ON SOLUTION TO THE INVERSE PROBLEM OF *DROSOPHILA* SEGMENTATION

Alexander M. Samsonov[1]

in collaboration with

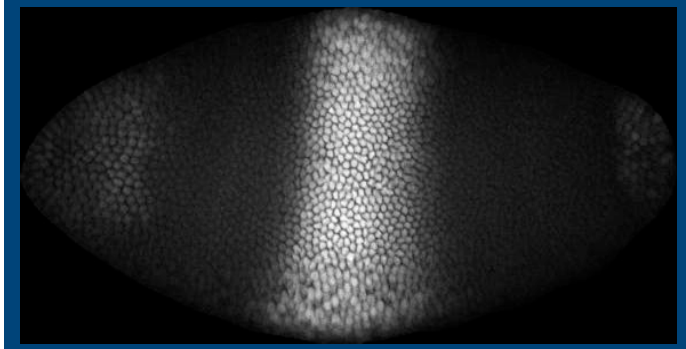
V.V. Gursky[1], J.Jaeger[3] , K.N. Kozlov[2], J. Reinitz[3]

[1] The Ioffe Physico-Technical Institute of the Russian Academy of Sciences, St. Petersburg, 194021 RUSSIA

[2] St. Petersburg State Polytechnical University, St. Petersburg, 195251 RUSSIA

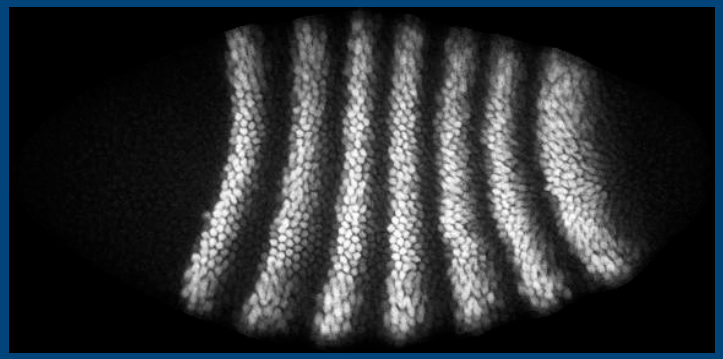
[3] The University at Stony Brook, Stony Brook, NY 11794–3600, USA

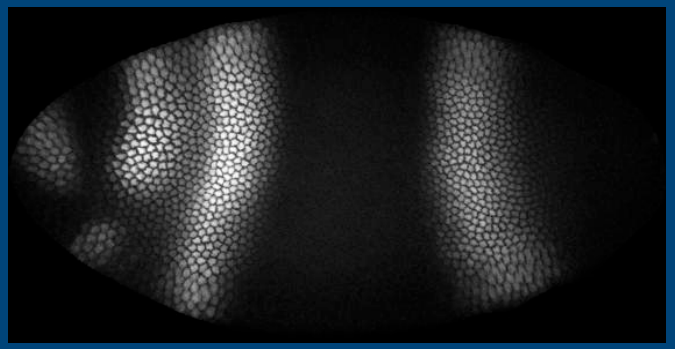
Segmentation gene expression images as in ***FlyEx*** *data base*: http://urchin.spbcas.ru/flyex



Kruppel (*Kr*)

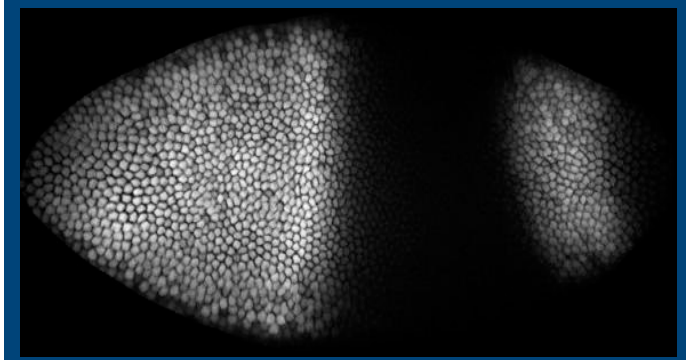
Even-skipped (*eve*)





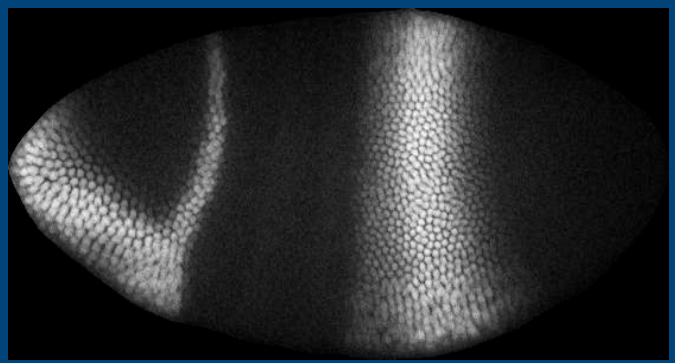
Giant (*gt*)

Hunchback (*hb*)

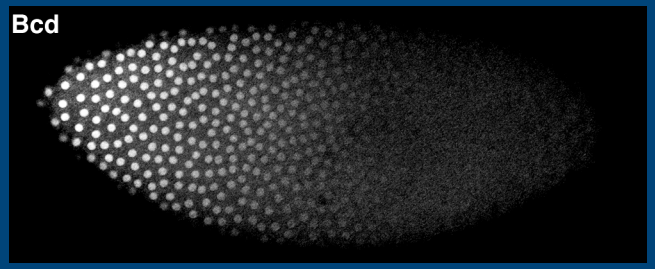


Insect body has a periodic structure called segments. Determination stage will be considered.

Knirps (*kni*)

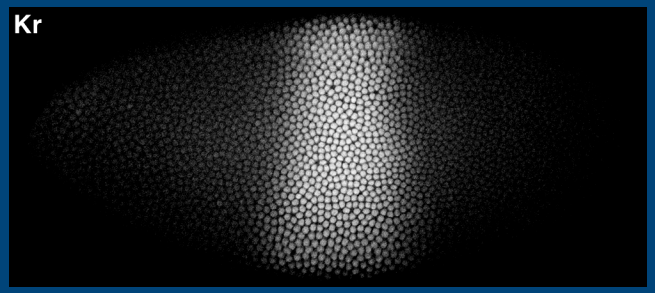


Maternal Coordinate Genes



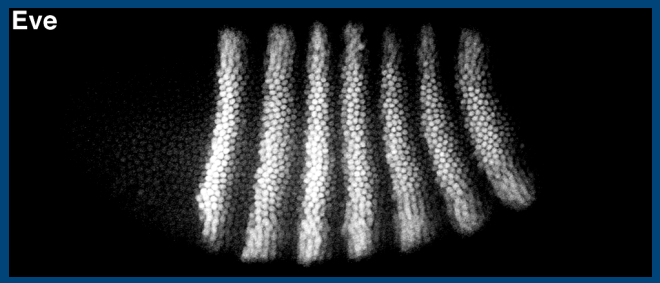
bicoid (bcd)
caudal (cad)
hunchback (hb)

Gap Genes



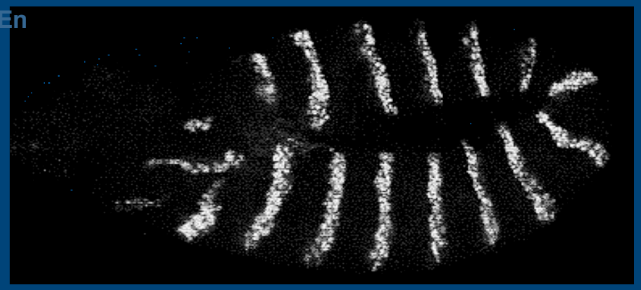
hunchback (hb)
Krüppel (Kr)
knirps (kni)
giant (gt)
tailless (tll)
huckebein (hkb)

Pair-Rule Genes



even-skipped (eve)
odd-skipped (odd)
hairy (h)
runt (run)
fushi-tarazu (ftz)
paired (prd)

Segment Polarity Genes



engrailed (en)
wingless (wg)

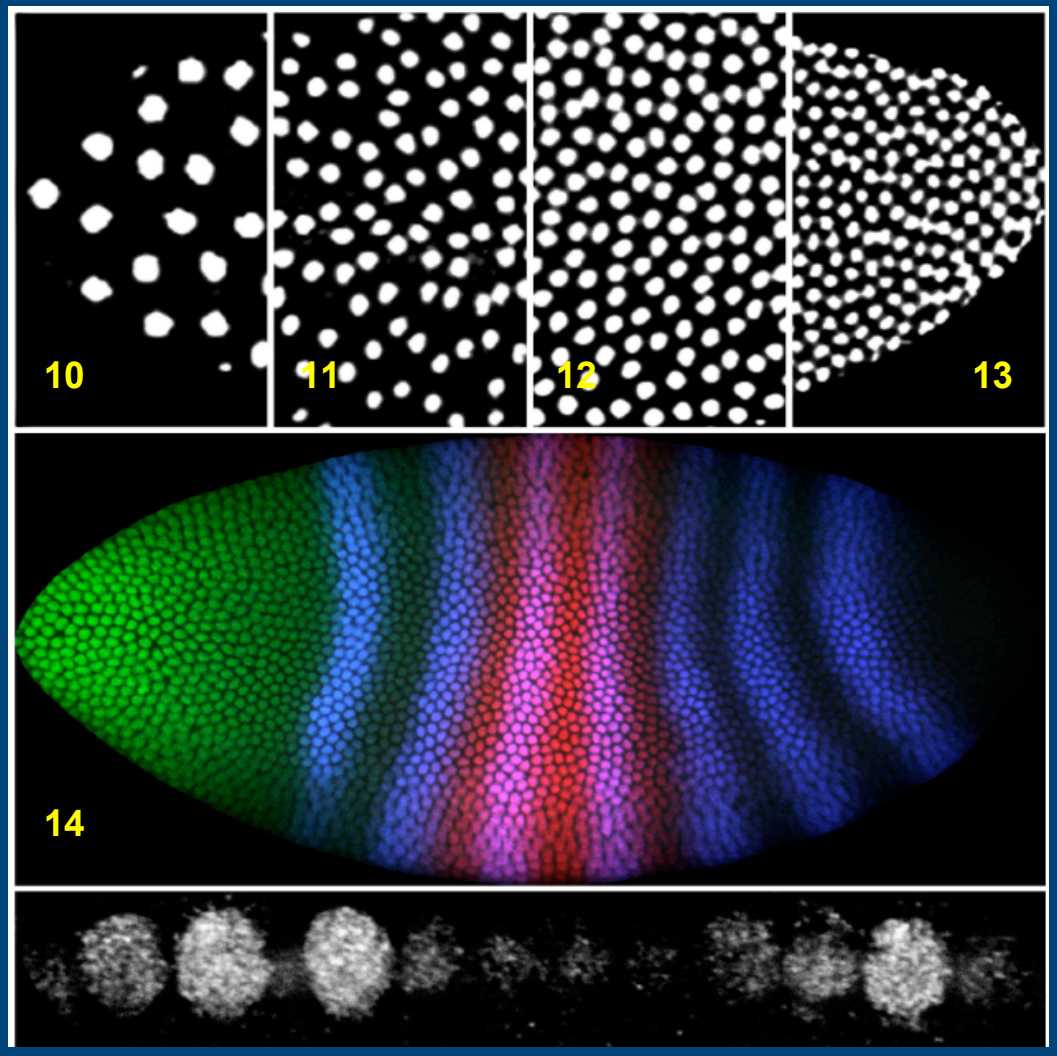
14 segmentation genes are concerned with initial DETERMINATION event.
Each has a distinct pattern of expression.

