Modeling and simulation of gene regulatory networks 5

Hidde de Jong
IBIS
INRIA Grenoble – Rhône-Alpes
Hidde.de-Jong@inria.fr

January16, 2013

INRIA Grenoble - Rhône-Alpes and IBIS

- IBIS: systems biology group at INRIA/Université Joseph Fourier/CNRS
  - Analysis of bacterial regulatory networks by means of models and experiments
  - Biologists, computer scientists, mathematicians, physicists, …

http://ibis.inrialpes.fr

Overview

1. Gene regulatory networks in bacteria

2. Deterministic modeling of gene regulatory networks

3. Qualitative modeling of gene regulatory networks

4. Stochastic modeling of gene regulatory networks

5. **Some current issues and perspectives**

   - **Global physiological effects on the dynamics of gene expression**

   - **Strategies for dealing with incomplete information: the case of Drosophila development**

Bacterial growth and adaptation

- Bacteria are geared towards growth and division

    E. coli cells have doubling times up to 20 min
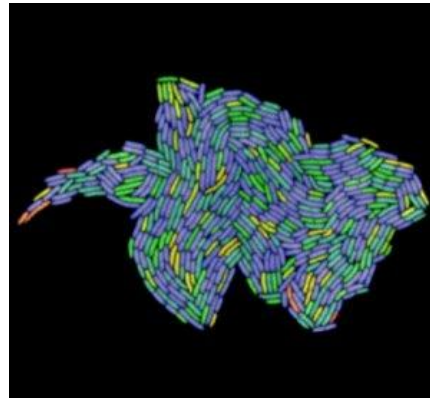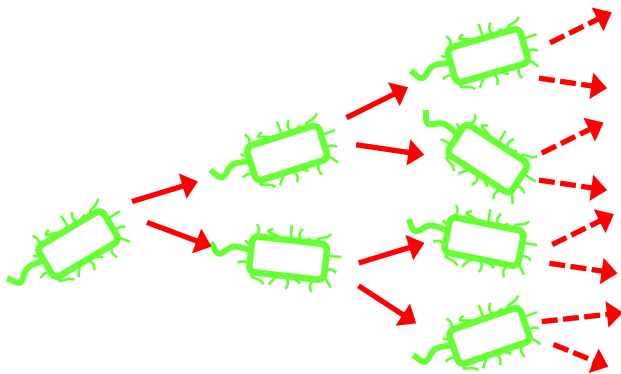

Stewart *et al.* (2005), *PLoS Biol.,* 3(2): e45

❖ External perturbations may cause adaptation of growth rate, and more generally, change physiology of bacterial cell

    Nutrient starvation, heat shock, osmotic stress, high population density,…

Growth transition and gene expression

- Genome-wide **reorganization of gene expression** following growth transitions in bacteria
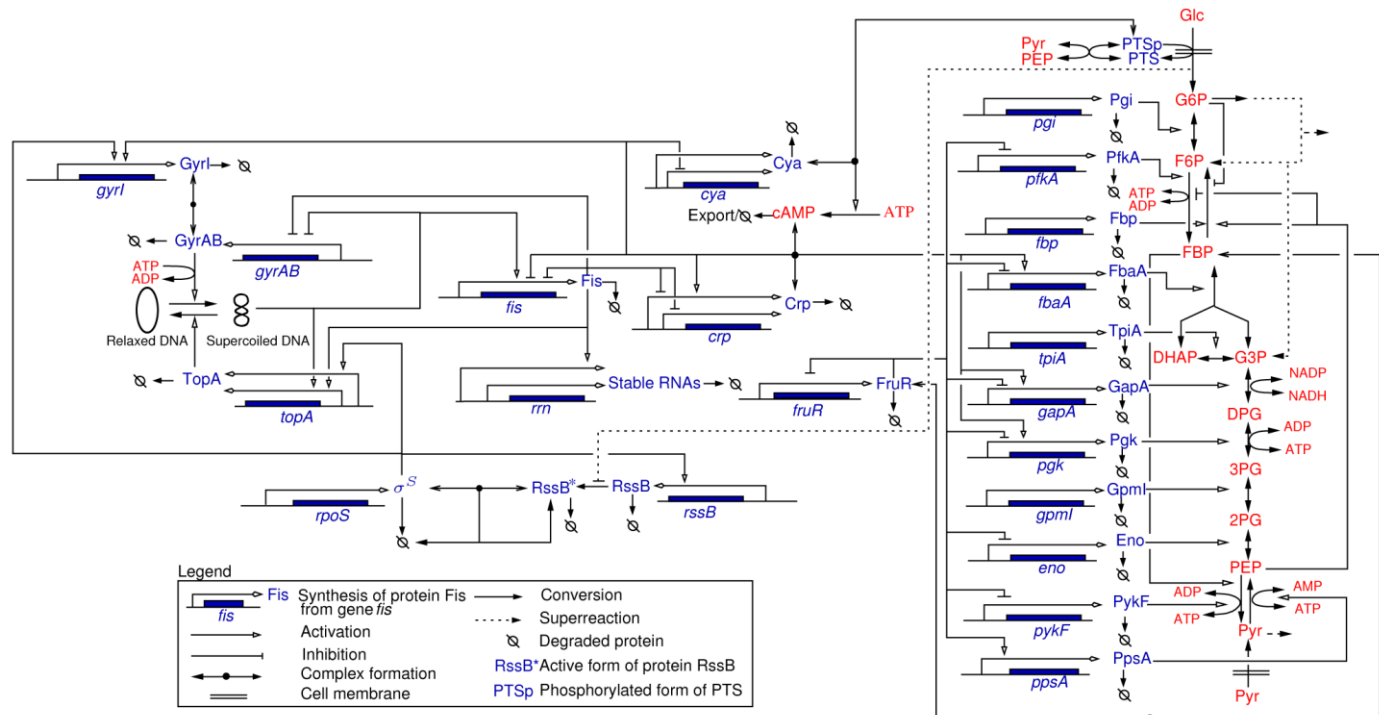
  Gene expression during glucose-lactose diauxie in *E. coli*
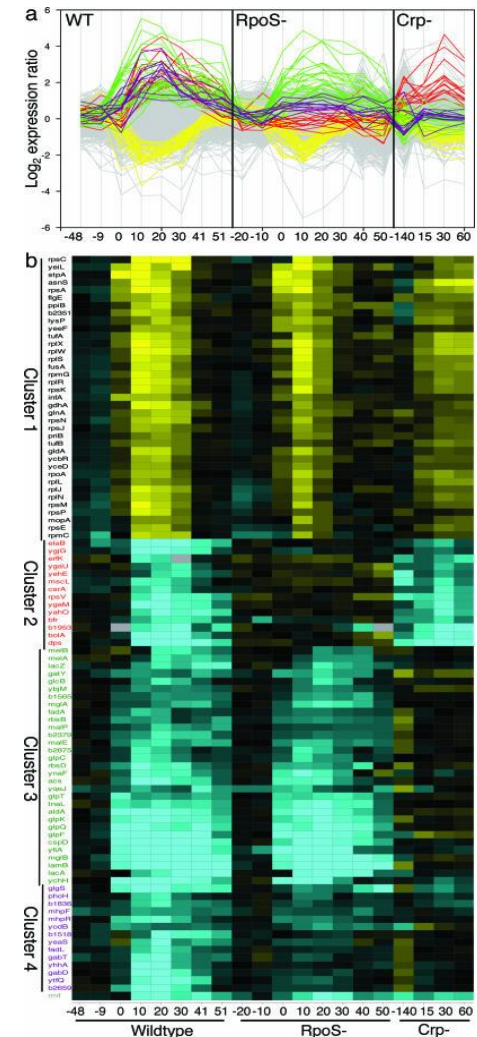
Traxler *et al.* (2006), *Proc. Natl. Acad. Sci. USA*, 103(7):2374–9

Growth transition and gene expression

- Adjustment of gene expression involves variety of specific regulators

    Transcription factors, small regulatory RNAs, …

- Complex regulatory networks control adaptive responses of cell

Baldazzi *et al.* (2010), *PLoS Comput. Biol.*, 6(6):e1000812

Growth transition and gene expression

- Adjustment of gene expression also involves global physiological effects

  Activity of transcriptional and translational machinery, size of metabolic pools, gene copy number, …

TABLE 2   Macromolecular composition of exponentially growing *E. coli* B/r as a function of growth rate at 37°C[a]

Parameter	Symbol	Units	τ, 100 μ, 0.6	τ, 60 μ, 1.0	τ, 40 μ, 1.5	τ, 30 μ, 2.0	τ, 24 μ, 2.5	Observed parameter(s)
Protein/mass	P_M	10^{17} aa/OD_{460}	6.5	5.8	5.2	5.1	5.0	P, M
RNA/mass	R_M	10^{16} nucl./OD_{460}	4.3	4.9	5.7	6.6	7.8	R, M
DNA/mass	G_M	10^{8} genomes/OD_{460}	18.3	12.4	9.3	8.0	7.6	G, M
Cell no./mass	C_M	10^{8} cells/OD_{460}	11.7	6.7	4.0	2.7	2.0	Cells/OD_{460}
$(P + R + G)/M$	PRD_M	μg/OD_{460}	149	137	129	131	136	
Protein/genome	P_G	10^{8} aa residues	3.5	4.7	5.6	6.3	6.6	P_M, G_M
RNA/genome	R_G	10^{7} nucl. residues	2.3	4.0	6.1	8.2	10.3	R_M, G_M
Origins/genome	O_G	Dimensionless	1.25	1.32	1.44	1.58	1.73	C
Protein/origin	P_O	10^{8} aa residues	2.8	3.6	3.9	4.0	3.8	P_G, O_G
Protein/cell	P_C	10^{8} aa residues	5.6	8.7	13.0	18.9	25.0	P_M, C_M
	P_C (μg)	μg/10^{9} cells	100	156	234	340	450	
RNA/cell	R_C	10^{7} nucl. residues	3.7	7.3	14.3	24.4	39.0	R_M, C_M
	R_C (μg)	μg/10^{9} cells	20	39	77	132	211	
DNA/cell	G_C	genome equiv./cell	1.6	1.8	2.3	3.0	3.8	C, D
	G_C (μg)	μg/10^{9} cells	7.6	9.0	11.3	14.4	18.3	
Mass/cell	M_C	OD_{460} units/10^{9} cells	0.85	1.49	2.5	3.7	5.0	C_M
	M_C (μg)	μg dry weight/10^{9} cells	148	258	433	641	865	μg/OD_{460}
Sum $P + R + G$	PRD_C	μg/10^{9} cells	127	204	322	486	679	P_C, R_C, G_C (in μg)
Origins/cell	O_C	no./cell	1.96	2.43	3.36	4.70	6.54	C, D
Termini/cell	T_C	no./cell	1.23	1.37	1.54	1.74	1.94	D
Replication forks/cell	F_C	no./cell	1.46	2.14	3.64	5.92	9.19	C, D
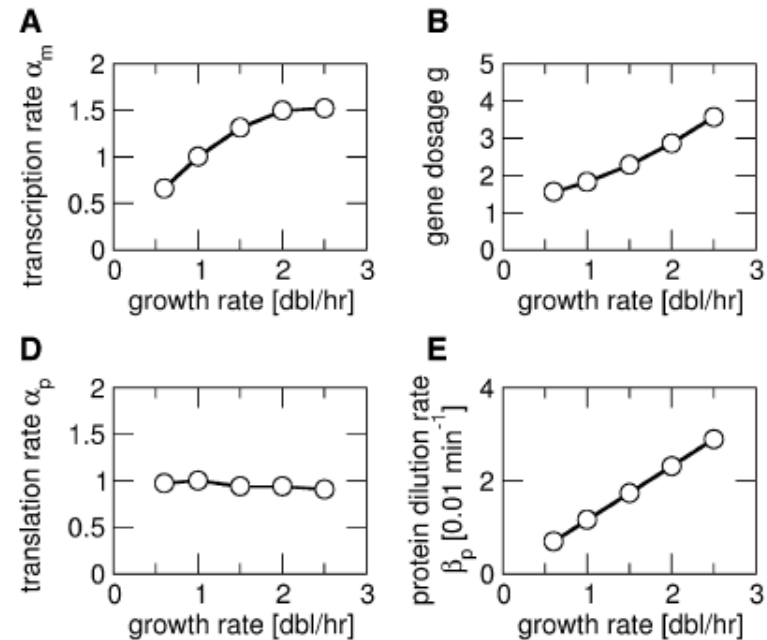
Klumpp *et al.* (2009), *Cell*, 139(7):1366-75

