Mathematical models of complexity

Ovidiu Radulescu, IRMAR and Symbiose project IRISA

Summary

CV

Brief state of the art: complex systems, systems biology

Contributions in biology:

- ✓ Markov processes in molecular biology
- ✓ Qualitative equations for functional genomics
- ✓ PDE models for pattern formation

Conclusion

CV

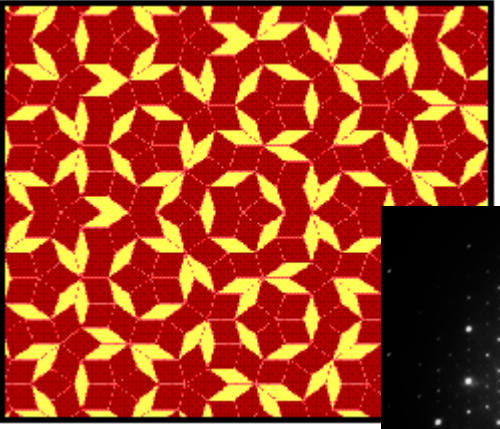
Education:
1989 Diplôme d'Ingénieur Physique des Solides, Bucarest
1994 Doctorat Physique des Solides, Orsay (félicitations)
1996 DEA probabilités, Marne-la-Vallée

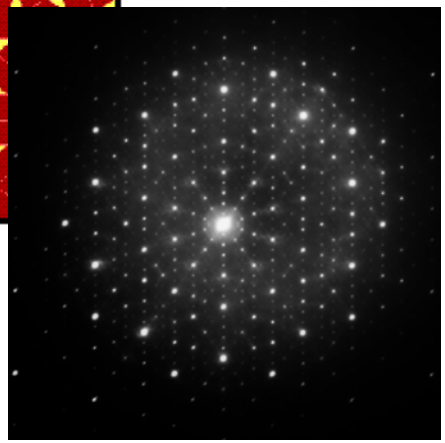
Recherche: interdisciplinarité, transversalité
2 post-docs (Pays Bas et Angleterre)
27 articles acceptés, 12 proceedings conf.

Enseignement:
1991-1993, moniteur physique Orsay, vacataire Ecole Centrale de Paris
1993-1996 ATER et PRAG physique, Marne la Vallée
depuis 1999 MC en mathématiques à Rennes 1
encadrement d'une thése (en mathématiques) et d'une dizaine de stages

Responsabilités:
membre commission informatique, animation d'un groupe de travail
coordinateur d'une ACI

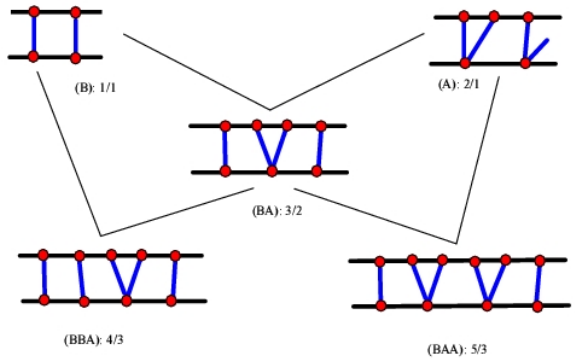
Complex systems

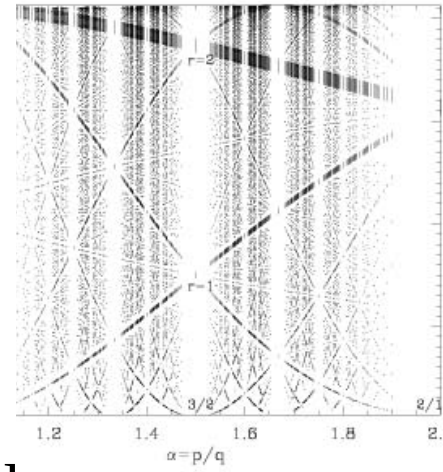




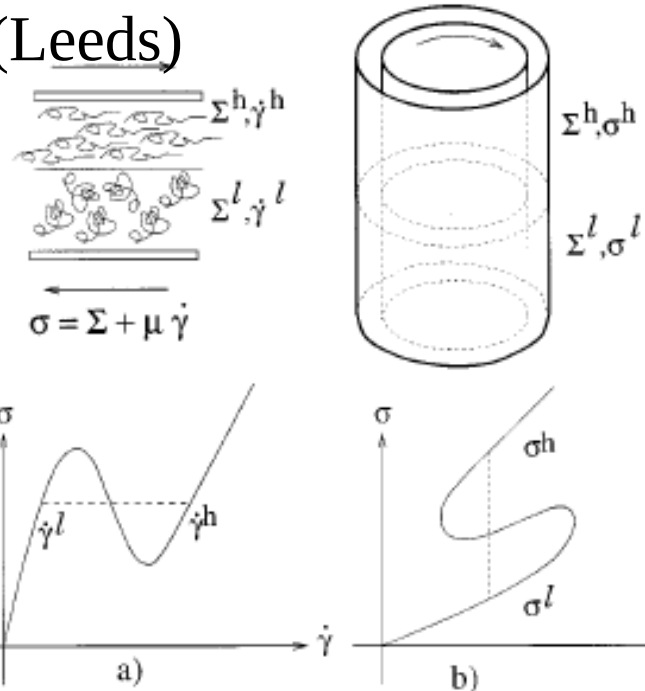
Quasicrystals
(Orsay)

Incommensurate composites
(Nimégue)

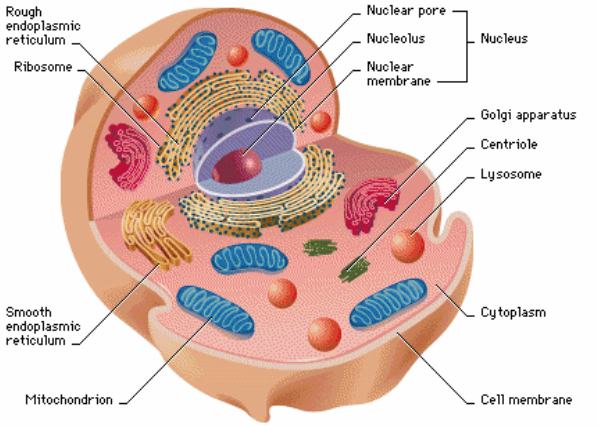




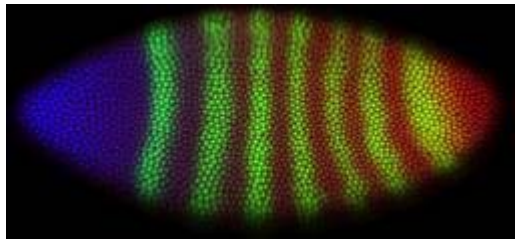
Wormlike micelles
(Leeds)



Cellular physiology
(Rennes)



Development
(Rennes)



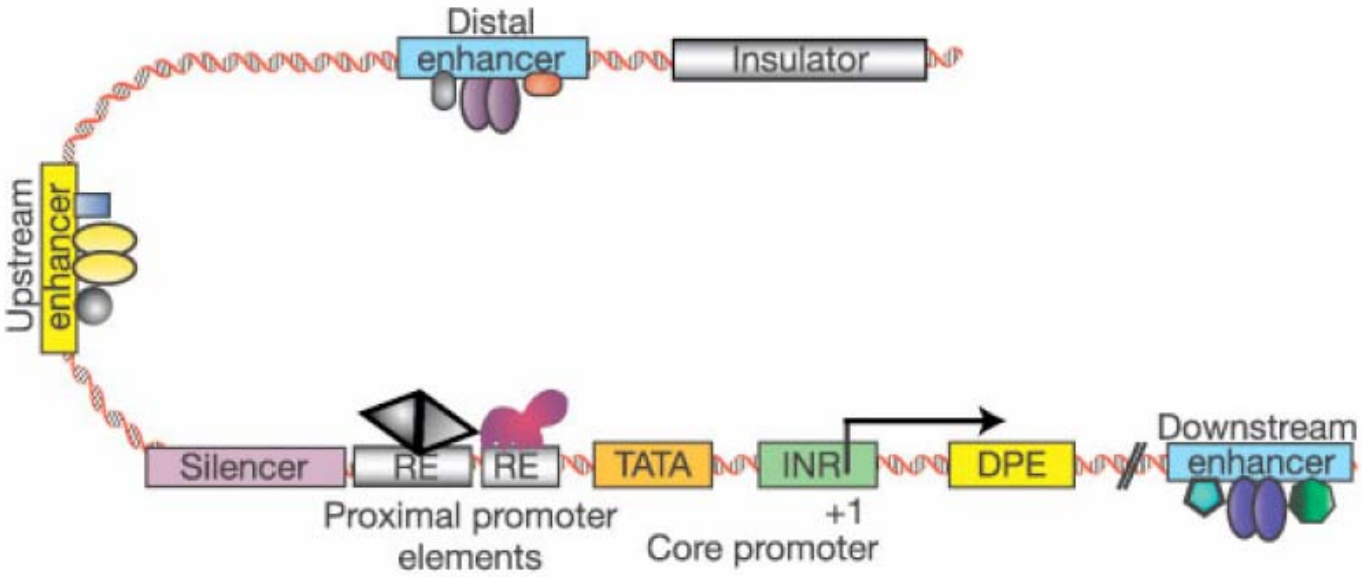
What complex systems have in common

- Order as framework for transformation: crystals, dissipative structures, patterns
- Defects as motors for transformation: points, lines, interfaces
- Hierarchical organisation
- Nonlinearity
- Stability, robustness
- Universality

Systems biology

- Mathematical modeling of physiology
- Transversal field, imports methods from physics, control theory, automata, chemical kinetics
- After rapid evolution, critical stage: obstacle raised by the complexity of higher organisms (models are scarce or weakly predictive)
- There is a need for new methods
  - analysis methods for massive data
  - model reduction
  - more realistic models using physico-chemistry