About 40 segmentation genes of the whole genome are active
in blastoderm, while others are involved in oocyte formation OR active after gastrulation

In addition, each expression pattern changes over time.

Each expression pattern changes over time.
Unlike as in MA we take the data
In space also.
Pair-rule gene (eve) late in blastoderm period resolves to 7 distinct stripes
4 nuclei wide

Formation and nuclear structure of *Drosophila* blastoderm



An embryo is only 0.5 mm long

Even in microscale
a nuclear structure exists on each cycle of an embryo development

Highly discrete structure ?
In what scale?

Can it be modelled with a continuum ?

Are the nuclear divisions = mitosis necessary for pattern formation ?

Motivation

1. Data from embryos or MA's both are series of snapshots.
2. They cannot describe temporal behaviour of gene expression with given accuracy up to date.
3. No spatial pictures of gene expressions provided, too.
4. Formal representation of experimental data for further analysis?

Motivation (cont'd)

5. Pattern formation is common for many irreversible (bio)chemical reactions (and physical processes).

6. Many p.d.e. provide patterns as suitable solutions.

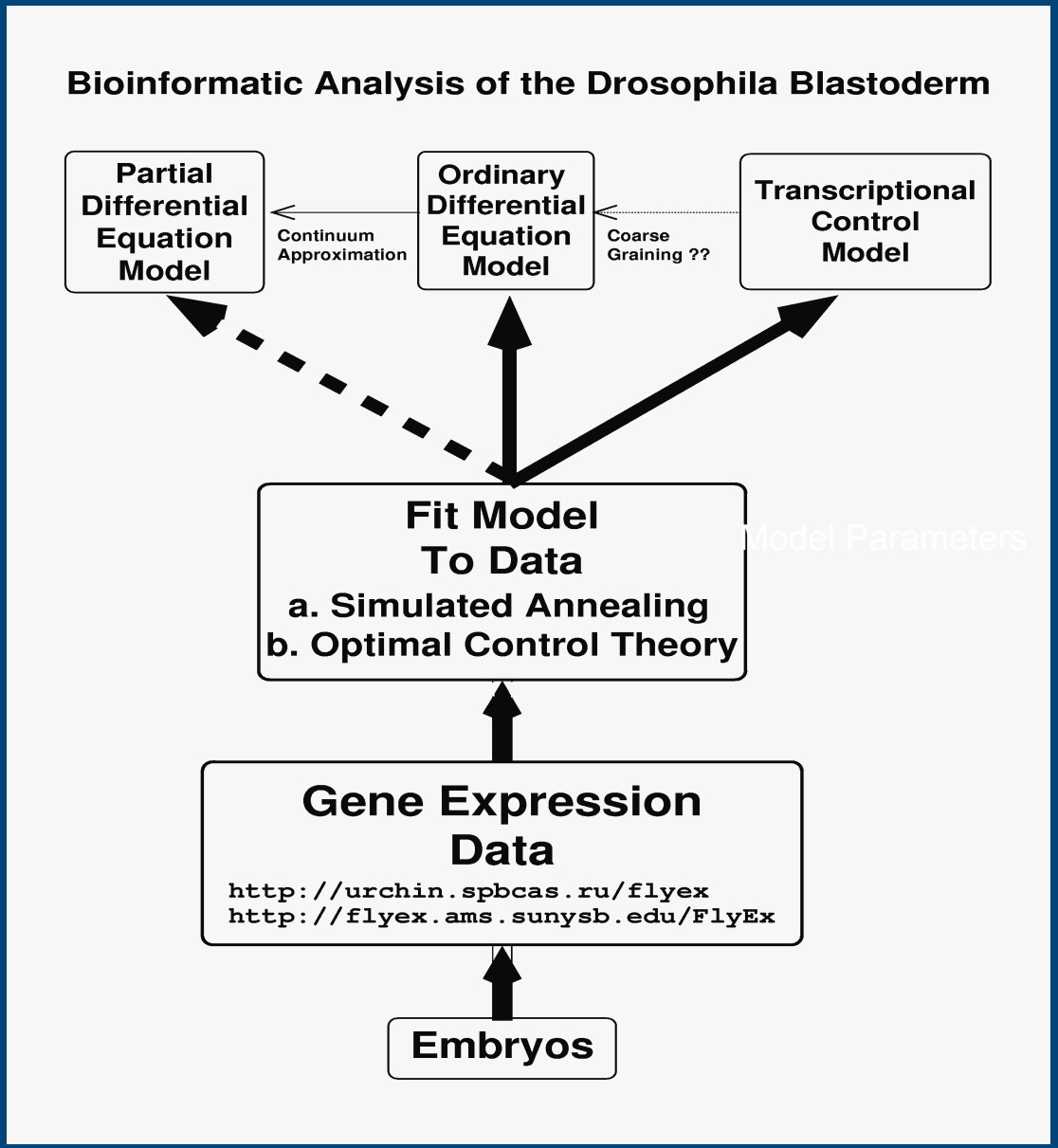
7. Drosophila is one of the simplest biological system for bio & math tests.

Math simulation is the cheapest one.

* How to develop a simplest possible mathematical model providing solutions with features given by biology ? Patterns?

* How to define coefficients in continuous model (p.d.e) in order to **fit data** of biological experiments ?

Analysis of the ***Drosophila*** blastoderm



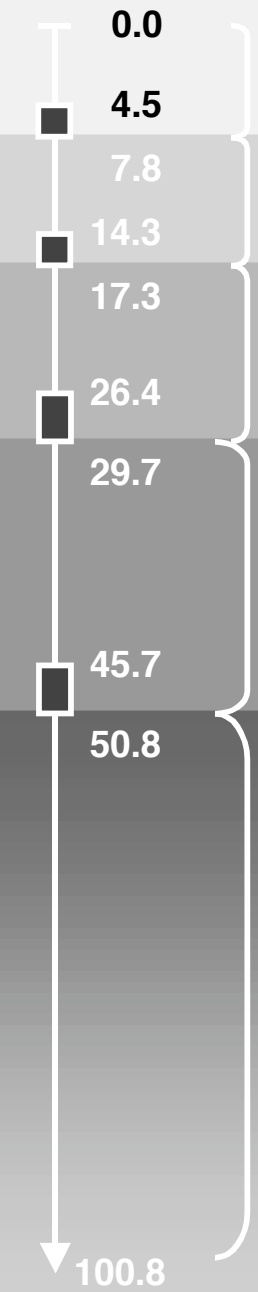
Model Parameters

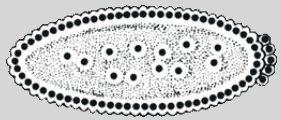
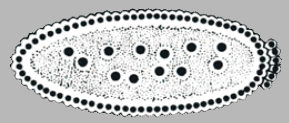
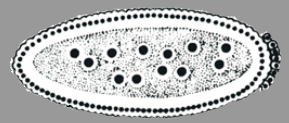
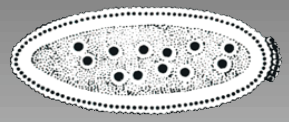
Aim to adopt continuum model instead of tracking the dynamics of each single nucleus ->modification of mitosis rule representation ?

Quality Functional

Database Records; Numerical Text Files

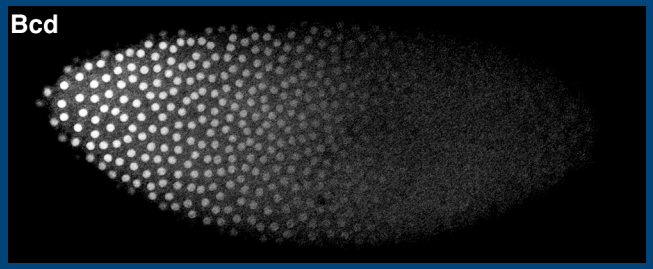
Images

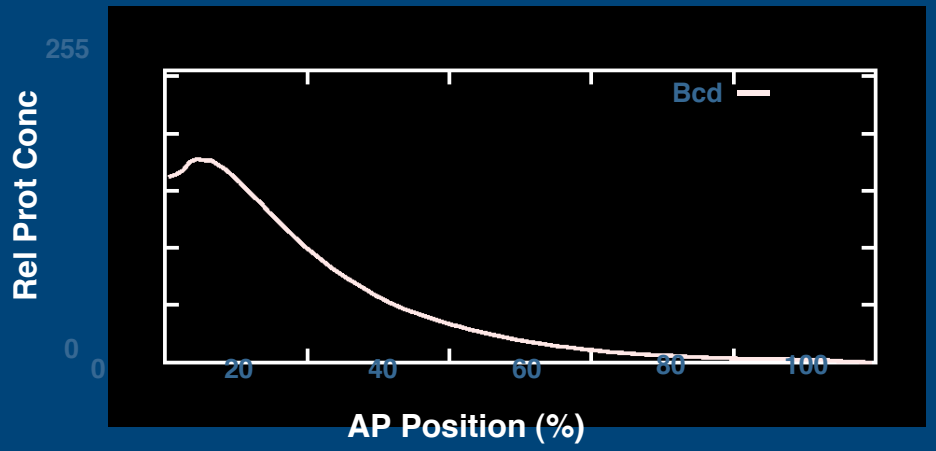


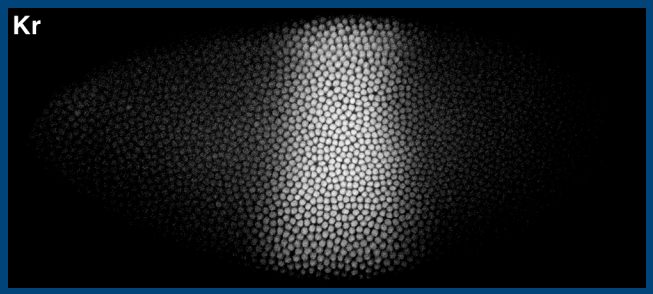
Time (min)	Cycle Number	Duration	
0.0 – 4.5	10	7.8 min	
7.8 – 14.3	11	9.5 min	
17.3 – 26.4	12	12.4 min	
29.7 – 45.7	13	21.1 min	
50.8 – 100.8	14	50.0 min	

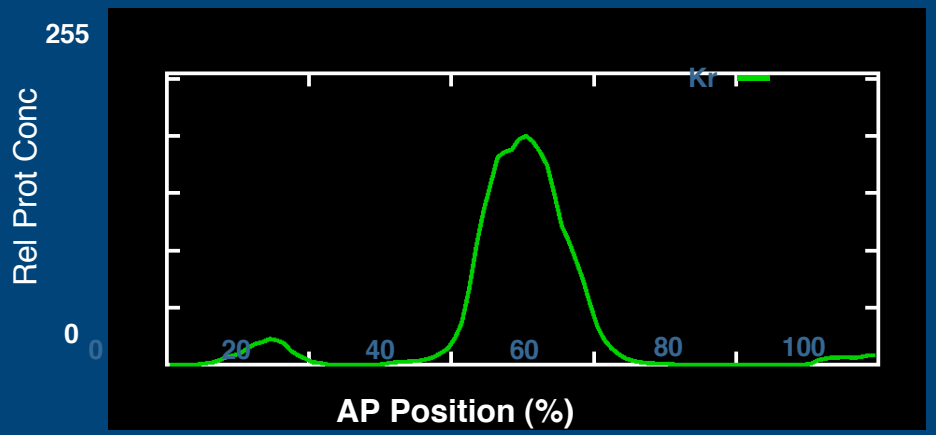
5 Cleavage Cycles

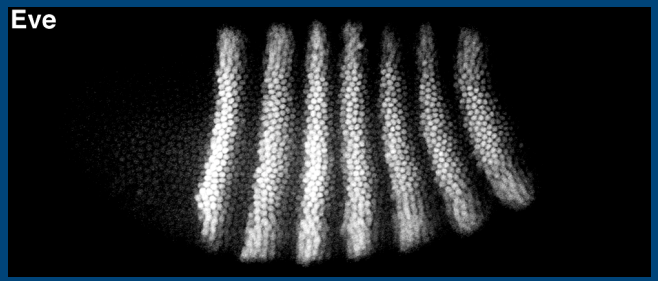
Foe & Alberts (1983). J Cell Sci 61: 31-70.