Bremer and Dennis (1996), *Escherichia Coli and Salmonella*, ASM Press, 1553-69

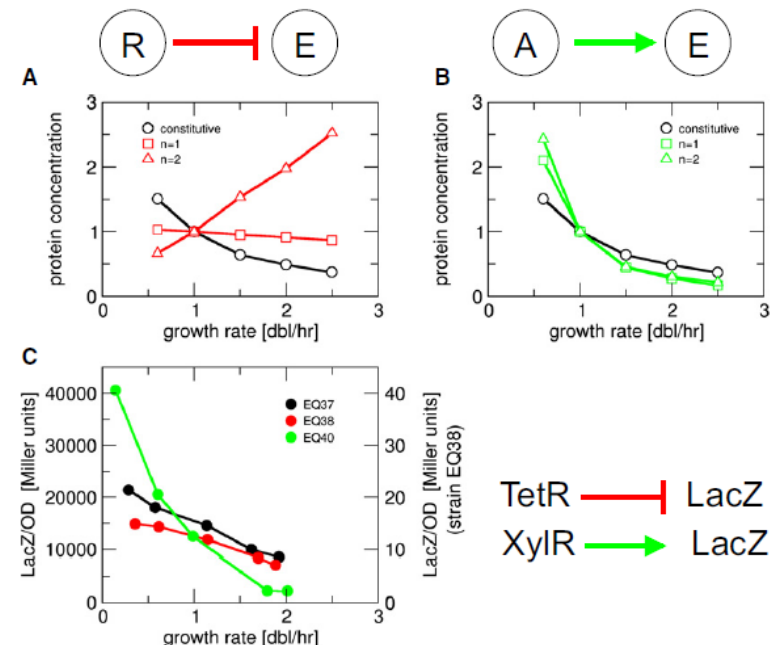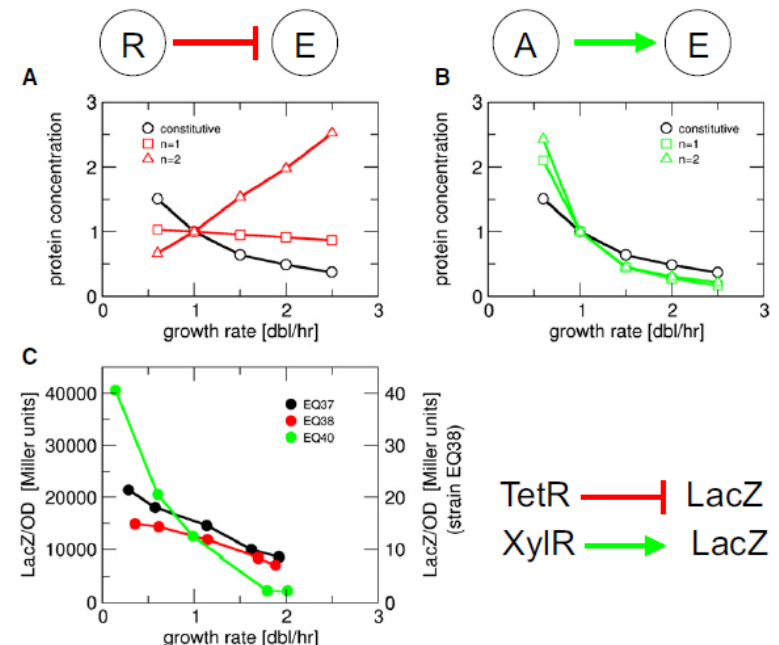# Growth transition and gene expression

- **Question:** what are relative contributions of specific regulators and global physiological effects in adaptation of gene expression during growth transitions?

Growth transition and gene expression

- **Question:** what are relative contributions of specific regulators and global physiological effects in adaptation of gene expression during growth transitions?

- Previous work on growth-rate dependent expression of constitutive and regulated genes

  - Expression of **constitutive** gene is growth-rate dependent

  - Weaker growth-rate dependence under **repression**, stronger growth-rate dependence under **activation**

    Klumpp *et al.* (2009), *Cell*, 139(7):1366-75

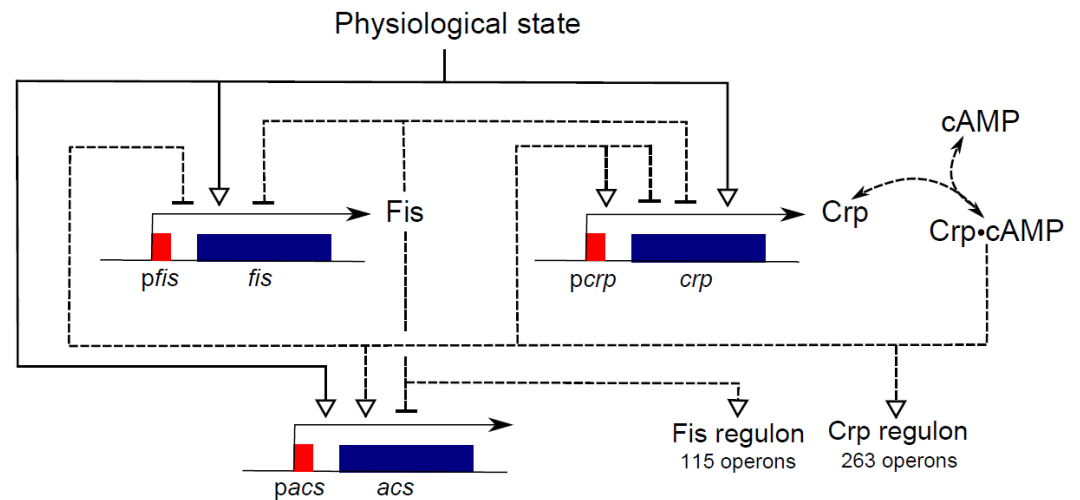# Growth transition and gene expression

- **Question:** what are relative contributions of specific regulators and global physiological effects in adaptation of gene expression during growth transitions?

- Previous work on growth-rate dependent expression of constitutive and regulated genes

  - Expression of **constitutive** gene is growth-rate dependent

  - Weaker growth-rate dependence under **repression**, stronger growth-rate dependence under **activation**

    Klumpp *et al.* (2009), *Cell*, 139(7):1366-75

Growth transition and gene expression

- **Question:** what are relative contributions of specific regulators and global physiological effects in adaptation of gene expression during growth transitions?

  Dynamics instead of steady-state, **network** instead of single gene

Berthoumieux *et al.* (2013), *Mol. Syst. Biol.*, in press

Growth transition and gene expression

- **Question:** what are relative contributions of specific regulators and global physiological effects in adaptation of gene expression during growth transitions?

  Dynamics instead of steady-state, **network** instead of single gene

- Question addressed in context of central regulatory circuit of carbon metabolism in *E. coli*
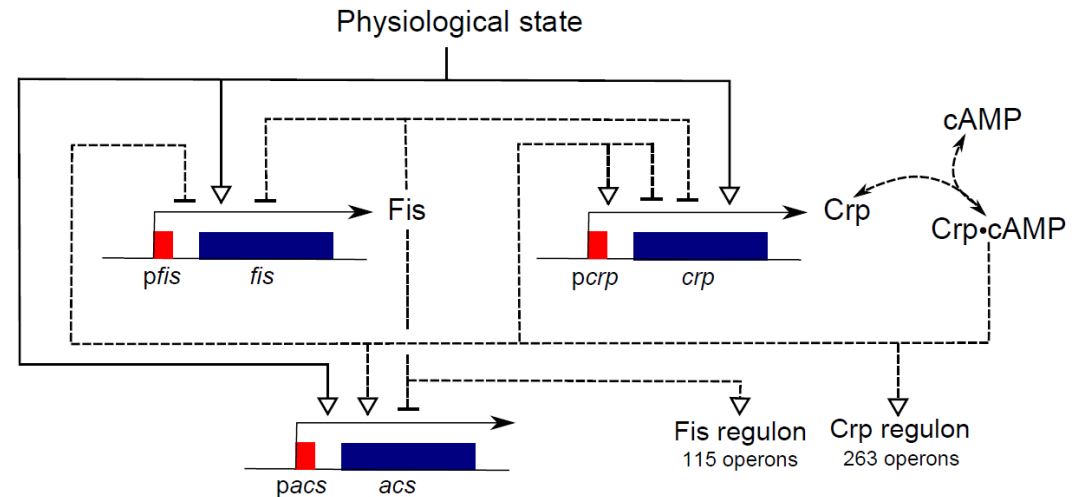
Berthoumieux *et al.* (2013), *Mol. Syst. Biol.*, in press

Approach

- **Real-time monitoring of dynamic response** of network to depletion of carbon source (glucose):
  - Growth rate
  - cAMP concentration
  - Promoter activity of network genes
  - Global physiological state through use of constitutive phage promoter

Approach

- **Real-time monitoring of dynamic response** of network to depletion of carbon source (glucose):
  - Growth rate
  - cAMP concentration
  - Promoter activity of network genes
  - Global physiological state through use of constitutive phage promoter
- **Simple models of promoter activities** of network genes
  > Models represent different hypotheses on contributions from global and specific effects
- **Validation of models** using experimental data

Real-time monitoring of gene expression
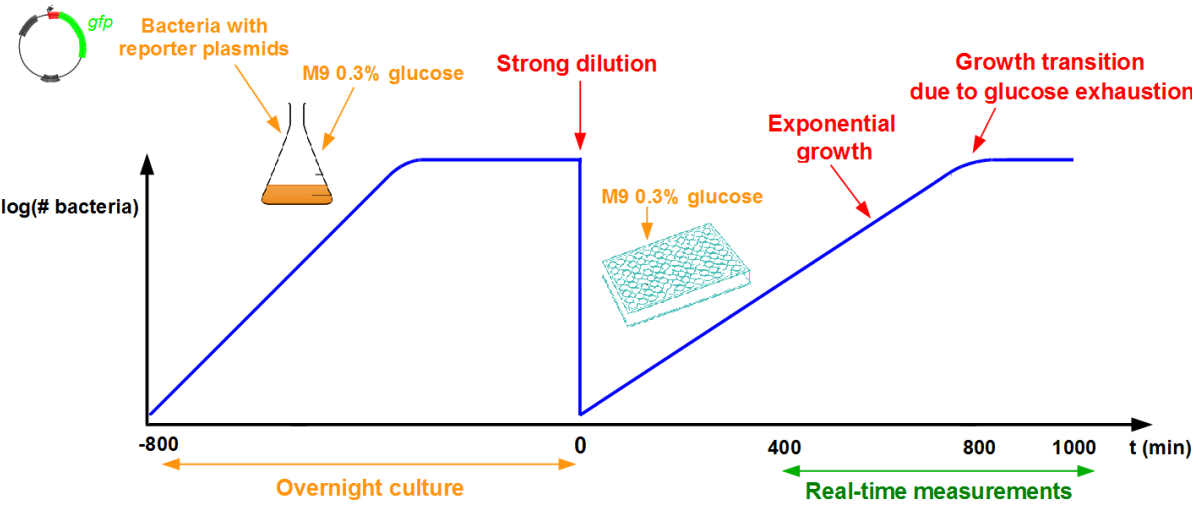
- Transcriptional fusion of promoters with *gfp* reporter genes on plasmid
- Measurement of absorbance and fluorescence signals, thermostated automated microplate reader
- Model-based derivation of promoter activities
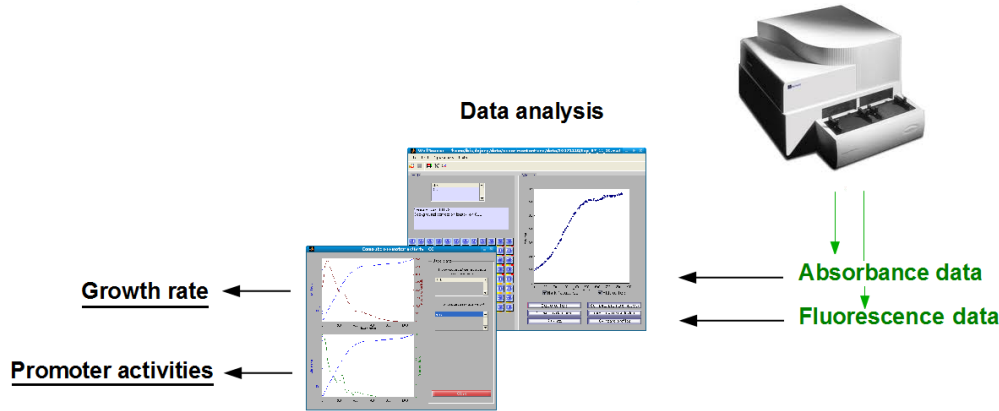
de Jong *et al.* (2010), *BMC Syst. Biol.*, 4:55
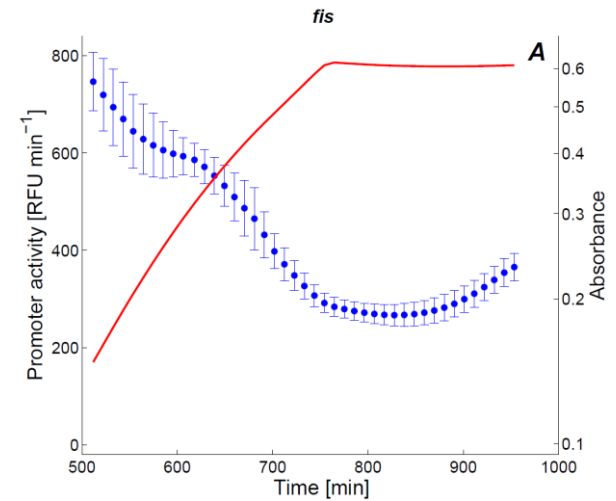
Real-time monitoring of gene expression

