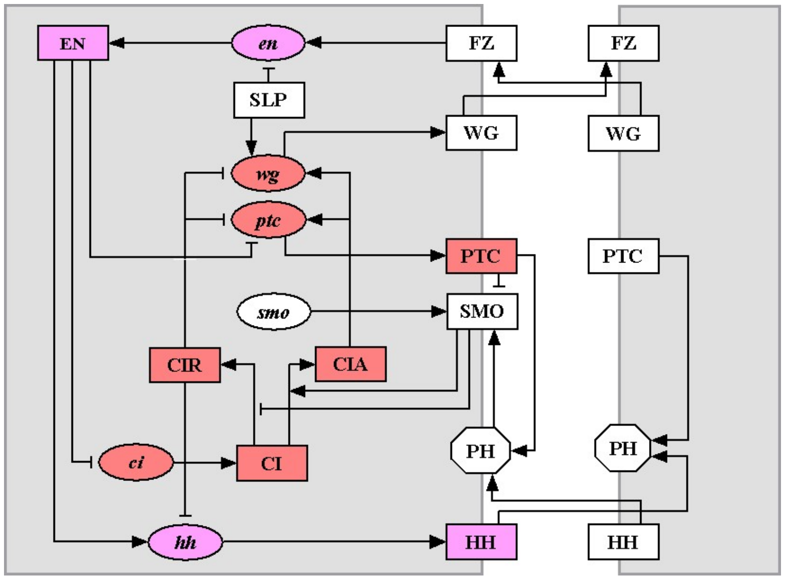
Assume that mRNA synthesis/decay timescale is longer than protein timescale, $\gamma_{mRNA} > \gamma_{prot}$, $\alpha_{mRNA} < \alpha_{prot}$

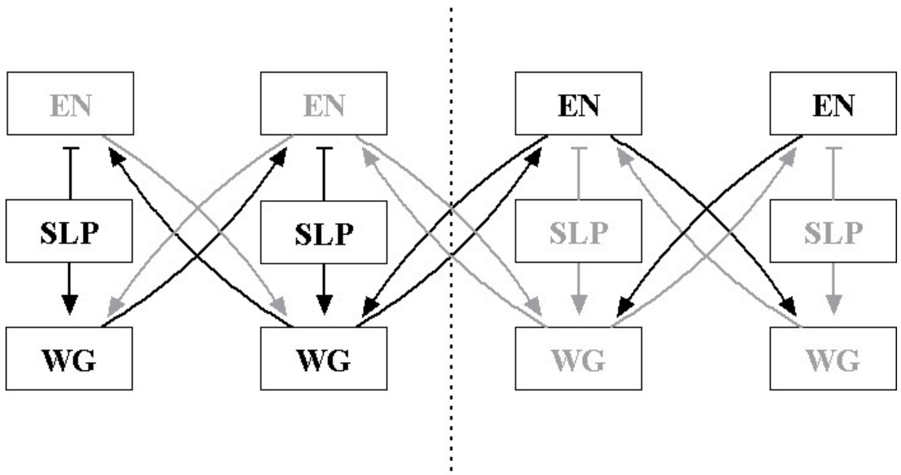
Then incidence of the WT steady state > 93%



Interplay between topology and function

- The network contains two activating clusters that inhibit each other in each cell, *en, hh* and *ci, wg, ptc*
- At the same time *en* and *wg* enforce each other in neighboring cells through the secreted proteins HH and WG





- SLP is a regulatory source that maintains asymmetry and limits *en* and *wg* to different halves of the parasegment.

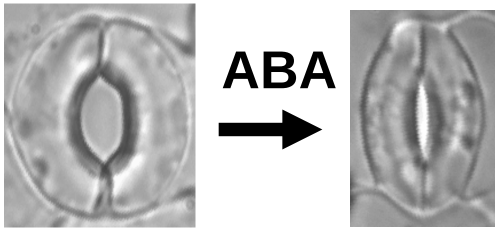
Modeling drought signaling in plants

Phenomenon: abscisic acid induced closure of plant stomata

Hypotheses: network inference from indirect information
protein activity is switch-like

Validation: reproduces known wild type and disrupted behavior.

Explored: disruptions
changes in initial conditions
changes in timing



Insight: variability in timing and initial conditions does not matter
65% of perturbations have no negative effect
identified critical perturbations

S. Li, S. Assmann and R. Albert, PLoS Biology 4, e312 (2006).

Stomata and guard cells

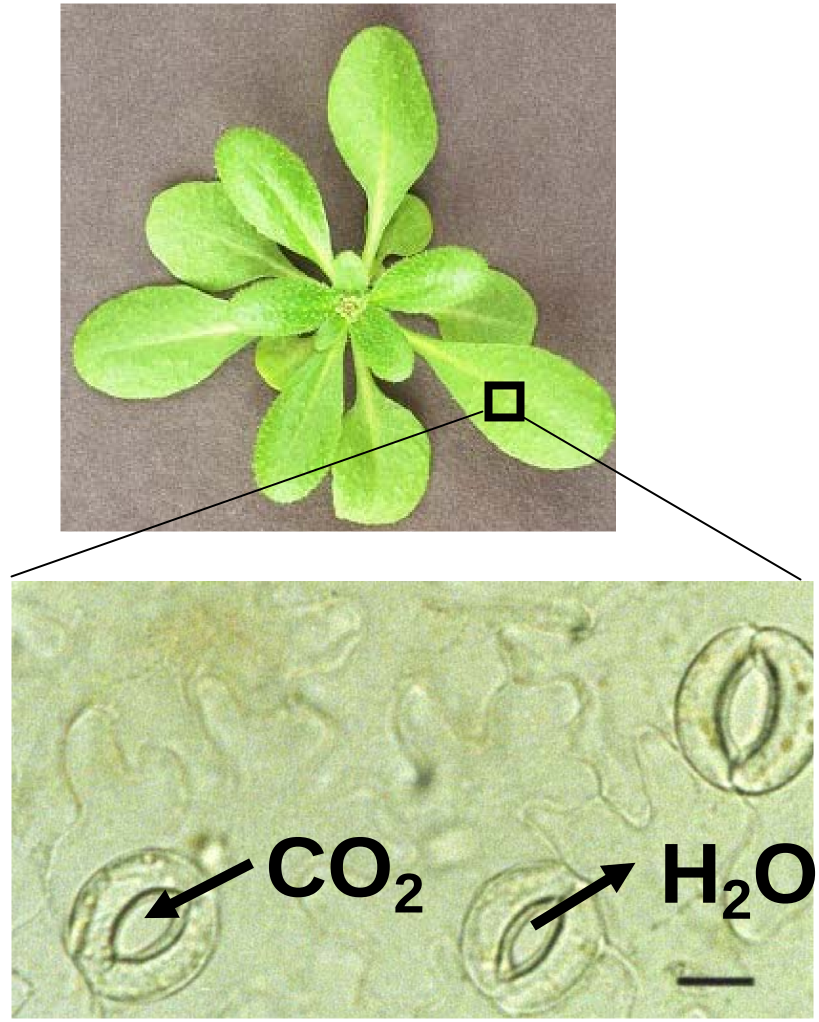
The exchange of O_2 and CO_2 in plants occurs through **stomata**.

90% of the water taken up by a plant is lost in transpiration.

Stomatal sizes are determined by the turgor (fullness) of the **guard cells**.

During drought conditions plants synthesize a hormone called abscisic acid (ABA) that initiates a **signal transduction network** to **close stomata**.