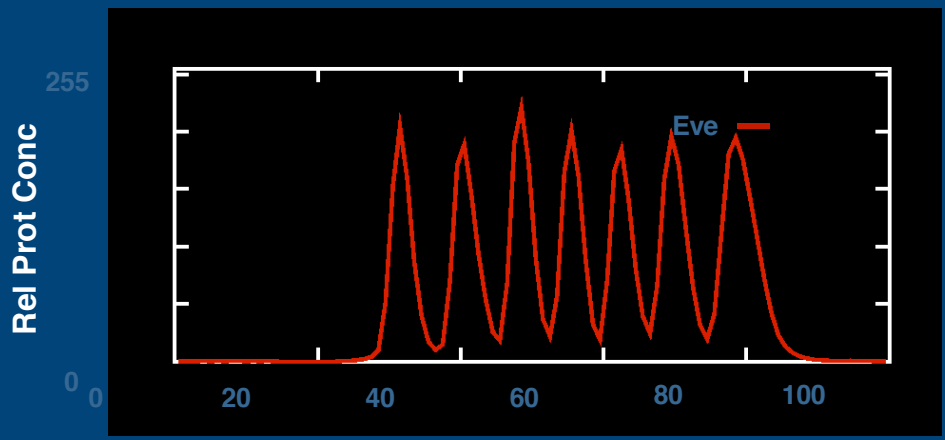

Bcd


255
Rel Prot Conc
0
0
20
40
60
80
100
Bcd
AP Position (%)

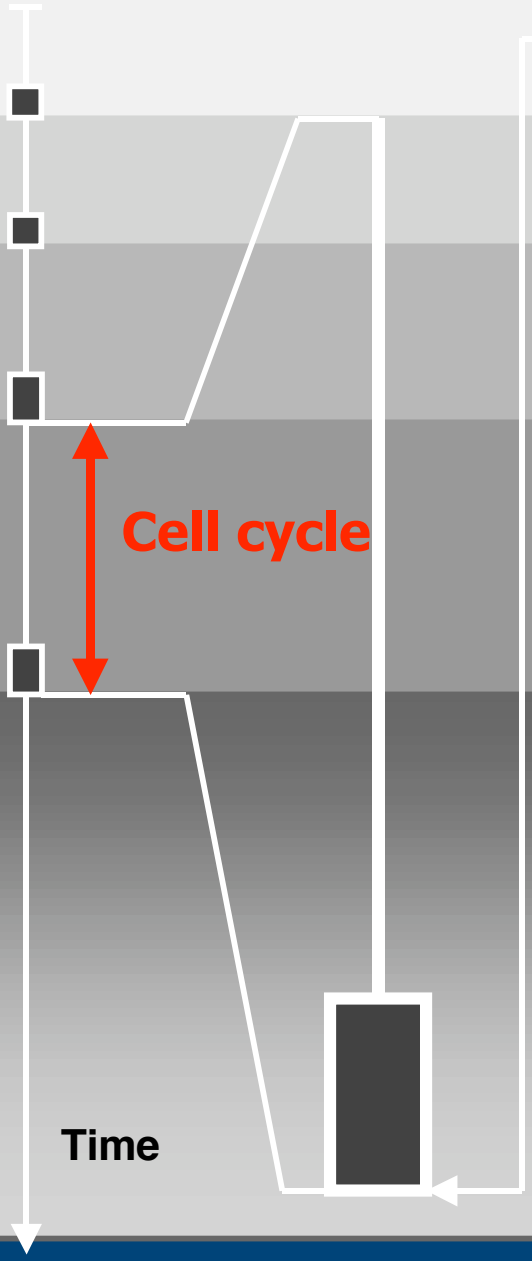

Kr

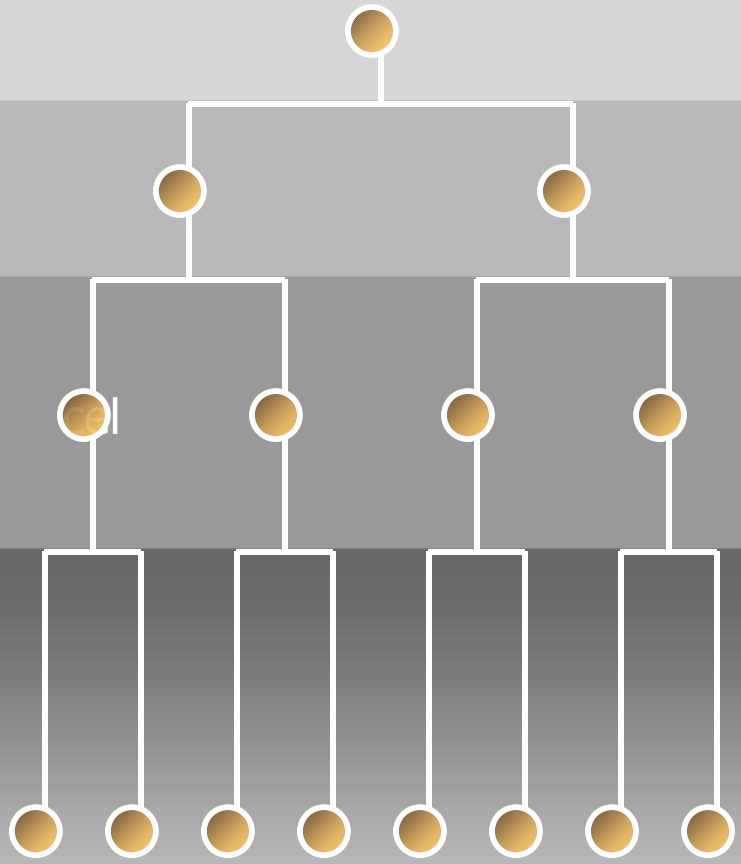

255
Rel Prot Conc
0
0
20
40
60
80
100
Kr
AP Position (%)


Eve


255
Rel Prot Conc
0
0
20
40
60
80
100
Eve
AP Position (%)

Nuclear Division

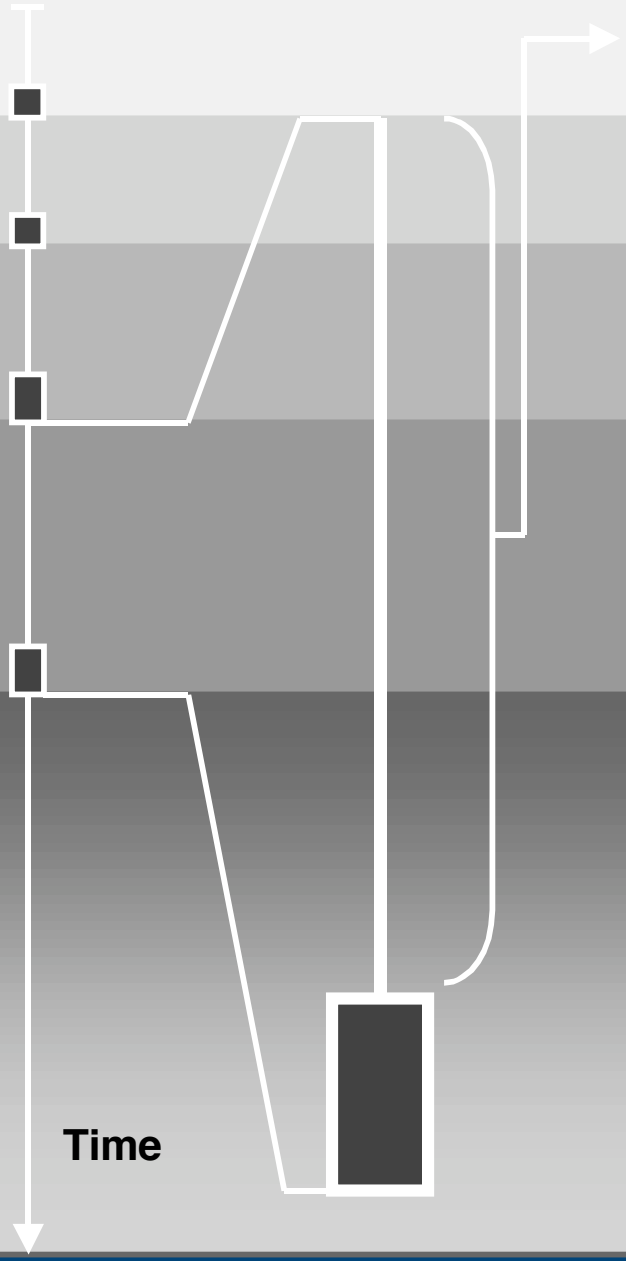




Mjolsness, Sharp & Reinitz (1991). JTB 152: 429-453.

