Modeling and simulation of gene regulatory networks 5

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

December 18, 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

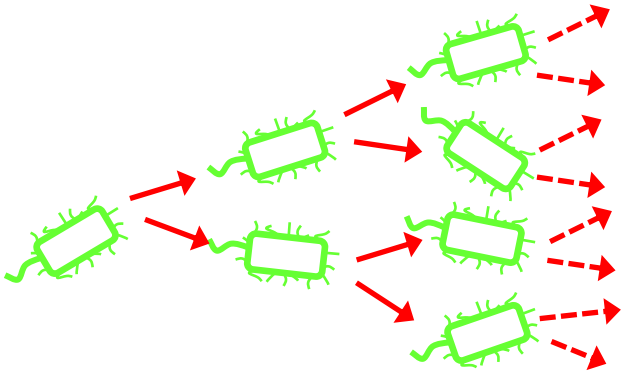
Overview

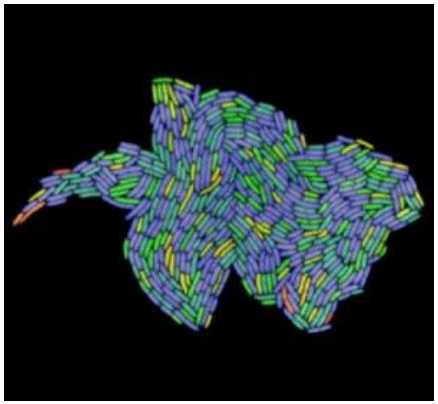
1. Gene regulatory networks in bacteria
2. Deterministic modeling of gene regulatory networks
3. Qualitative modeling of gene regulatory networks
4. Stochastic modeling of gene regulatory networks
5. **Some current issues and perspectives**
   - **Global physiological effects on the dynamics of gene expression**
   - **Strategies for dealing with incomplete information: the case of Drosophila development**

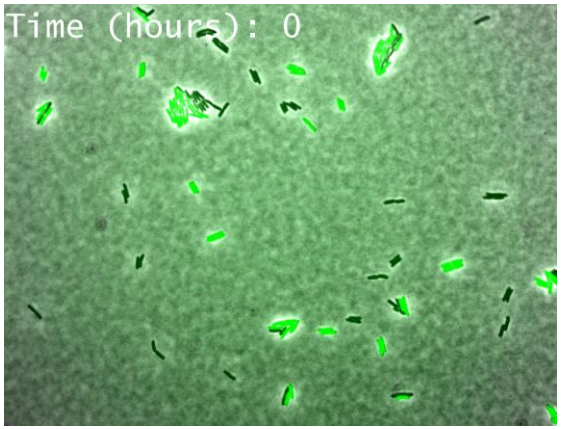
Bacterial growth and metabolism

- Bacteria are geared towards **growth** and **division**

  Escherichia coli cells have doubling times up to 20 min







G. Baptist

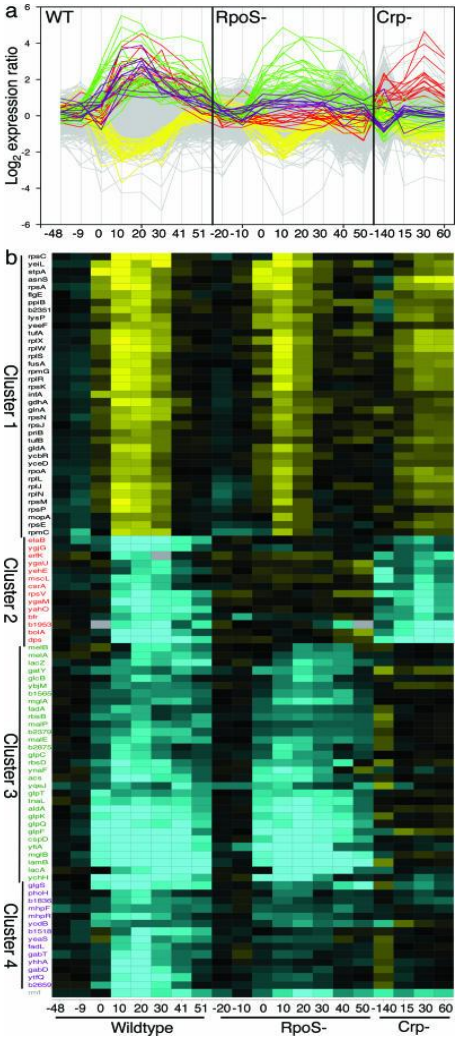
- **Metabolism** fuels growth by production of energy and building blocks for macromolecules, using nutriments from environment

  ATP, amino acids, nucleotides, …

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*



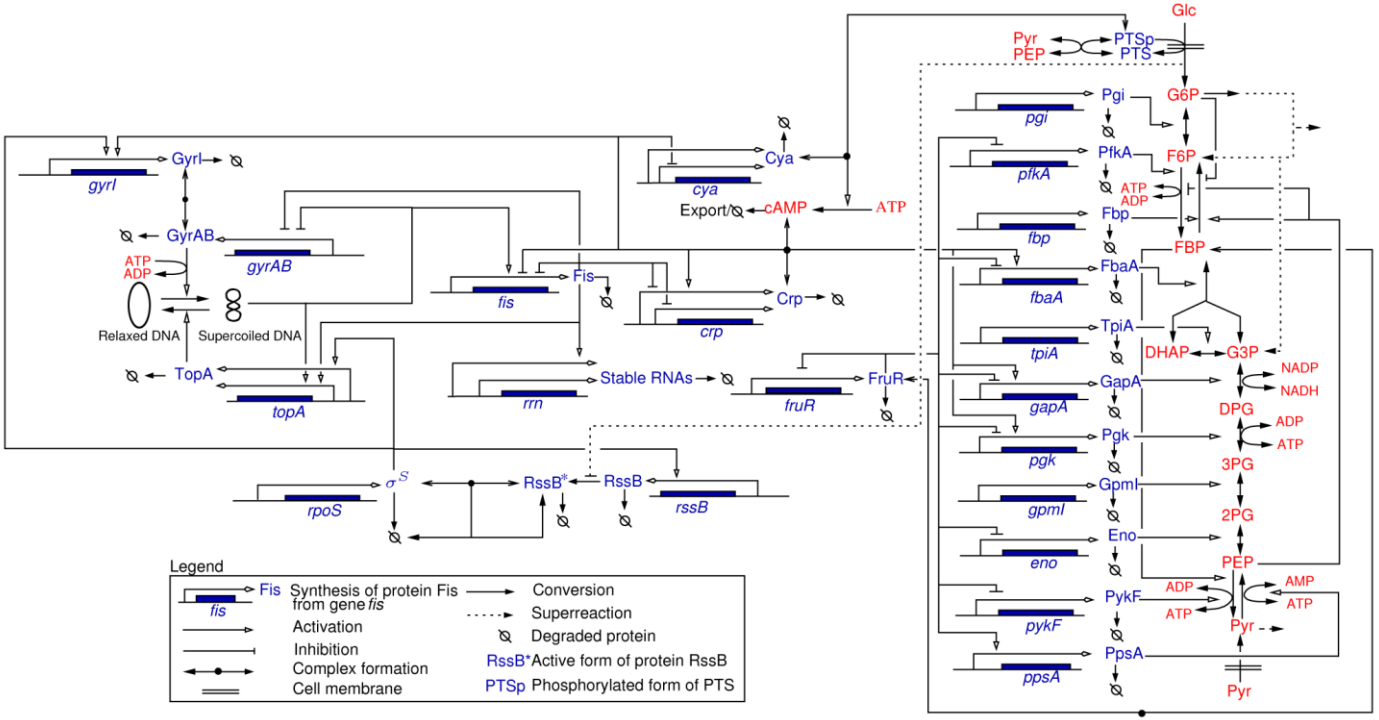
Traxler *et al.* (2006), *Proc. Natl. Acad. Sci. USA*, 103(7):2374–9

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

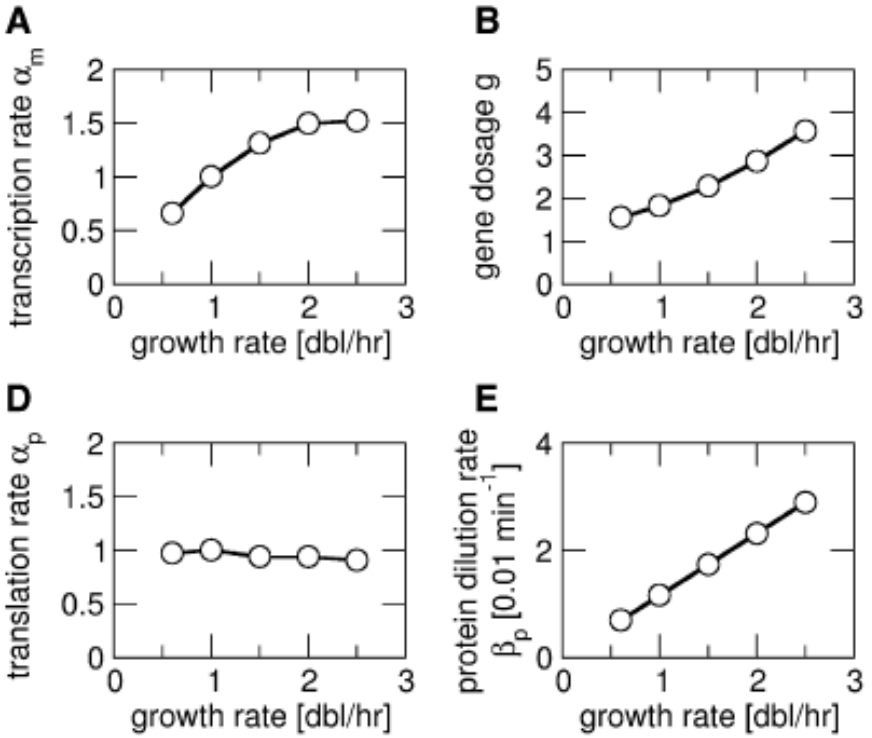


Baldazzi *et al.* (2010), *PLoS Comput. Biol.*, 6(6):e1000812

Growth transition and gene expression

- Adjustment of gene expression also involves global physiological effects

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



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

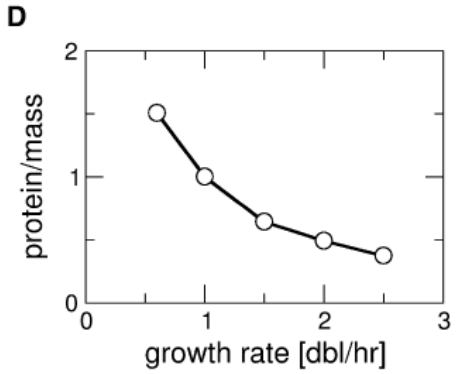
Bremer and Dennis (1996), *Escherichia Coli and Salmonella*, ASM Press, 1553-69

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

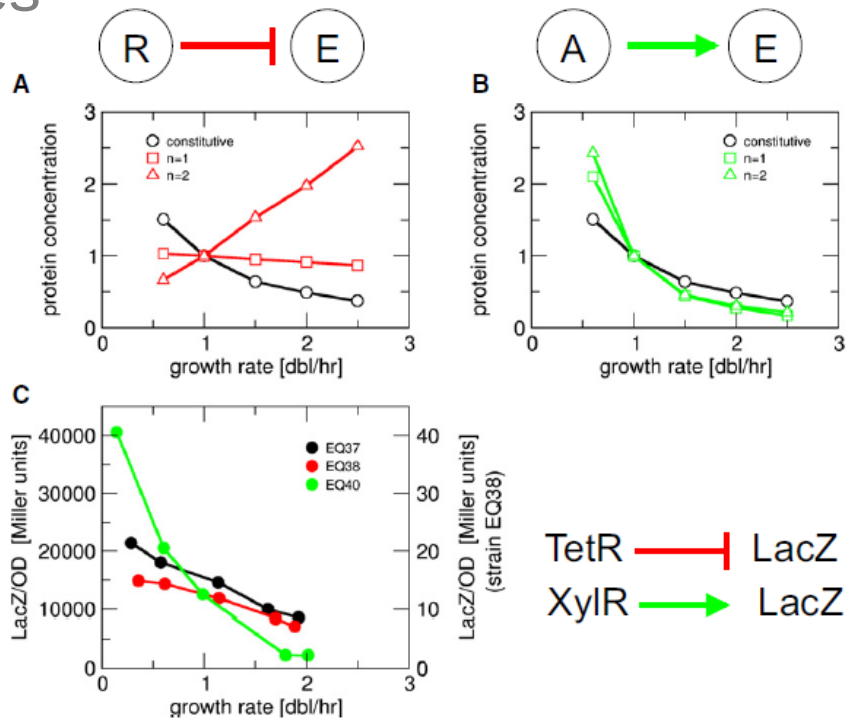
- **Question:** what are relative contributions of specific regulators and global physiological effects in adaptation of gene expression during growth transitions?
- Previous work on growth-rate dependent expression of **constitutive** and regulated genes
  - Constitutive gene: expression is controlled by global physiology, but not by specific transcription factors
  - Expression of constitutive gene is growth-rate dependent