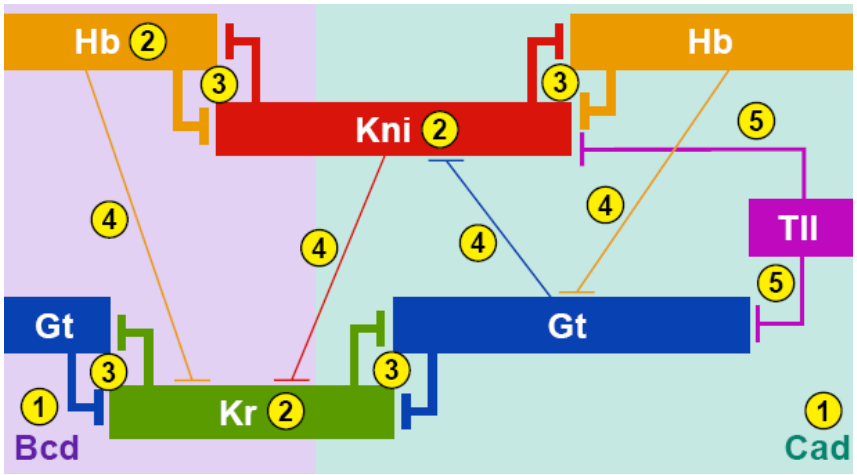
Generic complex metazoan transcriptional control modules



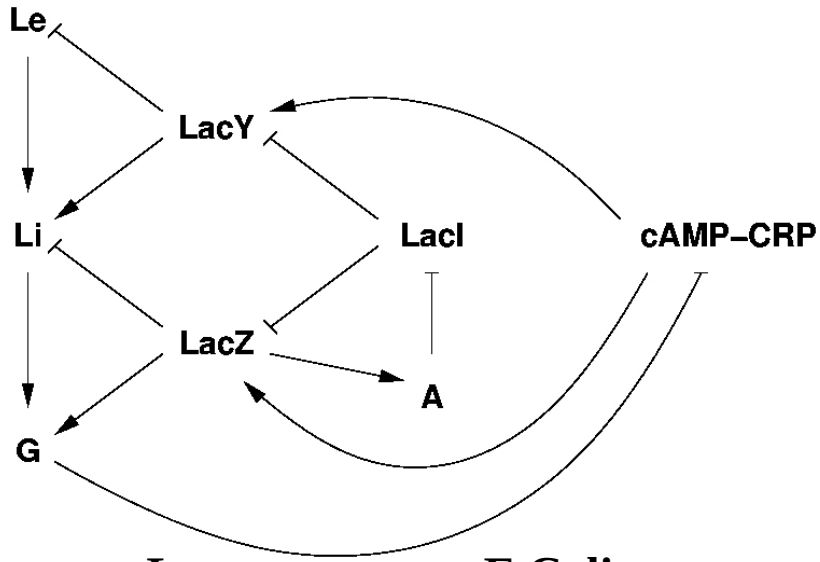
INR = initiator

DPE = downstream promoter element

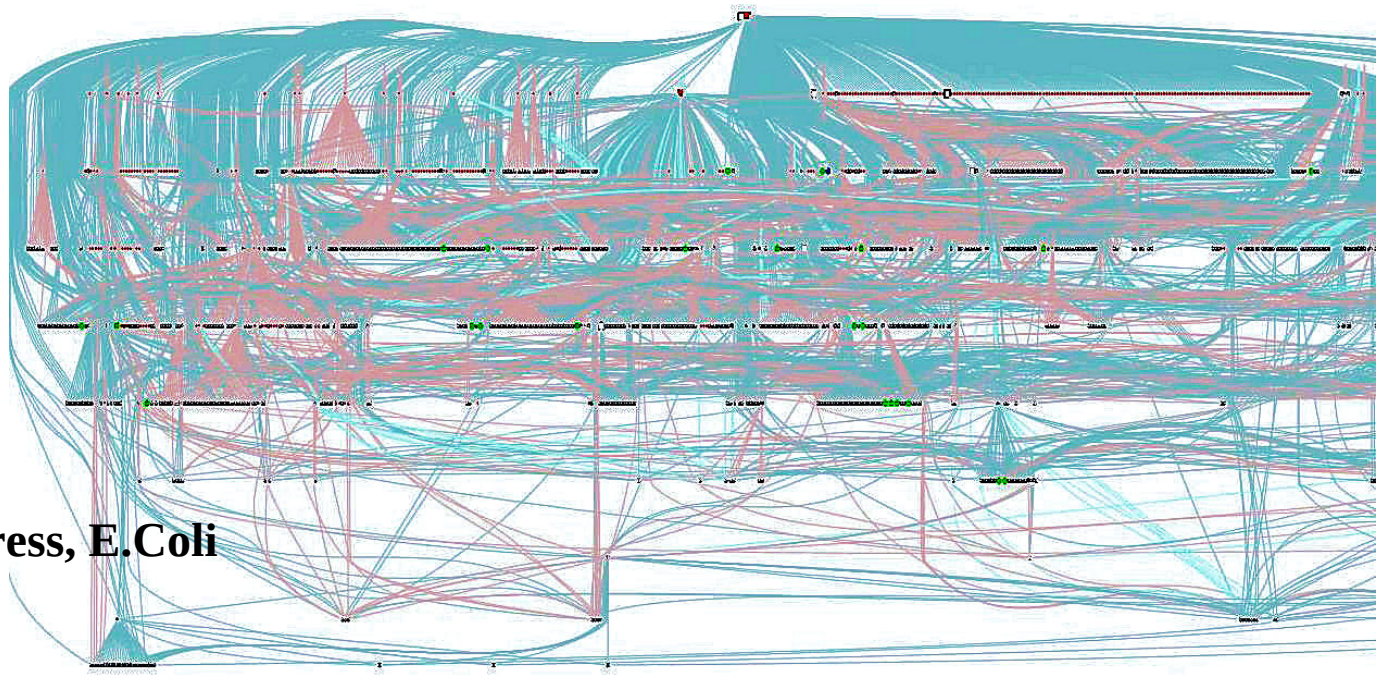
Gene regulation is the result of many interactions



Gap genes, first 3 hours of Drosophila



Lactose operon, E.Coli

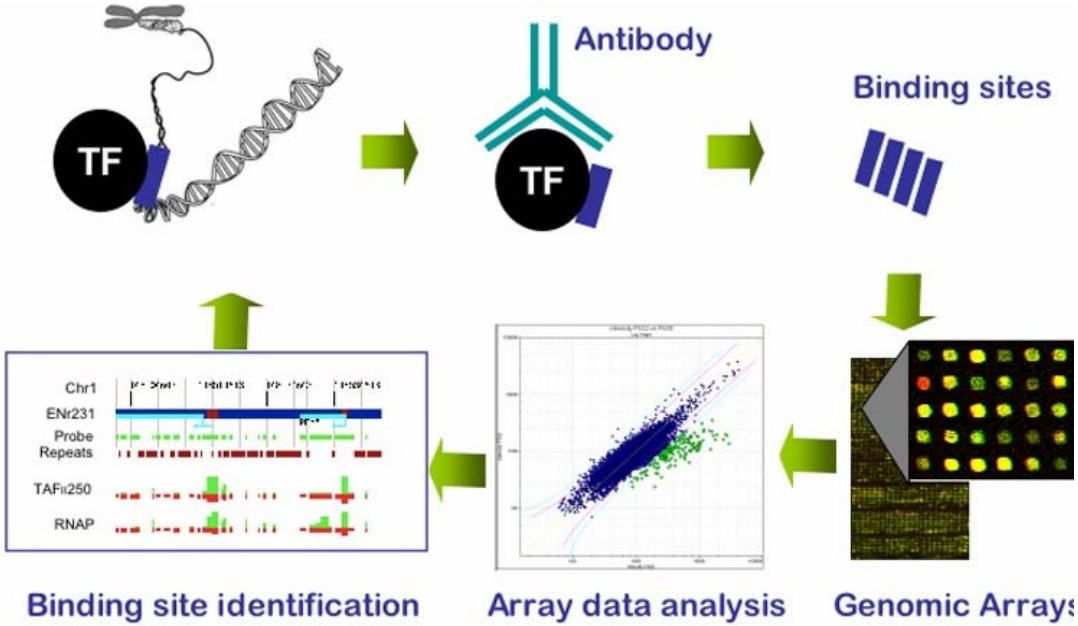


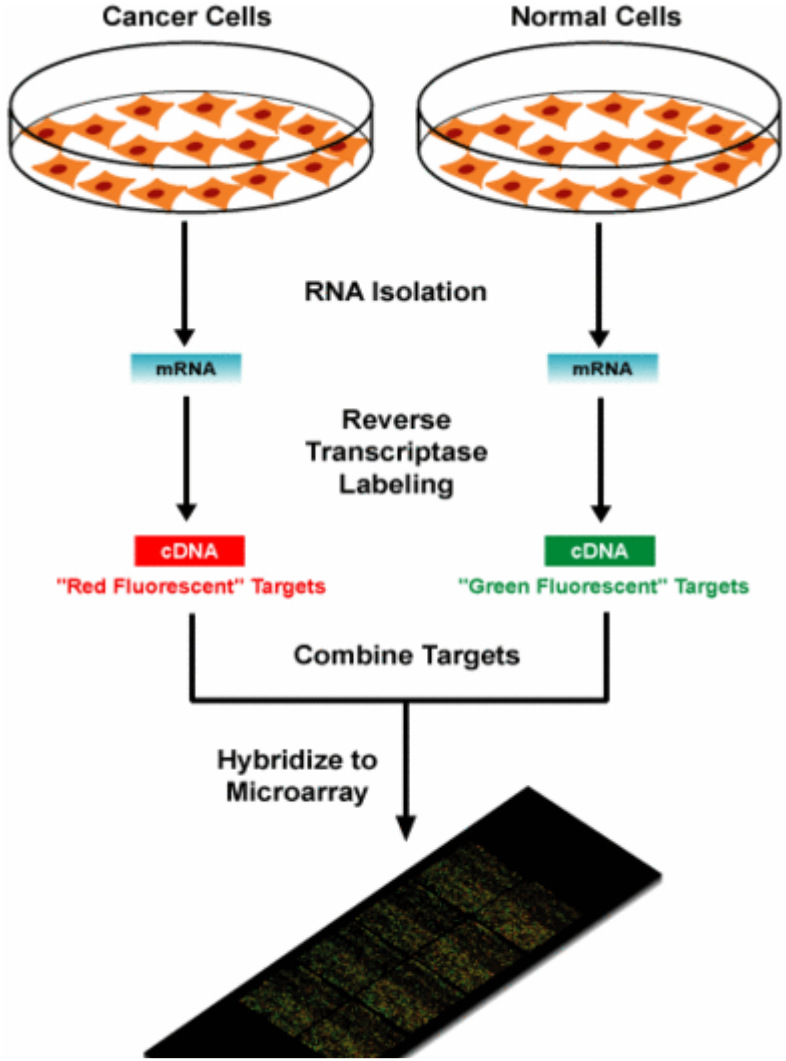
Nutritional stress, E.Coli

Network models unify various processes

DNA Chip

Chromatin ImmunoPrecipitation on Chip





Various kind of data: differences of concentrations, direct test of qualitative interaction