3 Rules

Interphase

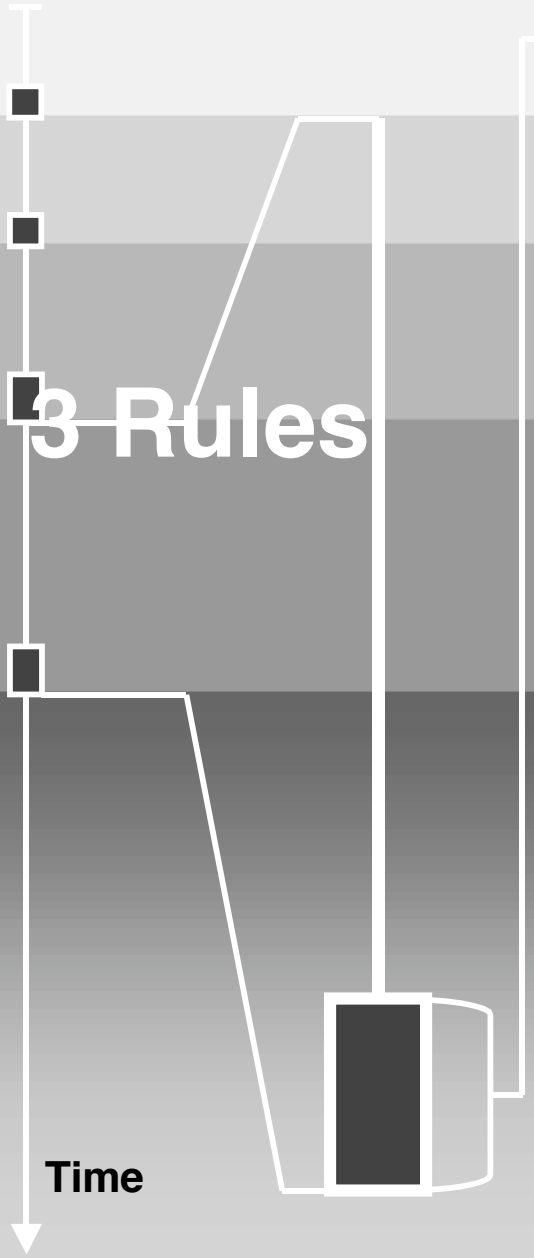


1. Protein Synthesis
2. Protein Transport
3. Protein Decay

3 Rules

Mjolsness, Sharp & Reinitz (1991). JTB 152: 429-453.

Mitosis



1. Protein Synthesis

2. Protein Transport

3. Protein Decay

Mjolsness, Sharp & Reinitz (1991). JTB 152: 429-453.

Fits to data

INPUT: gene expression patterns, random parameters

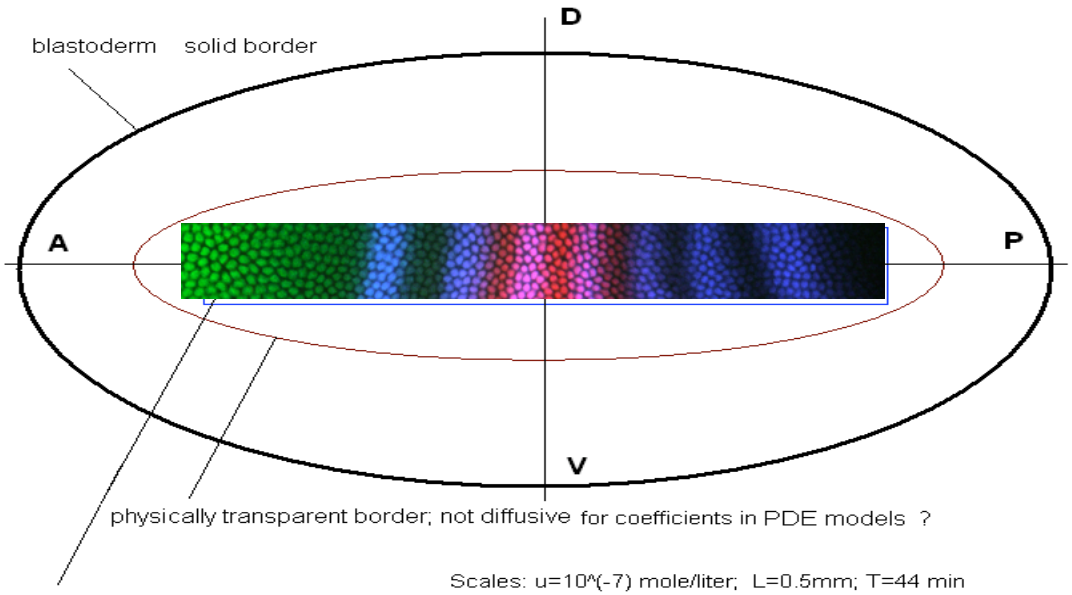
In natural sciences *the data* have priority ->
let us change parameters until the model given produces patterns that will be as similar as possible to gene expression data.

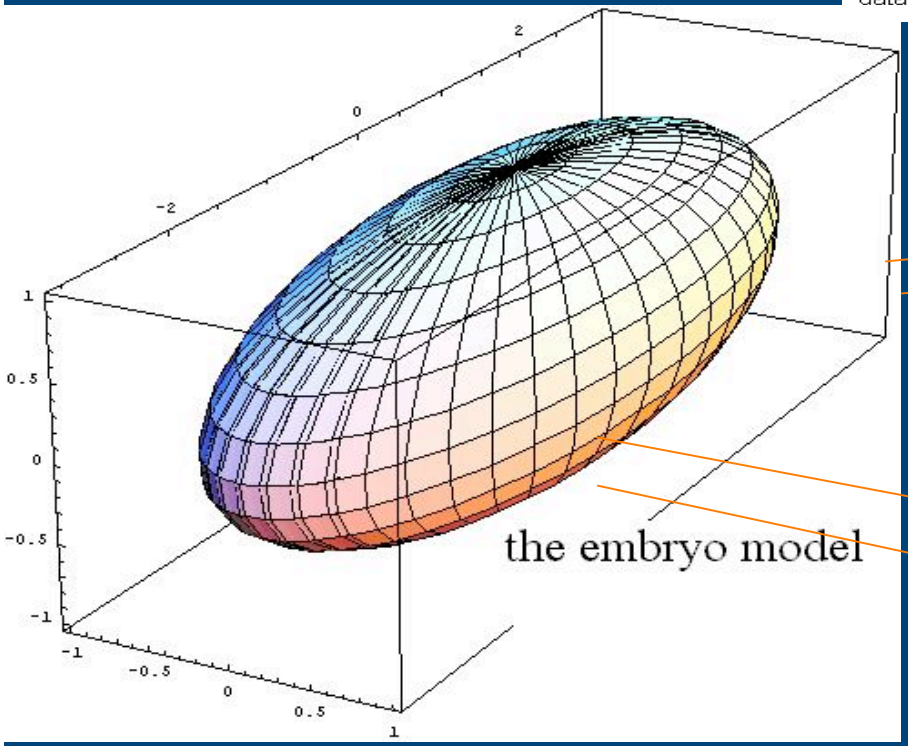
OUTPUT:

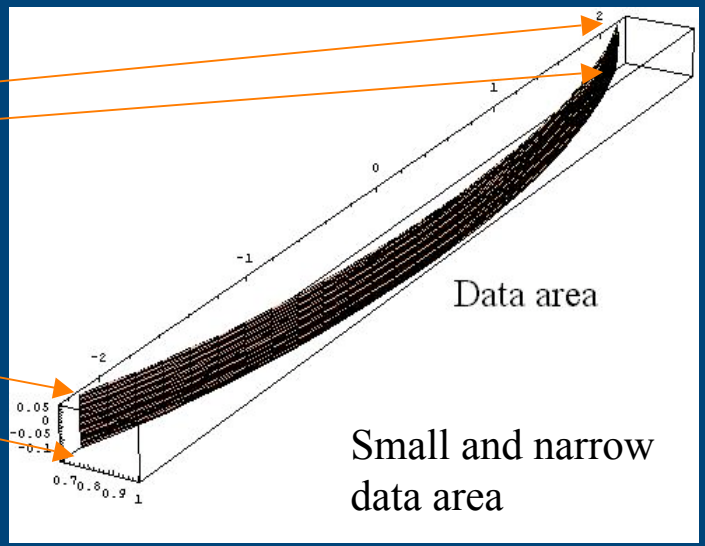
gene circuit=gene network=particular set of parameters

Formal geometry of the blastoderm model

2D projection







Data fitting

Usually we seek parameters for the differential-difference equations, governing the gene network, that minimize a functional:

$$E = \sum_{\substack{\text{all } a,\ i, \\ \text{and } t \\ \text{for which} \\ \text{data exists}}} \left(\mathrm{v}_i^a(t)_{\text{model}} - v_i^a(t)_{\text{data}} \right)^2$$

Minimize *E* using the method of *simulated annealing = Metropolis' method*

Simulated Annealing

- **Advantage:** **The method will yield the *global minimum* of *E.***

- **Disadvantage:** **Computationally intensive.**

  - **serial simulated annealing:**

    Min 12 hrs - Max 52 days on a 2 Ghz Pentium P4

  - **parallel simulated annealing:**

    Min 4 hrs - Max 14 days on 10 1.5Ghz AMD processors

Simulated Annealing

Metropolis et al., 1953; Kirkpatrick et al., 1983

1. Compute $E = E_{old}$ from the variables x_i .
2. Make a change in one (or more) of the x_i (this is referred to as a "move".
3. Compute $E = E_{new}$ from the newly generated set of x_i .
4. Compute $\delta = \exp((E_{old} - E_{new}) / T)$.
5. If $\delta >$ *a random number* γ : $0<\gamma<1$,
   keep the new x_i's ("accept the move").

Otherwise, restore the old x_i's ("reject the move").

Repeat while allowing T to decrease slowly from a large value to zero. Typically this entails 10^5 to 10^9 iterations.

Why look at the P.D.E's?

There is a large amount of papers concerning "Turing Models". Many useful results have been derived mathematically, but they use the following assumptions:

1. Spatially uniform initial state, in which:
2. The chemical reaction network is at a stable point,
3. The initial state becomes unstable because of diffusion,
4. An autoactivating slow diffusing substance is paired with a fast diffusing inhibitor.

These assumptions are false for most (not all!) biological systems. Can one build theoretical understanding from a biologically realistic set of PDE's? → The first step is to construct them!

NRD equations determine patterns? **VS** central role of cells in developmental biology.

Continuous model

Mass conservation in a volume *V*:

$$\frac{d}{dt}\int_V v^a(x,t)dx = -\int_{\partial V} \vec{j}_a(x,t)\cdot\vec{ds} + \int_V F_a(\vec{v})dx$$

$$= \int_V ((-\nabla\cdot\vec{j}_a) + F_a)dx; \qquad F - \text{“reaction”};$$

Fick’s law for diffusion: $\vec{j}_a(x,t) = -D_a\nabla v_a(x,t)$

$$\frac{d}{dt}\int_V v^a(x,t)dx = \int_V [D_a\Delta v^a(x,t) + F_a(\vec{v})]dx \Longrightarrow$$

Equations for concentration (**Non-linear Reaction-Diffusion Eqs**) are

$$v_t^a(x,t) = D_a v_{xx}^a(x,t) + F_a(x,t,\vec{v}); a\in[1,M]$$

No nuclei considered, they are small enough in comparison with pattern size.

Model equations -> Rea-Diff-Decay eqs

$$\frac{\partial v^a}{\partial t} = R^a(t) g\left(\sum_{b=1}^{N} T^{ab} v^b + m^a v^{BCD}(x) + h^a \right) - \lambda^a v^a + D^a \frac{\partial^2 v^a}{\partial x^2}$$

$v^a(x,t)$ -concentration of *a*-th protein, $v^{BCD}(x)$ – that of *BCD* protein

T^{ab} - describes regulatory effects of gene *b* on gene *a;* (a matrix)

λ^a - the decay rate of *a*-th gene product; R^a-max rate of protein synthesis

h^a - describes an effect of uniformly distributed maternal transcription factors on gene *a;*

m^a, D^a are regulatory and diffusion coefficients, resp.