- Monitoring of *fis* promoter activity during growth transition
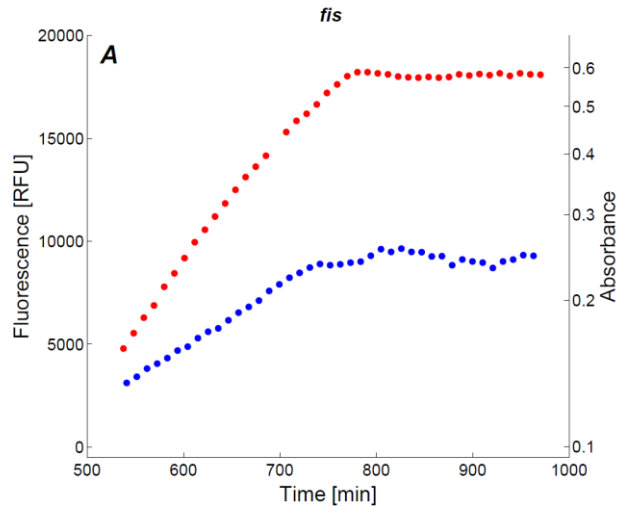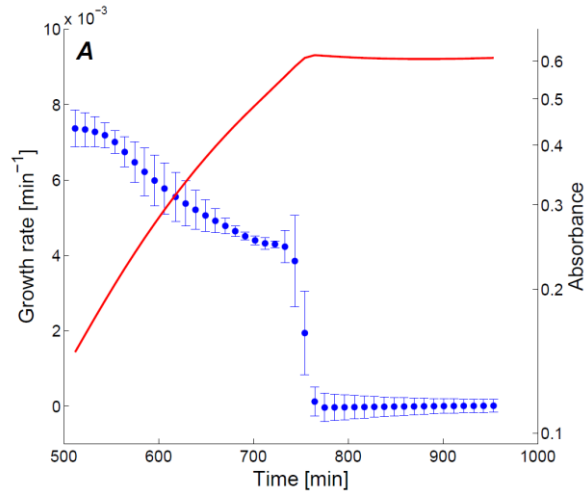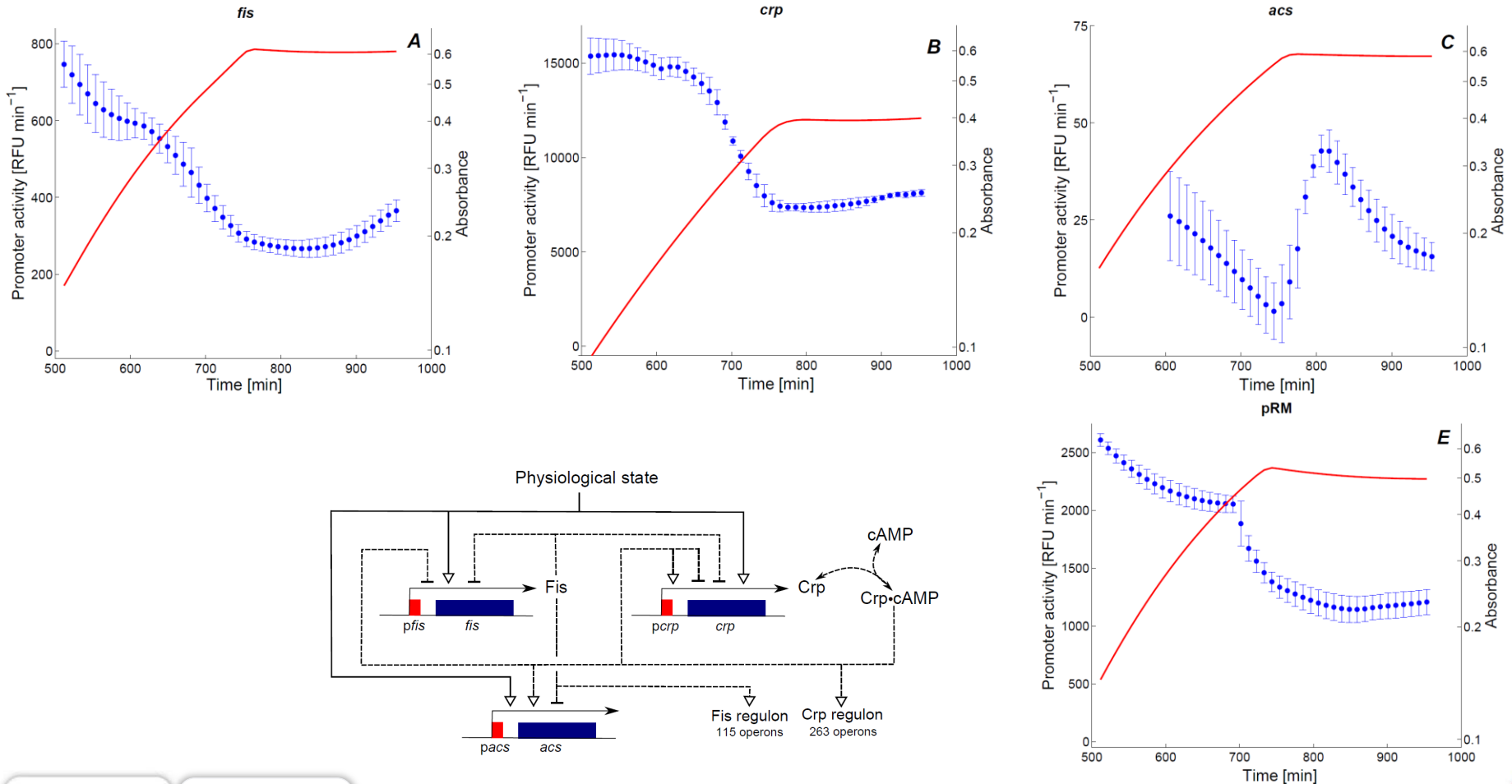

Absorbance and fluorescence   Growth rate   Promoter activity

Real-time monitoring of gene expression

- Monitoring of activity of *crp*, *fis*, *acs* and constitutive phage promoters during growth transition

# Bias introduced by plasmid copy number

- Plasmids are relatively easy to construct and have strong signal, but … **plasmid copy number** varies with growth rate

  Lin-Chao and Bremer (1986), *Mol. Gen. Genet.*, 203(1):143-9

- Measurement of relative plasmid copy number using qPCR

- Variation in plasmid copy number preserves qualitative shape of profiles, but introduces quantitative bias

- **Conclusion**: need for analysis method that corrects for growth-phase dependent variations of plasmid copy number

Measurement of cAMP

- Measurement of cAMP concentration during growth transition:
  - Measurement of extracellular cAMP concentration
  - Development of kinetic model accounting for cAMP import/export
  - Determination of intracellular cAMP concentration from measurements and model

External cAMP concentration

Internal cAMP concentration
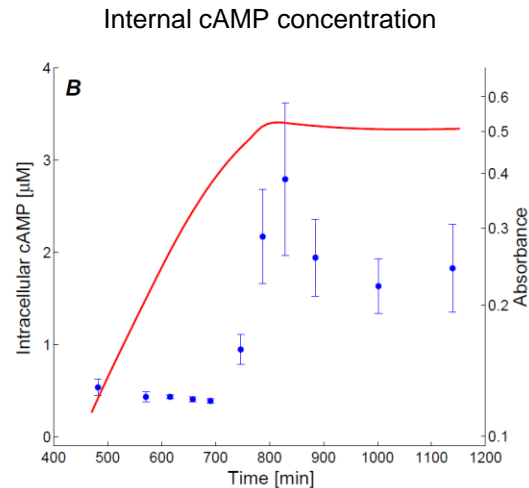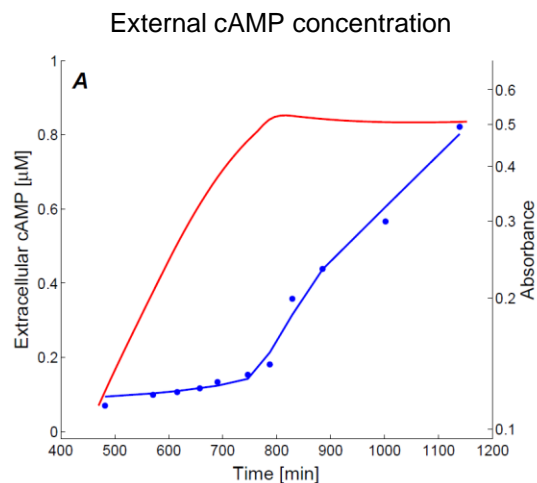
- Good correspondence with intracellular cAMP profiles published in literature

Kao *et al.* (2004), *Proc. Natl. Acad. Sci. USA*, 101(2):641-6

Approach

- **Real-time monitoring of dynamic response** of network to depletion of carbon source (glucose):
  - Growth rate
  - cAMP concentration
  - Promoter activity of network genes
  - Global physiological state through use of constitutive phage promoter
- **Simple models of promoter activities** of network genes

  Models represent different hypotheses on contributions from global and specific effects

- **Validation of models** using experimental data

Model of promoter activities

- Simple **model of promoter activity** separating specific effects of transcription factors from global effect of physiological state

$$p(t) = k\, p_1(t)\, p_2(t)$$

k : maximum promoter activity

$p_1(t)$: regulation by global physiological state

$p_2(t)$: regulation by specific transcription factors

$p_1(t)$ and $p_2(t)$ vary between 0 and 1

Model of promoter activities

- Simple **model of promoter activity** separating specific effects of transcription factors from global effect of physiological state

$$p(t) = k \, p_1(t) \, p_2(t)$$

- Normalization with respect to **reference state** at t^0 to get rid of unknown constant k and logarithmic transformation:

$$\log \frac{p(t)}{p^0} = \log \frac{p_1(t)}{p_1^0} + \log \frac{p_2(t)}{p_2^0}$$

Convenient choice of reference state: growth arrest (expression peak of *acs*) or steady state after growth transition

Model of promoter activities

- **Hypothesis 1**: effect of global physiological state (measured by phage promoter) is dominant and effect of specific regulators is negligible ($p_2(t) \approx p_2^0$):

$$\log \frac{p(t)}{p^0} = \log \frac{p_{RM}(t)}{p_{RM}^0}$$

- Advantages of model:
  - Straightforward to test by means of experimental data
  - Non-parametric, does not require model calibration
  - No effect of plasmid copy number variation if promoter activity is measured in same plasmid vector

Test of hypothesis 1

- Global effect is dominant for expression control of transcription factors (*crp* and *fis*), but not for metabolic gene (*acs*)
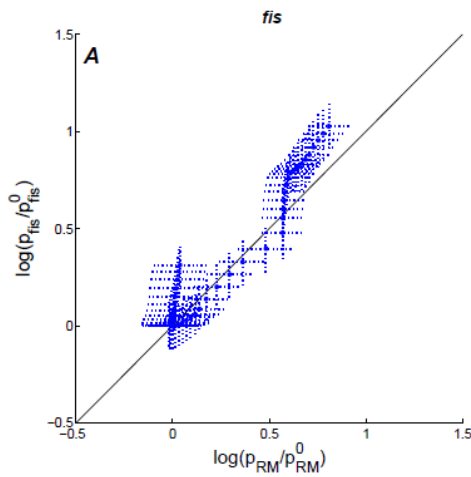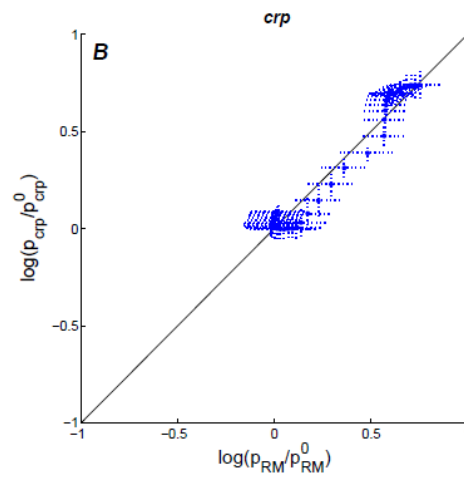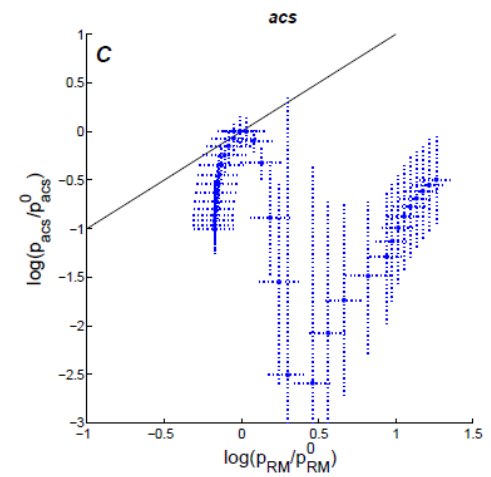
$R^2 = 0.93$ $R^2 = 0.96$ $R^2 = 0.08$