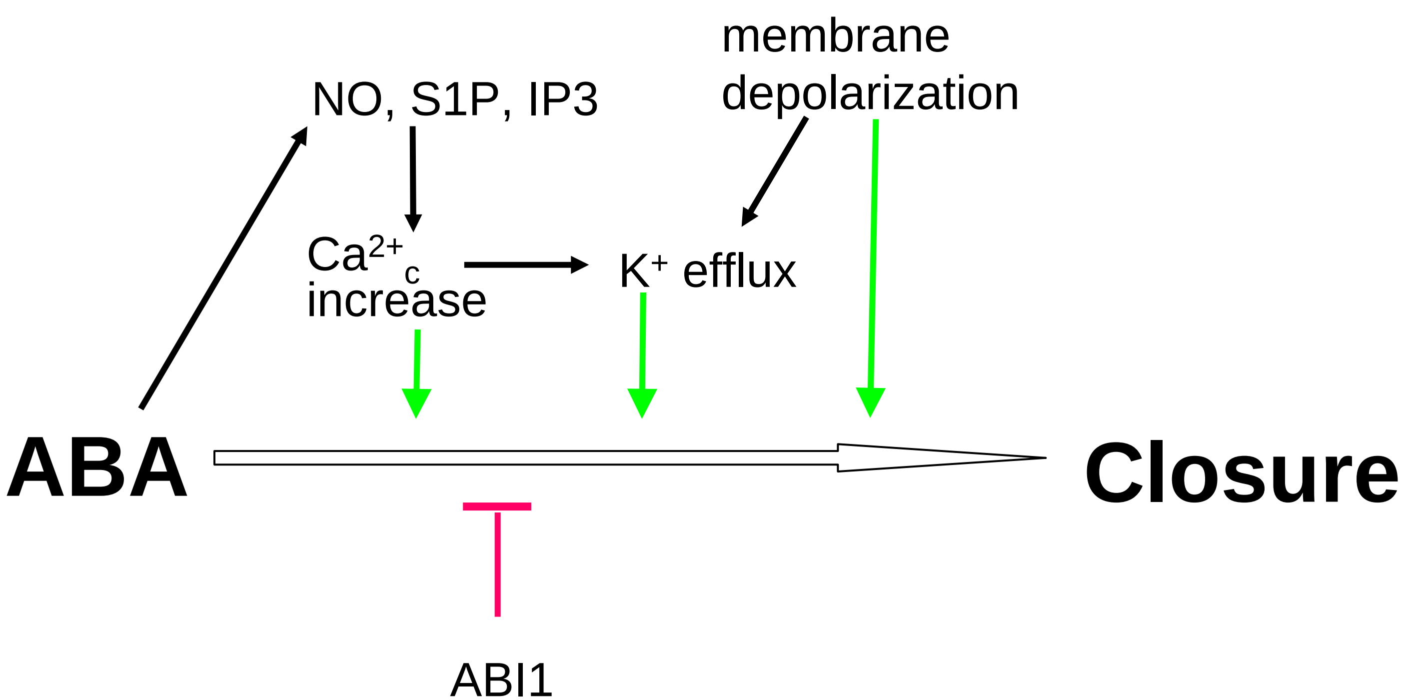
How is this crucial process being orchestrated, and how is its sensitivity and reliability maintained?



Experimental observations mainly indirect

- Genetic & pharmacological perturbations of putative mediators
  Compare input - output (ABA- aperture change) relationships
  in normal and perturbed plants

- Scarce biochemical evidence for interaction



Network construction from indirect evidence

- nodes: all proteins, molecules, ion channels implicated in the process
- compress biological information into activation or inhibition
- hypothesis: indirect causal relationships and processes correspond to paths ABA →→ ion flow, ABA →→ Sph kinase activity
- activating or inhibiting effects on processes represented as intersection of two paths SphK →→(ABA →→ closure)

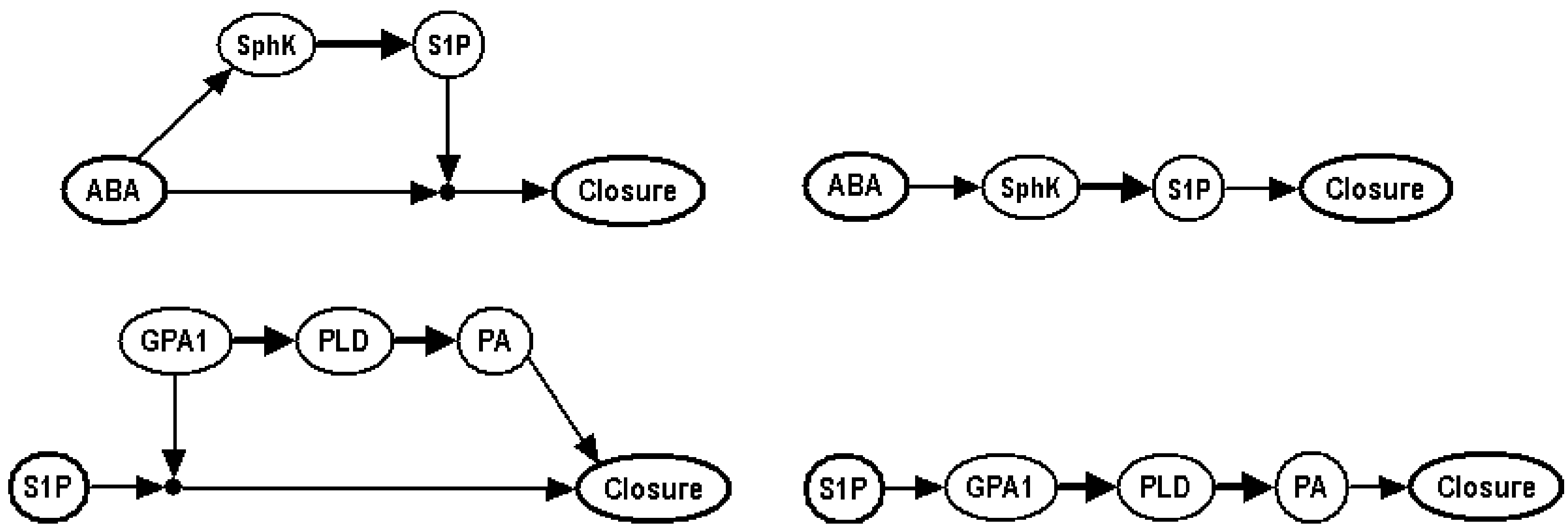
Node A	**interaction**	**Node/Process B**	**species**
ABA	promotes	SphK	*Arabidopsis*
PLC	promotes	ABA $\rightarrow$ closure	*Commelina communis*
SphK	partially promotes	ABA $\rightarrow$ AnionEM	*Arabidopsis*

Need to determine the closest regulator and target of each node

Network reduction

Find the most parsimonious (least redundant) network that incorporates all nodes and known processes.