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

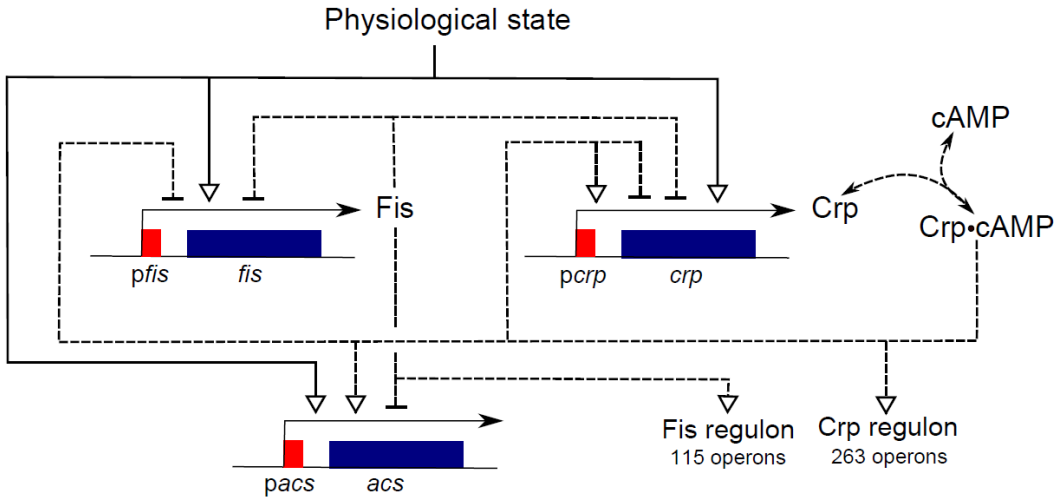
Berthoumieux *et al.* (2013), *Mol. Syst. Biol.*, 9:634

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

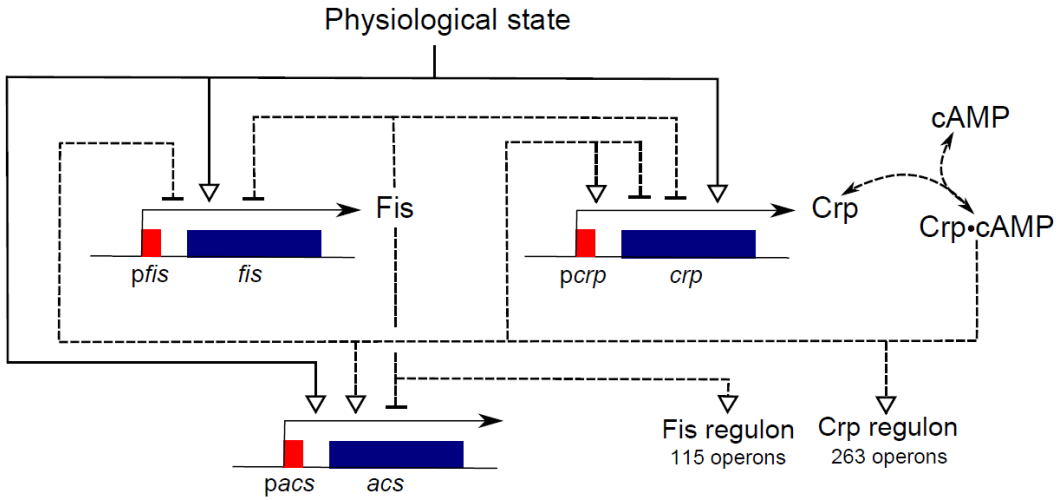
- Question addressed in context of central regulatory circuit of carbon metabolism in *E. coli*



Berthoumieux *et al.* (2013), *Mol. Syst. Biol.*, 9:634

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

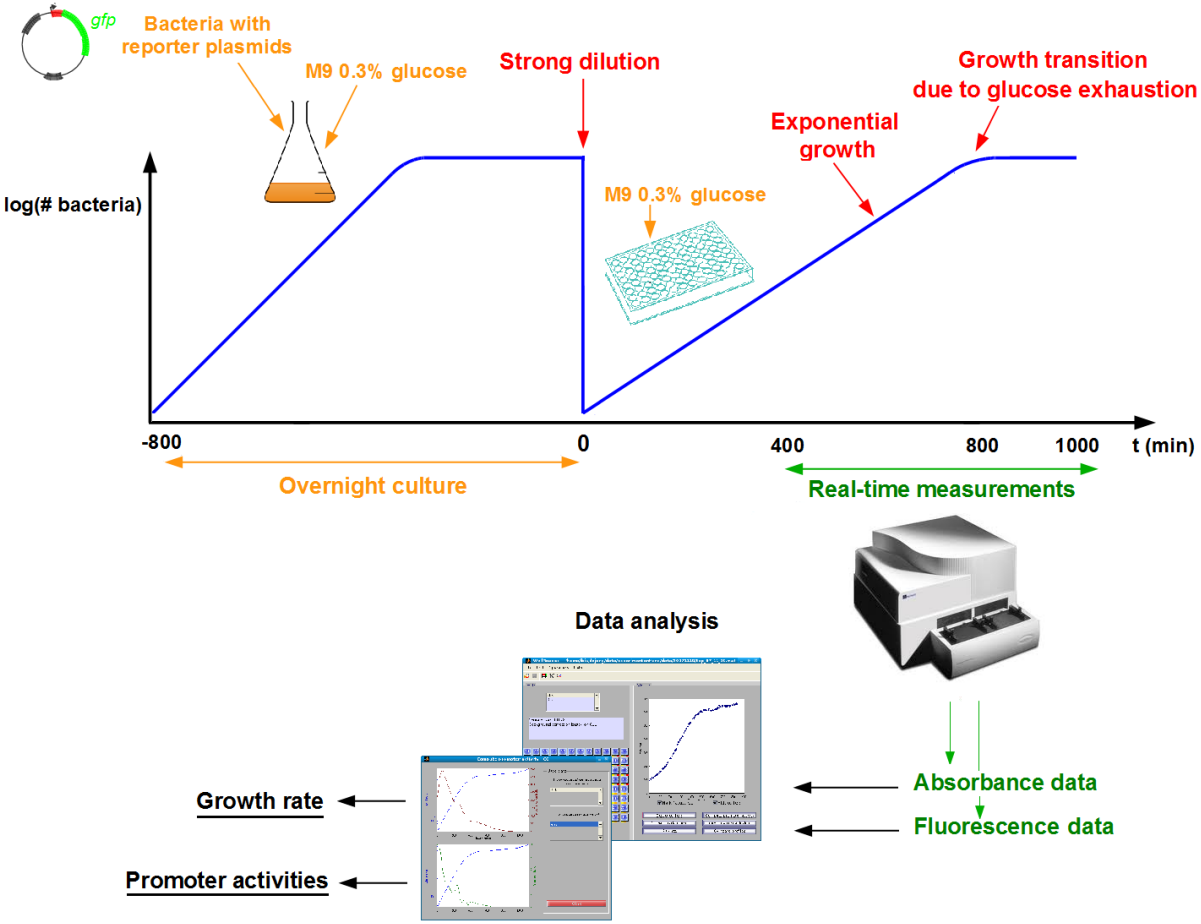


Approach

- **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

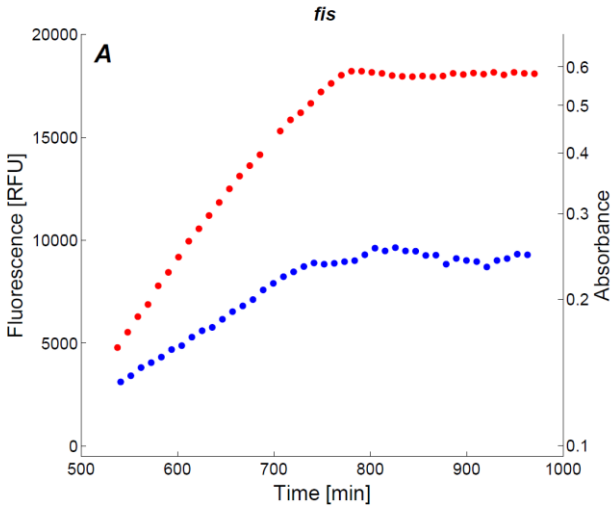


- Transcriptional fusion of promoters with *gfp* reporter genes on plasmid
- Measurement of absorbance and fluorescence signals, thermostated automated microplate reader
- Model-based derivation of promoter activities

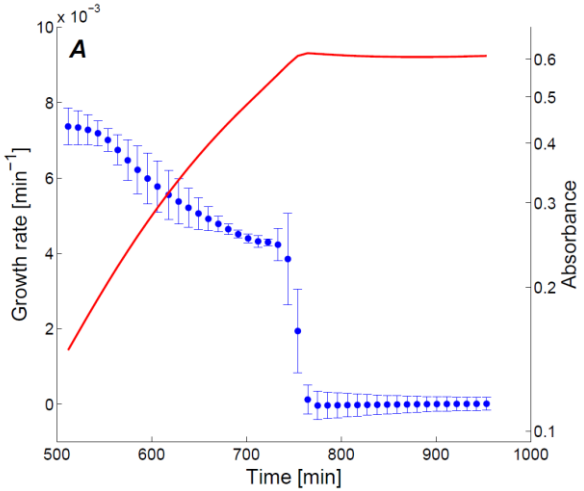
de Jong *et al.* (2010), *BMC Syst. Biol.*, 4:55

Real-time monitoring of gene expression

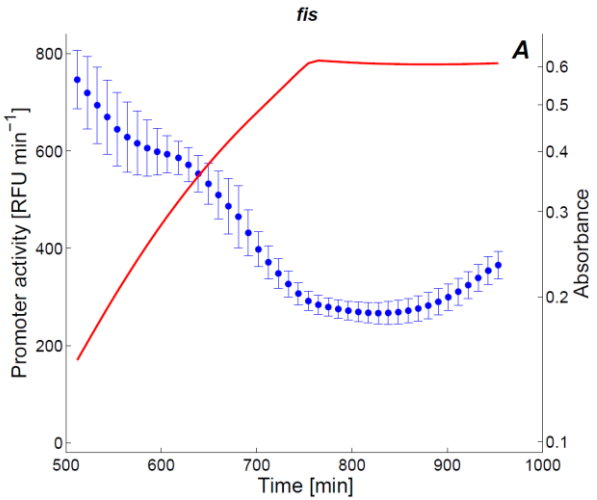
- Monitoring of *fis* promoter activity during growth transition



Absorbance and fluorescence



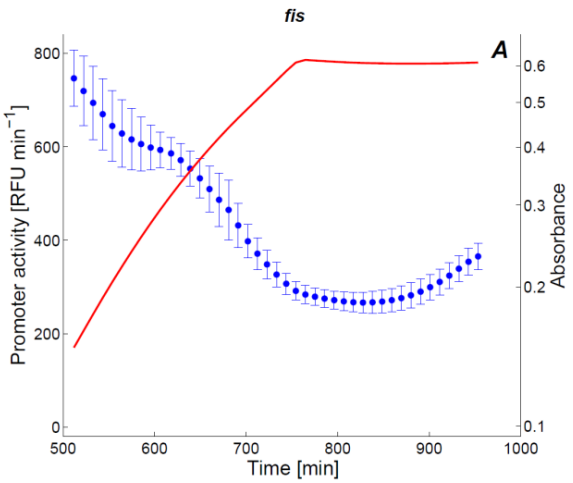
Growth rate

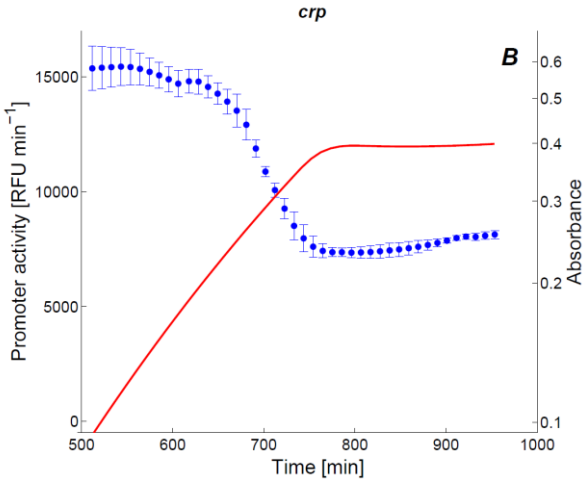


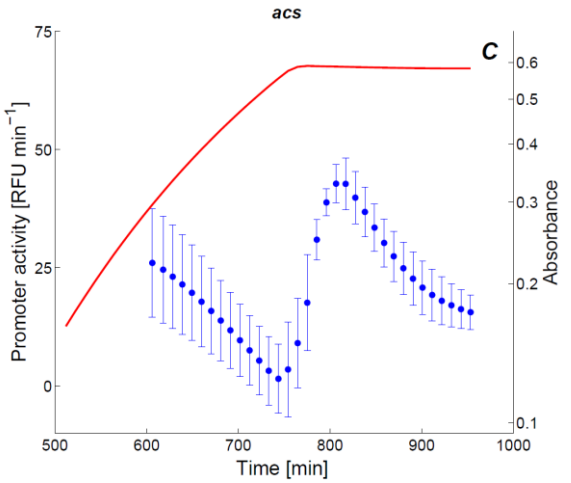
Promoter activity

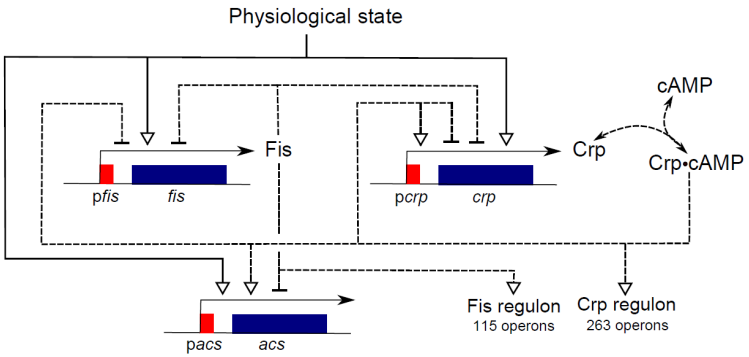
Real-time monitoring of gene expression

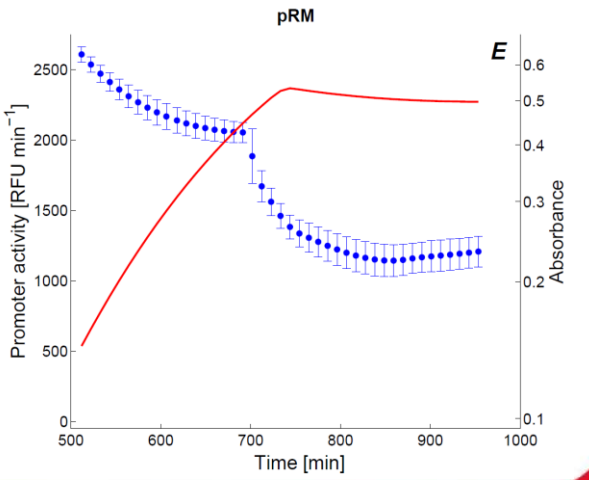
- Monitoring of activity of *crp*, *fis*, *acs* and constitutive phage promoters during growth transition