Model of promoter activities

- **Hypothesis 2**: effect of specific regulators is not negligible and can be reduced to effect of change in cAMP concentration $c(t)$:
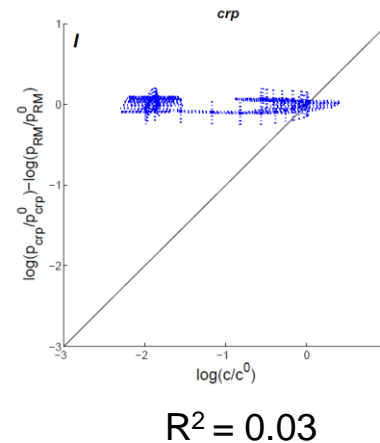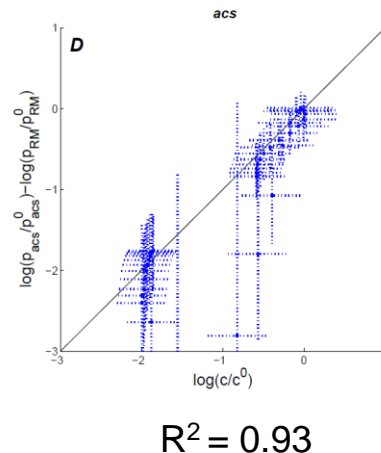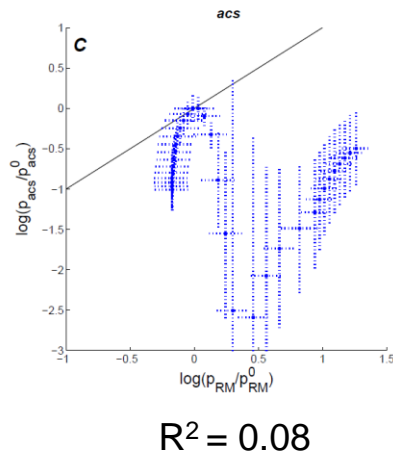
$$\log \frac{p(t)}{p^0} - \log \frac{p_{RM}(t)}{p_{RM}^0} = \log \frac{c(t)}{c^0}$$

- Hypothesis based on data, but biological assumptions underlying simplification can be explicitly formulated
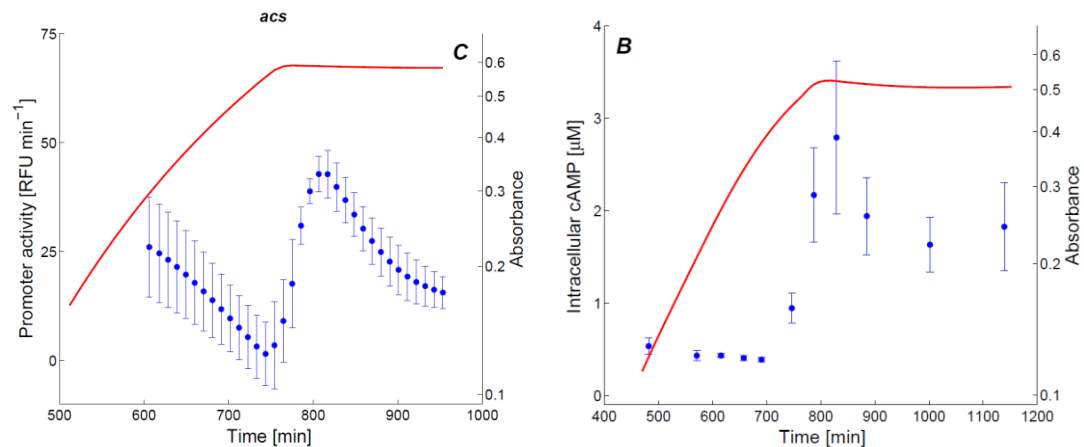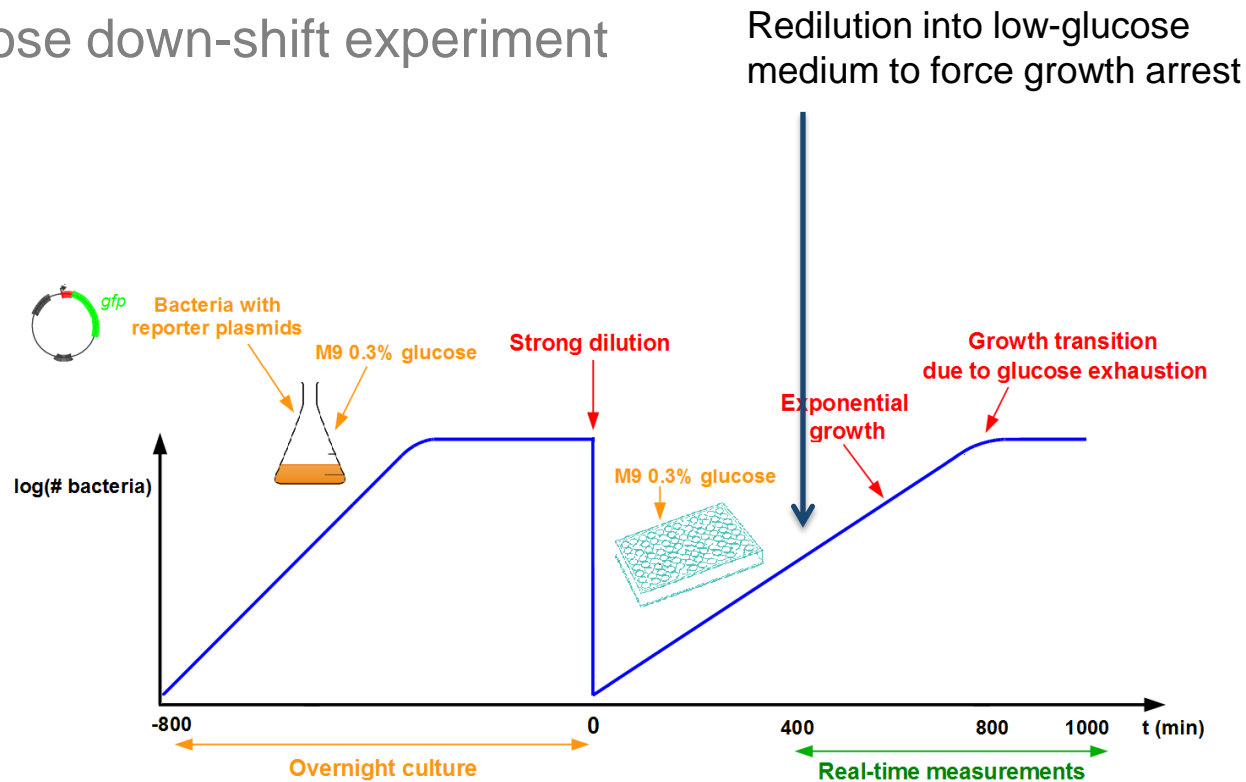
Test of hypothesis 2

- Combination of global effect and specific effect of cAMP explains variation in *acs* promoter activity

$R^2 = 0.08$ $R^2 = 0.93$ $R^2 = 0.03$

- Addition of cAMP as regulator yields bad fit for *crp* and *fis*: no improvement upon simpler hypothesis 1

Other experimental conditions

- Experiments and model tests were repeated in other conditions:
  - Glucose down-shift experiment

Redilution into low-glucose medium to force growth arrest

Bacteria with reporter plasmids

M9 0.3% glucose

Strong dilution

Growth transition due to glucose exhaustion

Exponential growth

log(# bacteria)

M9 0.3% glucose

-800   0   400   800   1000   t (min)

Overnight culture

Real-time measurements

Other experimental conditions

- Experiments and model tests were repeated in other conditions:
  - Glucose down-shift experiment
  - Deletion mutant *crp*
  - Deletion mutant *fis*
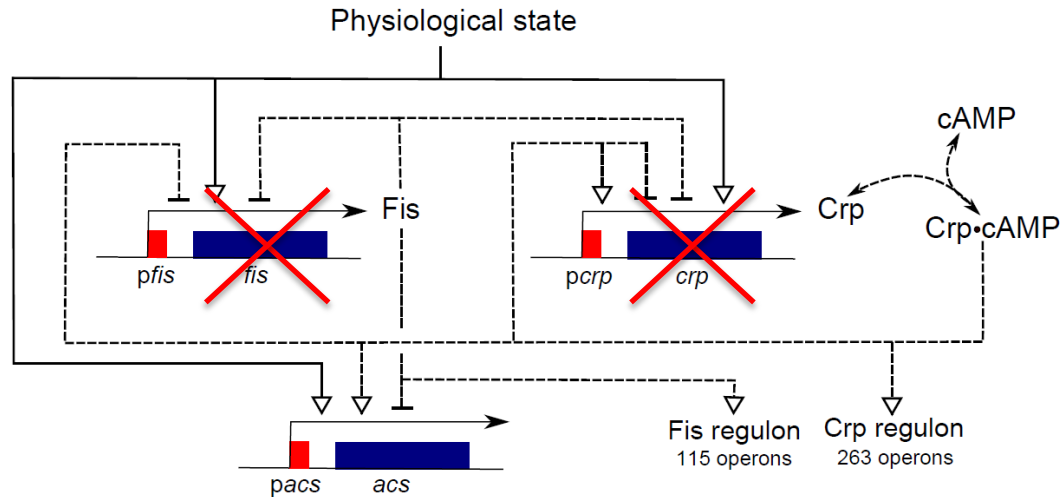
Other experimental conditions

- Experiments and model tests were repeated in other conditions:
  - Glucose down-shift experiment
  - Deletion mutant *crp*
  - Deletion mutant *fis*
- Additional data confirm conclusions:
  - Effect of global physiological state dominant for transcriptional control of genes encoding transcription factors Fis and Crp
  - Combined effect of global physiological state and cAMP accounts for variation of promoter activity of *acs*

Other regulators

- Is effect of global physiologal state also dominant in transcriptional control of other regulators?

  RpoS (σ^S), master stress regulator in *E. coli*, inhibited by Crp·cAMP
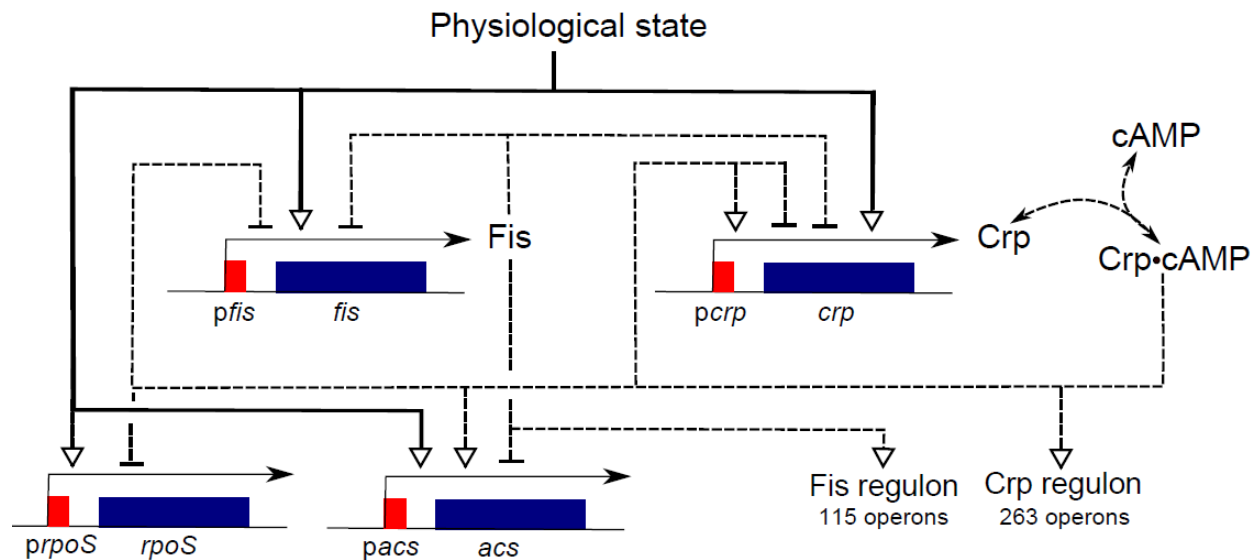
Other regulators

- Is effect of global physiologal state also dominant in transcriptional control of other regulators?