Strategy

Aims:

- Model construction
- Model analysis
- Biological predictions

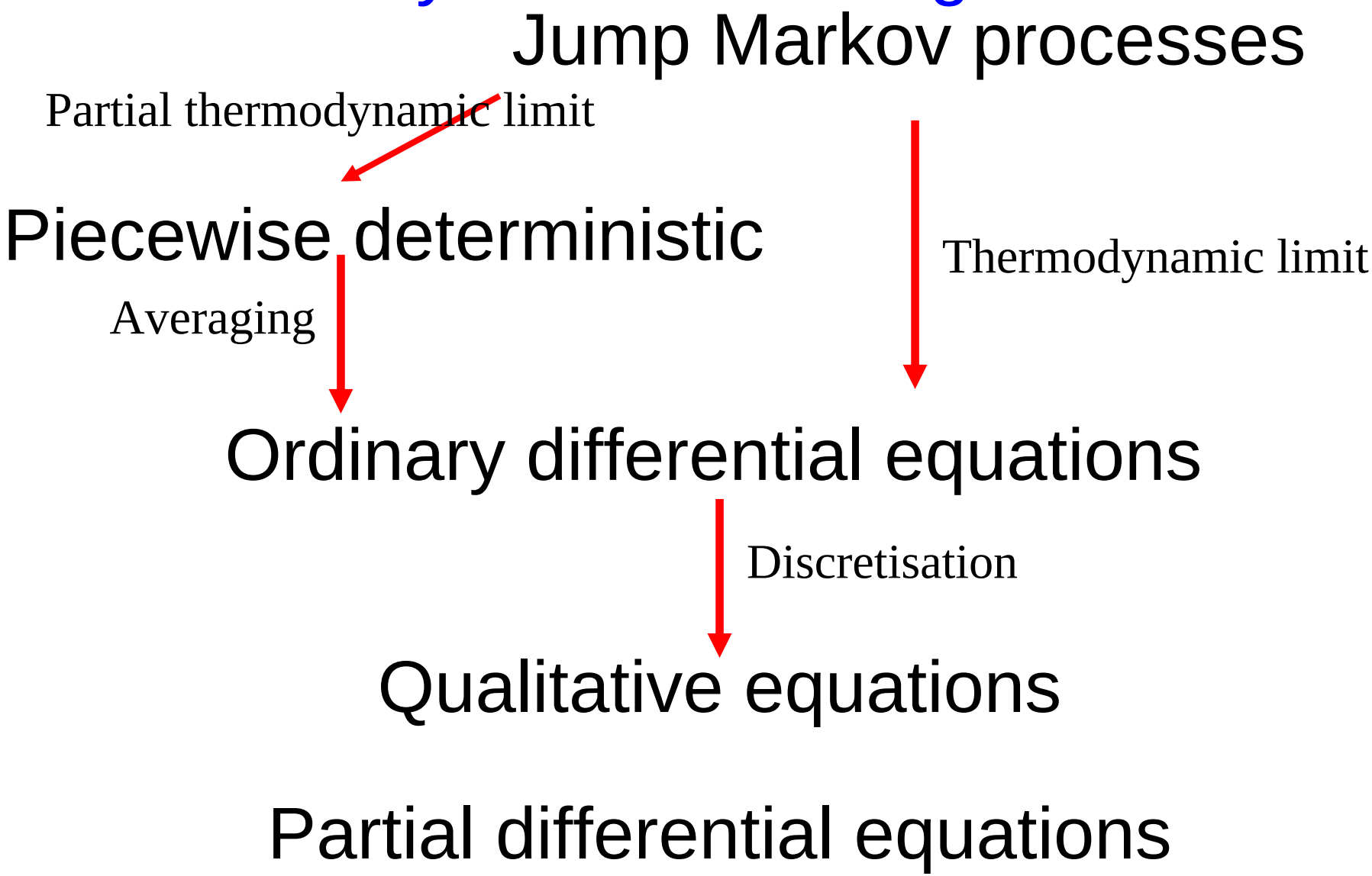
Difficulties:

- Data collection is massive but unguided
- Reverse engineering is difficult
- Models are non-linear and in very high dimension
- Interpretation of computer simulations is difficult

My solutions:

- Guide data collection (experiment design)
- Do not start reverse engineering from scratch (model correction)
- Develop new mathematical techniques for model analysis
- Look for network design principles

My mathematical garden



My contributions

My contributions to this field:

1) Modeling stochasticity of molecular biology processes by piecewise deterministic Markov processes
2) Qualitative equations for analysis of massive data
3) Carr-Pego type model reduction for pattern formation
4) Measure concentration as framework for robustness

Collaborations

Computer scientists: A.Siegel, M.LeBorgne (IRISA Symbiose), M.Samsonova(St.Petersburg)

Biologists: N.Theret (INSERM), S.Lagarrigue (INRA), A.Lilienbaum (CNRS), J.Reinitz (Stony Brook)

Mathematicians: S.Vakulenko(St.Petersburg), A.Gorban(Leicester), E.Pécou(Nice)

Research project MathResoGen (2003-2006)

Modeling stochasticity in molecular biology by Markov processes

Modeling stochastic effects

Markov jump processes: Renyi, Bartholomay, 50'

$A_1, \ldots, A_n$ **are n chemical species** $X \in Z^n$ **is the state**

$\alpha_{i1}A_1 + \ldots + \alpha_{in}A_n \rightleftarrows \beta_{i1}A_1 + \ldots + \beta_{in}A_n$ **biochemical reaction**

$\theta_i = \beta_i - \alpha_i \in Z^n,\ i = 1, n_r$ **jump vector**

$$\lambda(X) = \sum_{i=1}^{nr} [V_i(X) + V_{-i}(X)]$$ **intensity**

$$\mu(X,.) = \sum_{i=1}^{nr} [q_i(X)\delta_{X+\theta_i}(.) + q_{-i}(X)\delta_{X-\theta_i}(.)]$$ **distribution of jumps**

$$q_i(X) = V_i(X) / \sum_{j=1}^{nr} [V_j(X) + V_{-j}(X)]$$ **jump probability**

Thermodynamic (deterministic) limit

Suppose that the mass action law is satisfied

$$V_i(X)=\Omega v_i(X), \quad v_i(X)=k_i\prod_{s=1}^{n}x_s^{\alpha_{is}}$$

$$V_{-i}(X)=\Omega v_{-i}(X), \quad v_{-i}(X)=k_{-i}\prod_{s=1}^{n}x_s^{\beta_{is}}$$

Ω **: reaction volume**

Rescale the process $x_i=X_i/\Omega$

For $\Omega\to\infty$ **the Markov jump processes** x_i **converges in probability to the solution of a system of ordinary differential equations (Kurtz, 70)**

$$\frac{dx(s)}{ds}=F(x(s)), \quad F(x)=\sum_{i=1}^{nr}v_i(x)\theta_i$$

Piecewise deterministic limit

Some species are in small numbers!

$$\Omega \to \infty, \quad \varepsilon \to 0, \quad \Omega\varepsilon \to 1$$

concentration of rare species

use frequent/rare species decomposition

$$X = (X^f, X^r)$$

mass action law is not applicable and should be replaced by

$$V_i(X) = \tilde{V}_i\left(\frac{X^f}{\Omega}, \frac{X^r}{\varepsilon\Omega}\right), \quad \forall i, \theta_i^r \neq 0$$ **reactions acting on rare species**

$$V_i(X) = \Omega \tilde{v}_i\left(\frac{X^f}{\Omega}, \frac{X^r}{\varepsilon\Omega}\right), \quad \forall i, \theta_i^r = 0$$ **reactions not acting on rare species**

O.Radulescu, A.Muller, A.Crudu (TSI in press)

Piecewise deterministic limit result

For $\Omega \to \infty, \varepsilon \to 0, \Omega\varepsilon \to 1$ **the Markov jump process** $X = (X^f/\Omega, X^r)$ **converges to a piecewise deterministic process:**

$X^r(s)$ **is discrete and jumps with intensity** $\tilde{V}_i(x^f, X^r)$

Between two jumps $x^f(s)$ **is continuous and satisfies:**

$$\frac{dx^f(s)}{ds} = F^f(x^f(s), X^r(s)) = \sum_{\theta_i^r = 0} \tilde{v}_i \theta_i$$

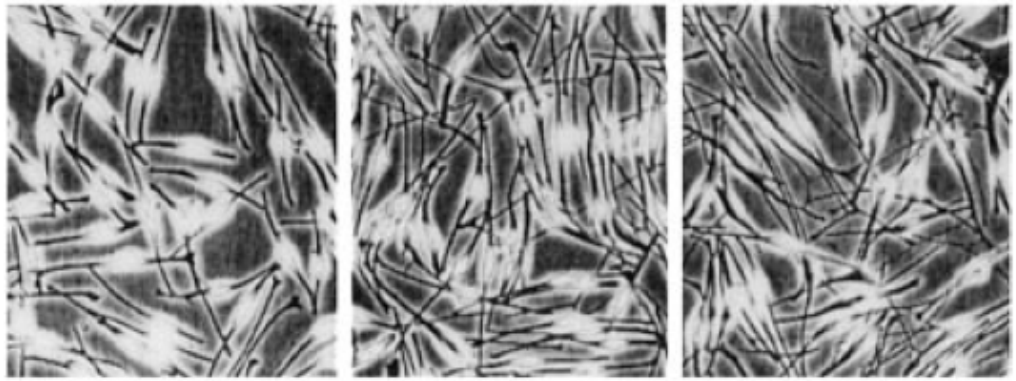
O.Radulescu, A.Muller, A.Crudu (TSI in press)