Bias introduced by plasmid copy number

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

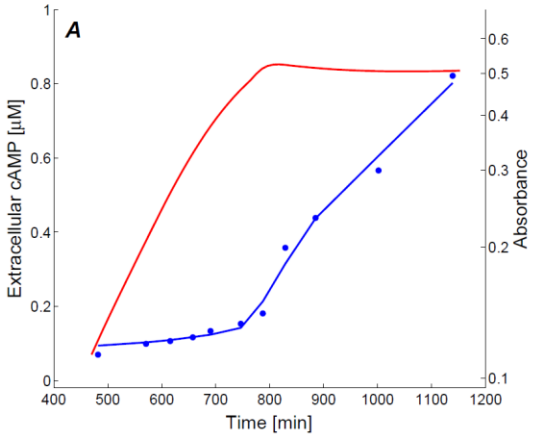
  Lin-Chao and Bremer (1986), *Mol. Gen. Genet.*, 203(1):143-9

- 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

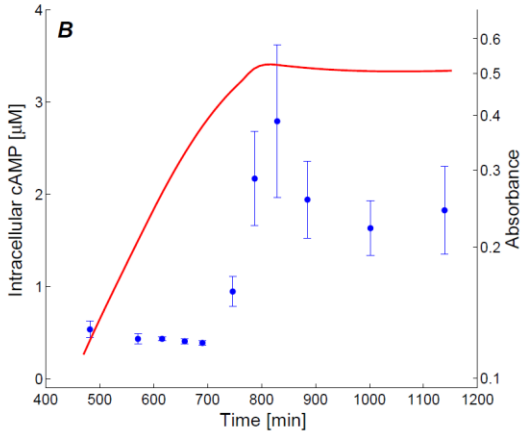
Measurement of cAMP

- 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

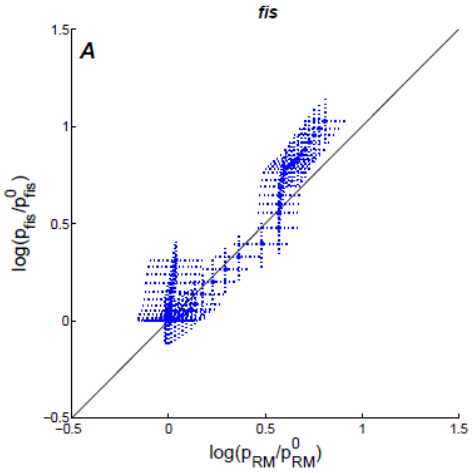
- **Hypothesis 1**: effect of global physiological state (measured by phage promoter) is dominant and effect of specific regulators is negligible ($p_2(t) \approx p_2^0$):

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

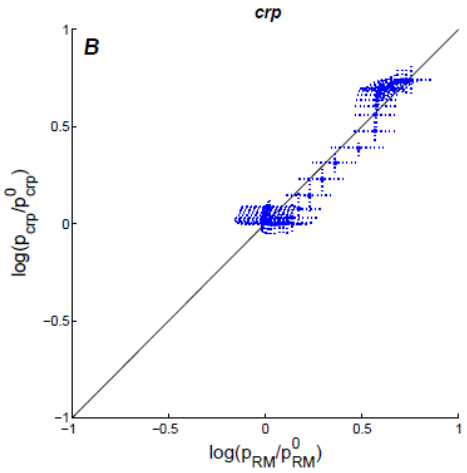
- Advantages of model:
  - Straightforward to test by means of experimental data
  - Non-parametric, does not require model calibration
  - No effect of plasmid copy number variation if promoter activity is measured in same plasmid vector

Test of hypothesis 1

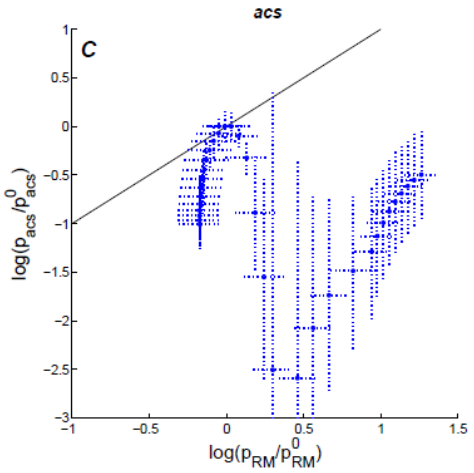
- Global effect is dominant for expression control of transcription factors (*crp* and *fis*), but not for metabolic gene (*acs*)



$R^2 = 0.93$



$R^2 = 0.96$



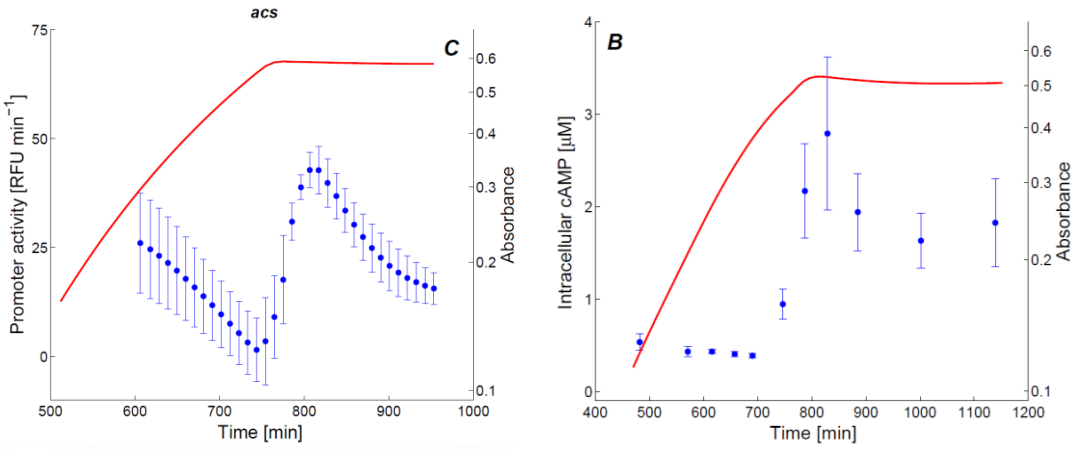
$R^2 = 0.08$

Model of promoter activities

- **Hypothesis 2**: effect of specific regulators is not negligible and can be reduced to effect of change in cAMP concentration $c(t)$:

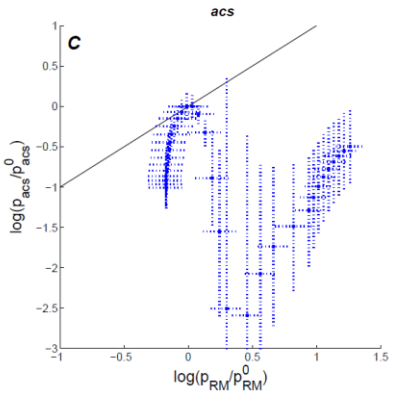
$$\log \frac{p(t)}{p^0} - \log \frac{p_{RM}(t)}{p^0_{RM}} = \log \frac{c(t)}{c^0}$$

- Hypothesis based on data, but biological assumptions underlying simplification can be explicitly formulated

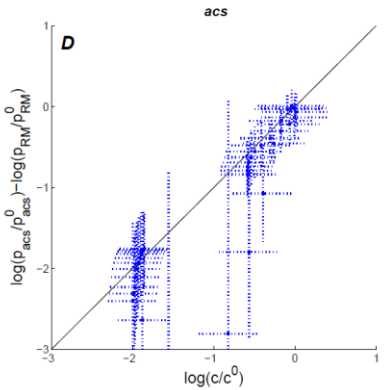


Test of hypothesis 2

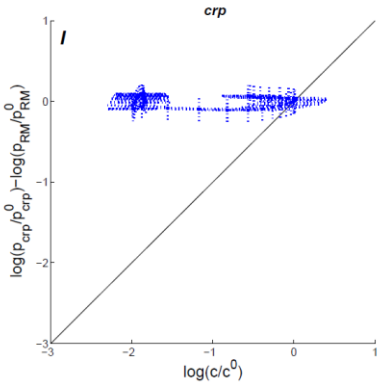
- Combination of global effect and specific effect of cAMP explains variation in *acs* promoter activity