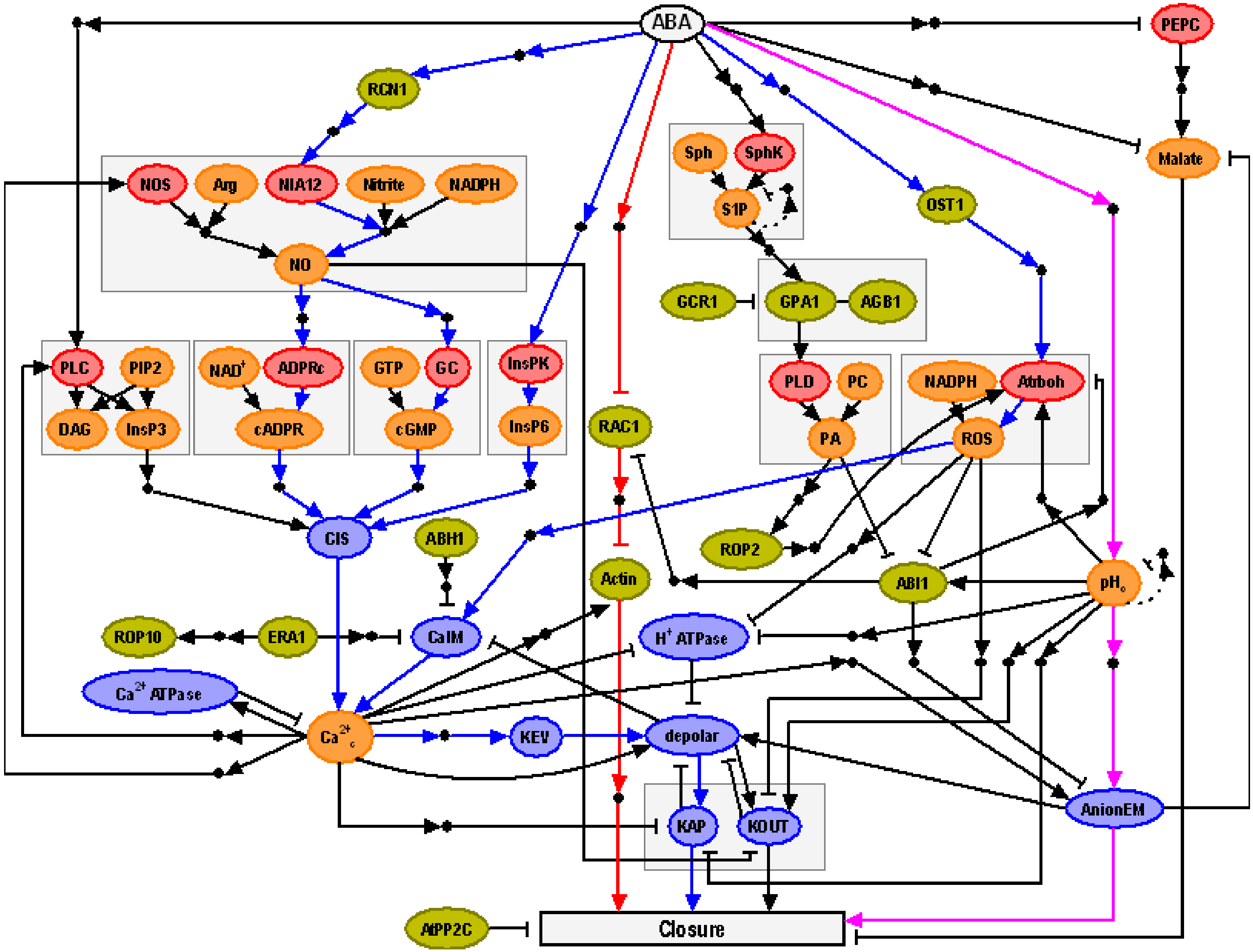
- Introduce intermediary nodes
- Contract intermediary nodes
- Review and revise



General algorithm - Bhaskar Dasgupta

binary transitive reduction with critical edges, pseudo-vertex collapse

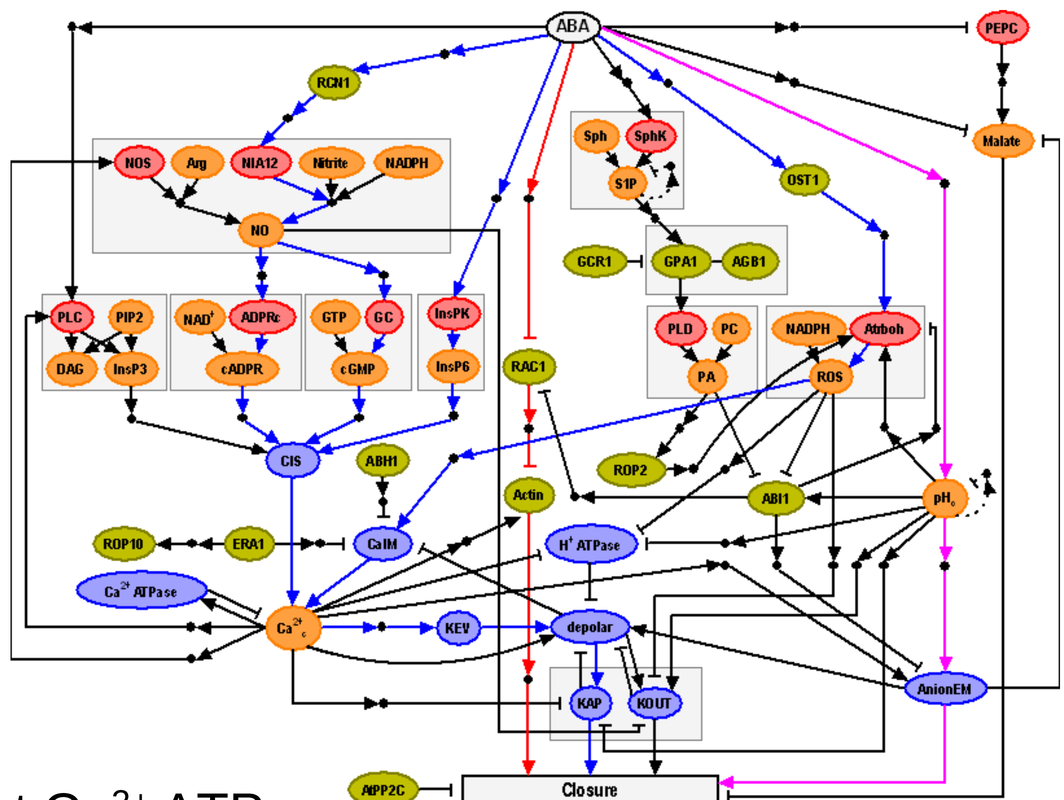
R. Albert, B. DasGupta et al, Journ. Comp Biology (in the press, 2007).



• intermediary nodes, enzymes, signal transduction proteins, transport, small molecules

What additional information is essential and available for describing the flow of information during this process?

Inhibitors – block the signal
Conditional activation
Independent activation



Boolean framework:
NOT, AND, OR
States: 1 = active/high/open,
0 = inactive/low/closed