$x \in L$ is a spatial domain on *A-P* axis of an embryo .

Source term is very complex due to regulatory matrix; decay is to be involved

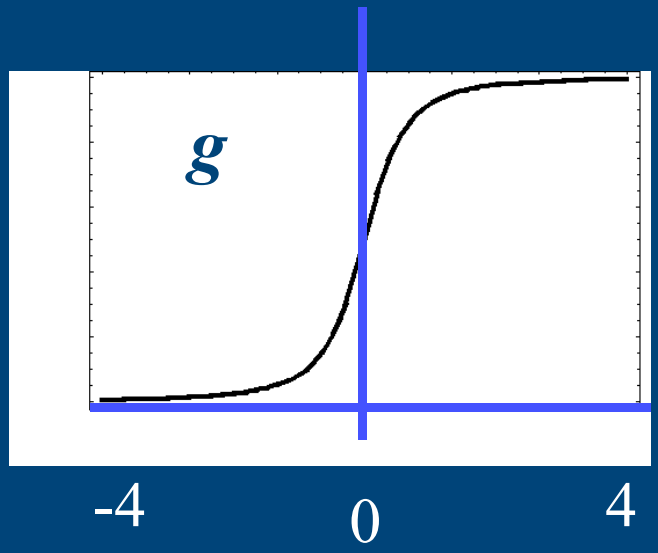
Discrete model: *Reinitz, Sharp. Mech. Dev. 49:133–158, 1995*

Regulation - expression function g is responsible for synthesis

g is to be of sigmoidal form.

Formal description is arbitrary.

It may be formalized as an algebraic function of its complex argument f:



$$g(f) = \frac{1}{2}\left(1 + \frac{f}{\sqrt{1 + f^2}}\right)$$

or the hyperbolic tangent :

$$g(f) = \frac{1}{2}(\tanh f + 1)$$

Near **g=1/2** a small variation of regulator v results in change of synthesis rate

Models of mitosis

R^a- the max rate of synthesis for protein a depends on time t due to mitosis.

Model **A** : $R^a(t) \equiv cons$

Model **B** : $$R^a(t) = \begin{cases} 0, & \text{during mitosis} \\ R_0^a - cons, & \text{otherwise} \end{cases}$$

Model **C** : Model (**B**) **+ doubling of** R_0^a after each division of nuclei

Aim to adopt continuum model instead of tracking the dynamics of each single nucleus ->modification of mitosis rule representation.

Fitting the model to expression data

There are $N(N+5)$ phenomenological parameters for N genes:

$$R^a, T^{ab}, m^a, h^a, \lambda^a, D^a$$

The functional to be minimized:

$$F = \sum_{\text{genotypes}} \sum_a \sum_k \int_L \left(v^a(x,t_k)_{\text{model}} - v^a(x,t_k)_{\text{data}} \right)^2 dx$$

Method: **a** new Optimal Steepest Descent Algorithm (OSDA).

Data are used at several time moments in cleavage cycles 13 and 14.

Formal problem statement

Let the vector function of protein concentrations

$v(t,q) = (v_0(t,q),...,v_{K-1}(t,q))^T$ describes the system in time ***t***, and

$q = (q_0,\ldots,q_{I-1})^T$ be the vector of parameters to be found.

The system of ODE's of the 1-st order written with respect to the independent variable ***t*** is:

$$\frac{\partial v}{\partial t} = f(v,q)$$

and the initial condition: $v(0,q) = V$

Quality functional and constraints

$$F(v,q) = \sum_{i=1}^{J} (v(t_i,q) - y(t_i))^T (v(t_i,q) - y(t_i)) \xrightarrow[q]{} \min$$

where all of ***J*** independent experimental observations are denoted as

$$y(t) = (y_0(t),\ldots,y_{K-1}(t))^T$$

Constraints in the form of inequalities are imposed on a subset of parameters

$$q_i^{low} \le q_i \le q_i^{up}$$

Lagrangian approach requires an expansion of a functional involving all constraints with add. Multipliers and a starting point.

Constraints

Constraints in the form of inequalities are imposed for the parameters

$$R^a, \lambda^a, D^a$$

as

$$R_{low} < R^a < R_{up} \Rightarrow R^a = \alpha_R + \beta_R \tanh(\gamma r_i),$$

$$\alpha_R = (R_{low} + R_{up})/2; \beta_R = (R_{low} - R_{up})/2$$

for each protein *a*.

To apply the Lagrange technique for optimization the constraints are to be transformed into equations for r.

Numerical algorithm

- Integrate the system of p.d.e. given
- Integrate system for the *Lagrangian* multipliers from $t(J)$ to $t(0)$ with the initial condition and conditions given in points, where data are available.

- Calculate the gradient $Z^k = Z\,(v^{(k)}, q^{(k)})$ where $v^{(k)}$ is the solution of the original p.d.e's obtained on k-th iteration using $q^{(k)}$ as parameters.

The new parameters $q^{(k+1)}$ are calculated as:

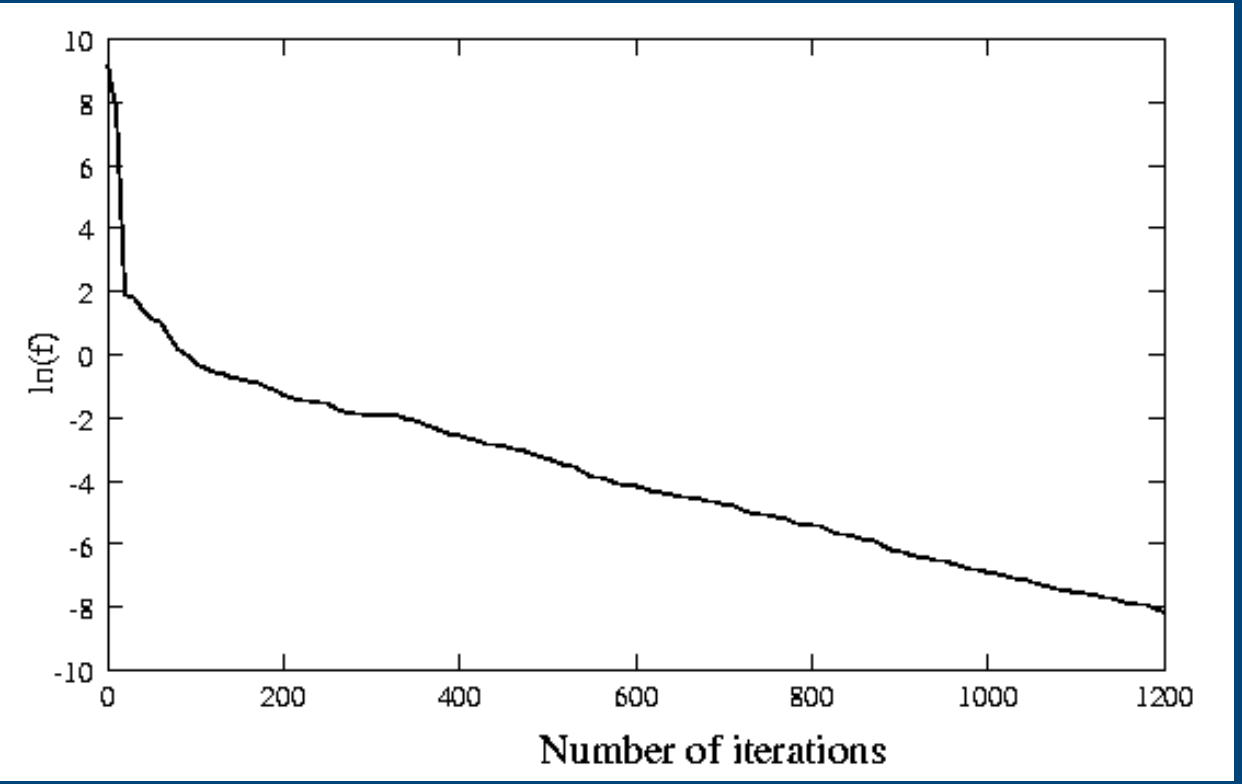
$$q^{k+1} = q^k - a_k Z^k$$

a^k is used to minimize the value of the functional on each step.

Repeat it until the value of the functional be less than a value given.

The set of q^N defines the solution of the problem and the optimal components of this vector.