$R^2 = 0.08$



$R^2 = 0.93$

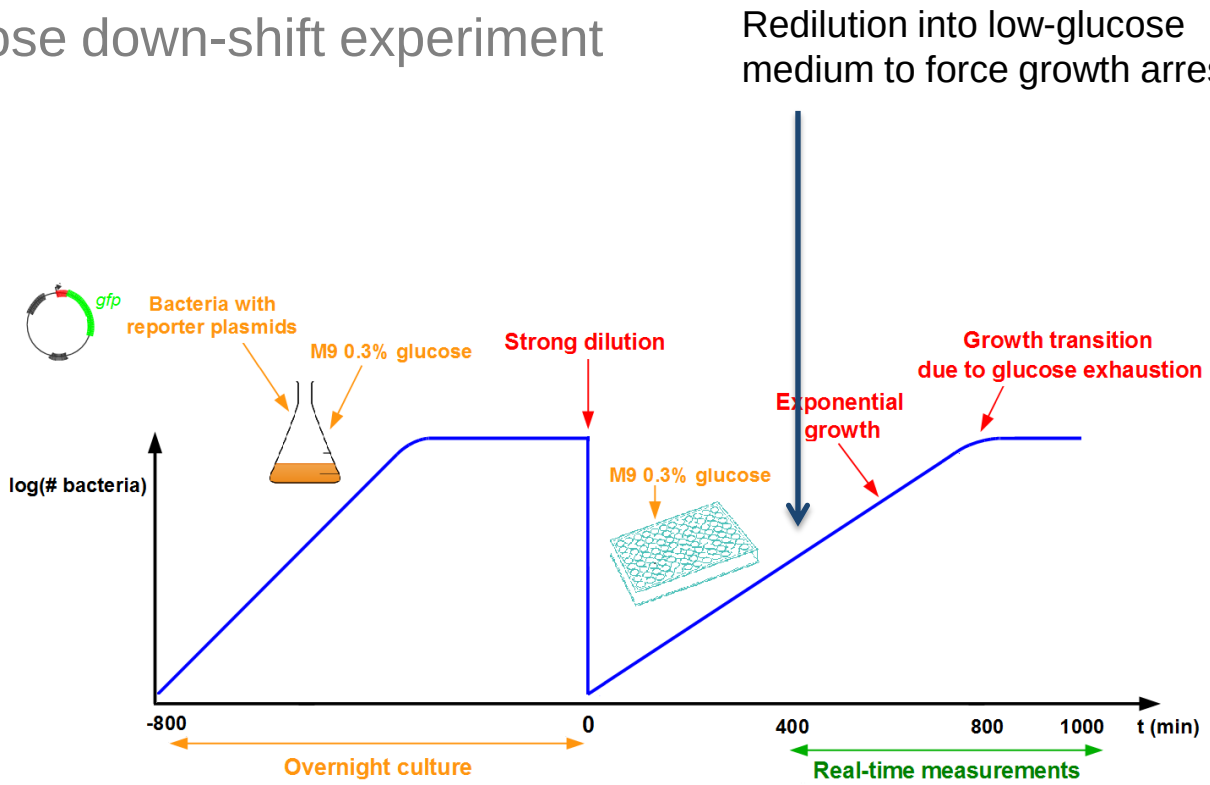


$R^2 = 0.03$

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

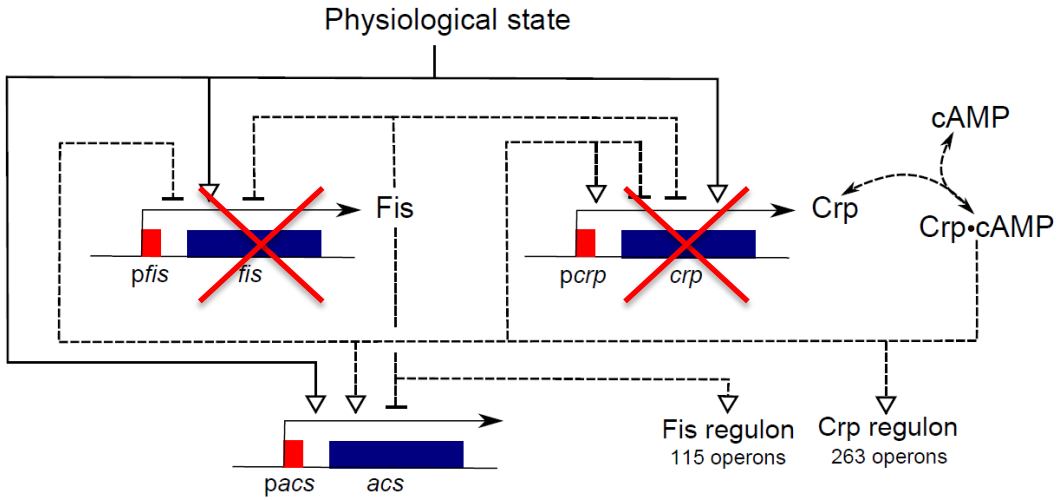
Other experimental conditions

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



Other experimental conditions

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



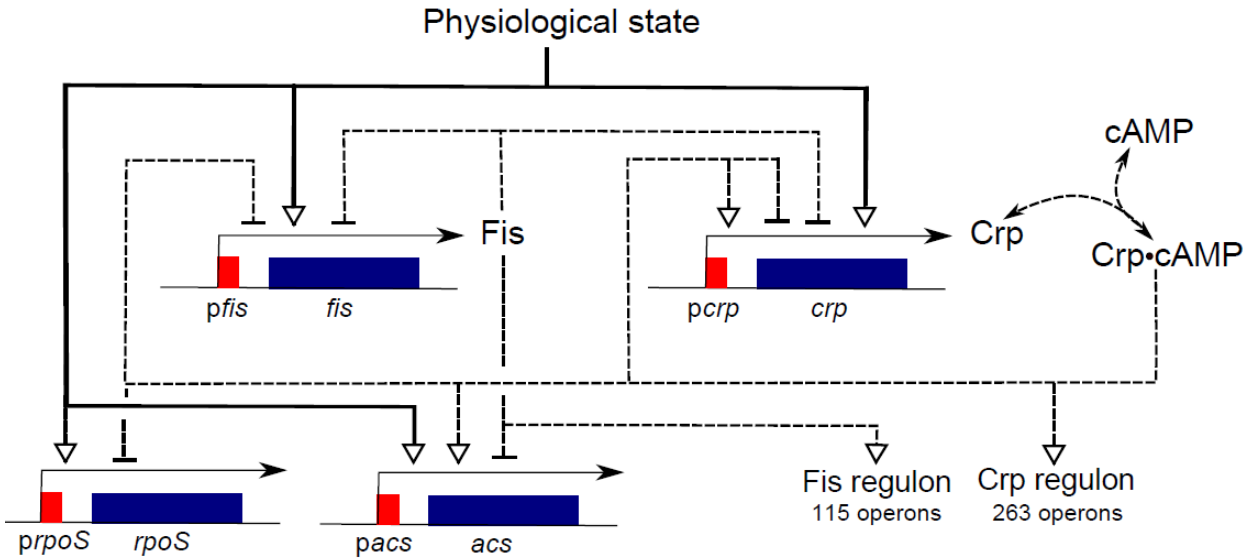
Other experimental conditions

- 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

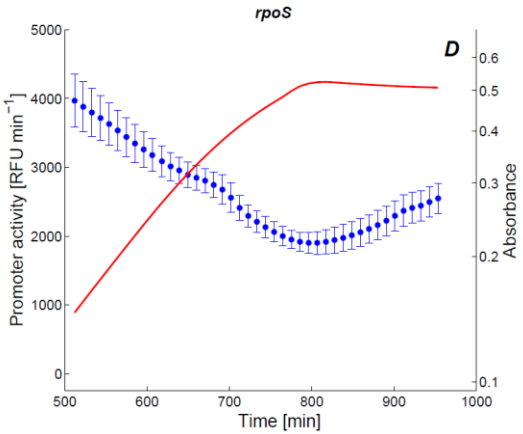


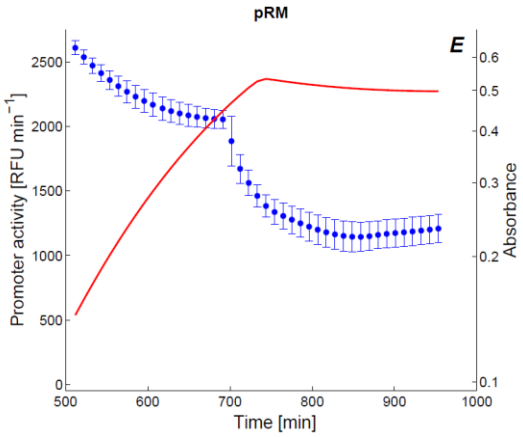
Other regulators

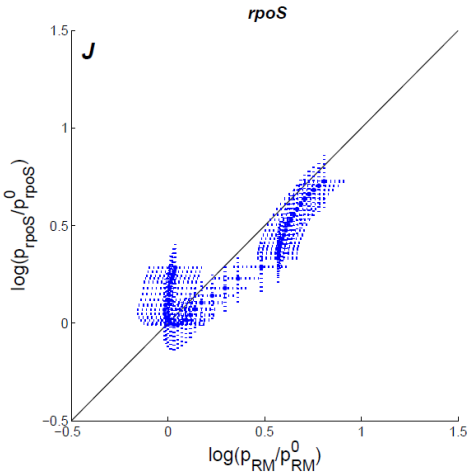
- Is effect of global physiologal state also dominant in transcriptional control of other regulators?

  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



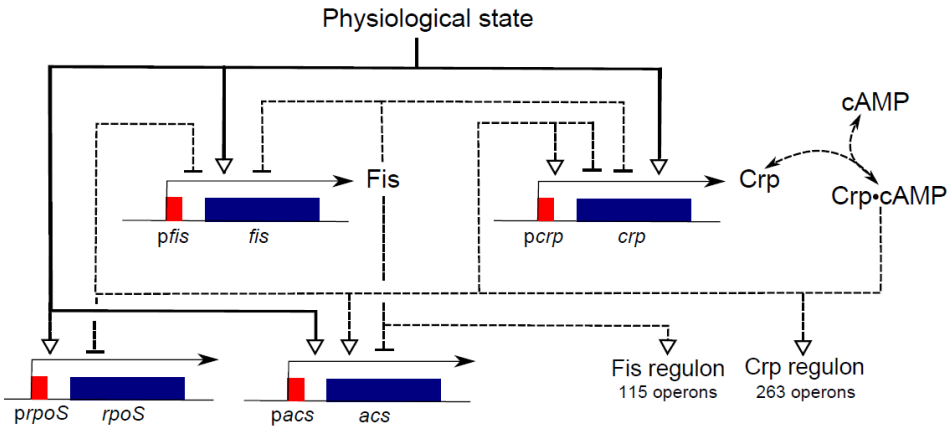




$R^2 = 0.84$

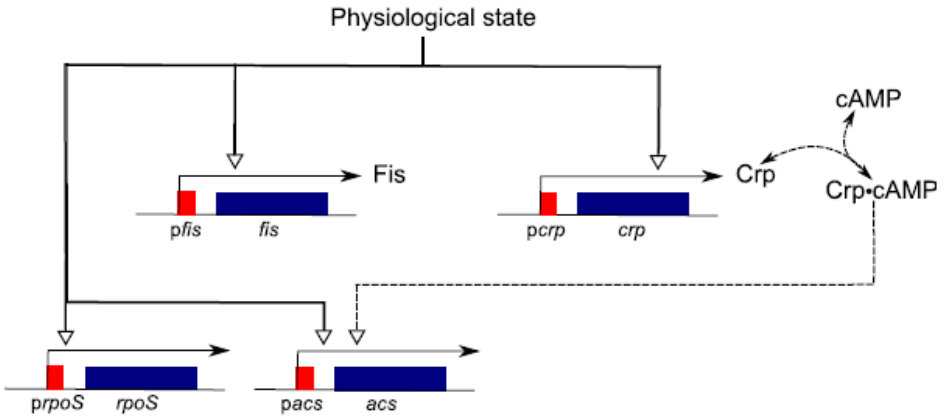
Conclusions

- Control of gene expression across growth phases is shared between global physiological state 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

- Transcriptional control of genes encoding 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
  - Choice of growth conditions? Weak effects?



Berthoumieux *et al.* (2013), *Mol. Syst. Biol.*, 9:634

Conclusions

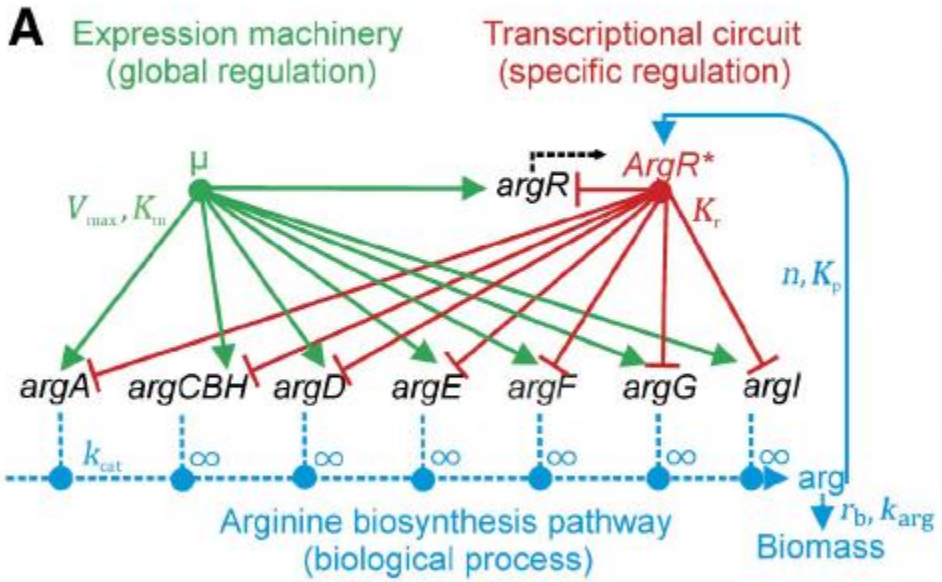
- Results question central role often attributed to transcriptional regulatory networks
- **Alternative view**: specific effects complement and finetune global control exerted by physiological state, notably gene expression machinery

Gerosa *et al.* (2013), *Mol. Syst. Biol.*, 9:658

- Consequences for interpretation of transcriptome data and design of synthetic circuits

Conclusions

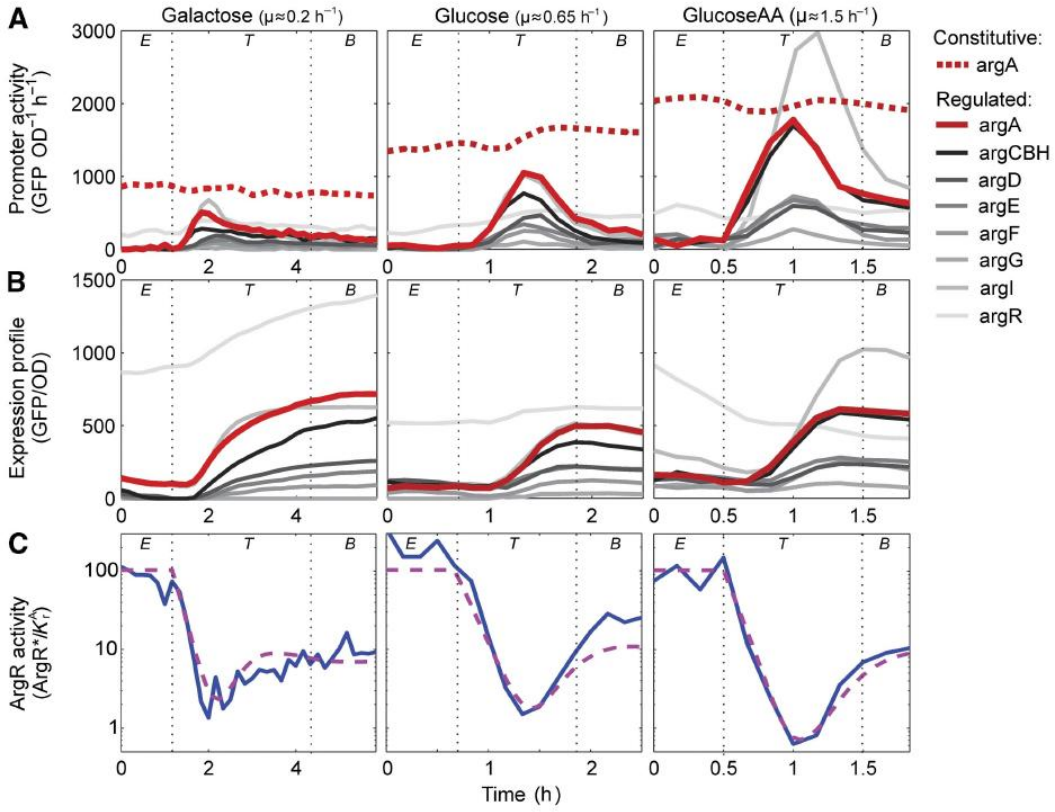
- Analysis generalizable to other networks?
  - Dissection of control of arginine biosynthesis genes in *E. coli*



Gerosa *et al.* (2013), *Mol. Syst. Biol.*, 9:658

Conclusions

- Analysis generalizable to other networks?
  - Dissection of control of arginine biosynthesis genes in *E. coli*
  - Control of gene expression is combination of global physiological state and specific effects (arginine concentration)

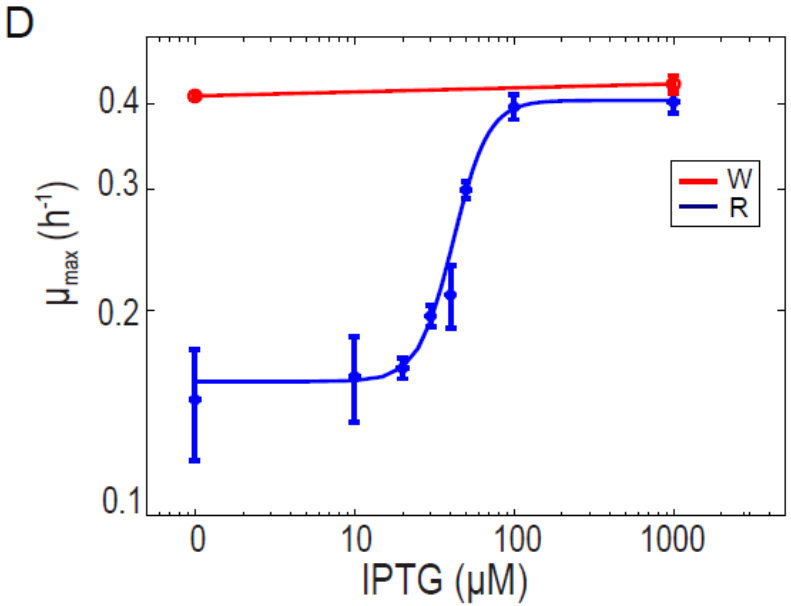


Gerosa *et al.* (2013), *Mol. Syst. Biol.*, 9:658

Perspectives

- Can we control global physiological state, and thus gene expression program of the cell?