Application: hybrid stochastic simulation algorithm

1. Initialize $x^f = x^f_0, \quad X^r = X^r_0, \quad t = 0$
2. Generate exponential random time $\tau \sim \exp[\ \lambda(x^f, X^r)]$
3. Use deterministic solver to propagate $x^f(t) \rightarrow x^f(t+\tau)$
4. Change X^r to a new discrete value
5. Increment time $t \rightarrow t+\tau$
6. If t<tmax goto 2

O.Radulescu, A.Muller, A.Crudu (TSI in press)

Application to haploinsufficiency



Biological problem:

Syndrom due to deficient genotype : insufficient copy number
Phenotype: heterogenous cell populations

Aim:

Find the simplest model that reproduces this situation

Model for haploinsufficiency

Markov jump model (Cook 99)

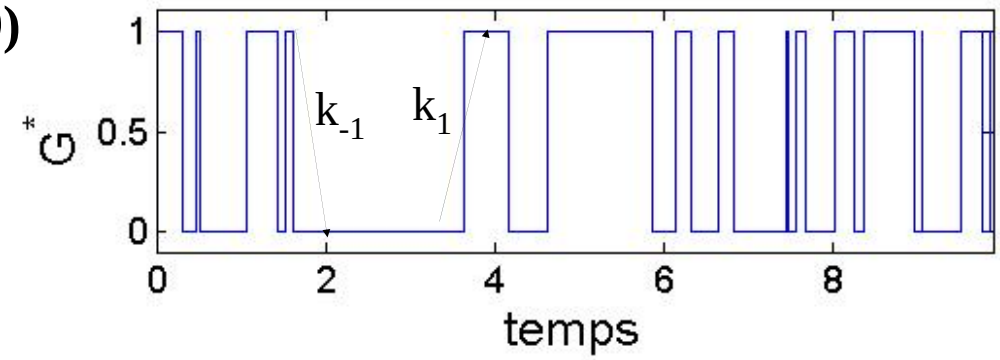
$$G \underset{k-1}{\overset{k1}{\rightleftarrows}} G^*$$

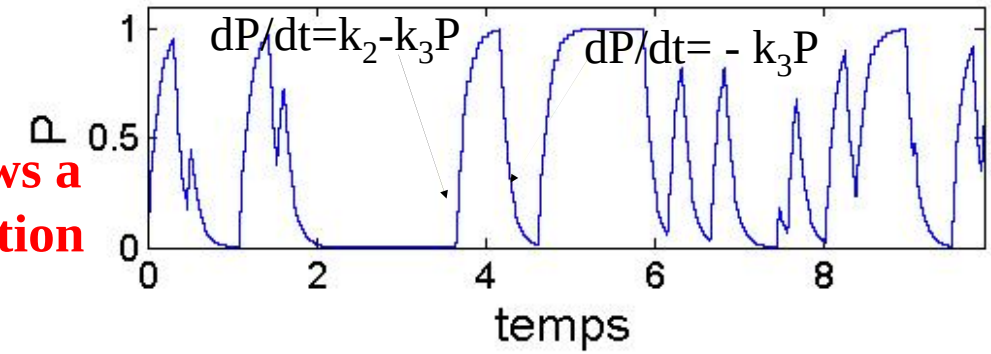
$$G^* \xrightarrow{k2} G^* + P$$

$$P \xrightarrow{k3}$$

$$G + G^* = 1$$

$$\varepsilon = 1/\Omega$$

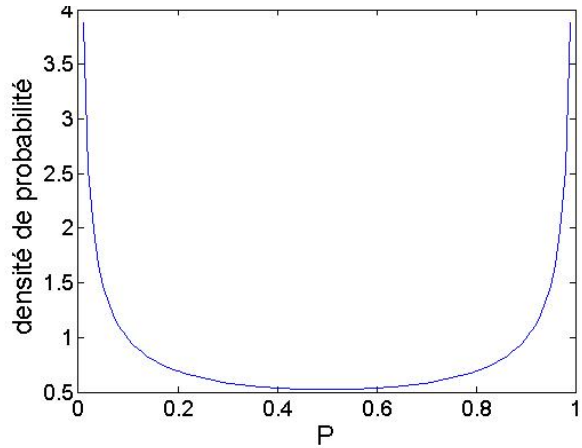




Result: If $k_2 = O(\Omega)$ **the model allows a piecewise deterministic approximation**

$$\frac{dP}{dt} = \begin{cases} -k_3 P + k_2, & \text{if} \quad G^* = 1 \\ -k_3 P, & \text{if} \quad G^* = 0 \end{cases}$$

Study intermittency of trajectories and the invariant distribution



O.Radulescu, A.Muller, A.Crudu (TSI in press)

Conclusion

Results:

- The protein production is intermittent
- The heterogeneity of the phenotype can be described by a Beta distribution

The same method will be applied to larger, more complex models; in project NFκB signaling

Qualitative equations

Qualitative equations

Biological problem:

Following a perturbation (stress, signal) the state of the cell changes. Variations of hundreds or thousands of variables can be monitored. How to use this information?

Steps:

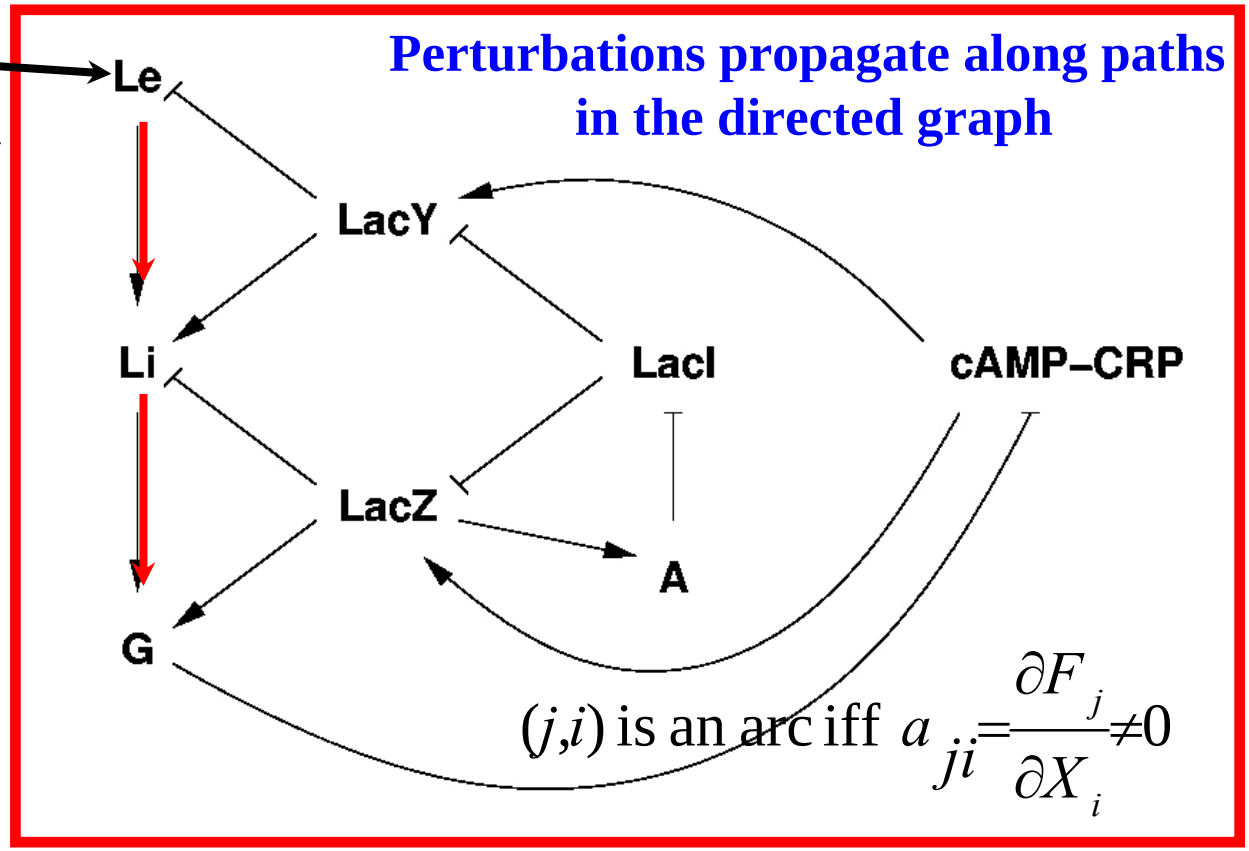
- develop an “elasticity” theory of graphs (O.Radulescu et al. J.R.Soc.Interface 2006)
- translate this theory into qualitative equations (with A.Siegel et al. Biosystems 2006)
- polynomial algorithms for solving systems of qualitative equations (with Ph.Veber, M.leBorgne et al. Complexus 2006)
- application to huge networks (with C.Vargas et al., proc. RIAMS 2006)

Elasticity of graphs

$\frac{dX}{dt}=F(X,P)$ **dynamics** $F(X,P)=0$ **Steady state equation**

Steady state is perturbed $\delta P \rightarrow \delta X$

∂G **graph boundary**



Dirichlet solution: $\delta X_i=\sum_{j\in\partial G}\sum_{j\rightarrow i}\frac{a_{j\rightarrow i}}{C_{j\rightarrow i}}\delta X_j$

(O.Radulescu et al. J.R.Soc.Interface 2006)