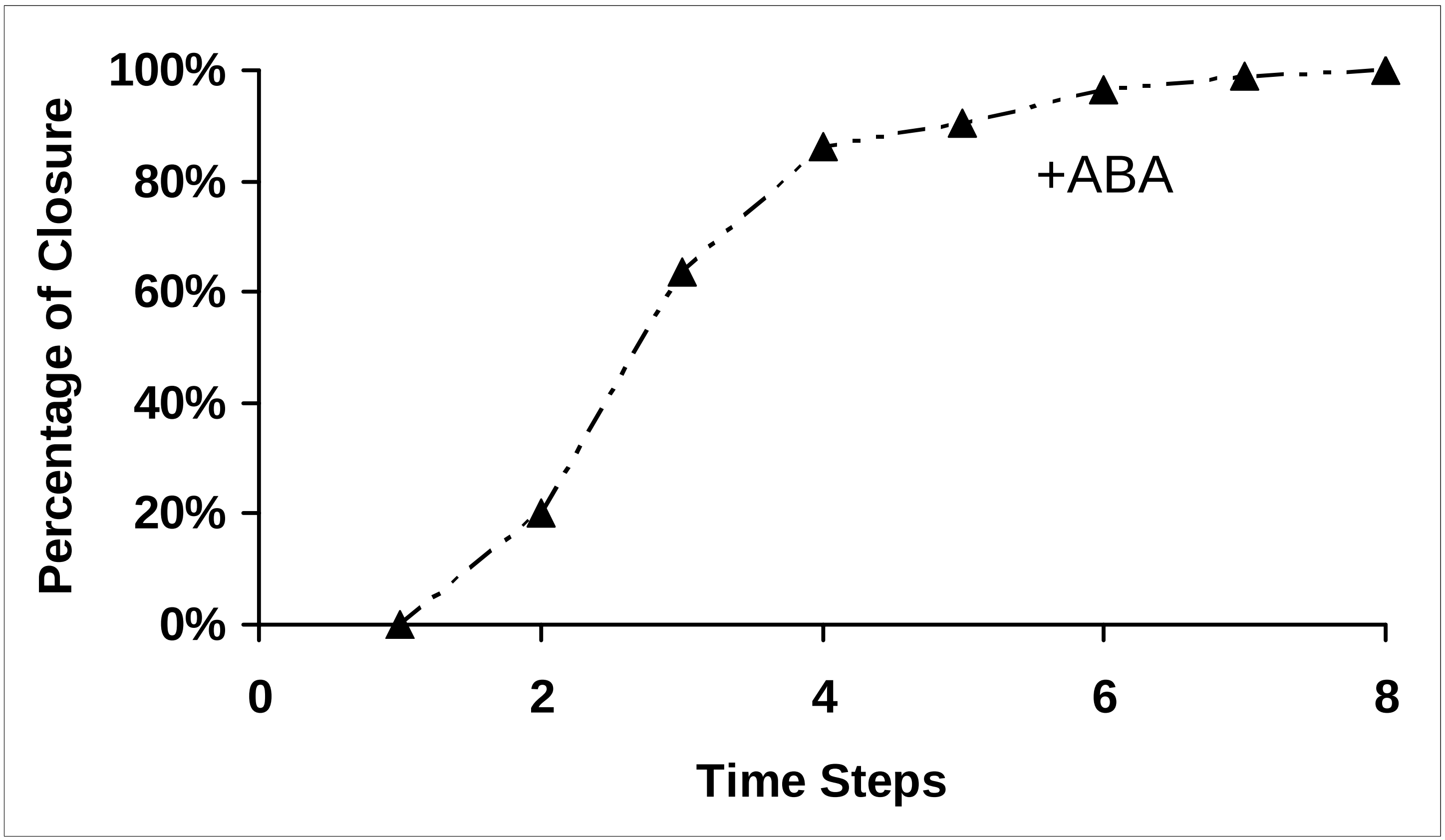
$Ca^{2+}{}_c$* = (CalM or CIS) and not Ca^{2+} ATPase

Closure* = (KOUT or KAP) and AnionEM and Actin and not Malate

Randomly selected initial condition (except for ABA), random timing

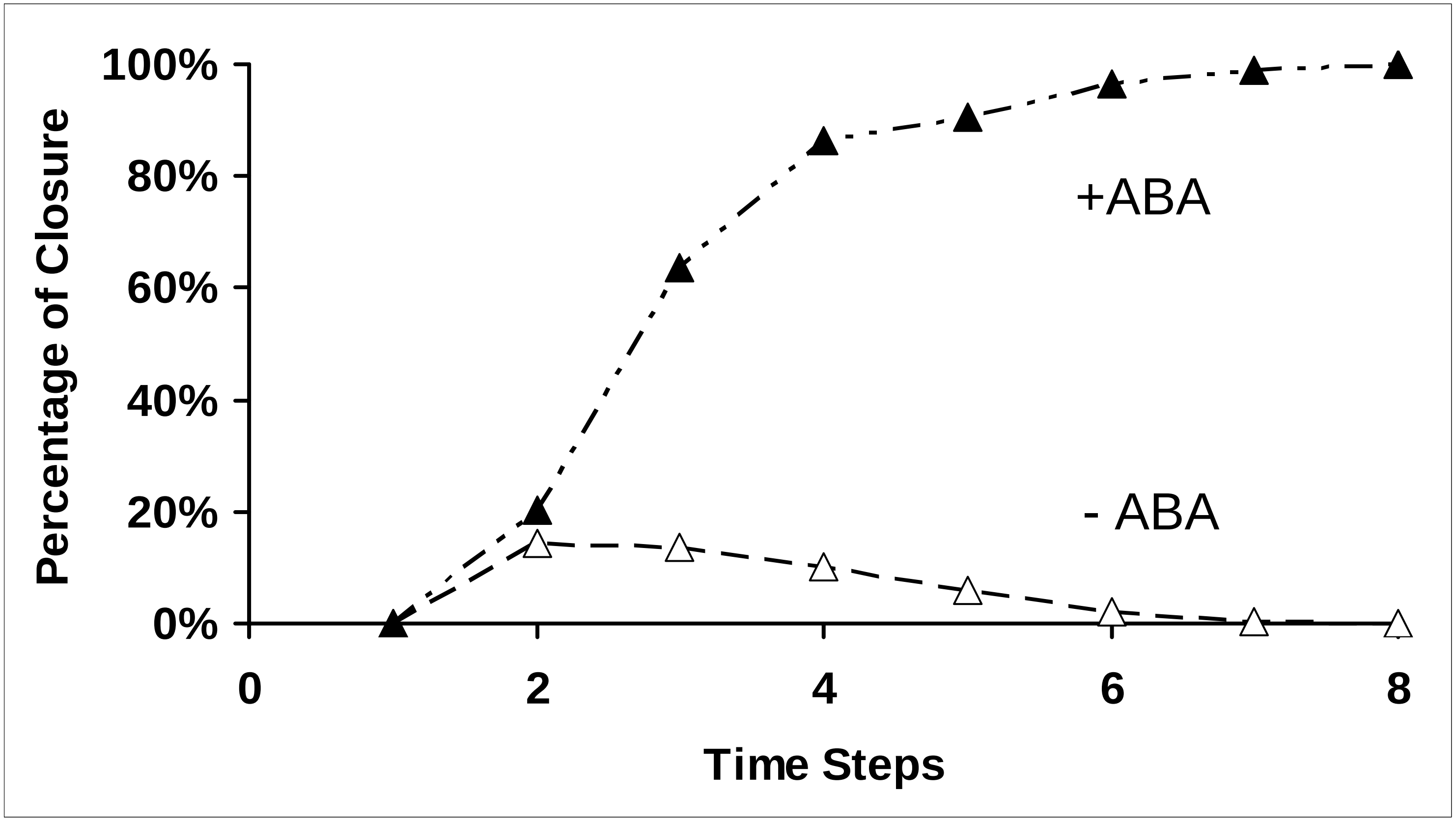
Results: normal behavior, +ABA

- Percentage of in silico stomata that are closed after ABA treatment reaches 100% by 8 timesteps.