  RpoS (σ^S), master stress regulator in *E. coli*, inhibited by Crp·cAMP

- Test of hypothesis 1 in different conditions confirms dominant role of global physiological state
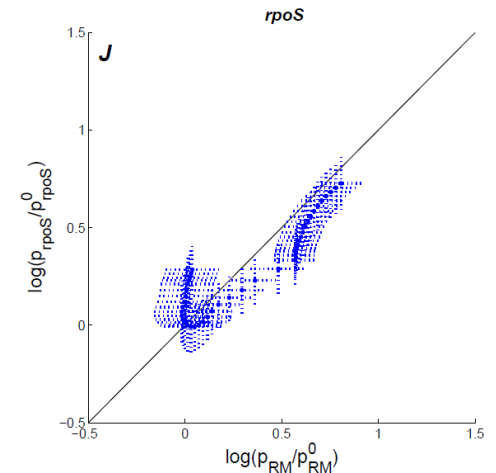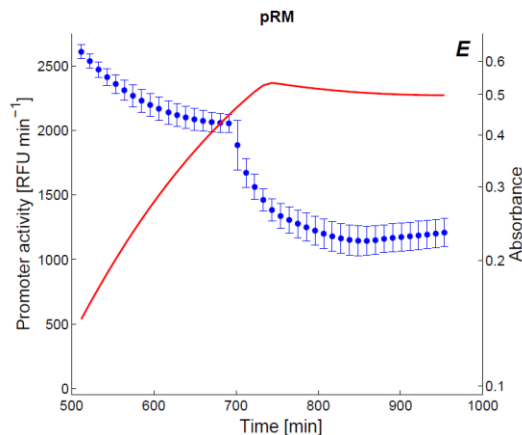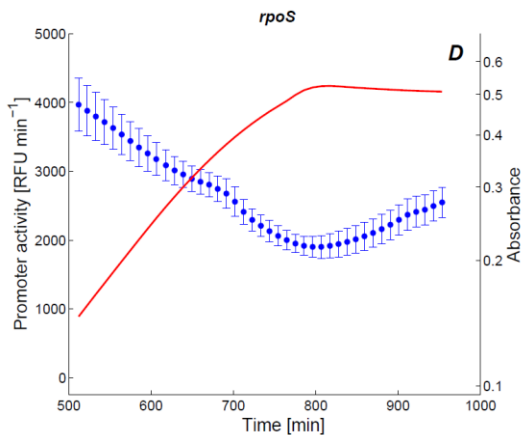

$R^2 = 0.84$

Conclusions

- Control of gene expression across growth phases is shared between physiological state of the cell and transcription factors

- Method to dissect shared control of promoter :
  - Simple mathematical model of promoter activity
  - Carefully designed data analysis procedures

- Application of method to analysis of regulatory circuit involving key regulators of carbon metabolism in *E. coli*
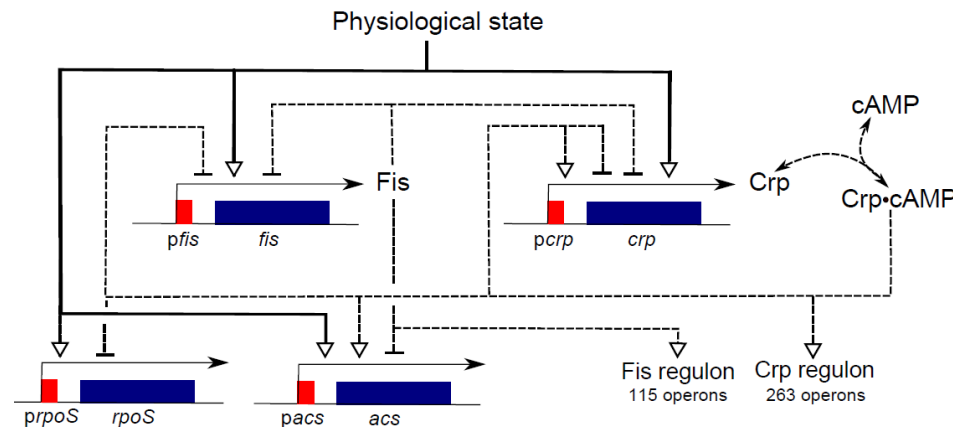
Conclusions

- Two surprising results:
  - Transcriptional control of genes encoding the transcription factors is dominated by growth-phase-dependent effect
  - Many regulatory interactions involving Crp·cAMP and Fis do not contribute to transcriptional control in our conditions. Other conditions? Weak effects?
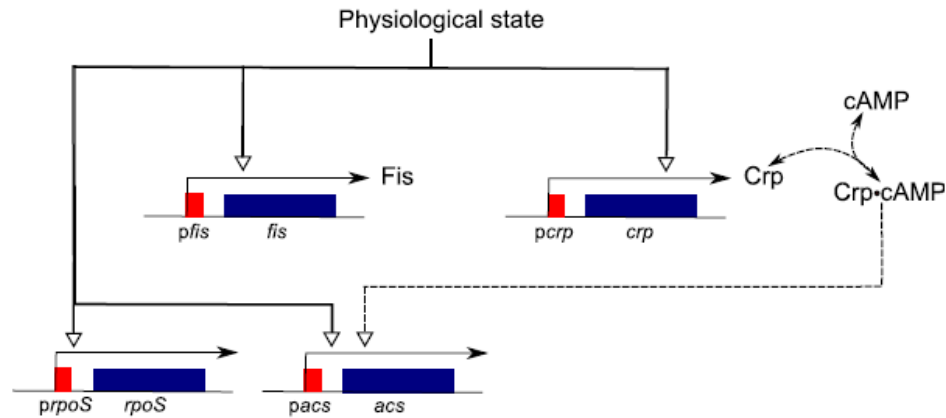
Berthoumieux *et al.* (2013), *Mol. Syst. Biol.*, in press

Conclusions

- Two surprising results:
  - Transcriptional control of genes encoding the transcription factors is dominated by growth-phase-dependent effect
  - Many regulatory interactions involving Crp·cAMP and Fis do not contribute to transcriptional control in our conditions. Other conditions? Weak effects?

- Results question central role often attributed to transcriptional regulatory networks

    Are results generalizable to entire regulatory network of *E. coli*?

- **Alternative view**: specific transcription factors complement and finetune global control exerted by physiological state

- Relevance for biotechnology and synthetic biology

Lack of quantitative information: strategies

- Three main strategies to deal with lack of quantitative data:
  - **Test of parameter sensitivity**
  - Model reduction and simplification
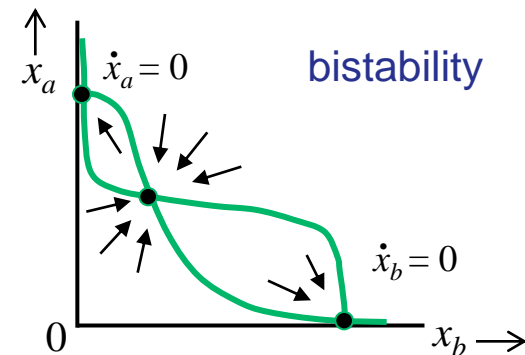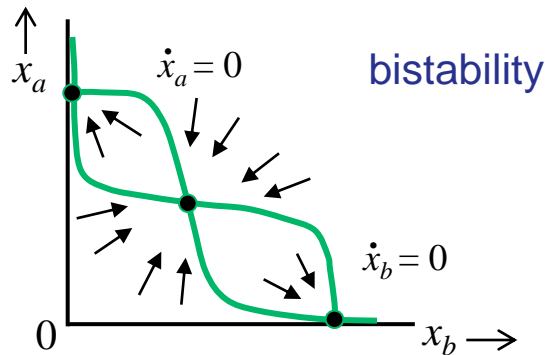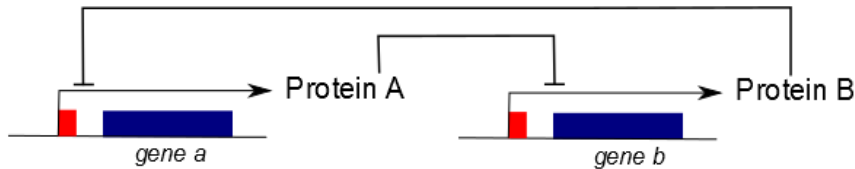 - Parameter estimation from time-series data

De Jong and Ropers (2006), *Brief. Bioinform.*, 7(4):354-363

Test of parameter sensitivity

- Important dynamic properties are expected to be **robust** over large ranges of parameter values

   Important dynamic properties should be insensitive to moderate variations in parameter values

Stelling *et al.* (2004), *Cell*, 118(6):675-685

Model reduction and simplification

- Use model reduction and simplification to obtain models that can be analyzed with less information on parameter values

  - Piecewise-linear instead of nonlinear models

  - Also: Boolean models

$$\dot{x}_a = \kappa_a f(x_b) - \gamma_a x_a$$
$$\dot{x}_b = \kappa_b f(x_a) - \gamma_b x_b$$
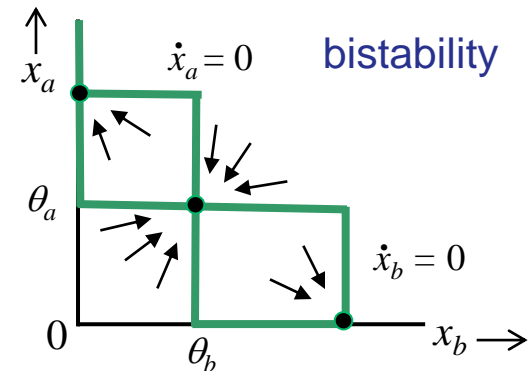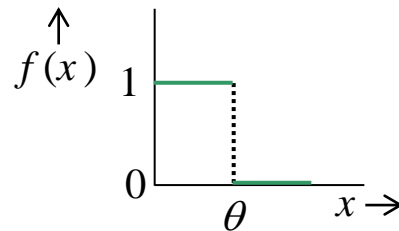
Glass and Kauffman (1973), *J. Theor. Biol.*, 39(1):103-29

de Jong *et al.* (2004), *Bull. Math. Biol.*, 66(2):301-40

Parameter estimation

- **Estimate** parameter values from experimental time-series data
  - **Systems identification** in control and engineering

  Ljung (1999), *System Identification: Theory for the User*, Prentice Hall

- Given model structure, search parameter values for which model predictions best fit experimental data
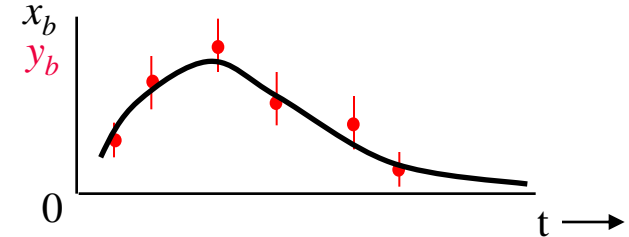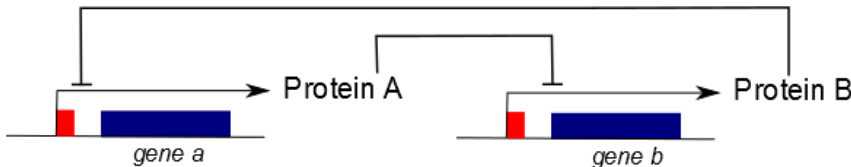
- Minimization of objective function, for instance sum of squared errors:

$$\sum_t (x(t,\theta) - y(t))^2$$

  Possibility to add constraint or penalty terms to restrict parameter space

Lack of quantitative information: strategies

- Three main strategies to deal with lack of quantitative data:
  - **Test of parameter sensitivity**
  - Model reduction and simplification
  - Parameter estimation from time-series data

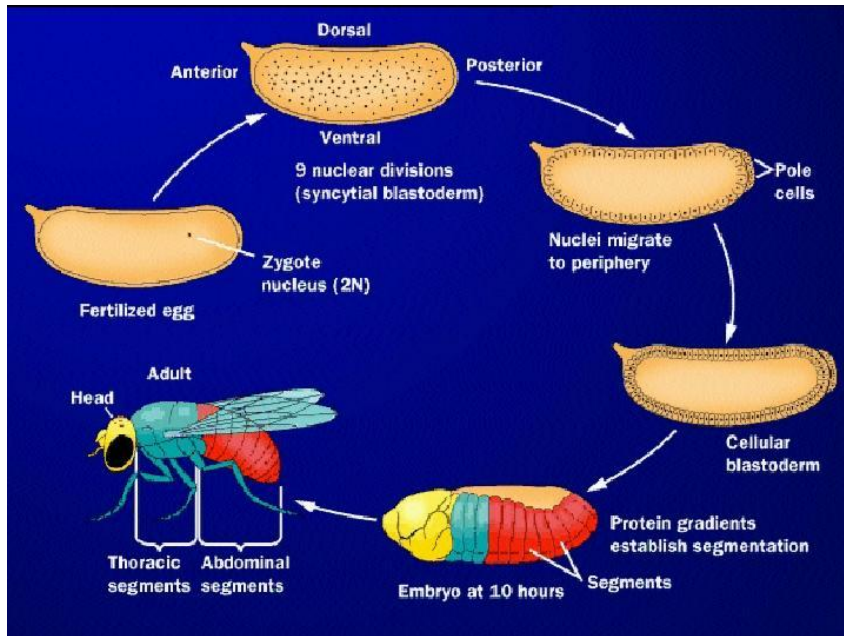 De Jong and Ropers (2006), *Brief. Bioinform.*, 7(4):354-363