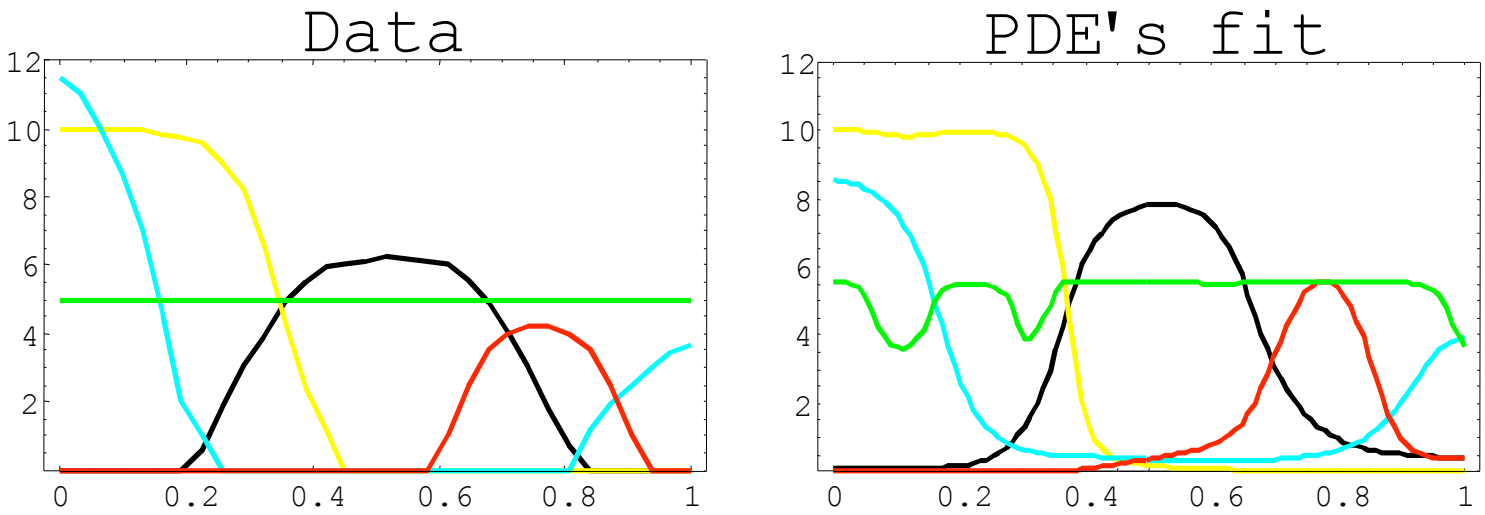
Convergence diagram



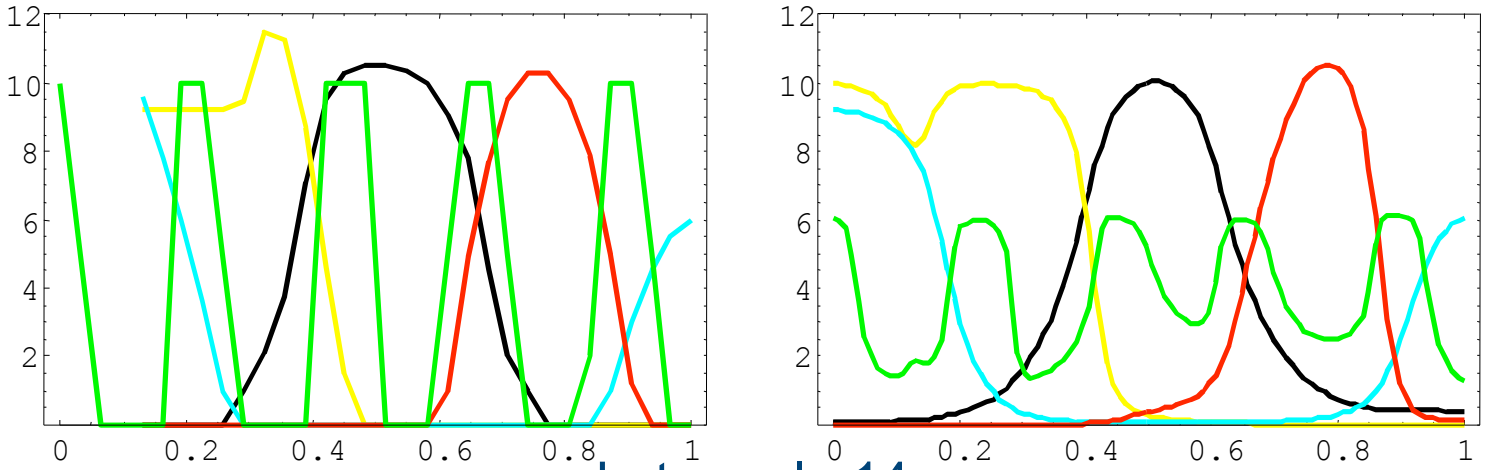
Log F vs Number of iterations is shown

Functional achieves a value of 10^{-8} in just 1000 steps.

Best patterns: mitosis model A, rms=1.04



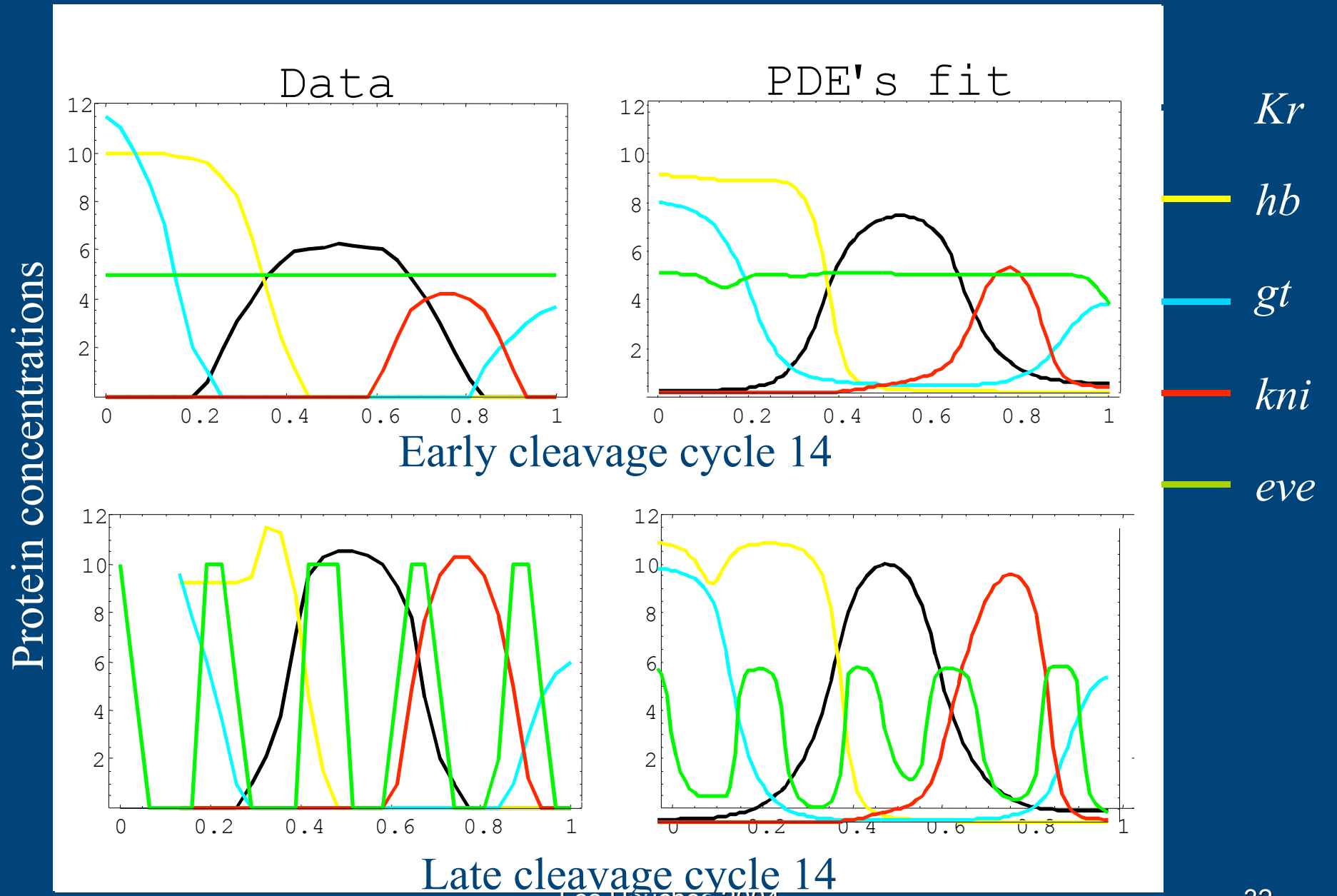
Early cycle 14



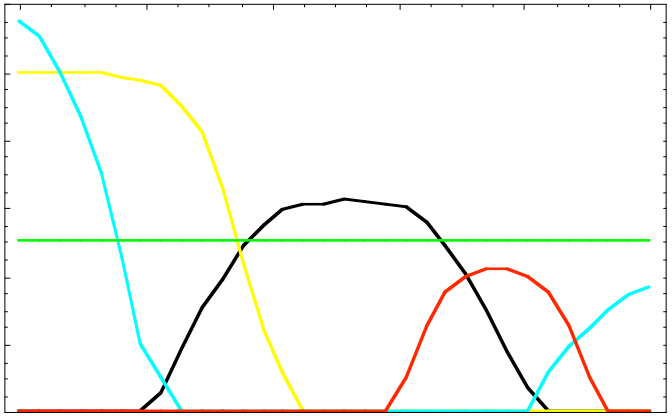
Late cycle 14

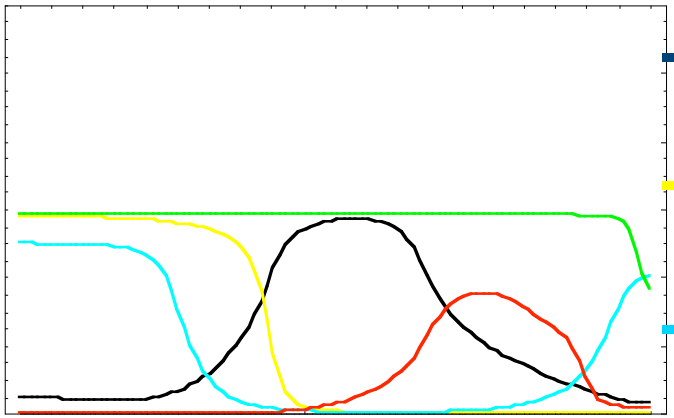
Data and Solution are compared.

Best patterns: mitosis model B, rms=1.003



Best patterns: mitosis model C, rms=1.28

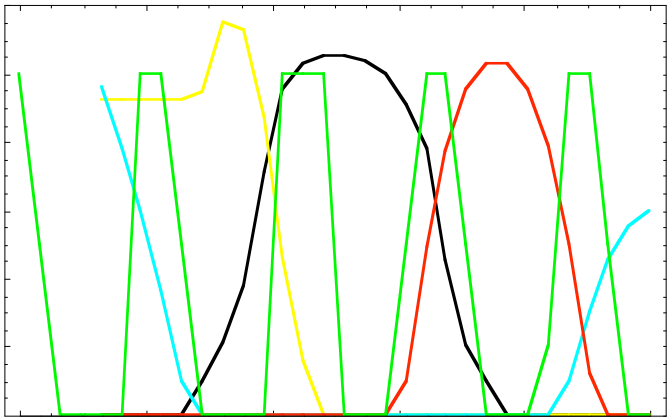


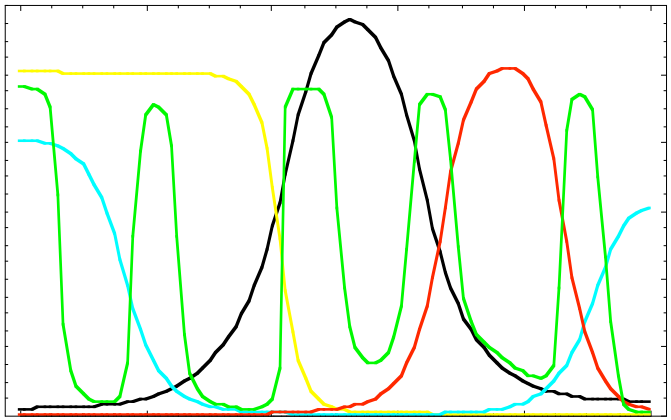


Kr
hb
gt
kni
eve

Protein concentrations

Early cleavage cycle 14





Late cleavage cycle 14

Model C -Poorest fit!

x/L

Advantages of OSDA

- OSDA can be successfully applied for finding the minimum of the quality functional in the problem of analysis of big arrays of experimental data.

- Given a good initial guess the OSDA needs two orders less steps than the random search.

Details are in :

Kozlov, Samsonov. A novel approach to experimental data fitting by means of optimal control theory. Techn. Physics, 2003, 48, 11, 1-8.

- Optimal parameters can be defined with the accuracy of >10 %, and *h* (constant maternal transcription factor) recovered with OSDA has the biggest error. Simulated annealing showed the same behaviour ; see:

Chu, Deng, Reinitz. Parallel simulated annealing by mixing of states, J Comput. Physics, 148, 1999, 646-662.

Parametric stability of the model

Various perturbation schemes were used:

- all parameters,
- R^a and λ^a only
- T^{ab} only
- m^a and h^a only
- D^a only
- all parameters in the discrete model.

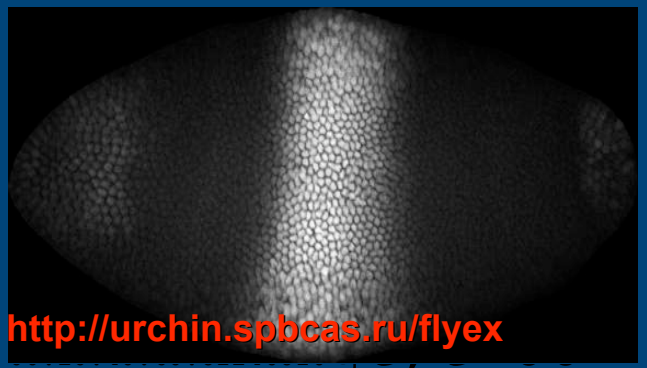
Results present a very close numerical approx. to the true minimum of functional for models A-C.