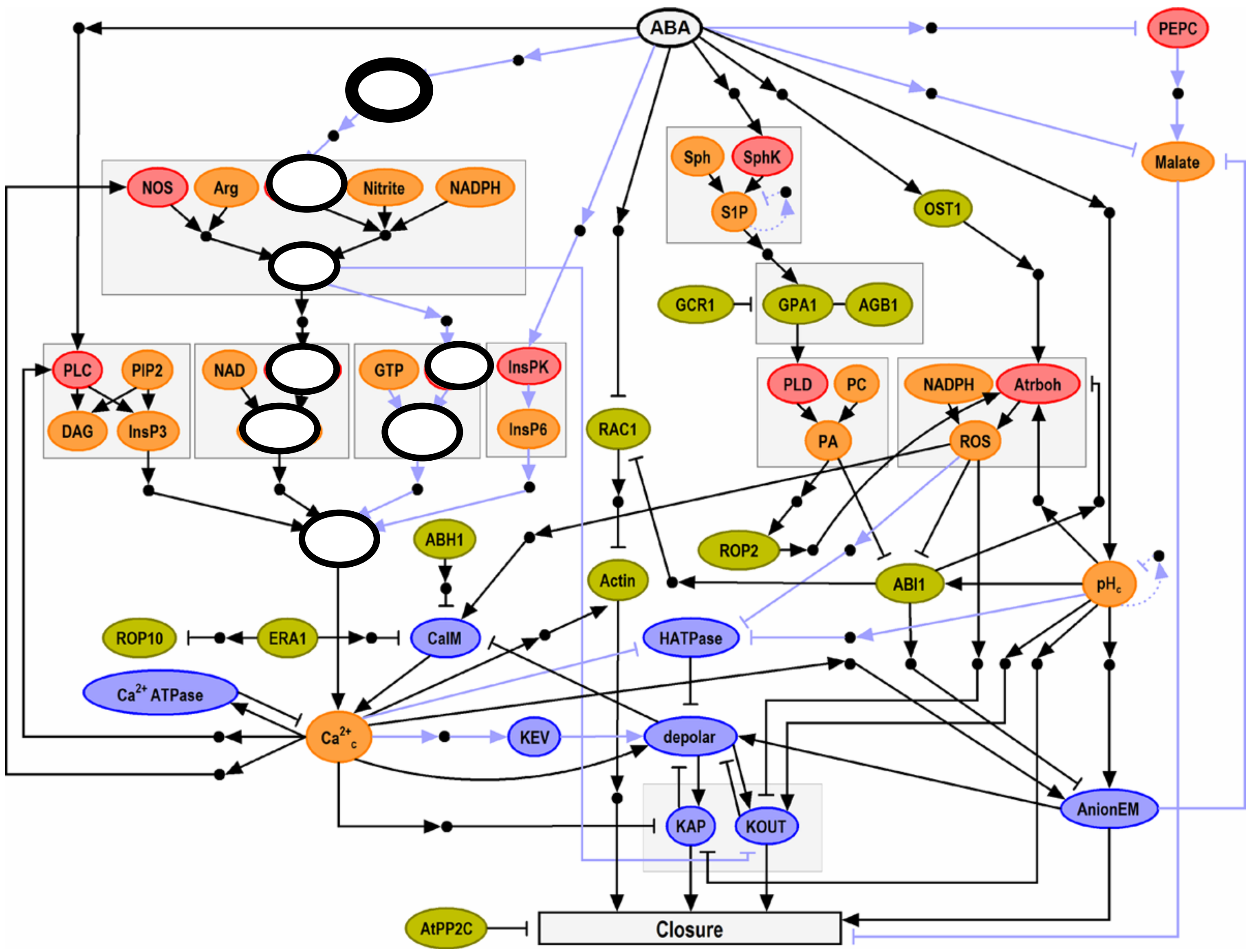


Results: normal behavior, -ABA

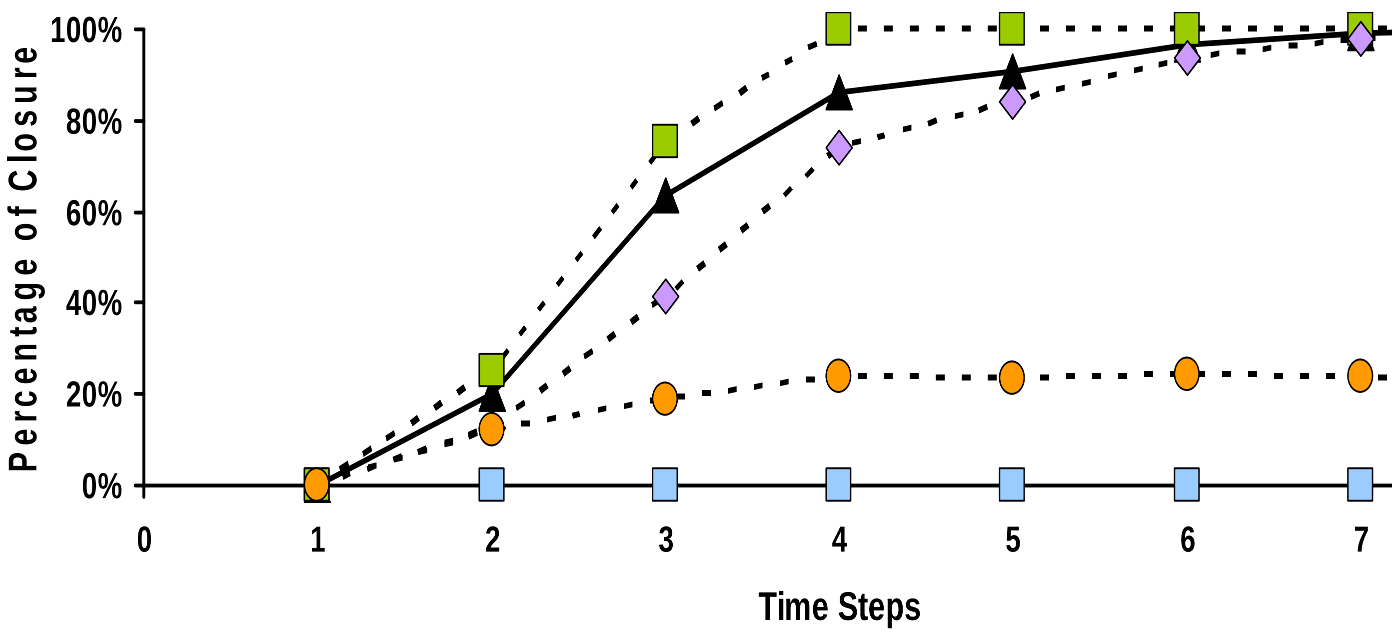
- Percentage of in silico stomata that are closed in the absence of ABA treatment reaches 0% by 6 timesteps.



Simulated knockouts



Results: dynamic effects of knockouts



▲ Normal response of simulated population to ABA stimulus.

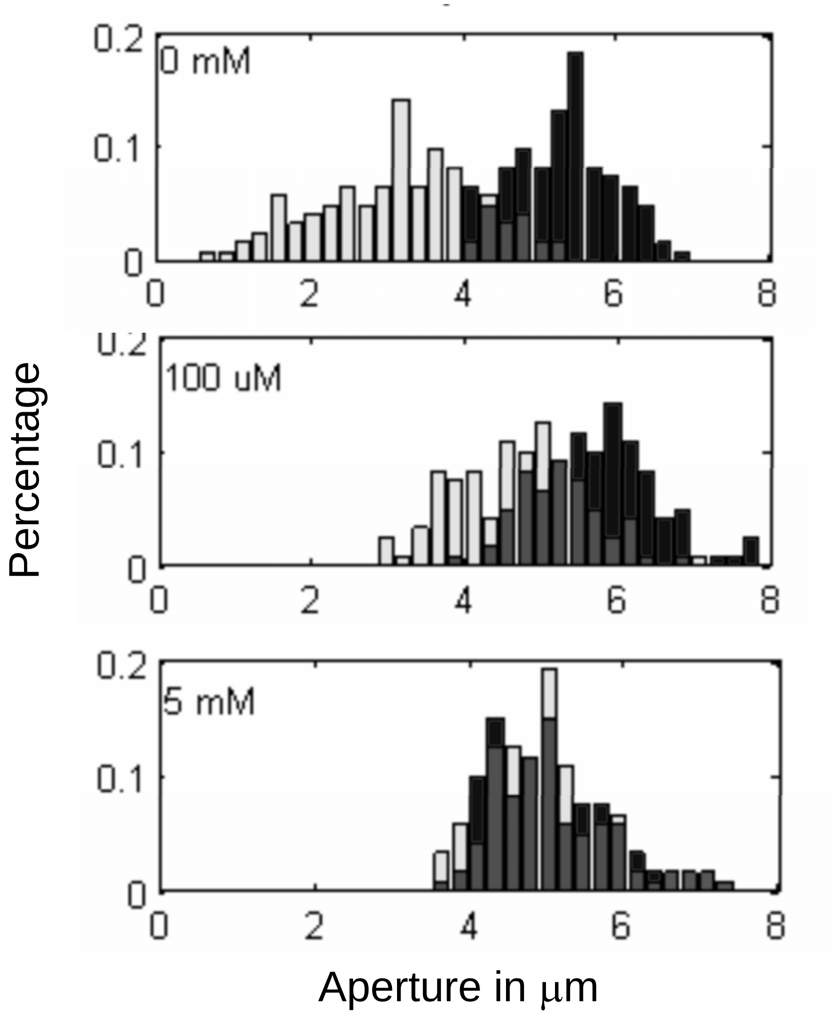
■ ABI1 knockout mutants have faster response at the population level.