Qualitative equations

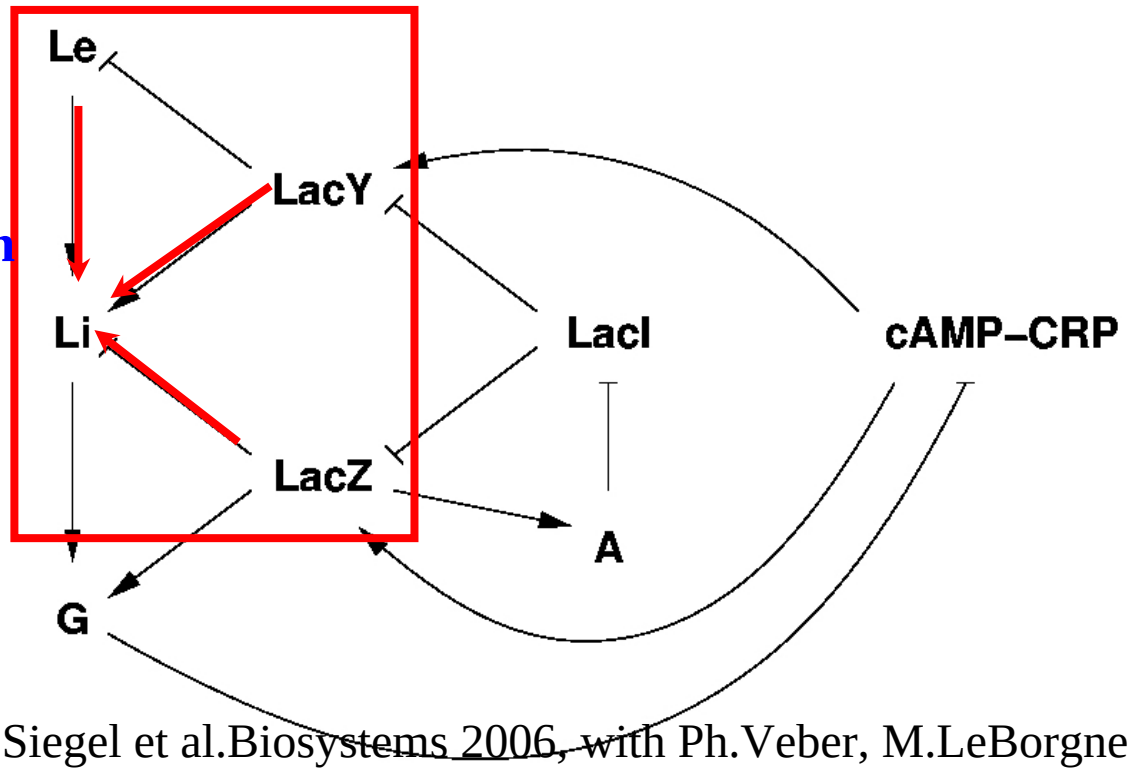
$$\delta X_i = -\left(\frac{\partial F_i}{\partial X_i}\right)^{-1} \sum_{j \in pred(i)} a_{ji}\delta X_j$$ **Dirichlet solution for subgraph**

$$sign(\delta X_i) = \sum_{j \in pred(i)} sign(a_{ji}) sign(\delta X_j)$$ **Qualitative equation**

$$sign \in \{-,+,?\}$$ **Sign algebra**

Li=Le+LacY-LacZ

subgraph



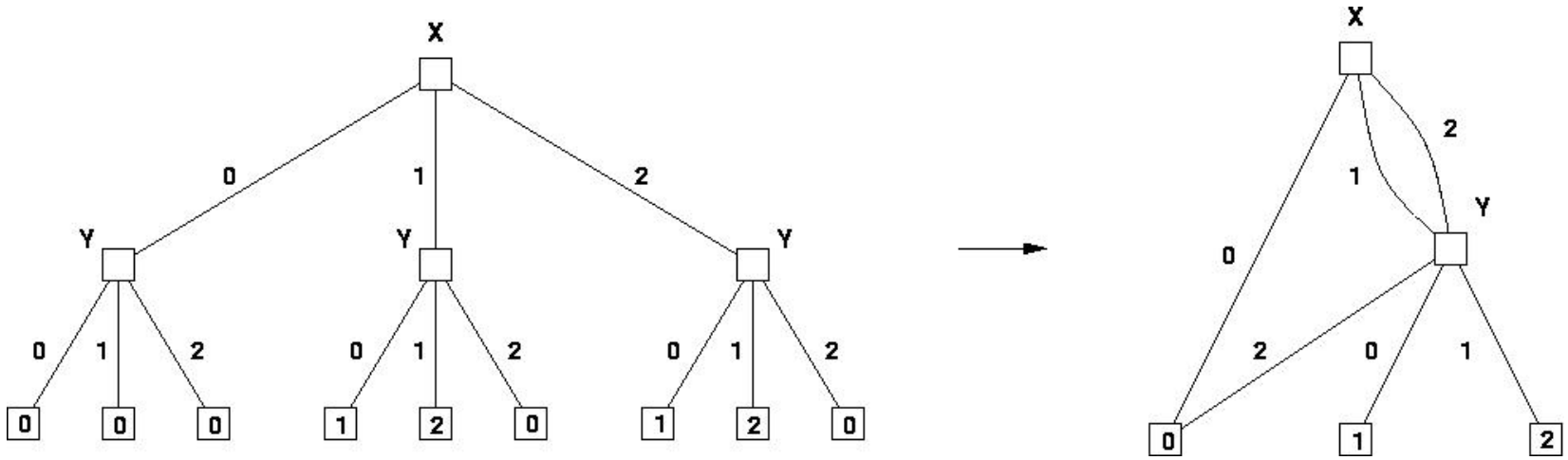
$$+ + - = ? \quad + + + = +$$
$$+ \times - = - \quad + \times + = +$$
$$? + - = ? \quad ? + + = ?$$
$$? \times - = ? \quad ? \times + = ?$$

$\approx$	$+$	$-$	$?$
$+$	T	F	T
$-$	F	T	T
$?$	T	T	T

(with A.Siegel et al.Biosystems 2006, with Ph.Veber, M.LeBorgne Complexus 2006)

Algorithm for solving qualitative equations

- **Map signs to elements of the finite field Z/3Z**
- **Map qualitative equations to polynomial equations over Z/3Z**
- **NP complete problem**
- **Ternary Decision Trees contracted to directed acyclic graphs and systematic use of cache memory for non-redundant computation**
- **Obtain exhaustive lists of solutions within minutes for 1000 nodes**

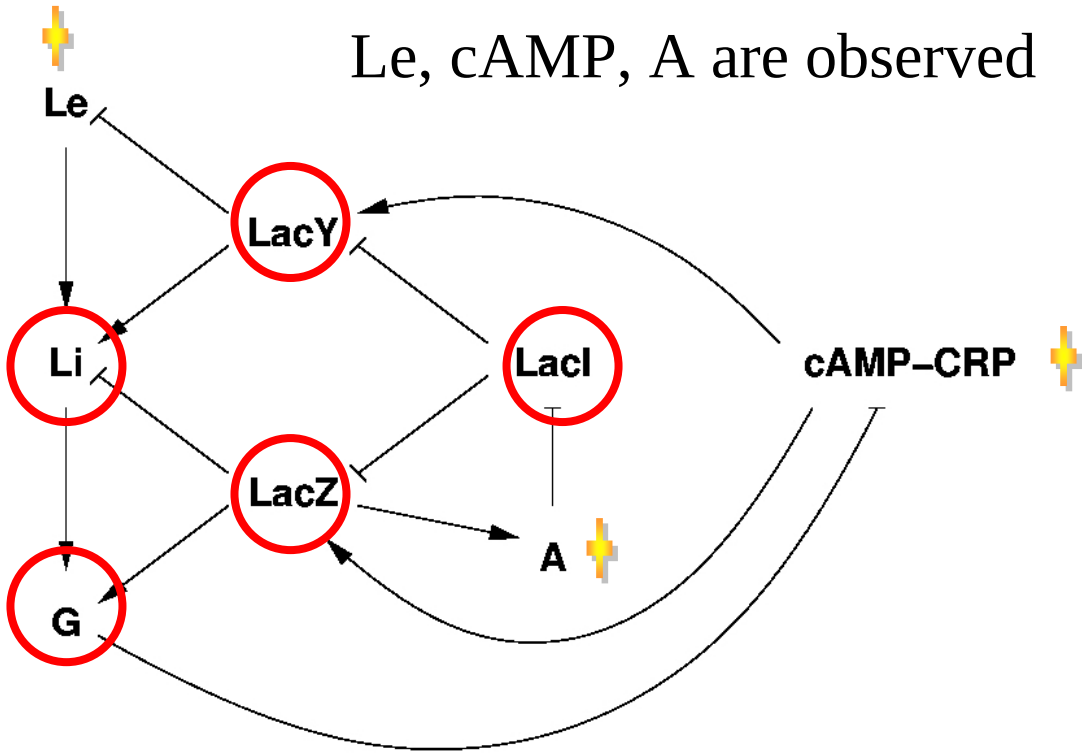


(with Ph.Veber, M.LeBorgne Complexus 2006)

Predictions of a model

hard components: variables whose values are the same (+ or -) in any solution

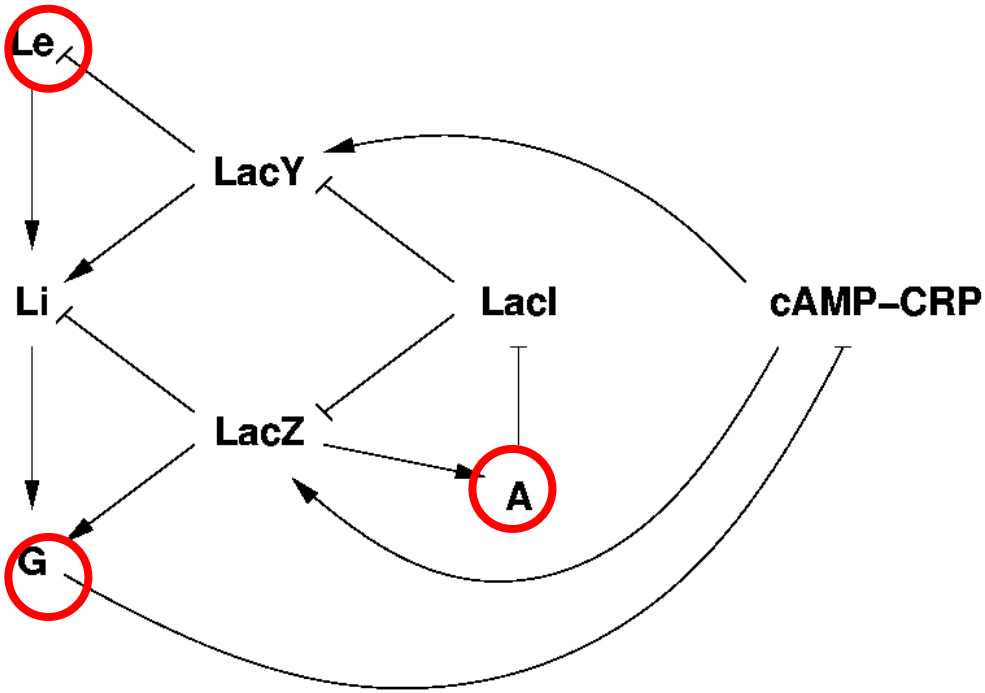
the hard components are the predictions of the model