- Illustration: models of developmental processes in multicellular organisms
  - **Development of *Drosophila* embryon**
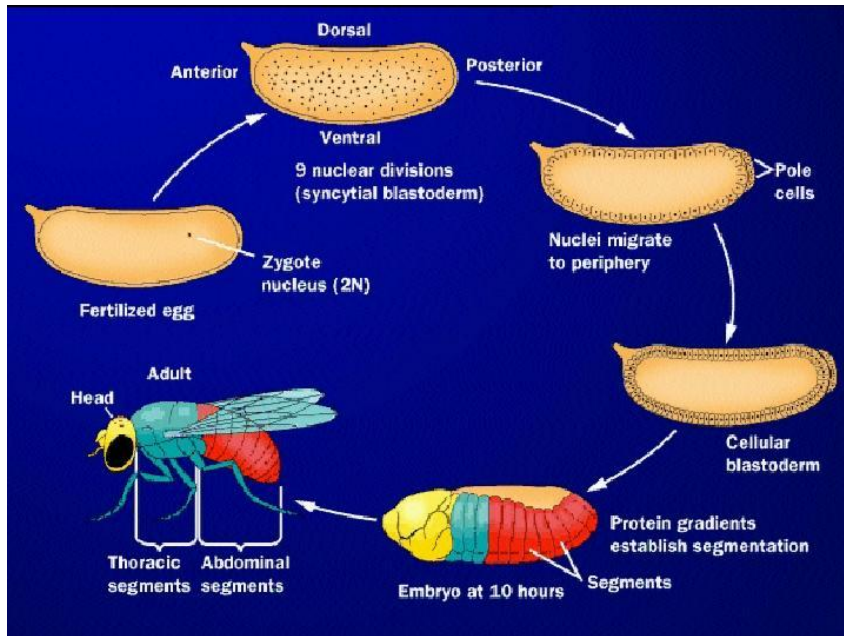
Development of *Drosophila* embryon

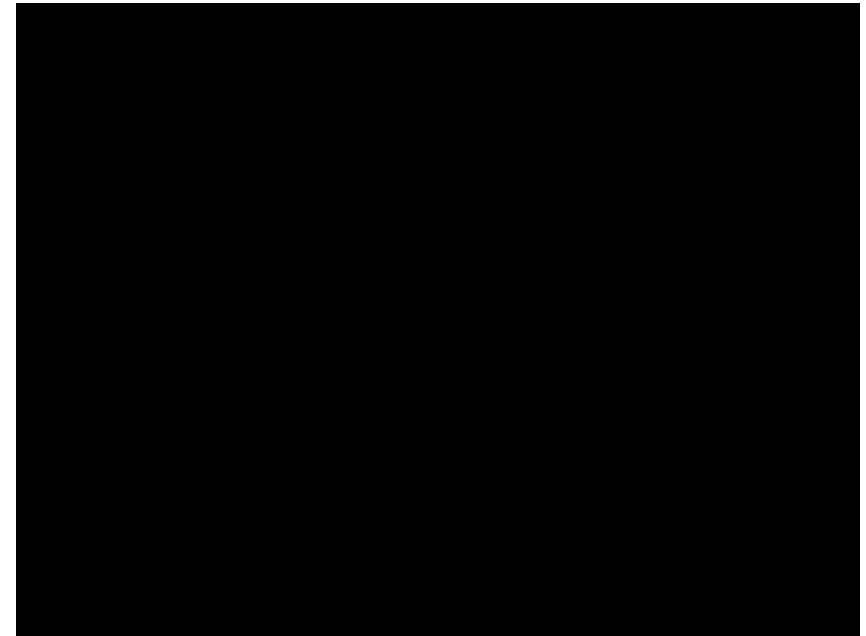- Development of *Drosophila melanogaster* (fruit fly)

Purves *et al.* (1998), *Life: The Science of Biology*, Sinauer

Development of *Drosophila* embryon

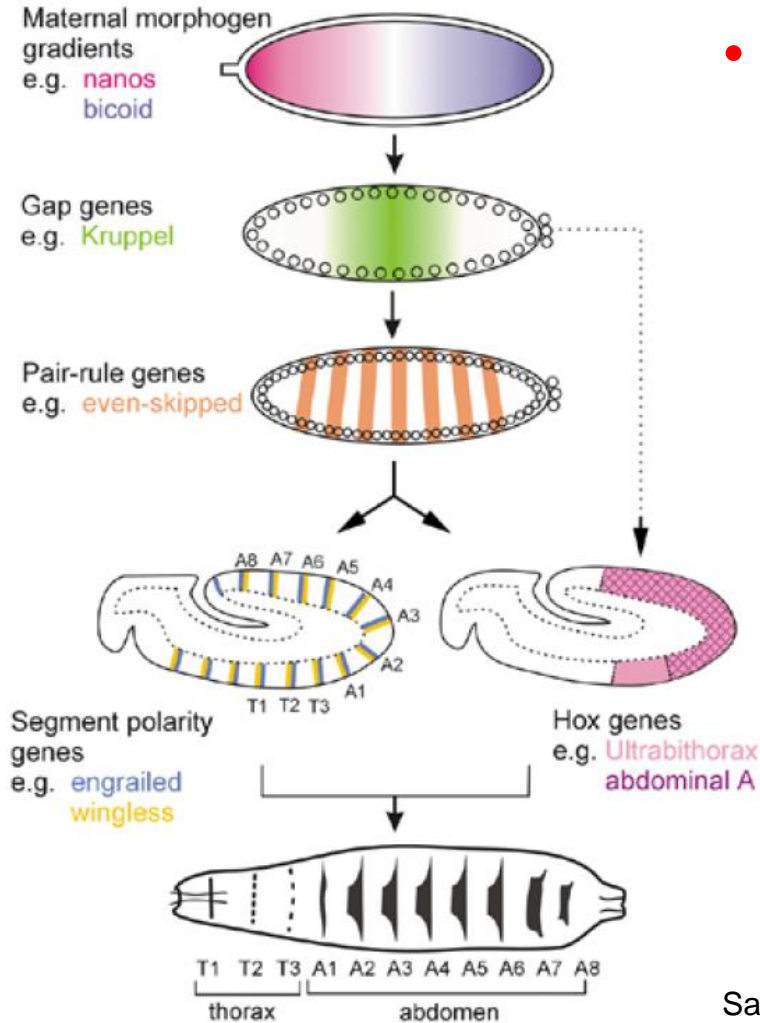- Development of *Drosophila melanogaster* (fruit fly)

Purves *et al.* (1998), *Life: The Science of Biology*, Sinauer

Tomer *et al.* (2012), *Nat. Methods*, 9(7):755–63

Development of *Drosophila* embryon

Maternal morphogen gradients
e.g. nanos, bicoid

Gap genes
e.g. Kruppel

Pair-rule genes
e.g. even-skipped

Segment polarity genes
e.g. engrailed, wingless

Hox genes
e.g. Ultrabithorax, abdominal A

A8 A7 A6 A5 A4 A3 A2 A1 T1 T2 T3

T1 T2 T3 A1 A2 A3 A4 A5 A6 A7 A8
thorax   abdomen

- Spatiotemporal gene expression patterns during development of *Drosophila melanogaster* (fruit fly)

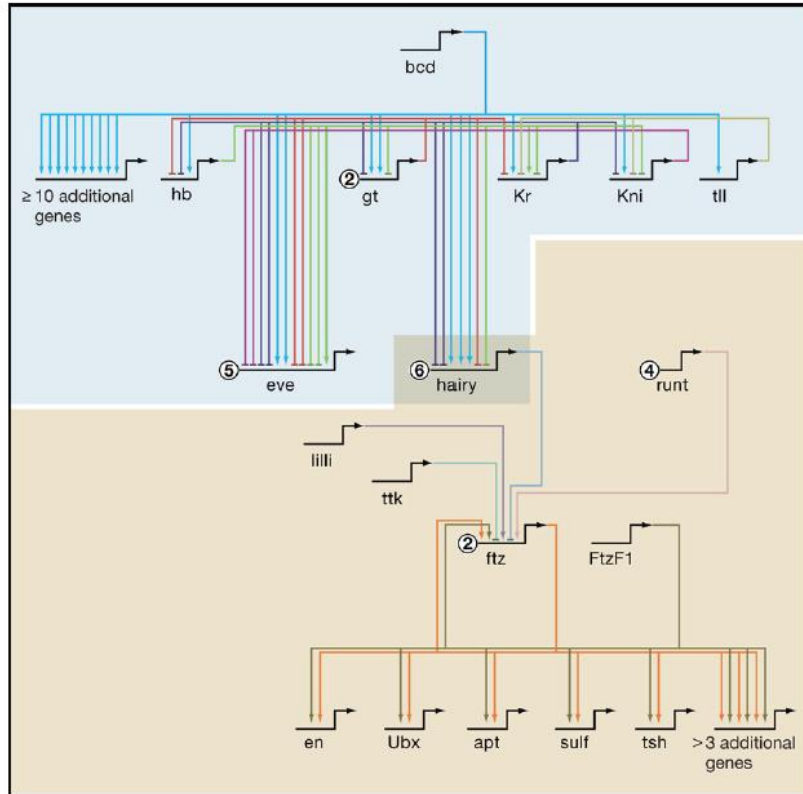Sanson (2001), *EMBO Rep.*, 2(12):1083–8

Development of *Drosophila* embryon
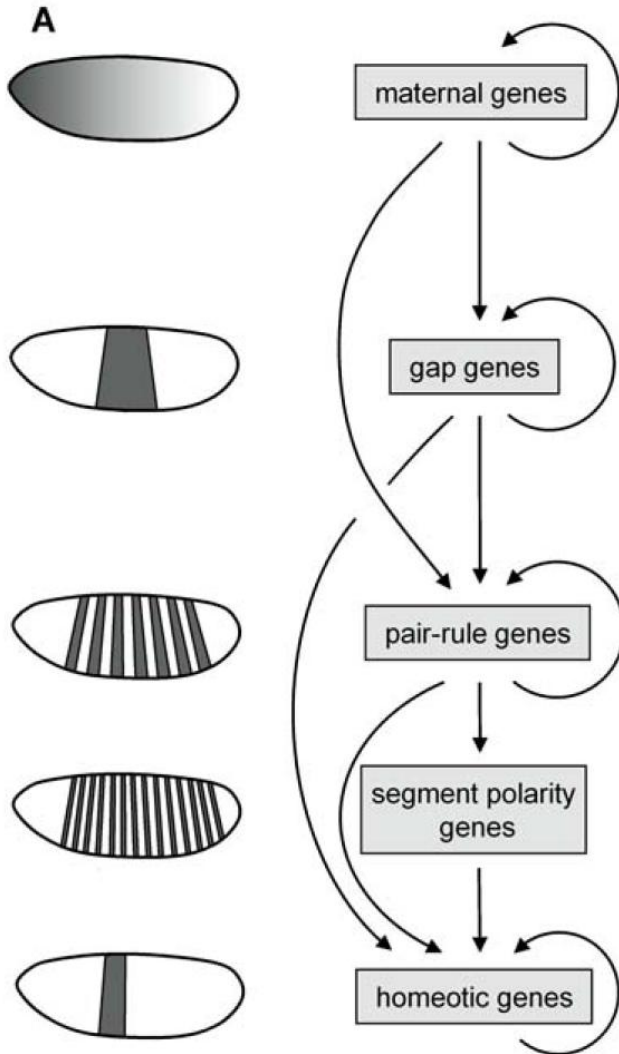
- Spatiotemporal gene expression patterns during development of *Drosophila melanogaster* (fruit fly)

- Gene classes and their interactions responsible for establishment of gene expression patterns

Schroeder *et al.* (2004), *PLOS Biol.*, 4(2):e271

Development of *Drosophila* embryon


Carroll (2008), *Cell*, 134(1):25-36

- Spatiotemporal gene expression patterns during development of *Drosophila melanogaster* (fruit fly)

- Gene classes and their interactions responsible for establishment of gene expression patterns