◆ Perturbations in Ca^{2+} lead to slower response at the population level.

● Perturbations in pH lead to decreased sensitivity at the population level.

■ Perturbations in anion flow lead to insensitivity.

Experimental validation of Ca^{2+} and pH prediction



Normal: "open" and "closed" state distinguishable

Ca^{2+} disrupted: "open" and "closed" state distinguishable

pH_c disrupted: "open" and "closed" state indistinguishable

Further analysis: quantitative degree of closure, transients

Modeling pathogen-immune system interactions

Phenomenon: respiratory infection of mice by two bacterial strains

Hypotheses: discrete states, switch-like state changes
finite decay times

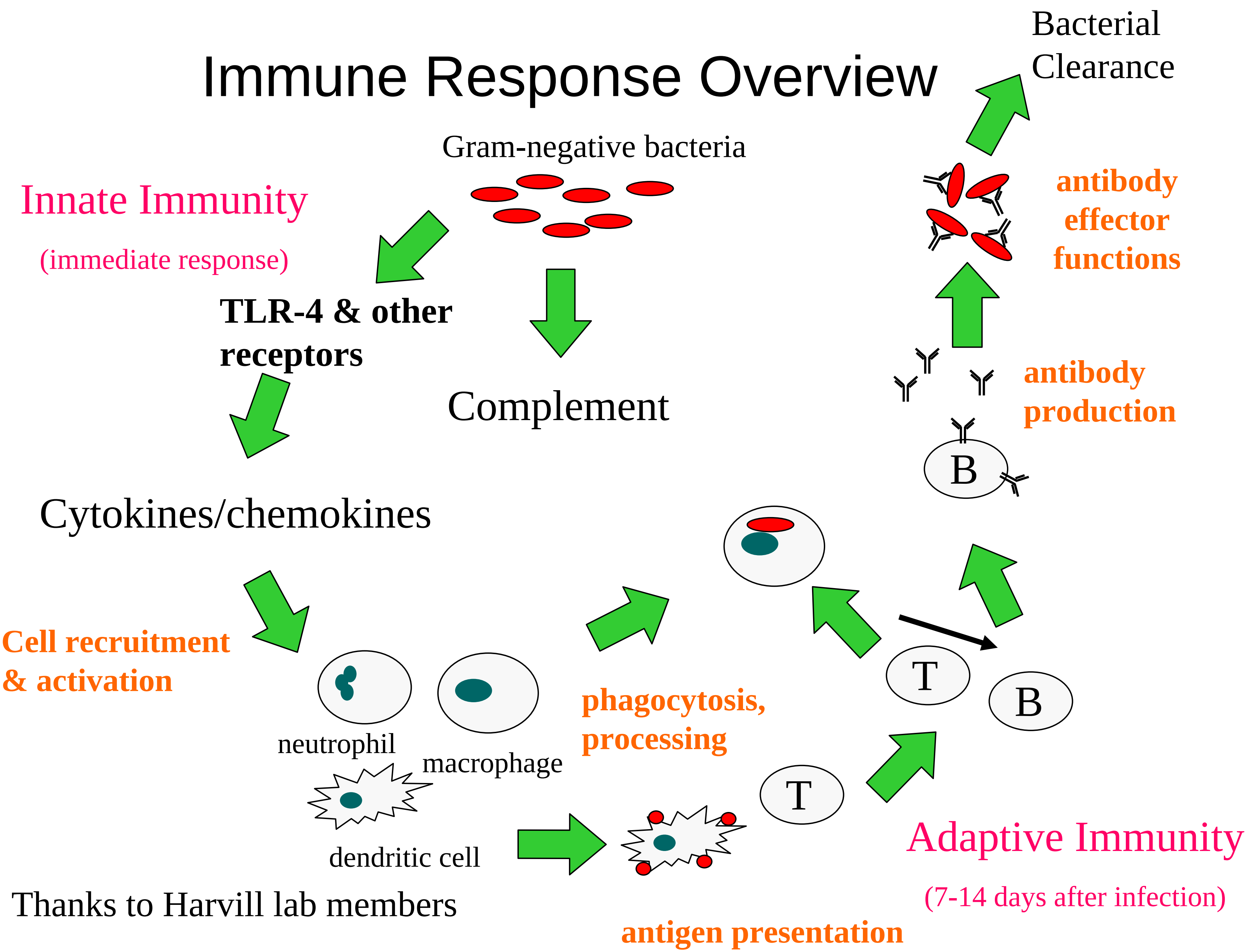
Validation: reproduces known wild type and disrupted infection timecourses.

Explored: disruptions in immune components or bacteria
changes in timing
different initial conditions
re-infections

Insights: discrete infection phases
relationships among cytokine timescales
differences among pathogens