Le, cAMP, A are observed

Li,G,LacZ,LacY, LacI are hard components

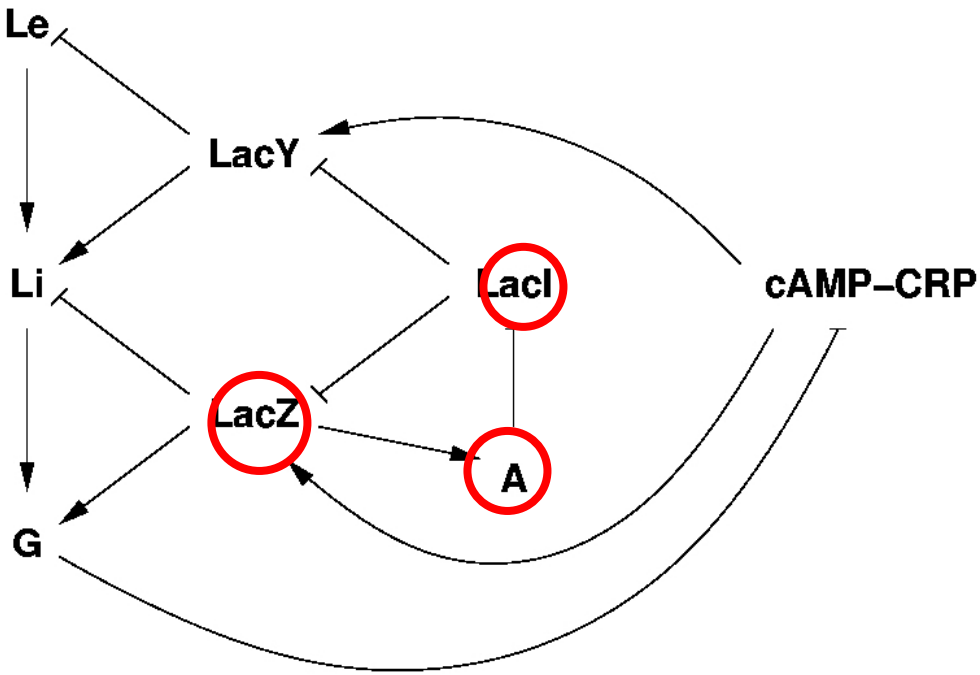
Experiment design



Any value of the triplet (Le,G,A) can be extended to a solution

These variables have no validation power

Use validation power for experiment design



Only 2 values (out of 8) of (LacI,A,LacZ), namely (+,–, –) (–, +,+) can be extended to a solution

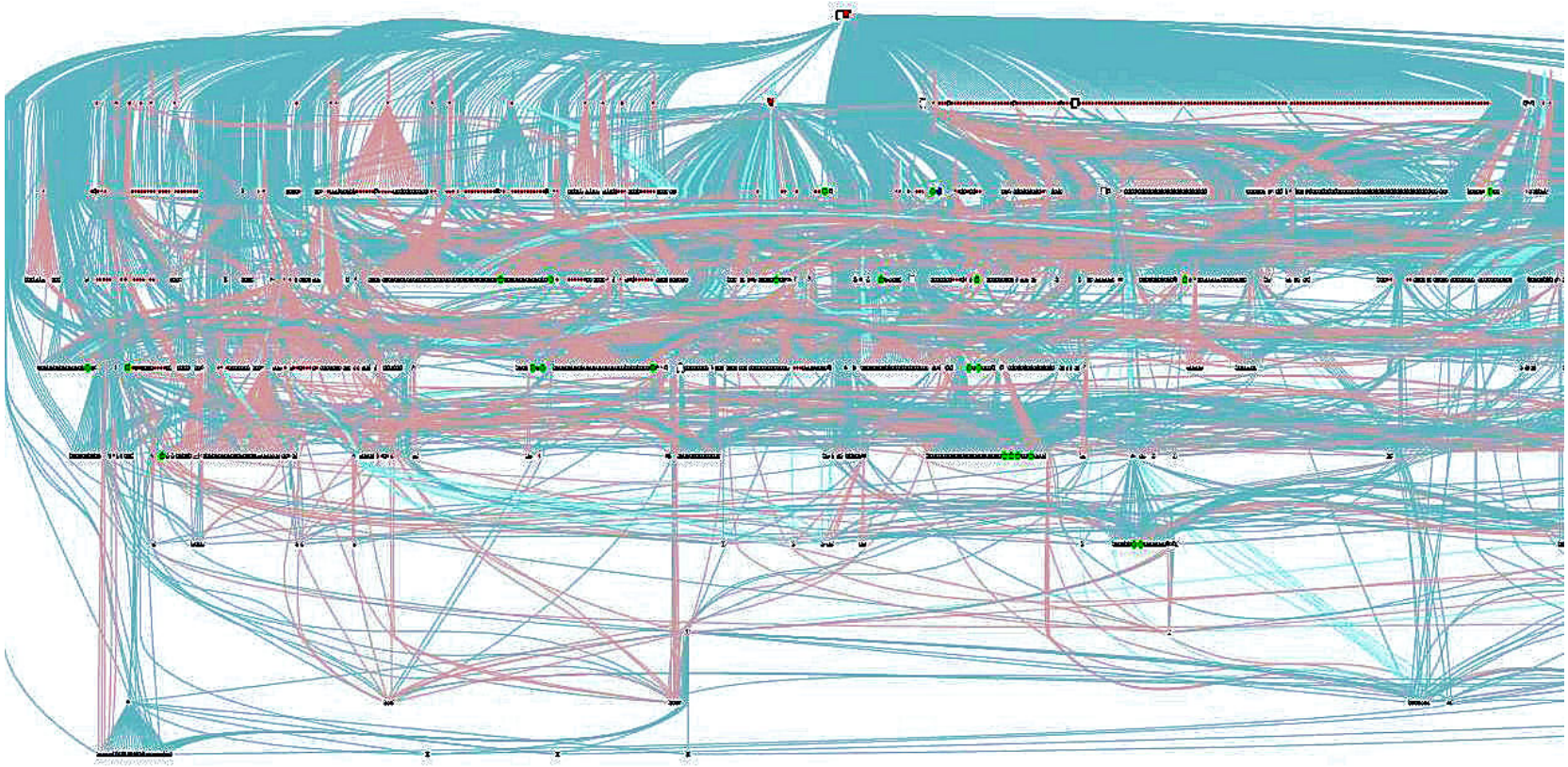
Define validation power as:

$$\tau(X_1, \ldots\ X_p) = 1 - \frac{\mathrm{val}(X_1, \ldots X_p)}{2^p}$$

Choose high validation power sets for optimal design

Large scale application: nutritional stress of E.Coli

1258 nodes, 2526 interactions, 10^{600} states, 10^{16} solutions



We have obtained both:

- **a set of predictions: from 40 observations in the stationary phase, 401 hard components, 26% of the network**
- **a set of corrections to the model: necessarily include σ factors**

Partial differential equations

Pattern formation

Problem:

- Patterns form in very different complex systems (Drosophila embryo before gastrulation, shear banding of complex fluids).
- The examples are of Wolpert type, less studied in mathematics. Can we find an unified approach?

Cornerstones:

- of complex fluids: understand the relation between structure and flow properties
- of developmental biology: understand canalization, stability of development

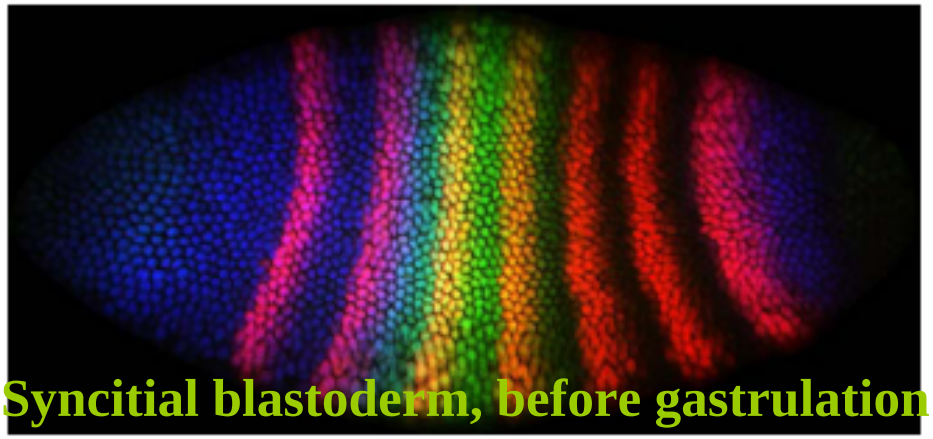
Collaborations:

P.D.Olmsted (Physics,Leeds), JP.Decruppe(Physics,Metz), JF.Berret, G.Porte (Physics,Montpellier) on wormlike micelles

S.Vakulenko (Maths,St.Petersburg), J.Reinitz(Appl.Maths and Biology, Stony Brook) on Drosophila

Problem 1: *Drosophila* segmentation genes

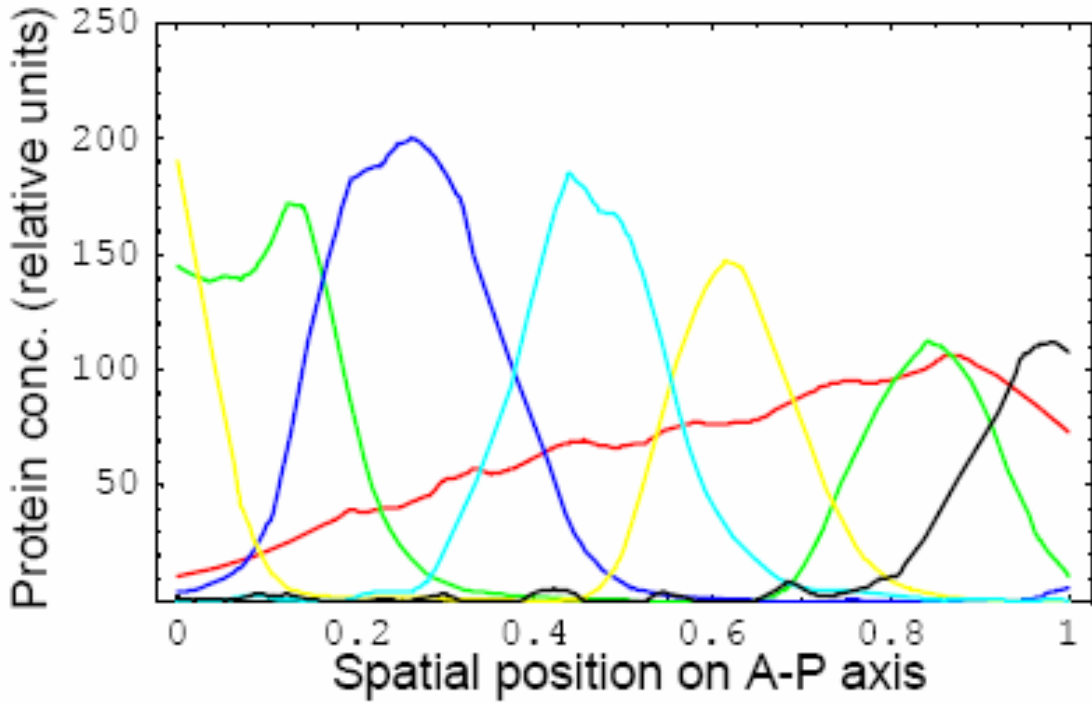
FlyEx Database: http://flyex.ams.sunysb.edu/FlyEx/