- Complex gene regulatory networks
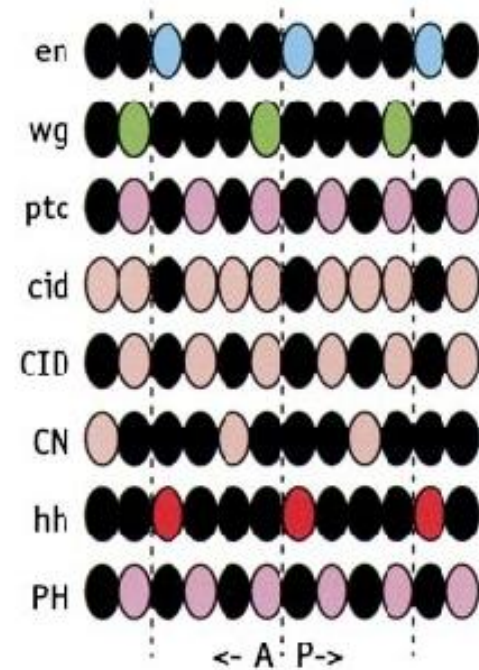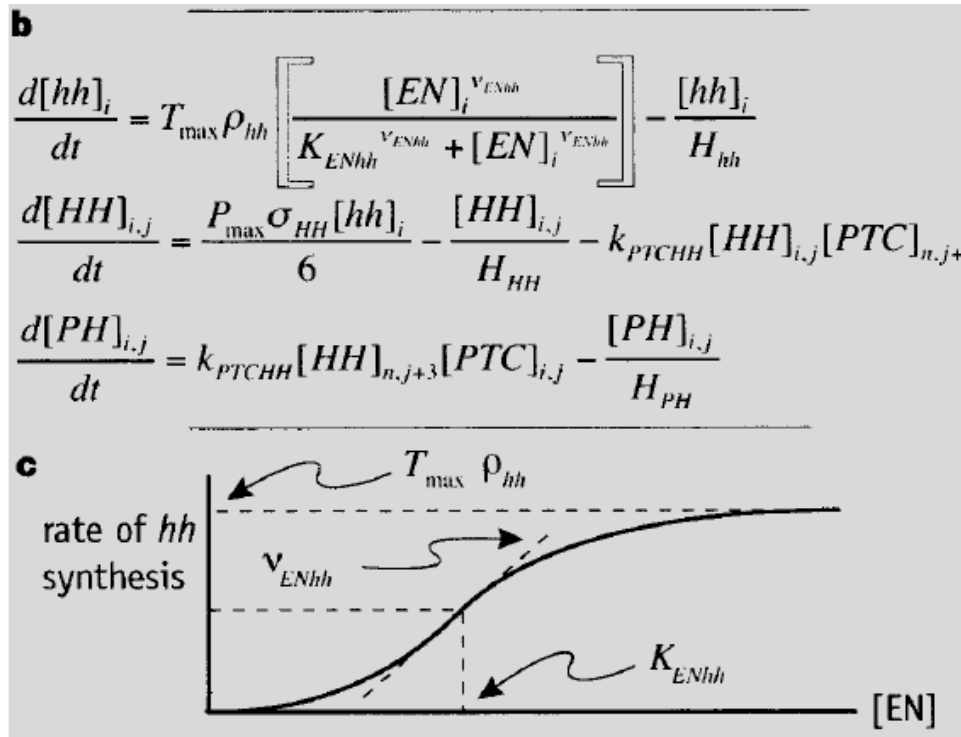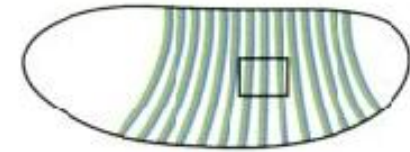
Model of *Drosophila* segmentation

- Model of network of **segment polarity** genes in early development of *Drosophila*
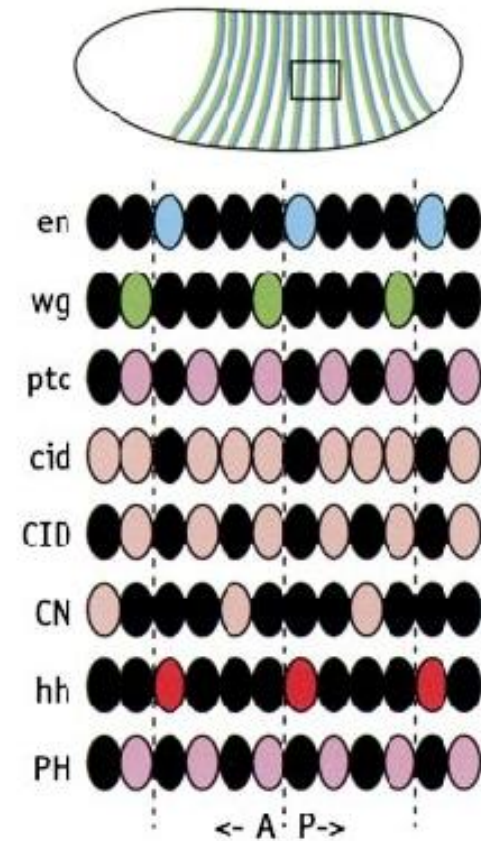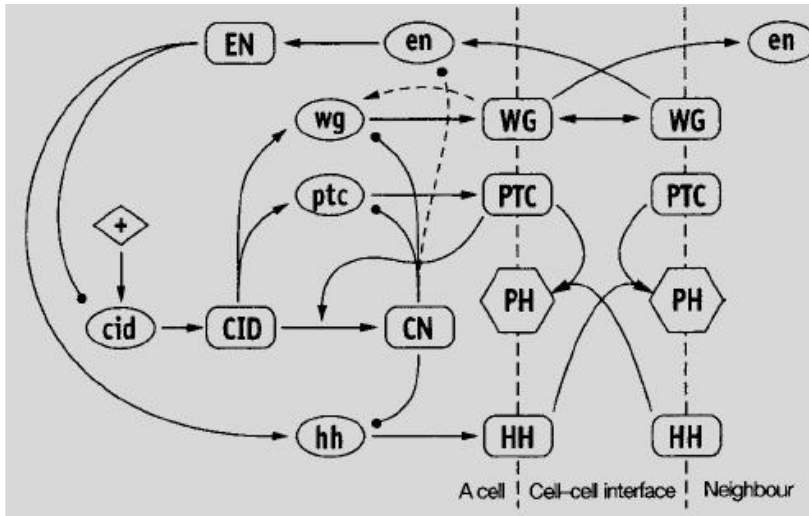
von Dassow *et al.* (2000), *Nature*, 406(6792): 188-92

Model of *Drosophila* segmentation

- Model of network of **segment polarity** genes in early development of *Drosophila*
  - 13 ODEs per cell and 48 parameters

$$\frac{d[hh]_i}{dt} = T_{max}\rho_{hh}\left[\frac{[EN]_i^{\nu_{ENhh}}}{K_{ENhh}^{\nu_{ENhh}} + [EN]_i^{\nu_{ENhh}}}\right] - \frac{[hh]_i}{H_{hh}}$$

$$\frac{d[HH]_{i,j}}{dt} = \frac{P_{max}\sigma_{HH}[hh]_i}{6} - \frac{[HH]_{i,j}}{H_{HH}} - k_{PTCHH}[HH]_{i,j}[PTC]_{n,j+}$$

$$\frac{d[PH]_{i,j}}{dt} = k_{PTCHH}[HH]_{n,j+3}[PTC]_{i,j} - \frac{[PH]_{i,j}}{H_{PH}}$$

von Dassow *et al.* (2000), *Nature*, 406(6792): 188-92

Robustness of gene expression patterns

- Spatial expression pattern of segment polarity genes **robustly reproduced** over large ranges of parameter values
  - 0.5% of sampled parameter combinations leads to solution compatible with data
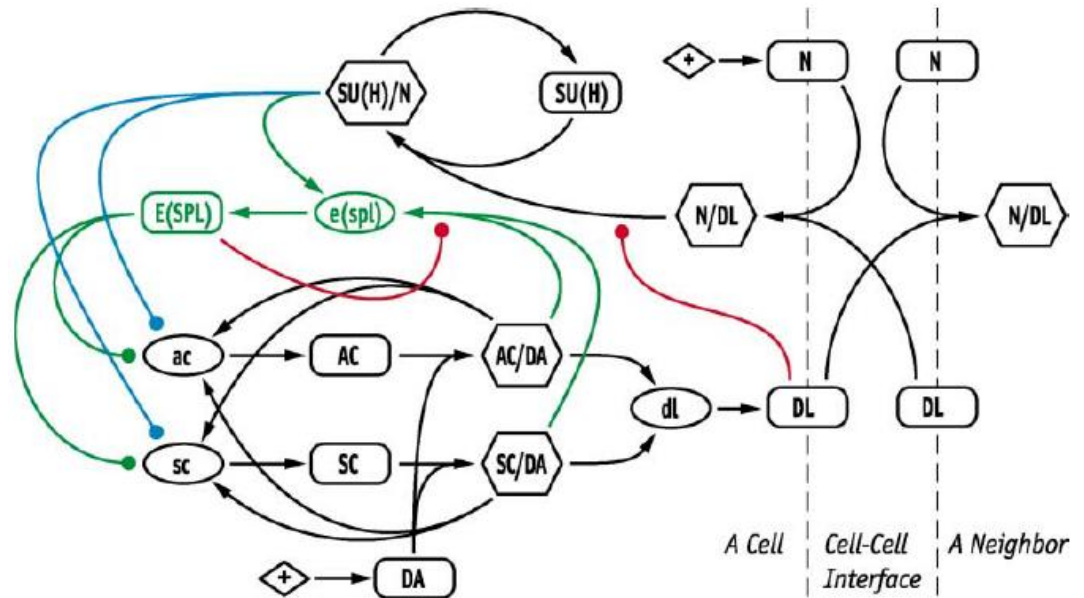

von Dassow *et al.* (2000), *Nature*, 406(6792): 188-92

Robustness of gene expression patterns

- Robustness of model predictions to variations in parameter values confirmed for other developmental networks
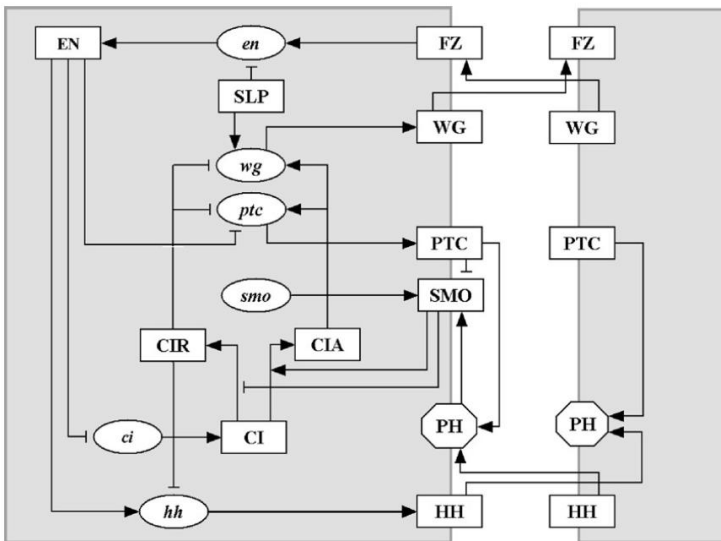
  Neurogenic network, determining neuroblasts in embryos and sensory organ precursor cells in imaginary disks

Meir *et al.* (2002), *Curr. Biol.*, 12(10): 778-86

Logical model of *Drosophila* segmentation

- **Logical model** of segment polarity network: variables take values 0/1 and Boolean functions to update variables

Albert and Othmer (2003), *J. Theor. Biol.*, 223(1):1-18

hh_i

HH_i

ptc_i

PTC_i

PH_i

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

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

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

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

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

Logical model of *Drosophila* segmentation

- **Logical model** of segment polarity network: variables take values 0/1 and Boolean functions to update variables

- Logical models are based on topology of network only (no parametrization), but are capable of reproducing experimental data: **robustness**

Observed initial state (before stage 8)