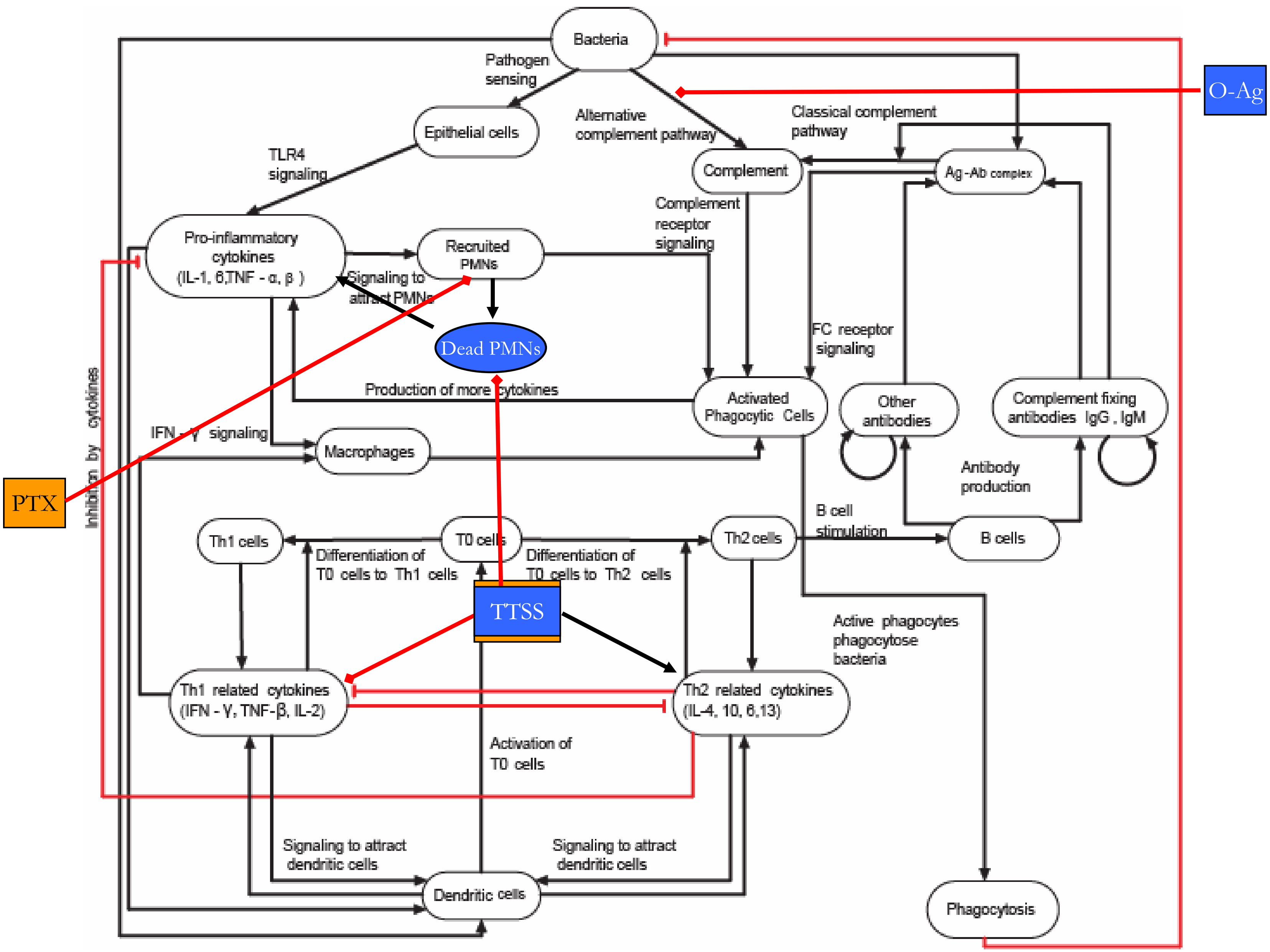
J. Thakar, M. Pilione,G. Kirimajeswara, E. T. Harvill, R. Albert,
PLoS Comp Biol. 3, e109 (2007).

Immune Response Overview



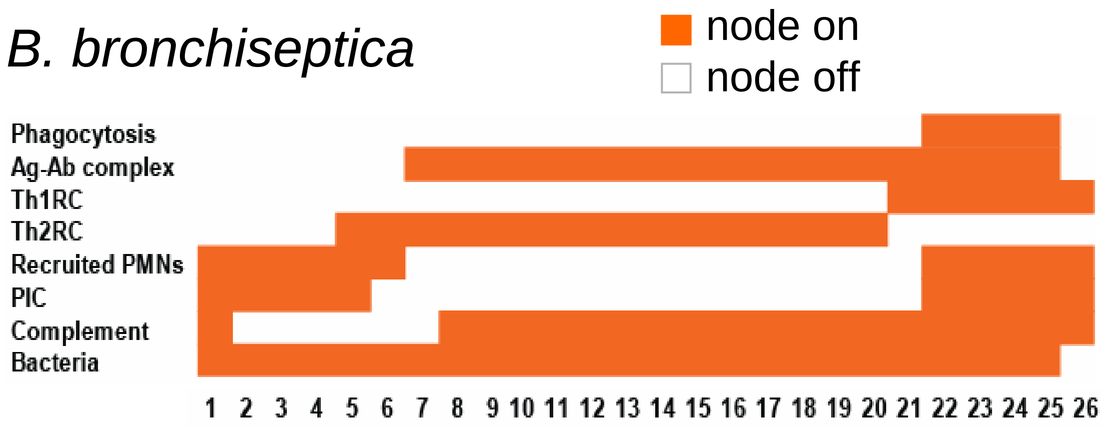
Thanks to Harvill lab members

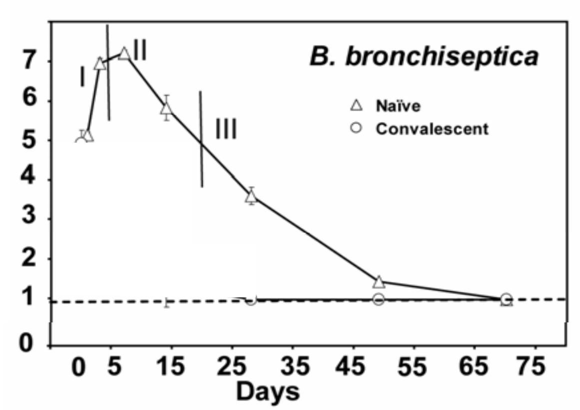
Interaction network of immune responses



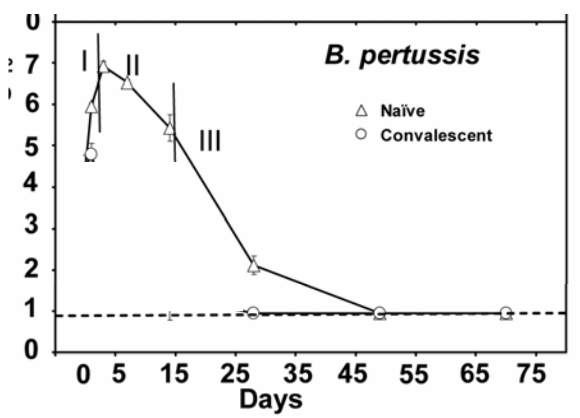
Infection timecourse in model and experiment