Syncitial blastoderm, before gastrulation

— data image of expression patterns for genes *eve*, *Kr*, and *hb*

1D approximation: Expression patterns for gap genes *hb*, *Kr*, *kni*, *gt*, *tll*, and *cad*



Model Reaction-diffusion equations

Synthesis

$$\frac{\partial u_a(x,t)}{\partial t} = R_a g_a\left(\sum_{b=1}^{N} T_{ab} u_b(x,t) + T_a m(x) + h_a\right)$$

Transport

$$+ D_a \nabla^2 u_a(x,t)$$

Decay

$$- \lambda_a u_a(x,t)$$

Synthesis

$$\frac{du_a(x,t)}{dt} = R_a g_a\left(\sum_{b=1}^{N} T_{ab} u_b(x,t) + T_a m(x) + h_a\right)$$

Genetic Interconnectivity Matrix (T):

Gene a \ b	1	2	⋯	N
1	T^{11}	T^{12}	⋯	T^{1N}
2	T^{21}	T^{22}	⋯	T^{2N}
⋮	⋮	⋮		⋮
N	T^{N1}	T^{N2}	⋯	T^{NN}

T parameters:

positive:	activation
negative:	repression
zero:	no interaction

Synthesis

$$\frac{du_a(x,t)}{dt} = R_a g_a\left(\sum_{b=1}^{N} T_{ab} u_b(x,t) + T_{am} m(x) + h_a\right)$$

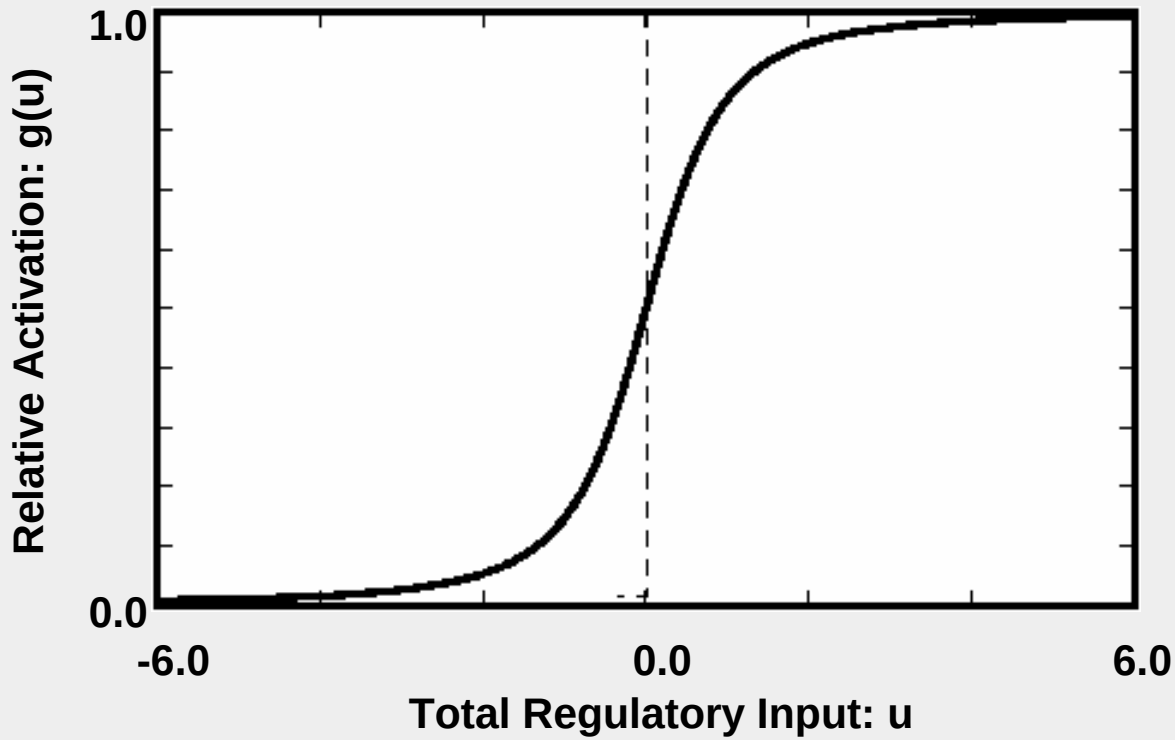
Action of maternal gradient (bicoid)

Bicoid profile m(x) develops in 1h after fertilization and remains constant during the blastoderm

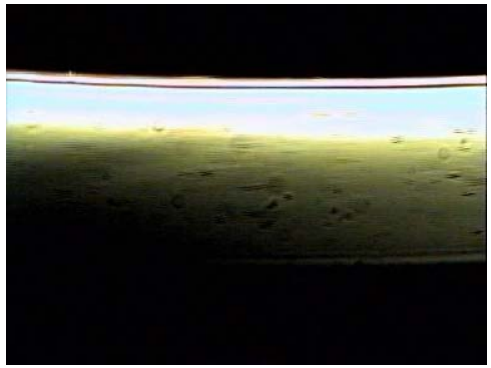
Synthesis

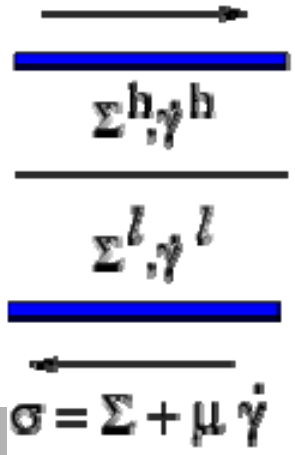
$$\frac{du_a(x,t)}{dt} = R_a g_a\left(\sum_{b=1}^{N} T_{ab} u_b(x,t) + T_a m(x) + h_a\right)$$

The regulation-expression function g(u):

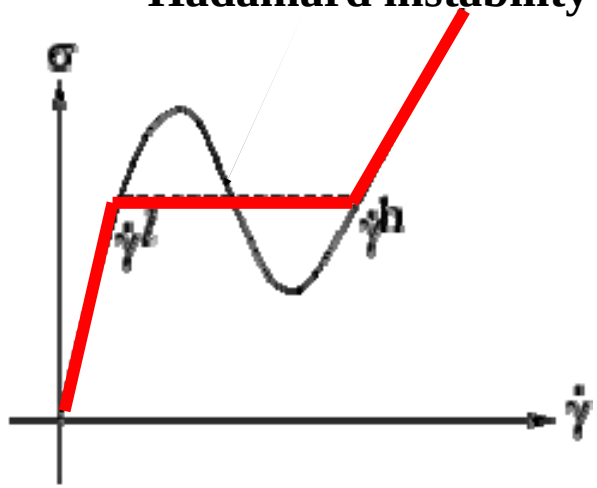


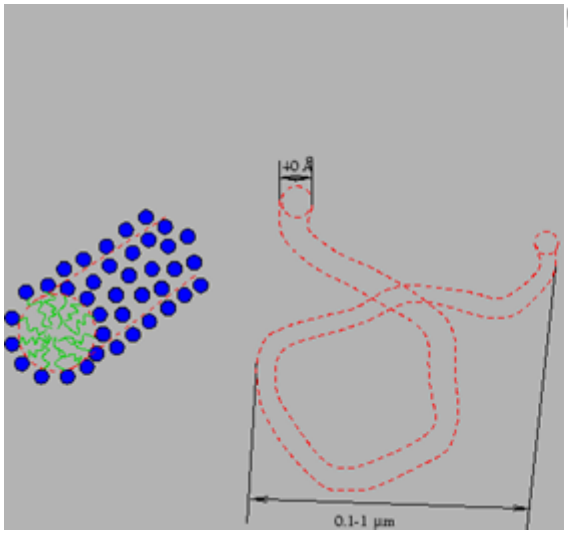
Problem 2: Shear banding of wormlike micelles

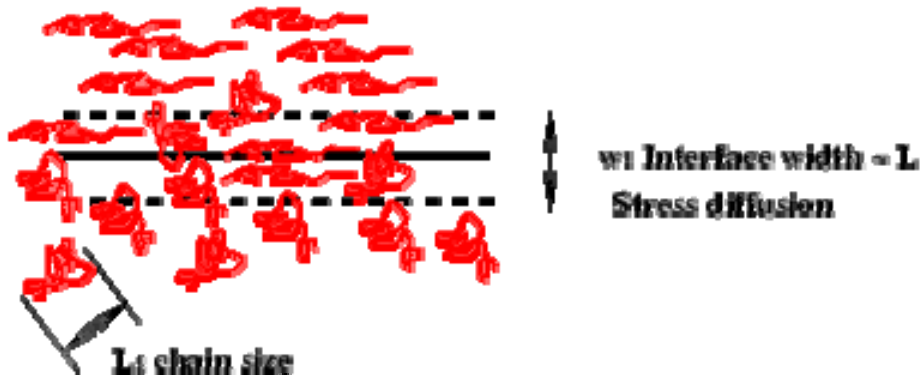




Hadamard instability







O.Radulescu et al. Rheol.Acta 1999, with PD.Olmsted J.Rheol. 1999

Model: Fluid-structure coupling

Navier-Stokes

$$\rho(\partial_t + \mathrm{v}.\nabla)\mathrm{v} = \nabla.\sigma$$

Re=0 approximation

$$\nabla.\sigma = 0, \quad \sigma = S + \varepsilon\dot{\gamma} = const.$$

Johnson-Segalman constitutive model + stress diffusion

$$(\partial_t + \mathrm{v}.\nabla)\Sigma - (\Omega\Sigma - \Sigma\Omega) - a(\Delta\Sigma + \Sigma\Delta) = D\nabla^2\Sigma + 2\mu\Delta/\tau - \Sigma/\tau$$

principal flow equations

$$\frac{\partial S}{\partial t} = D\frac{\partial^2 S}{\partial y^2} - \frac{S}{\tau} + \dot{\gamma}(1 - W)$$

$$\frac{\partial W}{\partial t} = D\frac{\partial^2 W}{\partial y^2} - \frac{W}{\tau} + \dot{\gamma}S$$