  Engineering of *E. coli* genome to put transcriptional machinery under control of inducible promoter



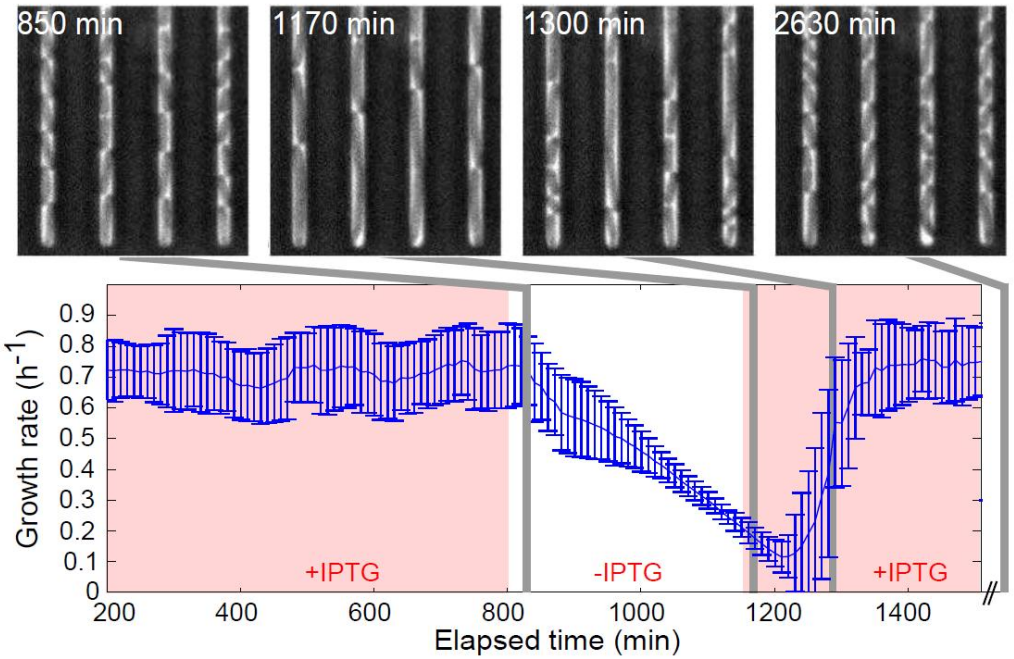
Izard *et al.*, submitted for publication

- Finetuning of growth rate…

Perspectives

- Can we control global physiological state, and thus gene expression program of the cell?

  Engineering of *E. coli* genome to put transcriptional machinery under control of inducible promoter



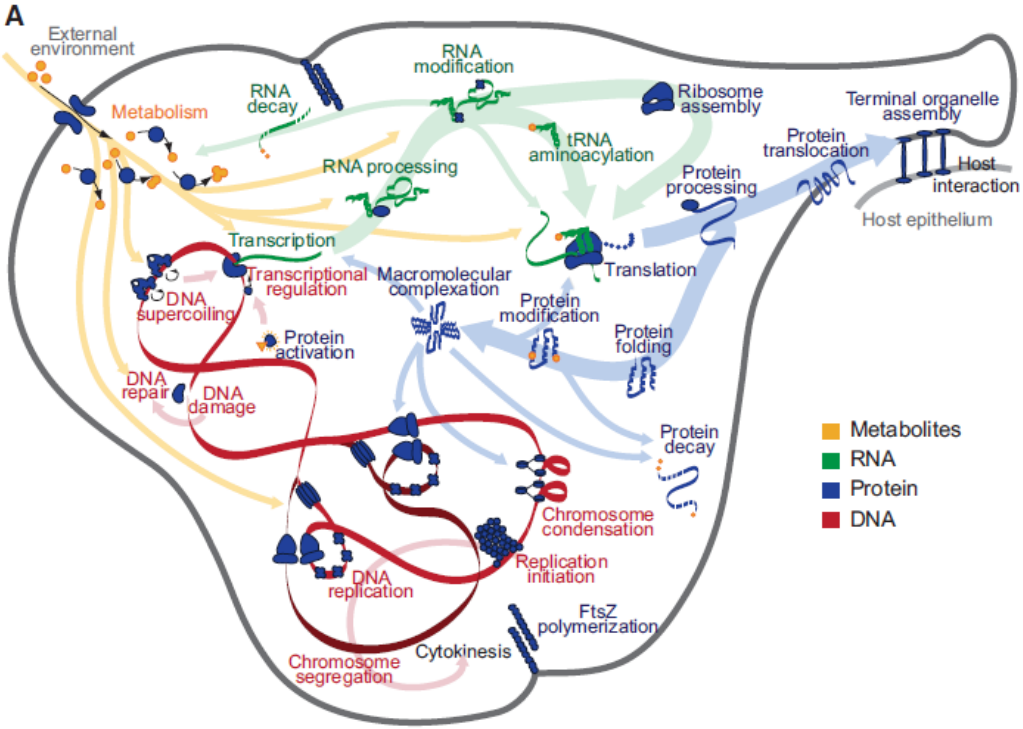
Izard *et al.*, submitted for publication

Collaboration with A. Lindner

- Finetuning of growth rate … in reversible way

Whole-cell model *M. genitalium*

- Metabolic networks are integrated with gene networks and signalling networks
  - Complex multi-level system with feedback across different time-scales

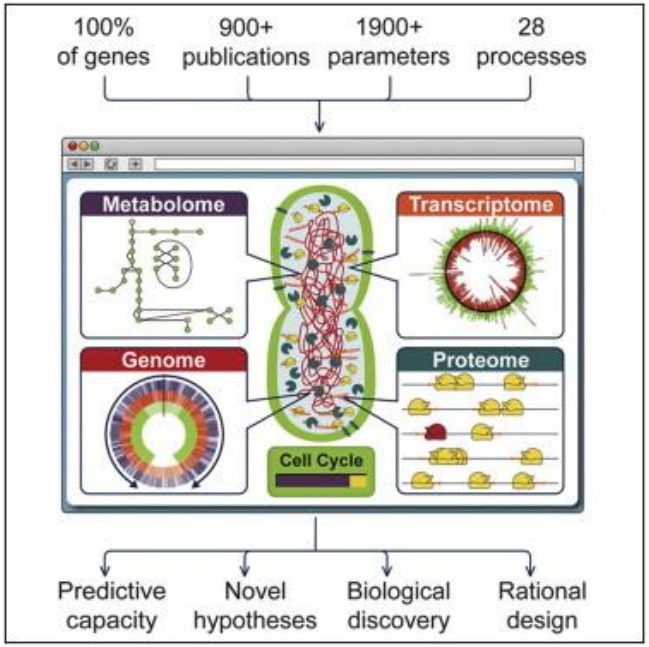


Whole-cell model of *Mycoplasma genitalium*

Karr *et al.* (2012), *Cell*, 150(2): 389-401

Whole-cell model *M. genitalium*

- Whole-cell model represents huge modelling effort:
  - Whole-genome model including **complete** known metabolic, gene, and signalling networks



Karr *et al.* (2012), *Cell*, 150(2): 389-401

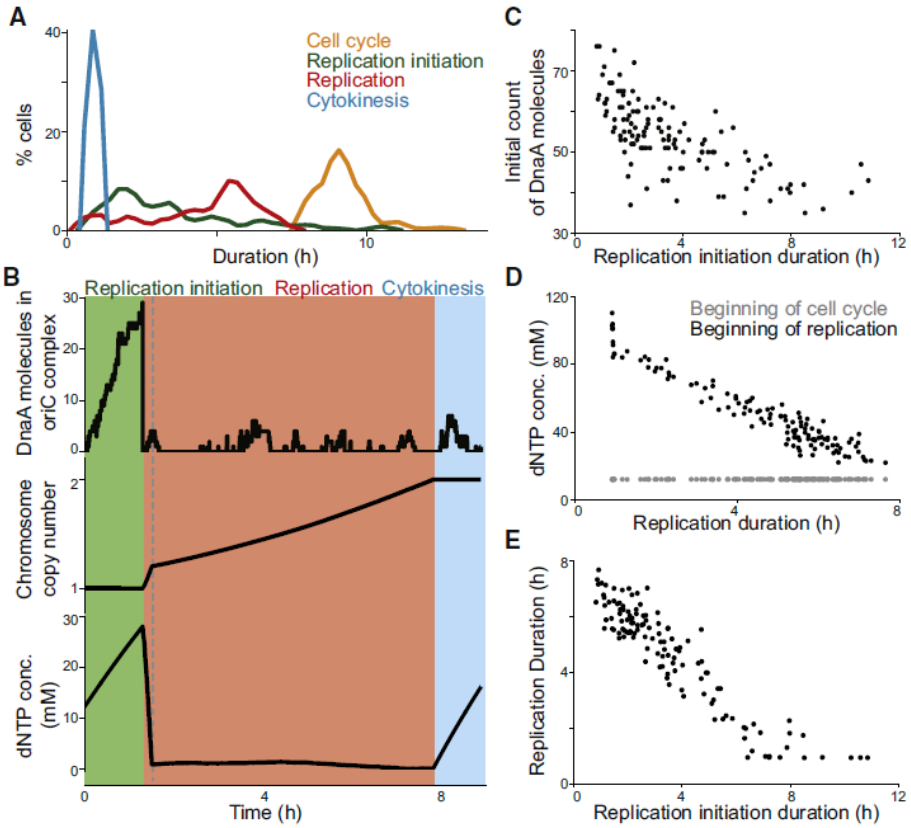
  - Variety of **formalisms** to model the 28 modules: FBA, kinetic ODE models, Boolean models, Markov chains, …
  - Cell cycle simulated for >100 cells, >30 mutants on 128-core machine

Whole-cell model *M. genitalium*

- Whole-cell simulation of *M. genitalium* cell cycle

Whole-cell model *M. genitalium*

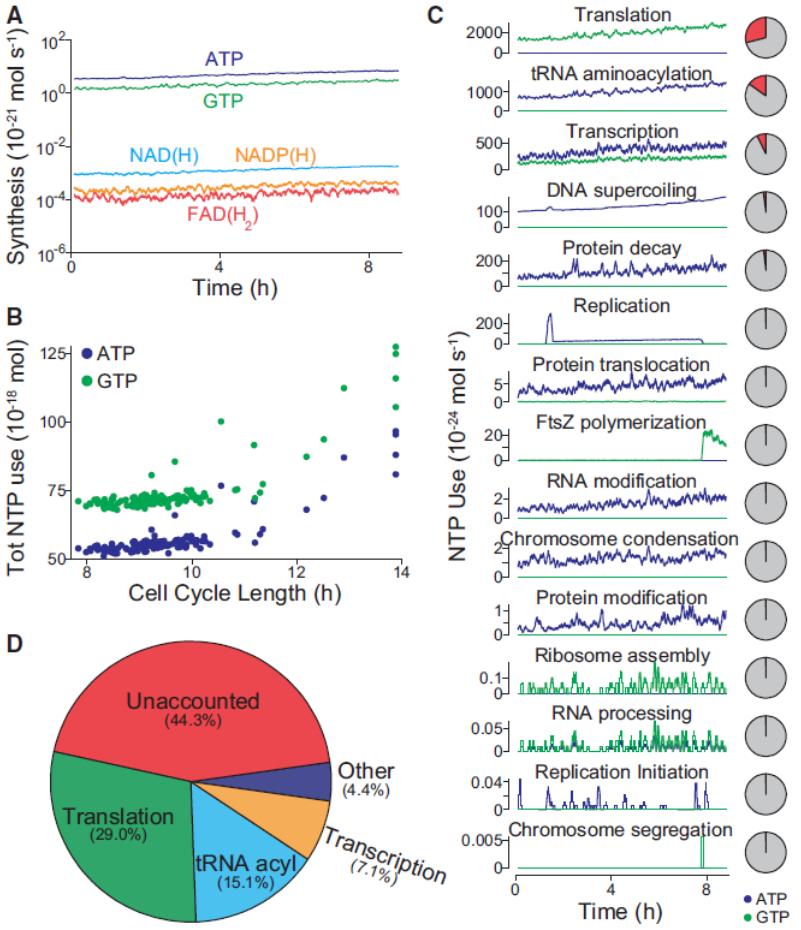
- Whole-cell simulations have provided new insights into **robustness of cell-cycle duration**
  - High variability of replication initiation buffered by dNTP-dependent duration of replication
  - This metabolic control of replication leads to decreased variability of cell-cycle length



Karr *et al.* (2012), *Cell*, 150(2): 389-401

Whole-cell model *M. genitalium*

- Whole-cell simulations have provided new insights into **global use and allocation of energy**
  - Transcription and translation most costly processes
  - Energy use largely independent of cell-cycle length
  - Usage of almost half of produced energy not accounted for!