To increase the confidence random perturbations were made.

10.000 numerical experiments were made in each vicinity of the optimal parameter set.

Vicinities considered: up to 1% of parameter values, from 1% to 5%, from 5% to 10%, … , from 35% to 40%.

Patterns perturbation: 50 random samples selected from the 1% vicinity of optimal parameters set

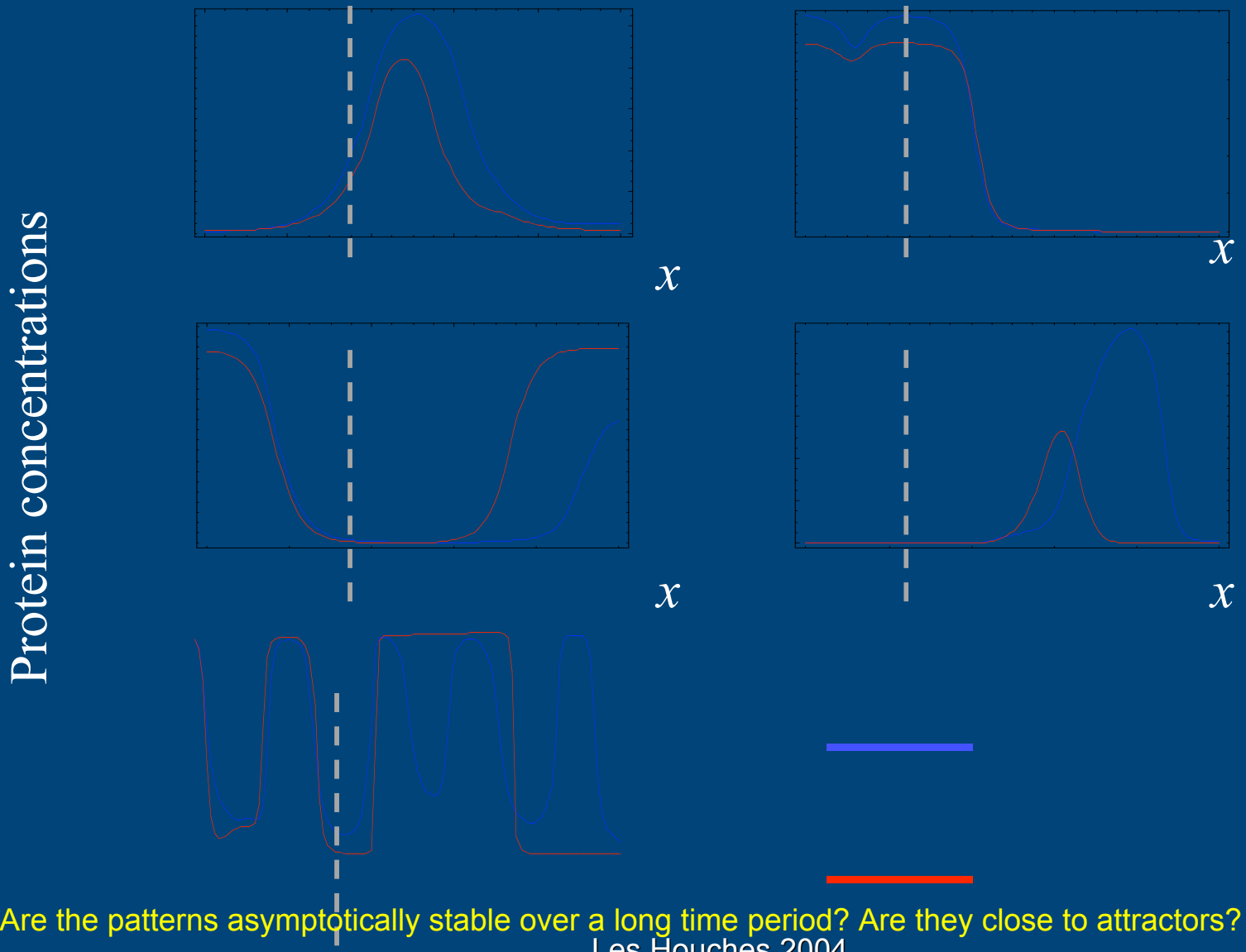


Results show a very close numerical approximation to the genuine minimum. Over 10^5 random perturbations of the optimal parameter values examined.

Patterns perturbation:
180 samples from 5% vicinity with rms < 2

In no case the functional was
Found smaller than the one
Provided by OSDA

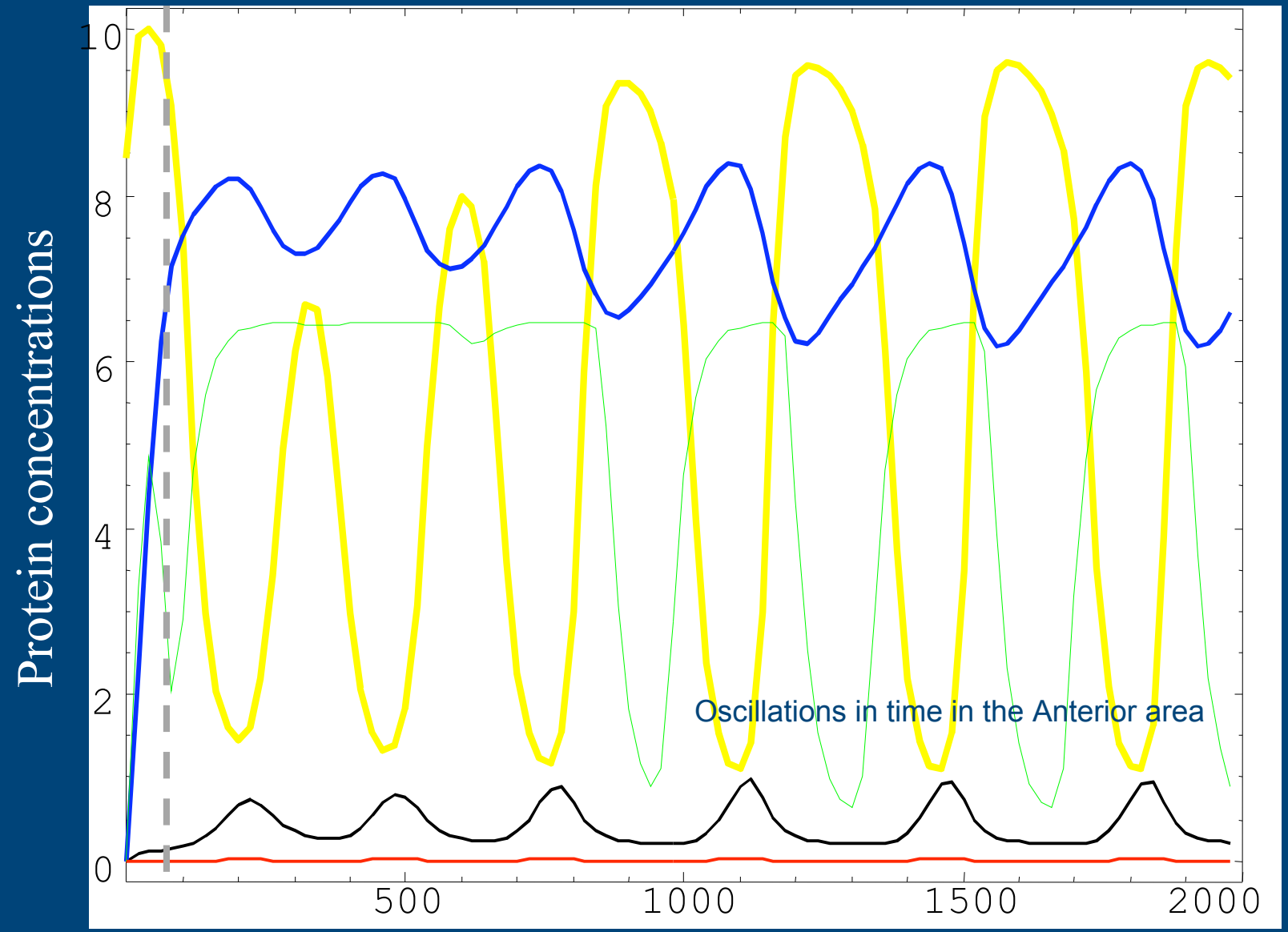
Long term patterns: mitosis model A



Are the patterns asymptotically stable over a long time period? Are they close to attractors? The answer is No.

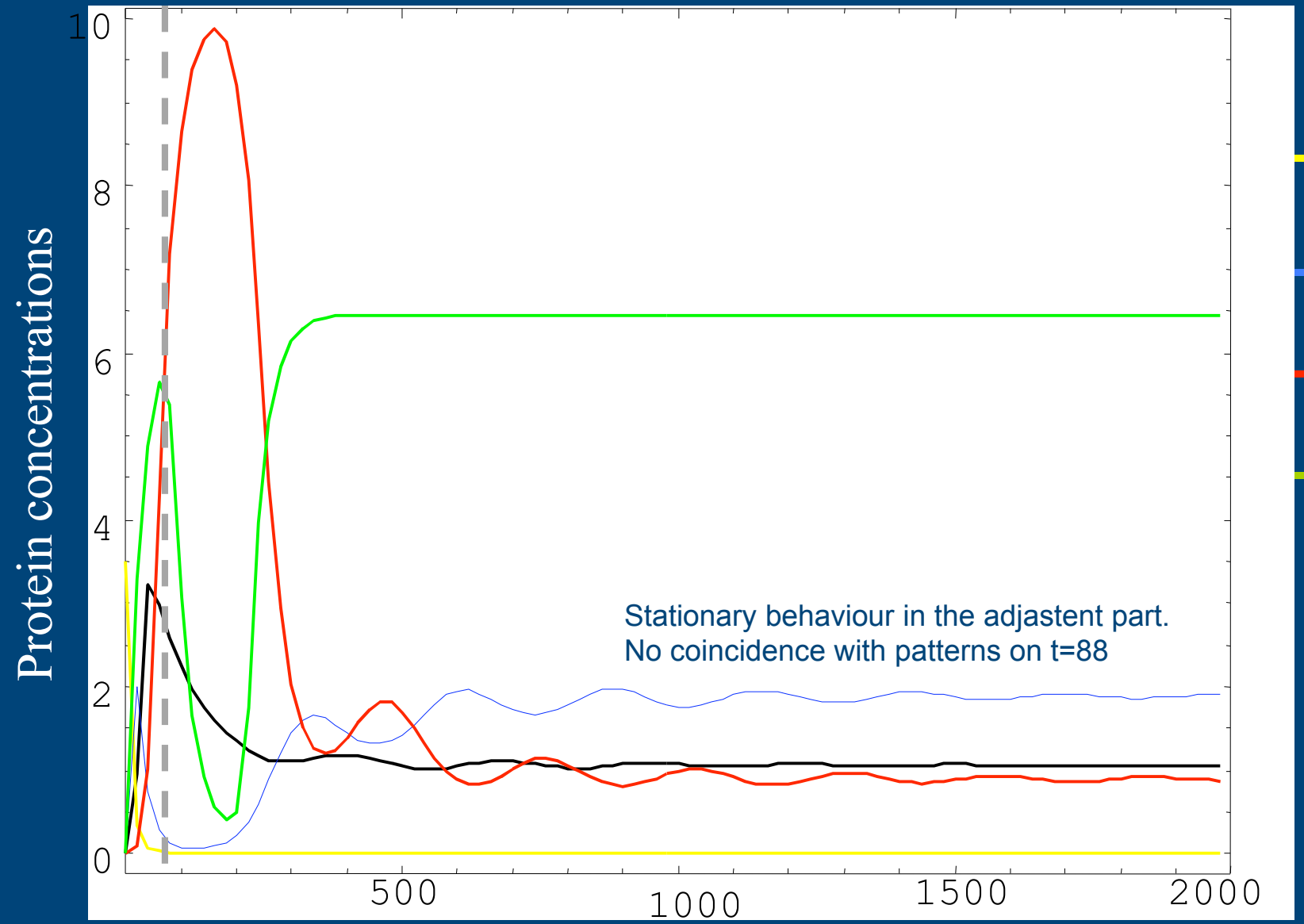
Long time dynamics: mitosis model A, x = 0.15L