Predicted steady state state (during stages 9-11)
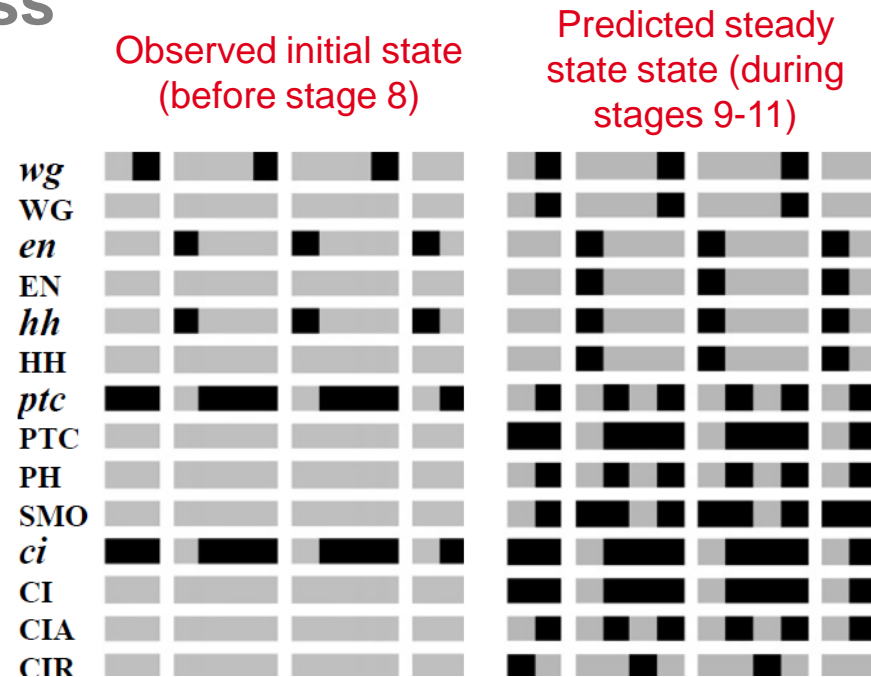
Albert and Othmer (2003), *J. Theor. Biol.*, 223(1):1-18

Logical model of *Drosophila* segmentation

- **Logical model** of segment polarity network: variables take values 0/1 and Boolean functions to update variables

- Logical models are based on topology of network only (no parametrization), but are capable of reproducing experimental data: **robustness**

- **Generalized logical models** allow variables with several discrete values (more complicated update rules)

Sánchez *et al.* (2008), *Int. J. Dev. Biol.*, 52(1):1059-75

Parameter estimation from *Drosophila* data

- Measurement of protein concentrations of gap genes during development of *Drosophila* embryon
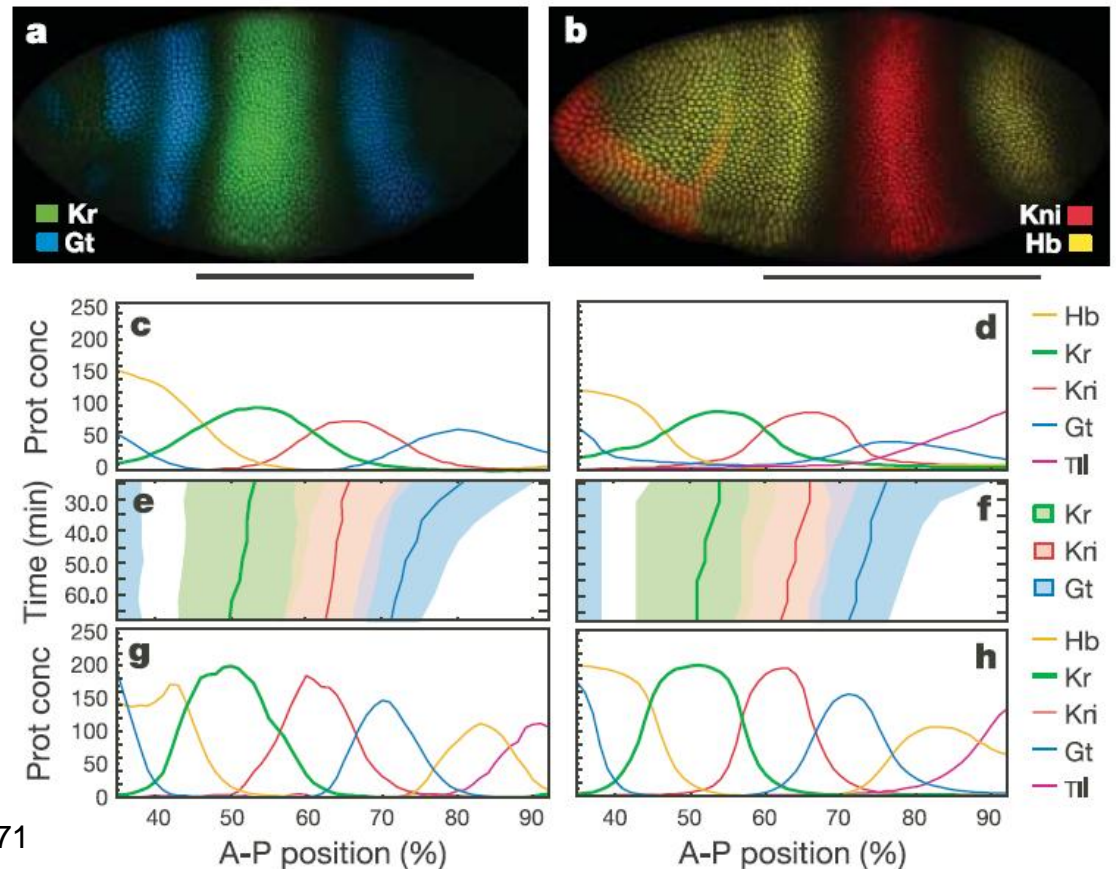
Jaeger *et al.* (2004), *Nature*, 430(6997):368-71

Parameter estimation from *Drosophila* data

- Neural-network-like model of connections between gap genes
  - Model with 58 nuclei and 7 variables (proteins) per nucleus
  - Free diffusion of proteins because at early stages of development embryon is syncytium (multinucleate cell)
  - Sigmodial response functions
  - Connectivity pattern encoded in parameter matrix *T*, so parametric **and** structural identification

$$\frac{dv_i^a}{dt} = R_a g(u^a) + D^a \left[(v_{i-1}^a - v_i^a) + (v_{i+1}^a - v_i^a) \right] - \lambda_a v_i^a$$
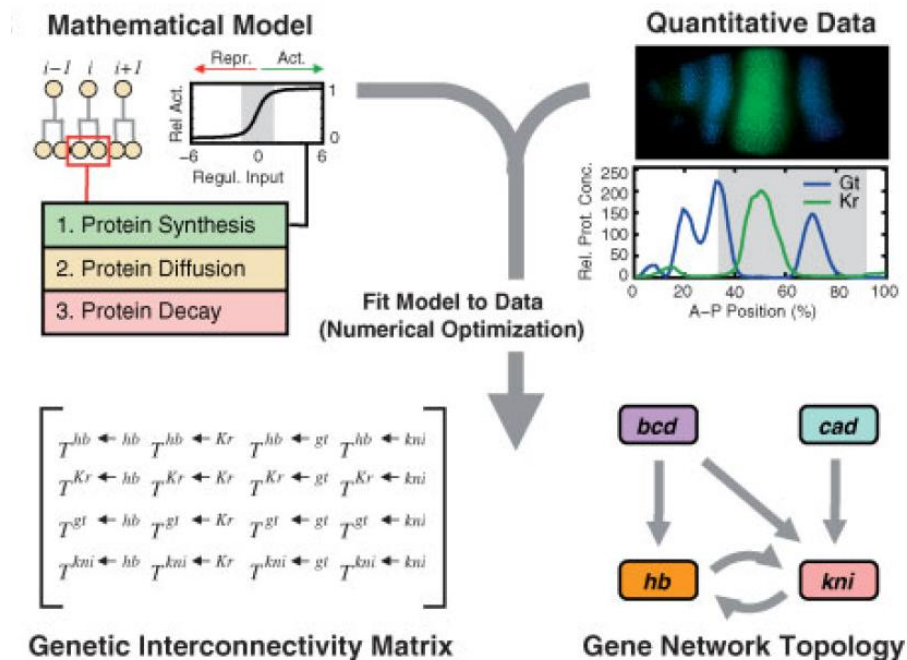
$$u^a = \sum_b T^{ab} v_i^b + m^a v_i^{Bcd} + h^a$$
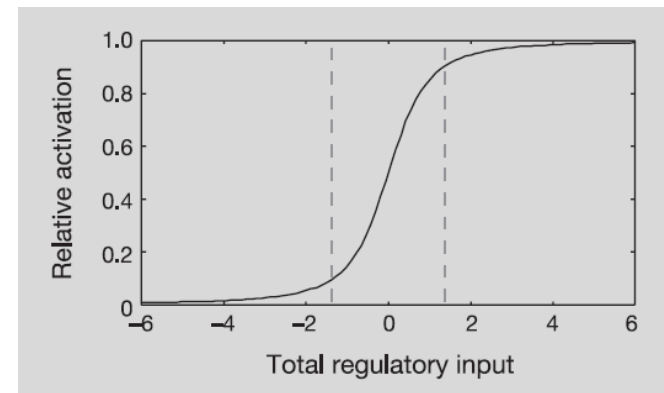
Jaeger *et al.* (2004), *Nature*, 430(6997):368-71

Parameter estimation from *Drosophila* data

- Neural-network-like model of connections between gap genes
- Brute-force parameter estimation by fitting model to data
  - Parallelized simulated annealing

Jaeger and Reinitz (2006), *BioEssays*, 28(11):1102-11

Shifts in gap gene domains

- What is function of **cross-inhibition between gap genes**?

  Model predicts that they are important for shift in gap gene domains after their initial establishment
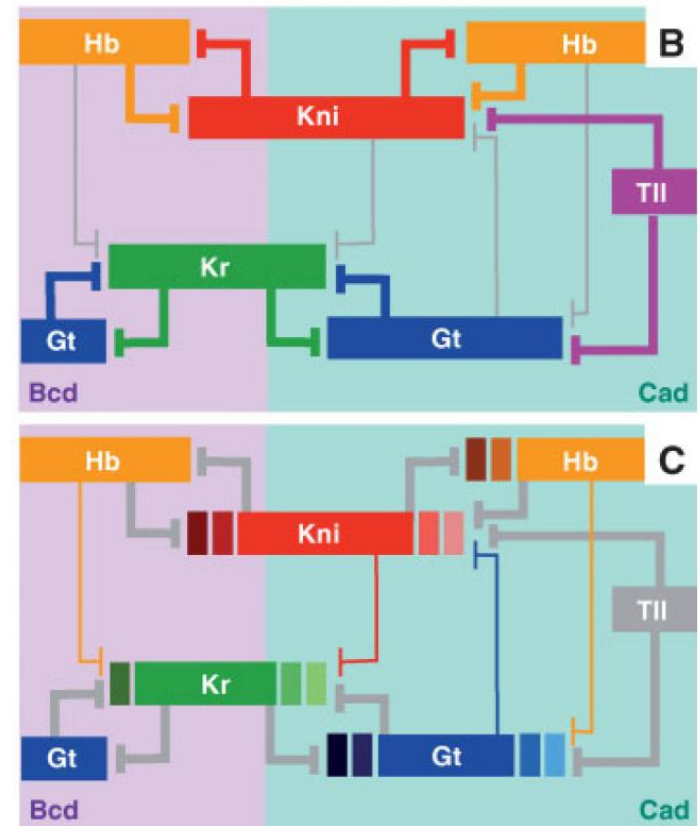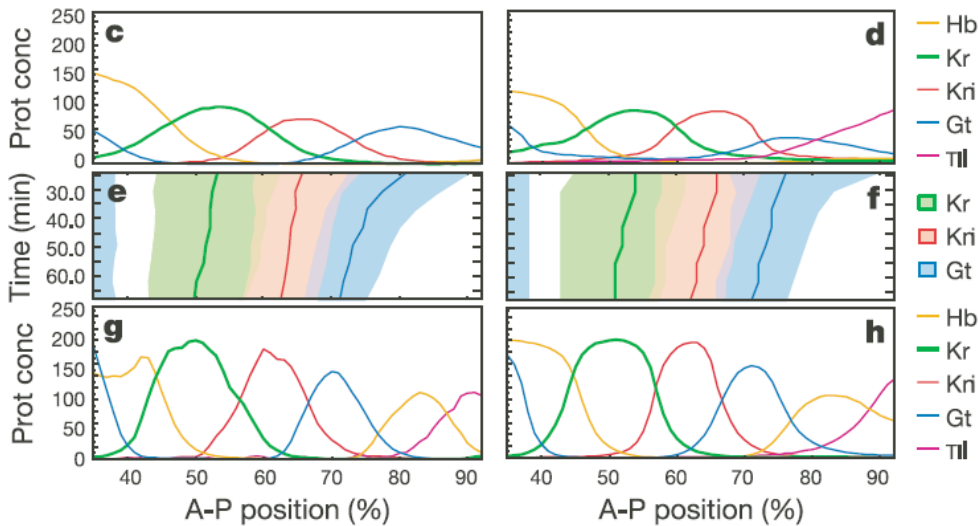
Conclusions

- Several strategies to deal with lack of quantitative information

- Model predictions often robust to changes in parameter values and to simplification/reduction of equations

  Model robustness reflects robustness of biological system?

- High-quality experimental data is becoming increasingly available, favoring estimation of parameter values from expression data

  Quantitative models can make precise predictions of subtle dynamic phenomena

Merci !

www.inrialpes.fr/ibis