Karr *et al.* (2012), *Cell*, 150(2): 389-401

Lack of quantitative information: strategies

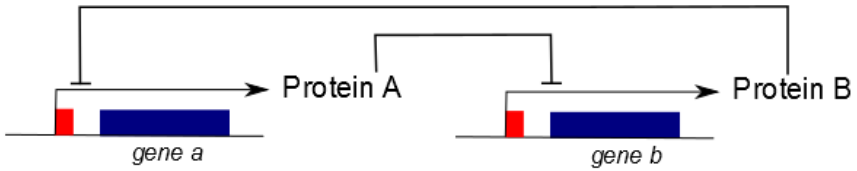
- Three main strategies to deal with lack of quantitative data:
  - **Test of parameter sensitivity**
  - Model reduction and simplification
  - Parameter estimation from time-series data

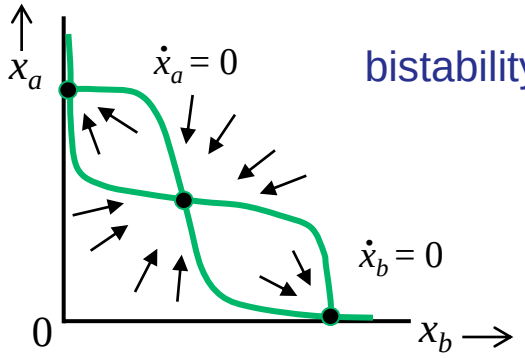
De Jong and Ropers (2006), *Brief. Bioinform.*, 7(4):354-363

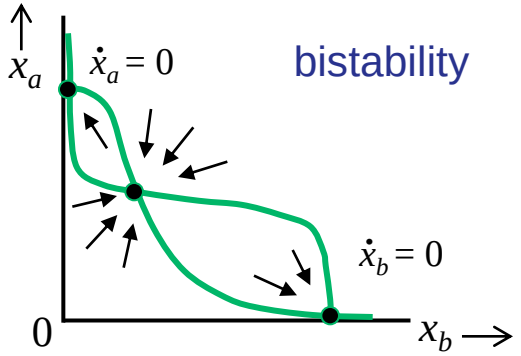
Test of parameter sensitivity

- 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



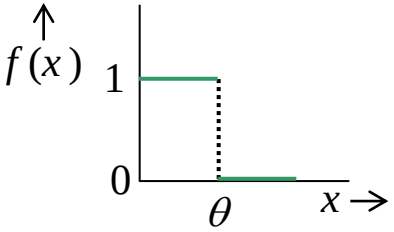


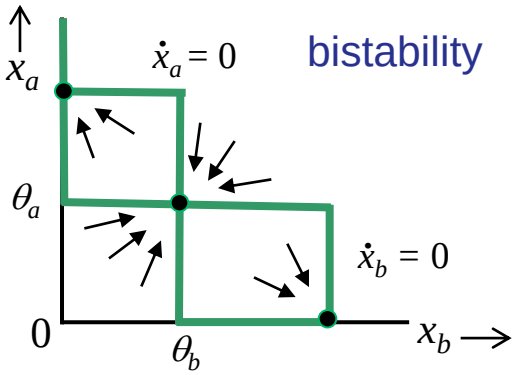


Model reduction and simplification

- 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

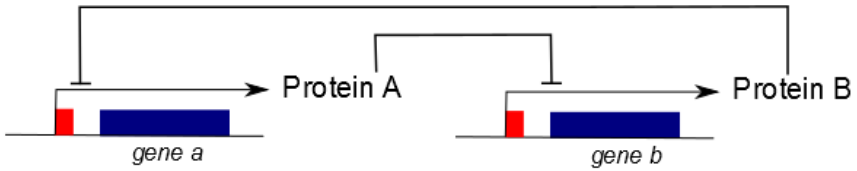
Parameter estimation

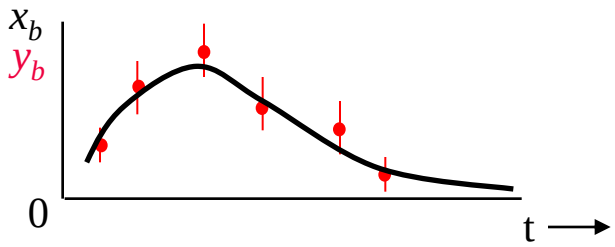
- **Estimate** parameter values from experimental time-series data

  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

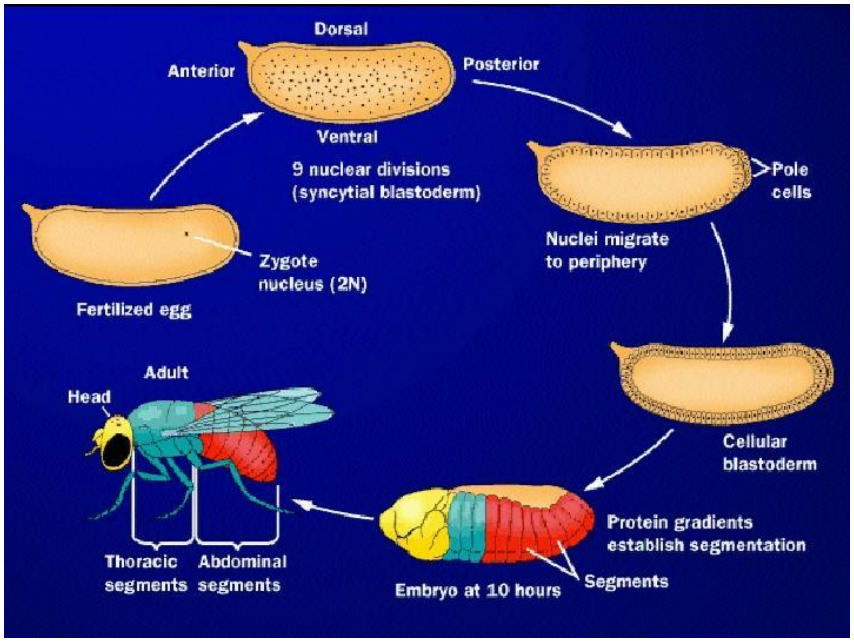
De Jong and Ropers (2006), *Brief. Bioinform.*, 7(4):354-363

- Illustration: models of developmental processes in multicellular organisms

  Development of *Drosophila* embryon

Development of *Drosophila* embryon

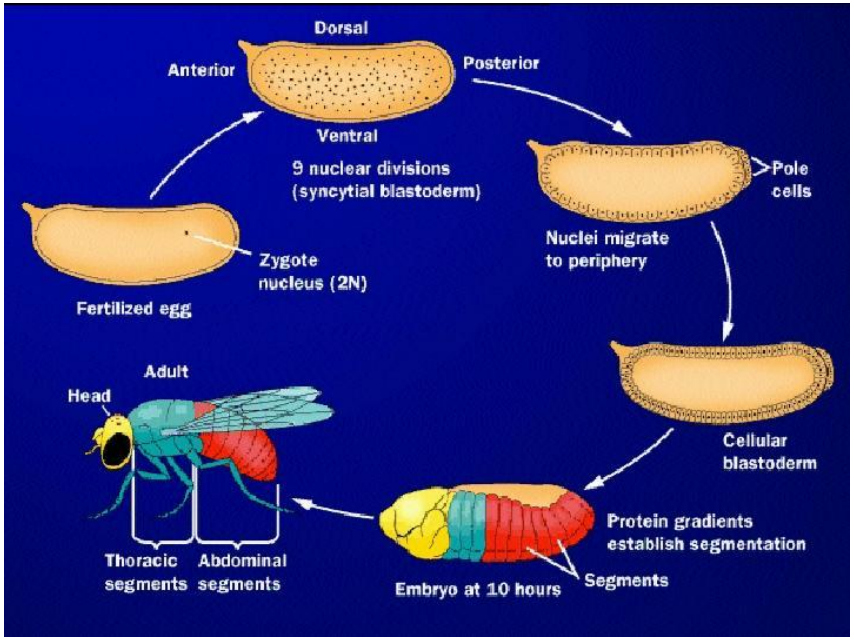
- Development of *Drosophila melanogaster* (fruit fly)



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

Development of *Drosophila* embryon

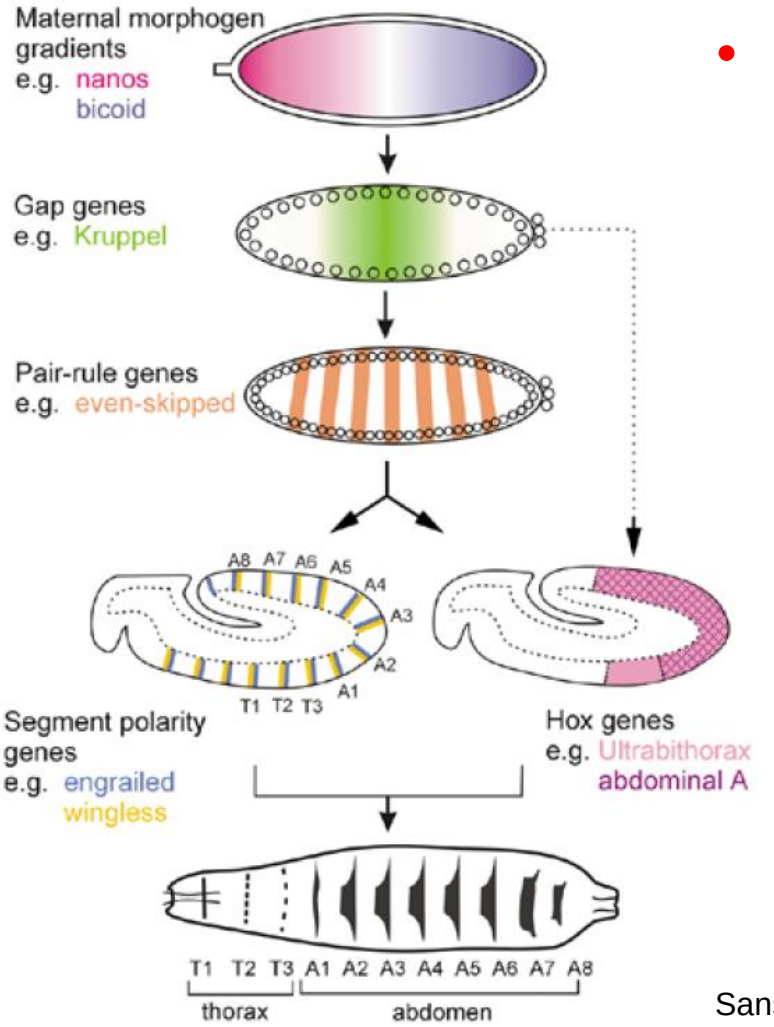
- Development of *Drosophila melanogaster* (fruit fly)