B. bronchiseptica



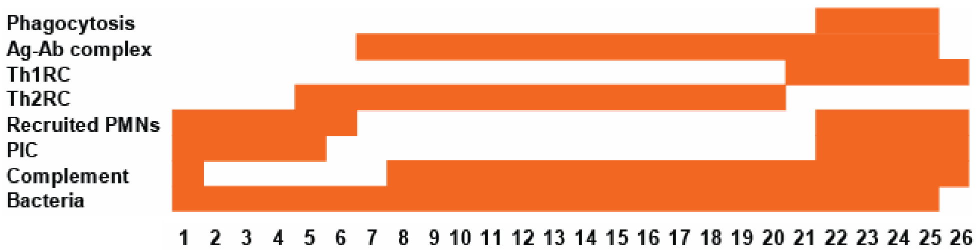


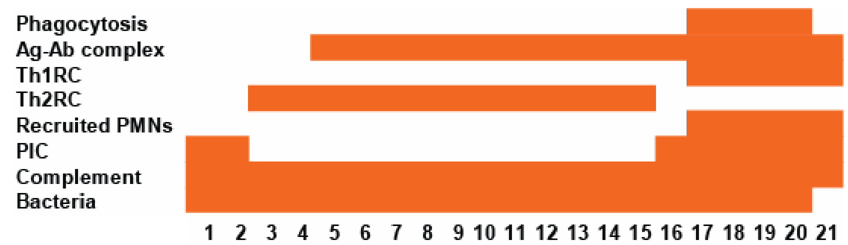
B. pertussis



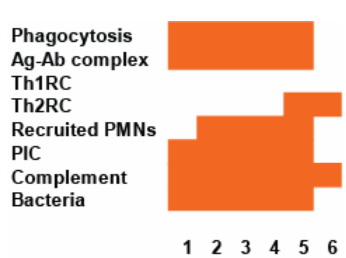
B. bronchiseptica *B. pertussis*

Wild Type

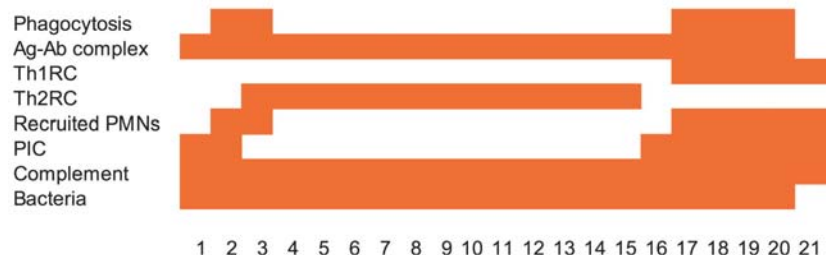


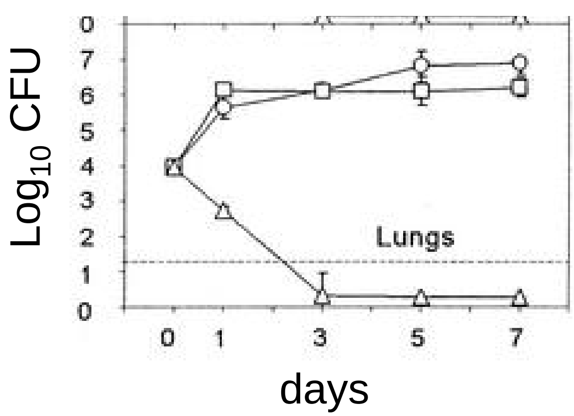


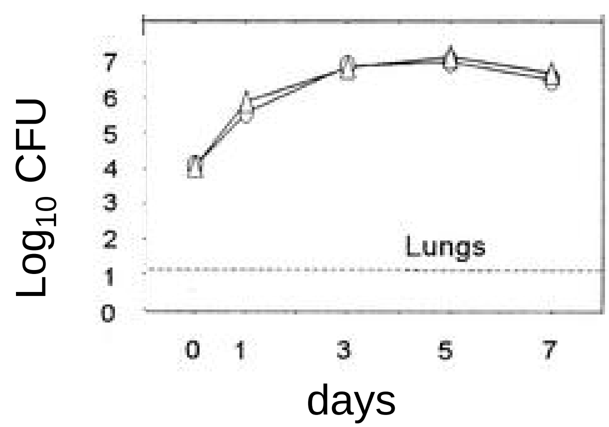
Antibody treatment



node on
node off



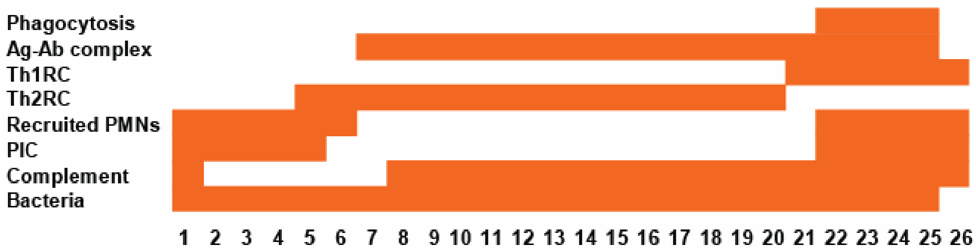


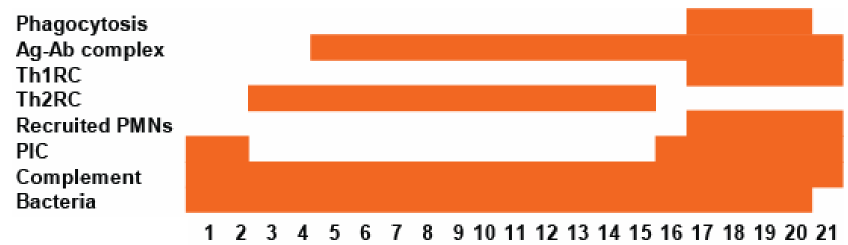


G. Kirimanjeswara et al. Infect Immun 71, 1719 (2003)

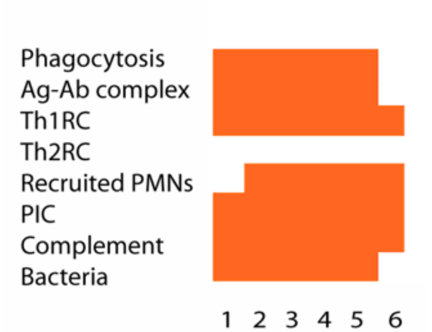
B. bronchiseptica *B. pertussis*

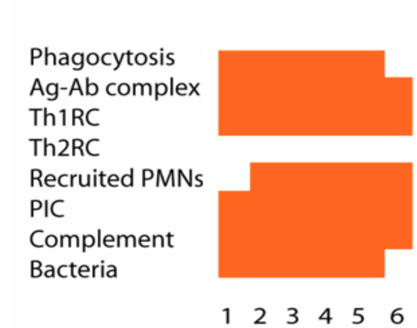
Wild Type

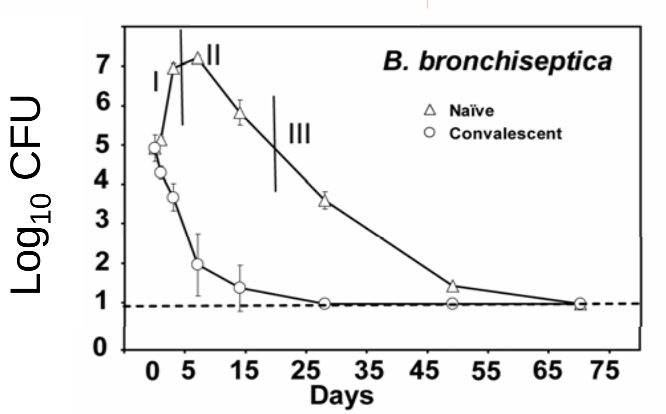


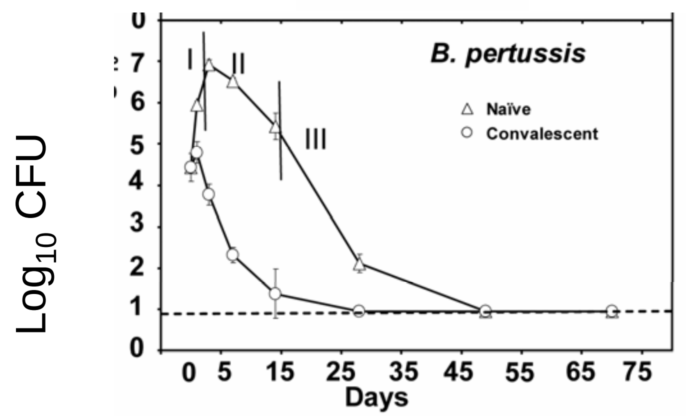


Re-infection of convalescent hosts









Conclusions

Network synthesis allows the logical organization of signaling components.

Network analysis and dynamic modeling:

- has predictive value
- has practical value for prioritization of experiments**.**
- allows discovery of new strategies
- will be useful for other incompletely known regulatory networks.

Methodology can be refined iteratively with experiments

Toward theory of biological circuits

- Robustness through redundancy & feedback
- Noise filtering through synergy
- Adaptability through cross-talk

Acknowledgements

Drosophila segmentation
Hans G. Othmer (Minnesota)
Madalena Chaves (INRIA)
Eduardo Sontag (Rutgers)

ABA signal transduction
Sarah M. Assmann
Song Li
Bhaskar Dasgupta (UIC)
Sema Kachalo (UIC)

Pathogen - host interactions
Juilee Thakar
Eric Harvill
Mylisa Pilione
Girish Kirimanjeswara

Alfred P. Sloan Foundation
National Science Foundation
US Dept. of Agriculture