(close to anterior)



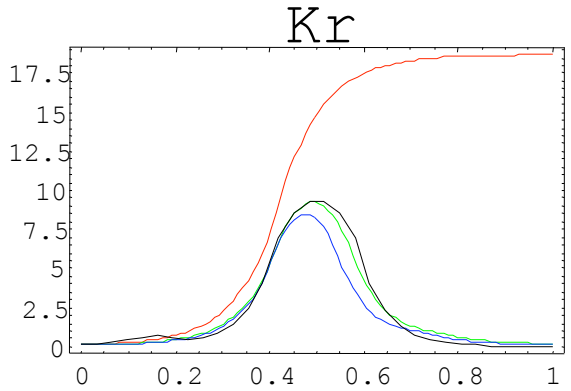
Long time dynamics: mitosis model A, x = 0.7L

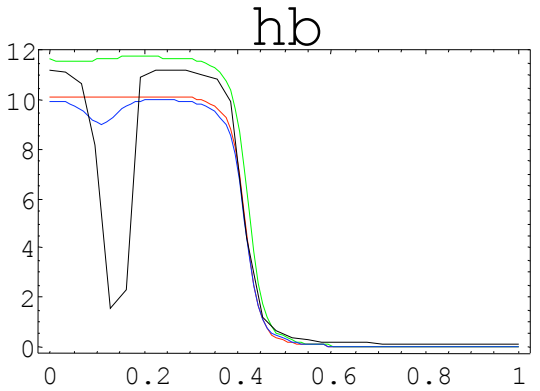
(close to posterior)

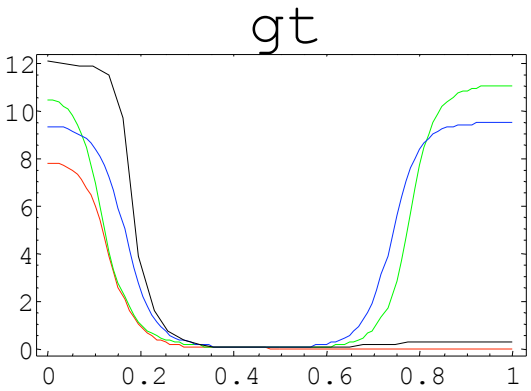


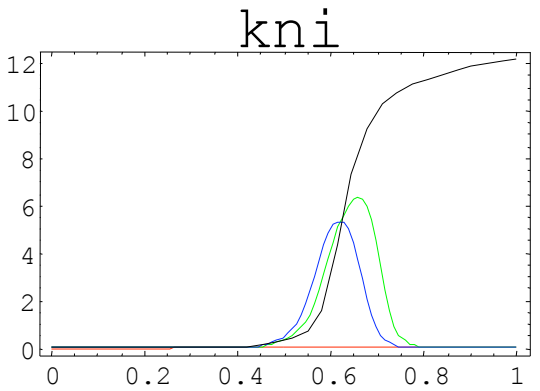
Long time patterns dynamics in different models

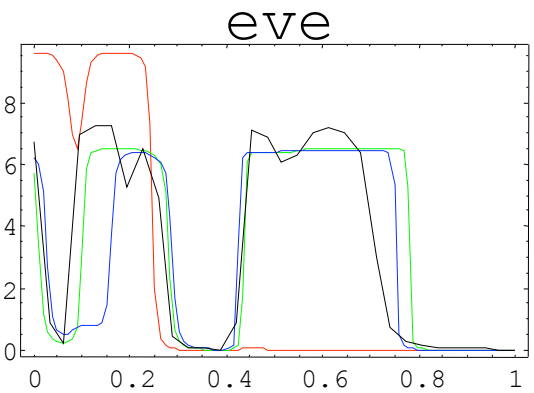
Protein concentrations











Nuclear divisions have no role in pattern formation
A – no mitosis
B – shutting off of synthesis only
C – shutting off +doubling of synthetic density

Mitosis A

Mitosis B

Mitosis C

Discrete model

Conclusions

The model is stable with respect to modification of the equations structure (various diffusion and mitosis).

The model is stable with respect to small (1–2%) perturbations of parameters, while the bigger perturbations may destroy the patterns calculated.

Nuclear Divisions are Not Required for Pattern Formation.

It could not have been obtained in experiments, since it is impossible to replace the blastoderm cellular structure with a continuum where segmentation genes act.

Turing (1952): *"most of organism most of time is going from one pattern into another rather than from homogeneity into a pattern".*

Acknowledgements

Stony Brook

John Reinitz
Carlos Alonso
Lucas Carey
King-Wai Chu
Lorraine Greenwald
Hilde Janssens
Dave Kosman
Manu
Yousong Wang
Marcel Wolf

St. Petersburg

Maria Samsonova
Vitaly Gursky
Konstantin Kozlov
Dimitry Malashonok
Ekaterina Myasnikova
Andrei Pisarev
Ekaterina Pustelnikova
Svetlana Surkova

Los Alamos

Dave Sharp
Shuling Hou

The support of the study by the NIH Grants RR07801, TW01147 via the CRDF GAP Awards RBO685, RBO1286 is gratefully acknowledged

References

•Turing, A. M. (1952) *Trans. Roy. Soc. London B 237, 37--72.*

•Chu, Deng, Reinitz. Parallel simulated annealing by mixing of states, *J Comput. Physics, 148, 1999, 646-662.*

Kozlov K.N., Samsonov A.M. A novel approach to experimental data fitting by means of optimal control theory. *Technical Physics J., 2003, 48, 11, 1-8.*

Gursky, V. V., Reinitz, J., & Samsonov, A. M. (2001) *Chaos 17, 3--12*.

Gursky, V. V., Jaeger J, Kozlov K.N., Reinitz, J., & Samsonov, A. M. (2004) Pattern formation and nuclear divisions are uncoupled in Drosophila segmentation: Comparison of spatially discrete and continuous models. *Physica D, to appear.*

Foe, V. A., & Alberts, B. M. (1983) *J. Cell Sci. 61, 31--70*.

•Thomas, R. (1991). *J. Theor. Biol. 153, 1--23.*

•Thom, R. (1975). **Structural Stability and Morphogenesis** , W. A. Benjamin, Reading MA.

•Mjolsness, E., Sharp, D. H., & Reinitz, J. (1991) *J. Theor. Biol. 152, 429--453.*

•Reinitz, J., & Sharp, D. H. (1995) *Mech. Dev. 49, 133--158.*

References (cont'd)

- Wieschaus, E., Nusslein-Volhard, C. & Jurgens,G. (1984) *Roux Arch. Dev. Biol. 193, 296--307.*
- Nusslein-Volhard, C., Frohnhofer, H. G., & Lehmann, R. (1987) *Science 238, 1675--1687.*
- Akam, M. (1987) *Development 101, 1--22.*
- Ingham, P. W. (1988) *Nature 335, 25--34*.
- Lawrence, P. A. (1992) ***The Making of a Fly*** , Blackwell Sci. Publ., Oxford, UK.
- Merrill, P. T., Sweeton, D., & Wieschaus,E. (1988) *Development 104, 495--509.*
- Wieschaus, E., & Sweeton, D. (1988) *Development 104, 483--493.*
- Carroll, S. B., & Scott, M. P. (1986) *Cell 45, 113--126*.
- Reinitz, J., Mjolsness, E., & Sharp, D. H. (1995) *J. Exp. Zool. 271, 47--56.*
- Reinitz, J., Kosman, D., Vanario-Alonso, C. E. & Sharp, D. H. (1998). *Dev. Gen. 23, 11--27.*

References (cont'd)

- Meinhardt, H. (1986) *J. Cell Sci. (Supp.) 4, 357--381.*
- Meinhardt, H. (1988) *Development 104, 95--110.*
- Nagorcka, B. N. (1988) *J. Theor. Biol. 132, 277--306*.
- Meinhardt, H. (1983) *J.Embryol. Exp. Morphol. 76, 115--137.*
- Bryson, A. E. & Ho, Y. C. (1975). **Applied Optimal Control** , Taylor and Francis.
- Sanchez, L. & Thieffry, D. (2001). *J.Theor. Biol. 211, 115--141.*
- Yasuda, G. K., Baker, J. & Schubiger, G. (1991). *Genes Dev. 5, 1800-1812.*
- Lacalli, T. C., Wilkinson, D. A., & Harrison,L. G. (1988) *Development 104, 105--113.*
- Meinhardt, H. (1982) **Models of Biological Pattern Formation**, Academic Press, New York.
- Hunding, A., Kauffman, S. A., & Goodwin, B. C. (1990) *J. Theor. Biol. 145, 369--384.*
- Lacalli, T. C. (1990) *J. Theor. Biol. 144, 171--194.*

References (cont'd)

- Parr, B. A., & McMahon, A. P. (1995) *Nature 374, 350--353.*
- Warrior, R., & Levine, M. (1990) *Development 110, 759--767*.
- Frasch, M., & Levine, M. (1987) *Genes Dev. 1, 981--995*.
- Kosman, D., Small, S., & Reinitz, J. (1998) *Dev. Genes Evol. 208, 290--294.*
- Lurie, K. A. (1993) **Applied Optimal Control Theory of Distributed Systems** , Plenum, NY
- Driever, W., & Nusslein-Volhard, C. A gradient of bicoid protein in Drosophila embryos. (1988) *Cell 54, 83--93.*
- Nusslein-Volhard, C., & Wieschaus, E. (1980) *Nature 287, 795--801.*
- Nusslein-Volhard, C., Wieschaus, E., & Kluding,H. (1984) *Roux Arch. Dev. Biol. 193, 267-282*.
- Jurgens, G., Wieschaus, E., Nusslein-Volhard, C., & Kluding, H. (1984) *Roux Arch. Dev. Biol. 193, 283--295.*