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

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

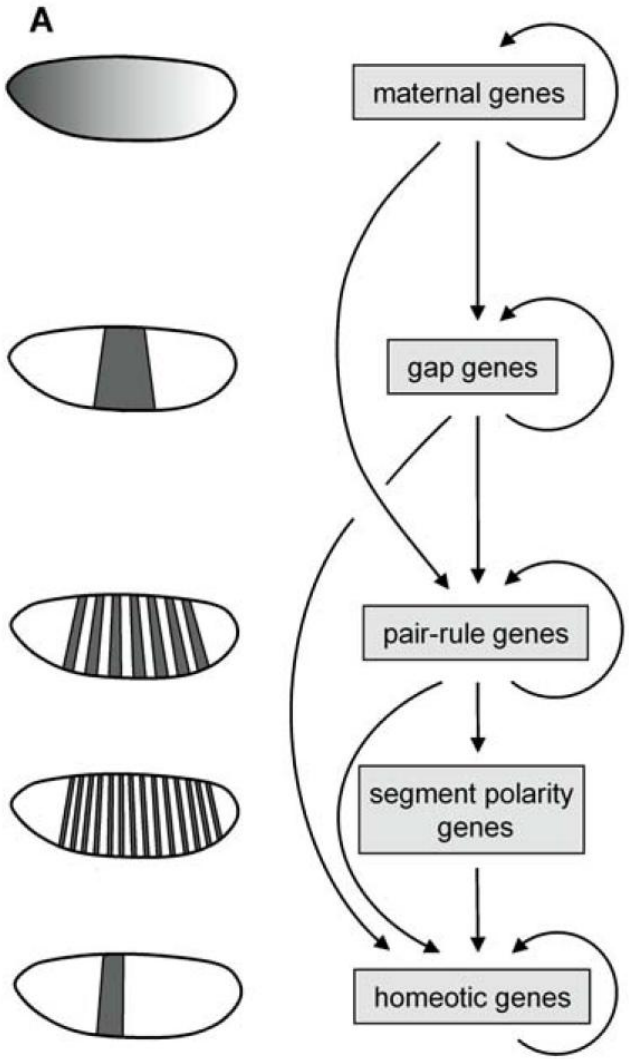
Development of *Drosophila* embryon



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

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

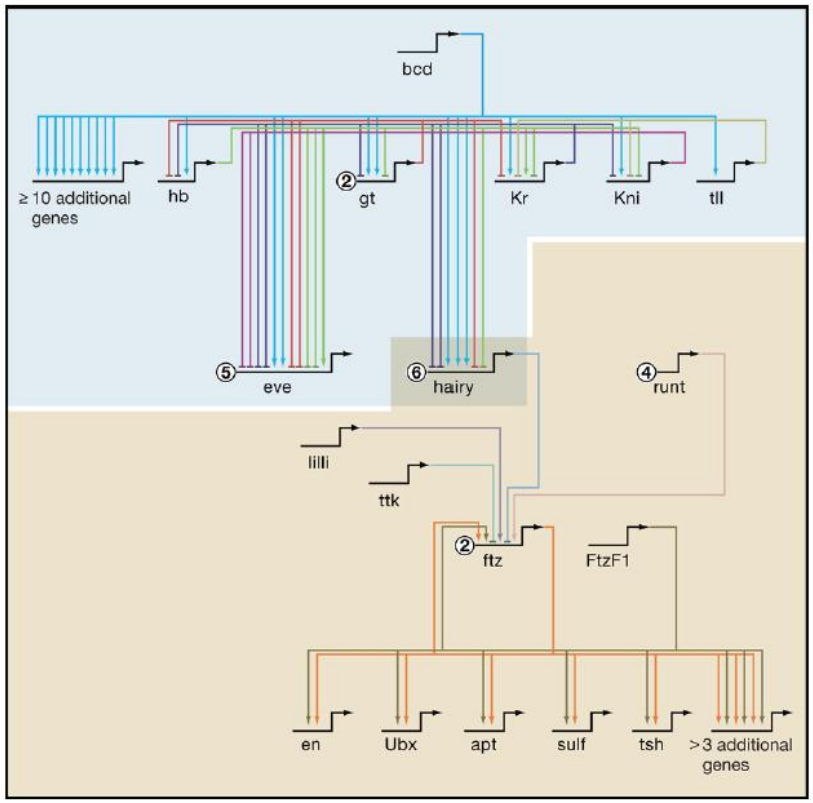
Development of *Drosophila* embryon



- Spatiotemporal gene expression patterns during early development of *Drosophila* (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

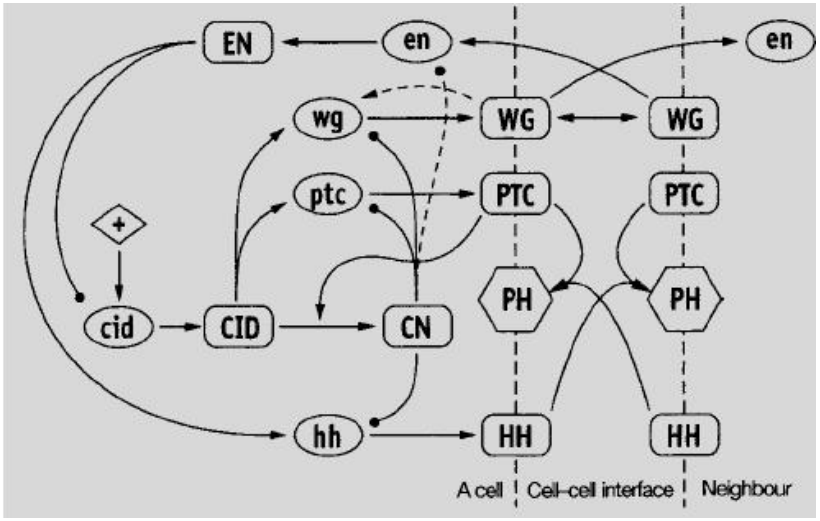


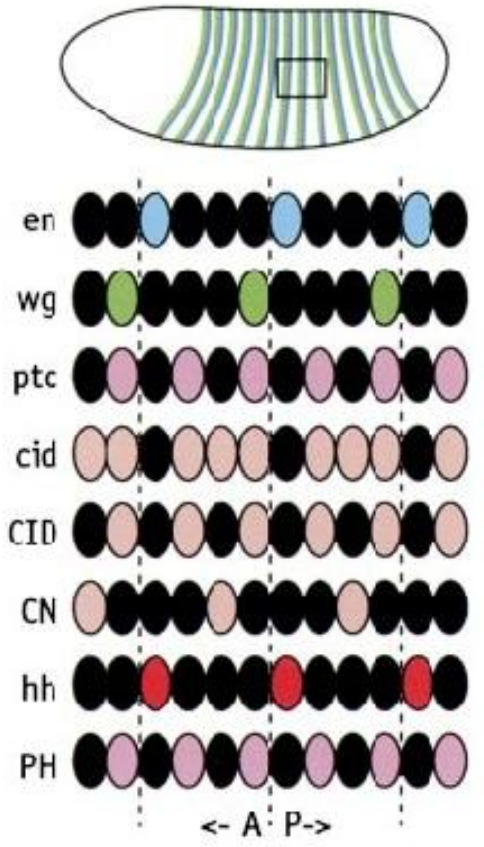
- Spatiotemporal gene expression patterns during early development of *Drosophila* (fruit fly)
- Gene classes and their interactions responsible for establishment of gene expression patterns
- Complex gene regulatory networks

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

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

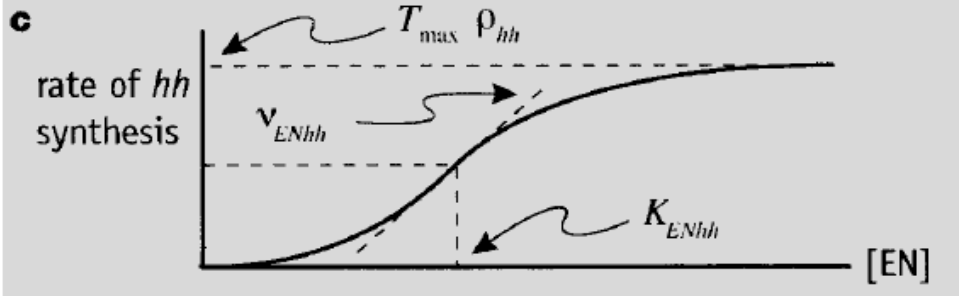
b

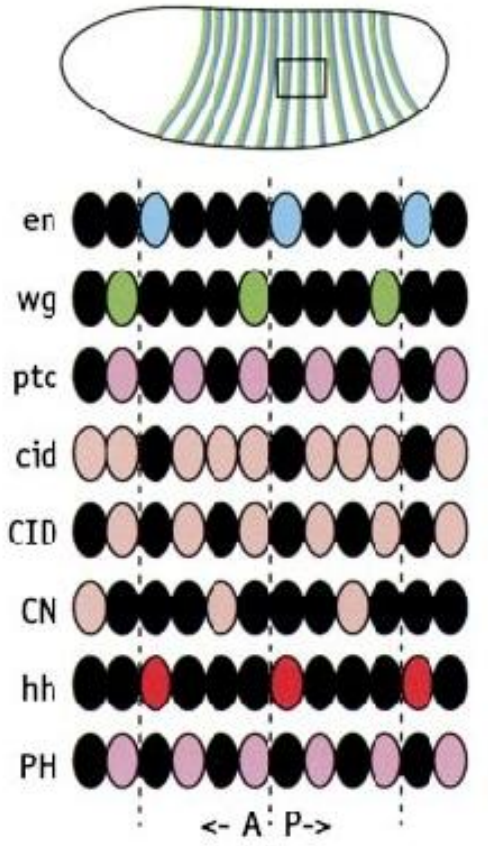
$$\frac{d[hh]_i}{dt} = T_{\max}\rho_{hh}\left[\frac{[EN]_i^{\nu_{ENhh}}}{K_{ENhh}^{\nu_{ENhh}} + [EN]_i^{\nu_{ENhh}}}\right] - \frac{[hh]_i}{H_{hh}}$$

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

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

c

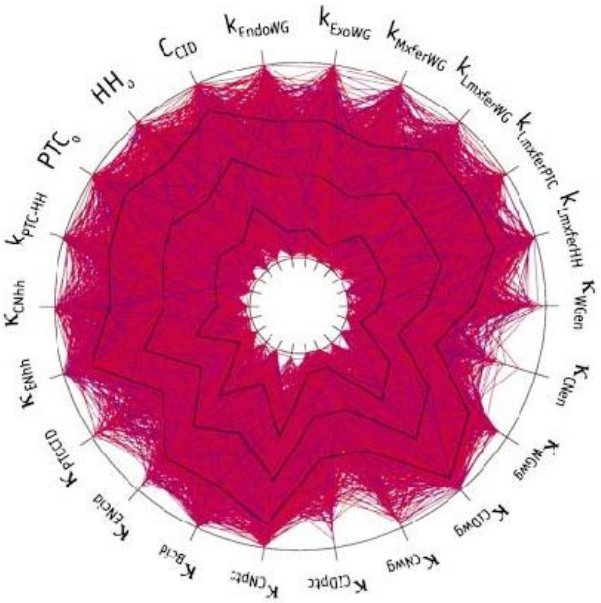


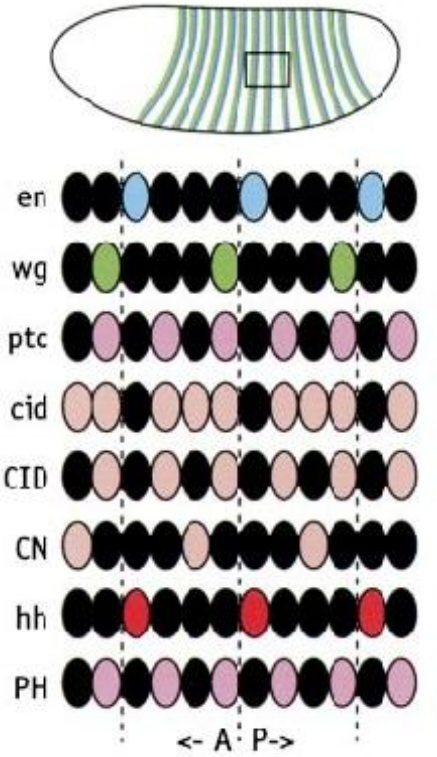


von Dassow *et al.* (2000), *Nature*, 406(6792): 188-92

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



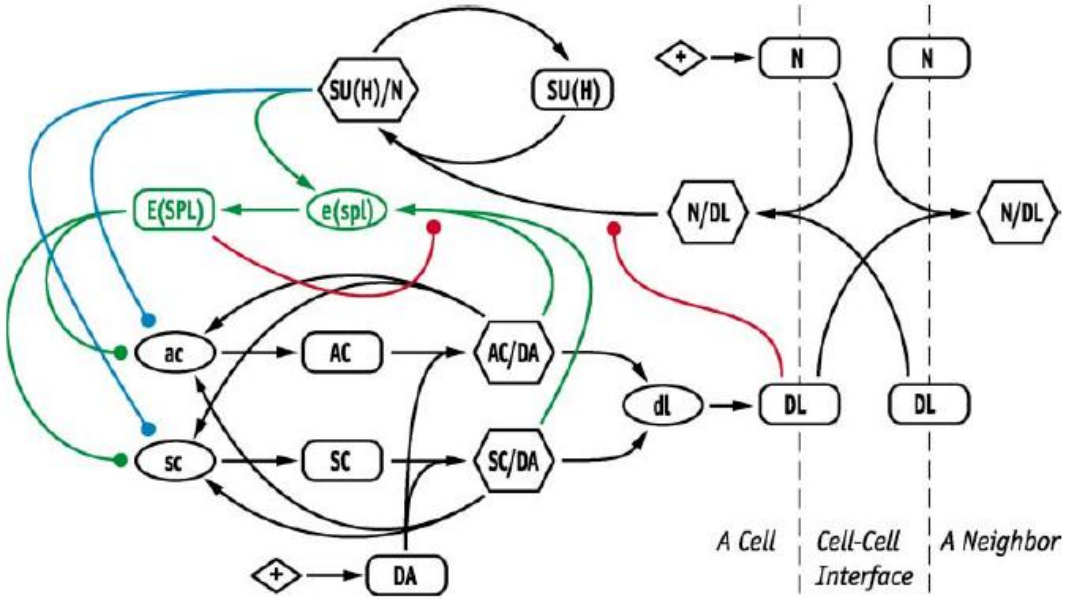


von Dassow *et al.* (2000), *Nature*, 406(6792): 188-92

Robustness of gene expression patterns

- Robustness of model predictions to variations in parameter values confirmed for other developmental networks

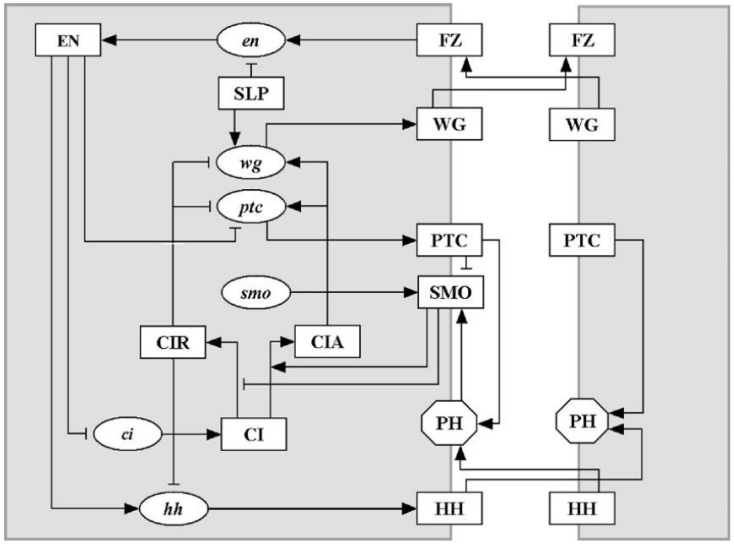
  Neurogenic network, determining neuroblasts in embryos and sensory organ precursor cells in imaginary disks



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



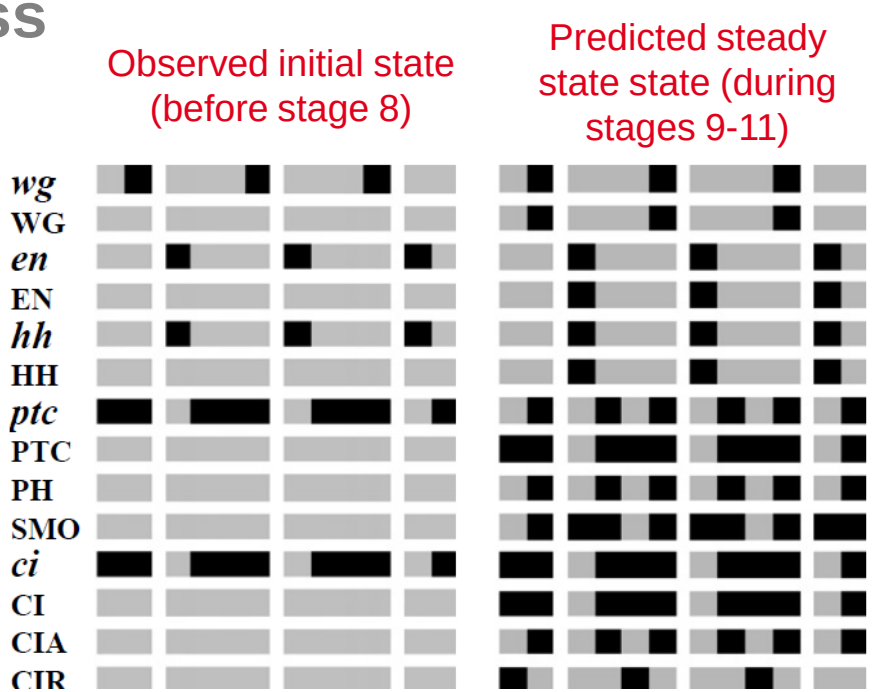
Albert and Othmer (2003), *J. Theor. Biol.*, 223(1):1-18

hh_i
HH_i
ptc_i
PTC_i
PH_i

$$hh_i^{t+1} = EN_i^t \text{ and not } CIR_i^t$$
$$HH_i^{t+1} = hh_i^t$$
$$ptc_i^t = CIA_i^{t+1} \text{ and not } EN_i^t \text{ and not } CIR_i^t$$
$$PTC_i^{t+1} = ptc_i^t \text{ or } (PTC_i^t \text{ and not } HH_{i-1}^t \text{ and not } HH_{i+1}^t)$$
$$PH_i^t = PTC_i^t \text{ and } (HH_{i-1}^t \text{ or } HH_{i+1}^t)$$

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**



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

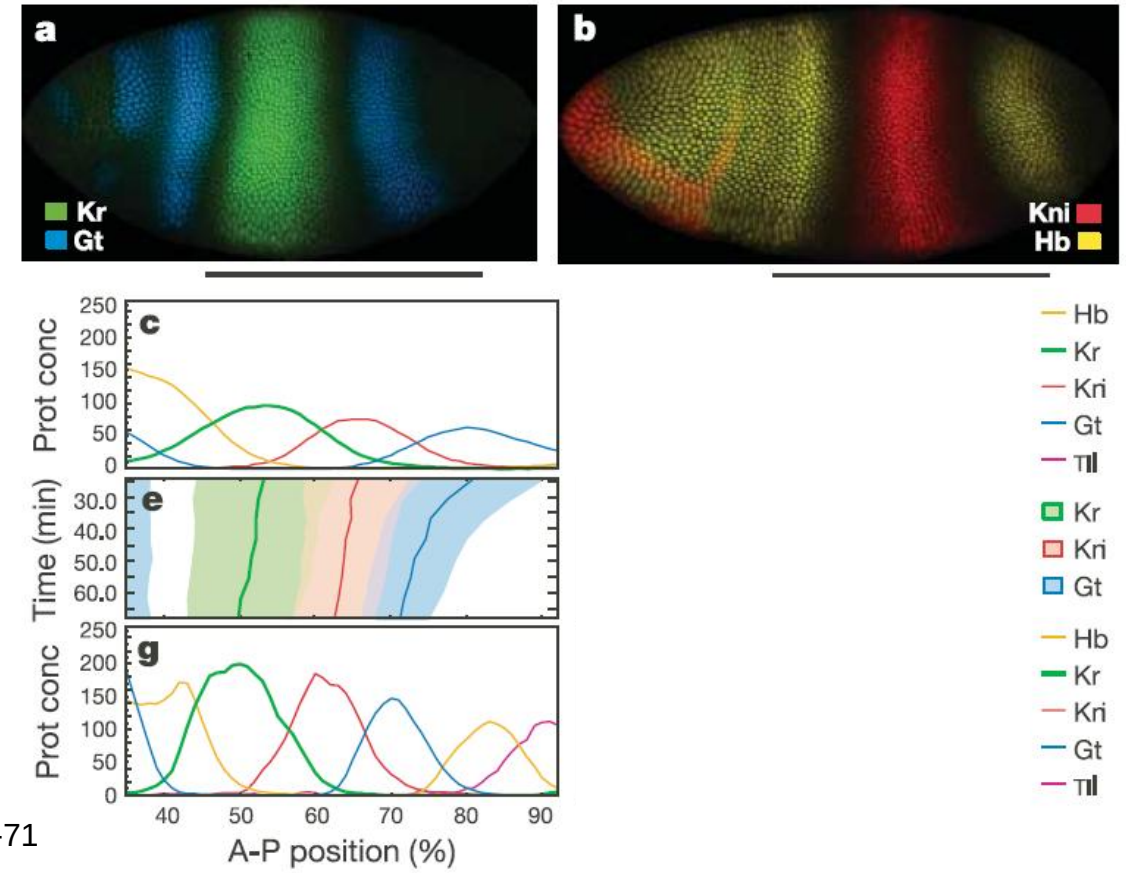
Logical model of *Drosophila* segmentation

- **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



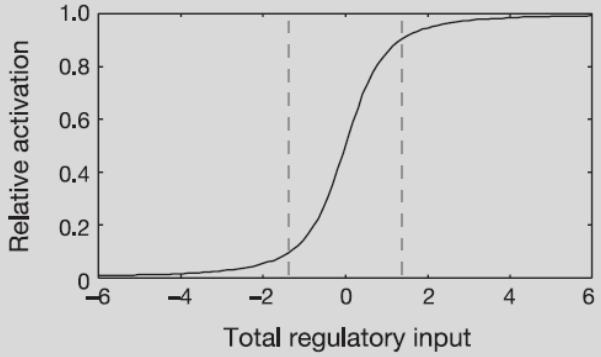
Jaeger *et al.* (2004), *Nature*, 430(6997):368-71

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

$$\frac{dv_i^a}{dt} = R_a g(u^a) + D^a\left[(v_{i-1}^a - v_i^a) + (v_{i+1}^a - v_i^a)\right] - \lambda_a v_i^a$$

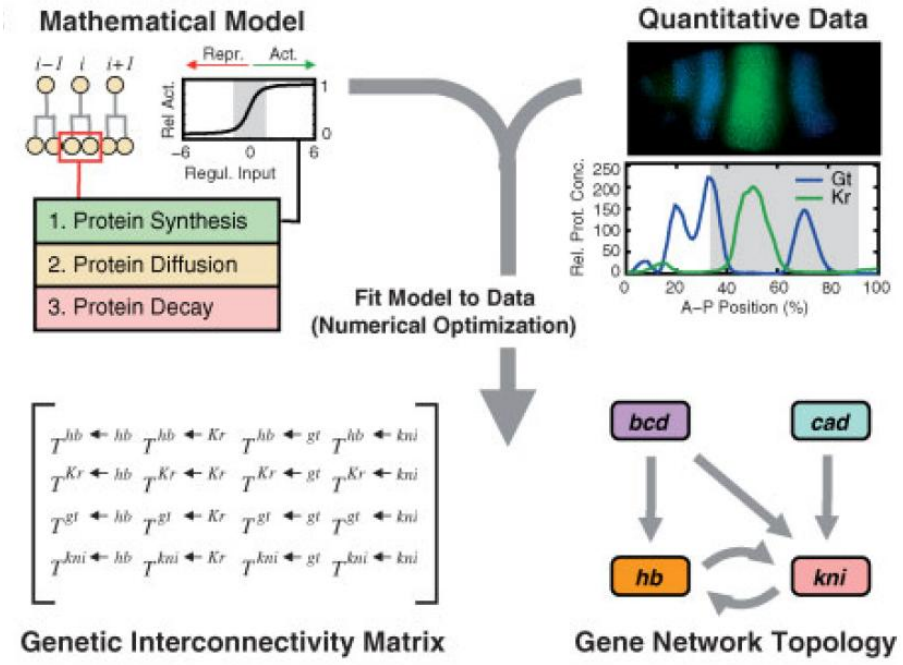
$$u^a = \sum_b T^{ab} v_i^b + m^a v_i^{\mathrm{Bcd}} + h^a$$



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

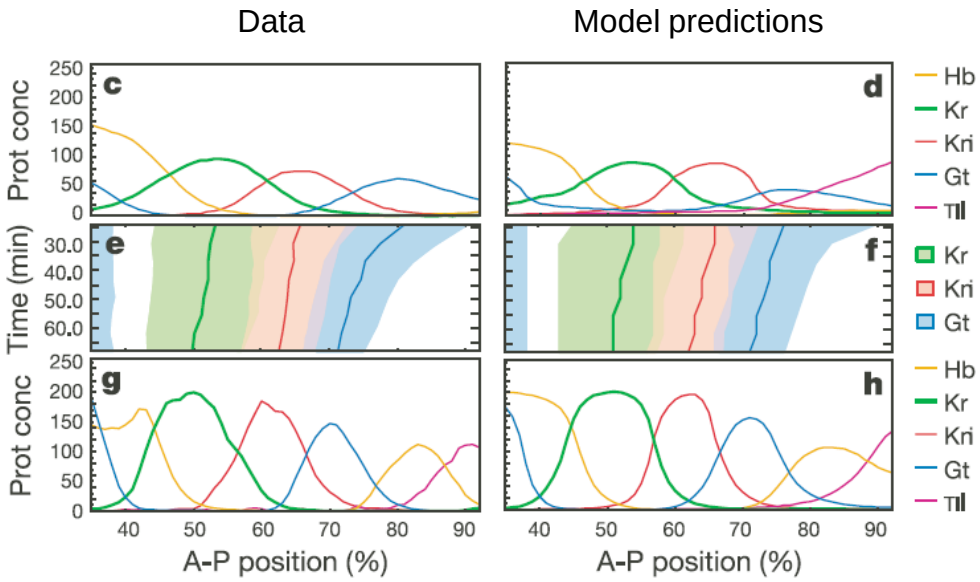


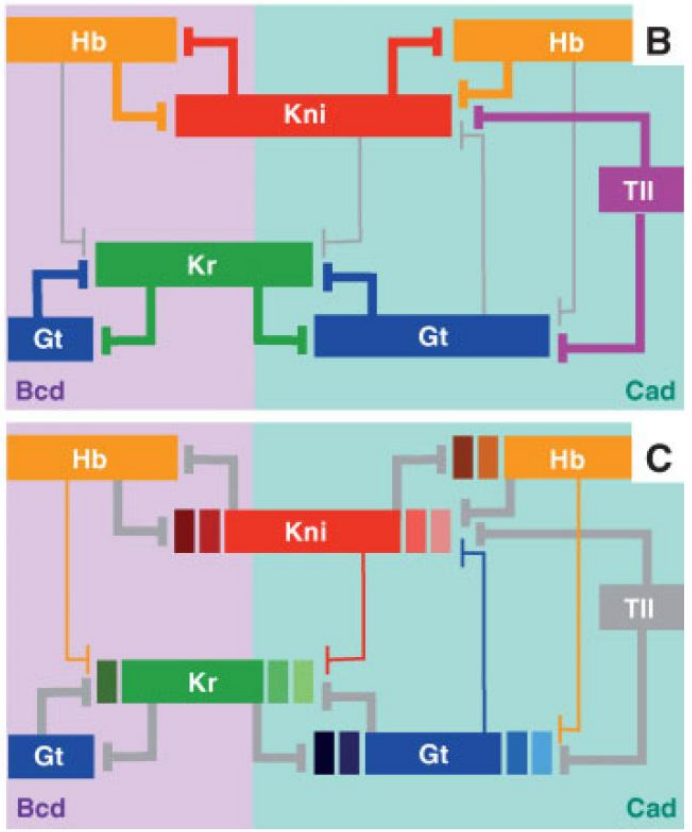
Jaeger and Reinitz (2006), *BioEssays*, 28(11):1102-11

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