Stress dynamics is described by a reaction-diffusion system

O.Radulescu, PDOlmsted J.Non-Newtonian.Fl.Mech. 2000

Common framework: R-D PDE with small diffusion

Cauchy problem for the PDE system

$$u_t = \varepsilon^2 D \nabla^2 u + f(u, x, \varepsilon t)$$

$u = u(x,t) \in R^n$ $\quad x \in \Omega \subset R^q$, Ω **is compact with smooth frontier**

$$D = diag\{d1, d2, \ldots, dn\}$$

$u(x,0) = u_0(x)$ **initial data**

$\nabla u(x).n(x) = 0, \quad x \in \partial\Omega$ **no flux boundary conditions**

idea : consider the following shorted equation

$$v_t = f(v, x, \varepsilon t)$$

Result 1: Classification of patterning mechanisms

Patterning is diffusion neutral if for vanishing diffusion, the solution of the full system converges uniformly to the solution of the shorted equation

$$\left|u^{\varepsilon}(x,t)-v(x,t)\right| \to 0, \text{ uniformly in } x\in\Omega,\ t>0,\ \text{when}\quad \varepsilon\to 0$$

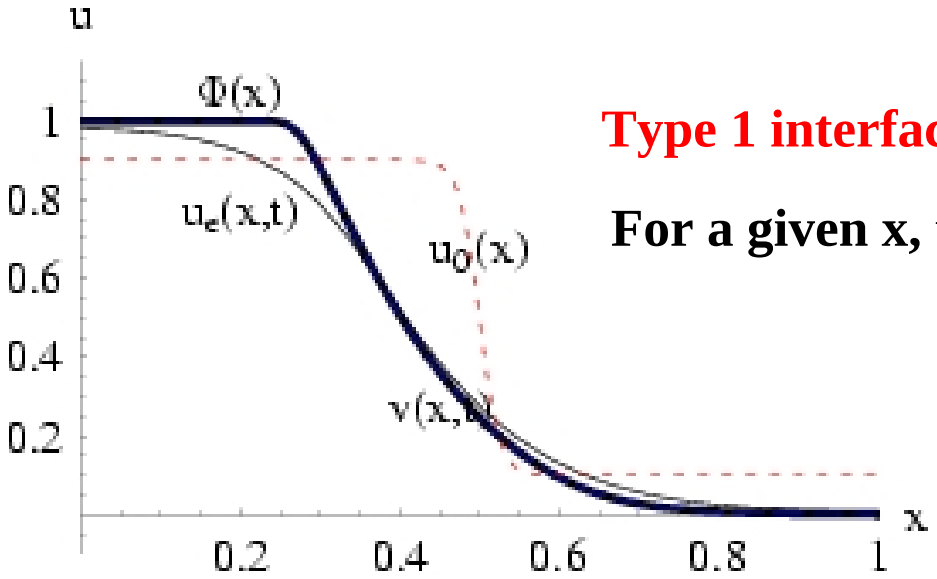
$u^{\varepsilon}(x,t)$ **solution of the full system**

v(x,t) solution of the shorted equation

If not, patterning is diffusion dependent

O.Radulescu and S.Vakulenko, arXiv qBio 2006

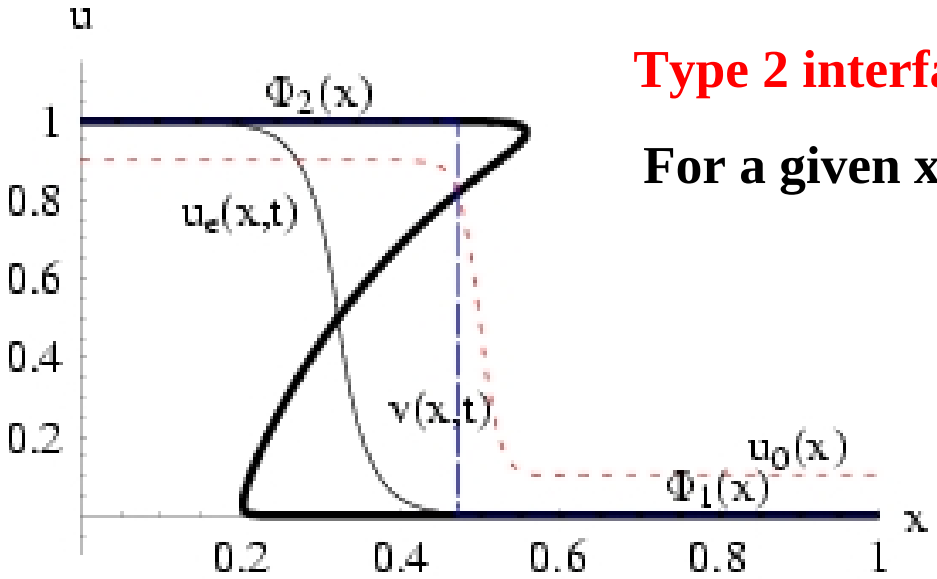
Result 2: Classification of interfaces



Type 1 interface

For a given x, the shorted equation has only one attractor $\phi(x)$

Patterning with type 1 interfaces is diffusion neutral



Type 2 interface

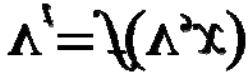
For a given x, the shorted equation has several attractors, here 2: $\phi_1(x),\phi_2(x)$

Patterning with type 2 interfaces is diffusion dependent
The width of type 2 interfaces can be arbitrarily small

O.Radulescu and S.Vakulenko, arXiv qBio 2006

Theorem on the diffusion neutral patterning

Consider the time autonomous situation 

and the shorted equation 

The patterning is diffusion neutral under the following conditions on the shorted equation:

i) uniform dissipativity

ii) strong linear stability

calculated at the attractor

iii) attraction basin condition

O.Radulescu and S.Vakulenko, arXiv qBio 2006

Theorem on the movement of type II interfaces in the bistable case

Invariant manifold decomposition for 

Travelling wave solution for the space homogeneous eq.

Equation for the position q(t) of the interface

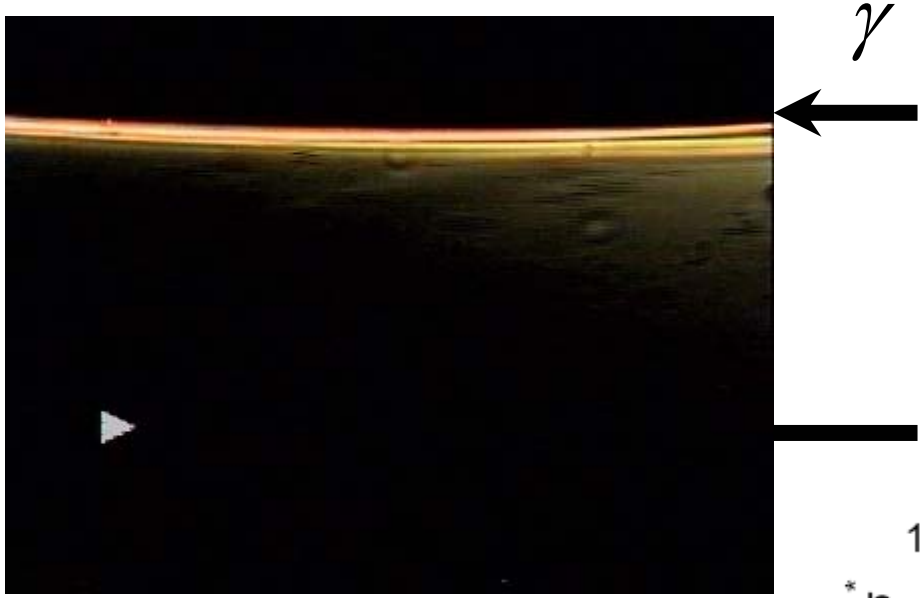
The solution of space inhomogeneous equation is of the moving interface type

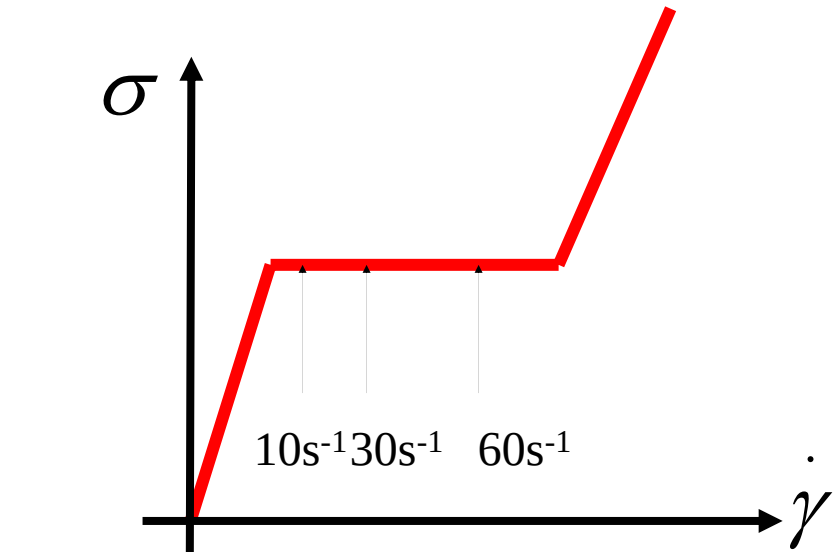
This extends results of Carr-Pego(90) and Fife (89)

The velocity of a Type II interface is proportional to the square root of the diffusion coefficient

O.Radulescu and S.Vakulenko, arXiv qBio 2006

Application1: stress diffusion coefficient from interface kinetics

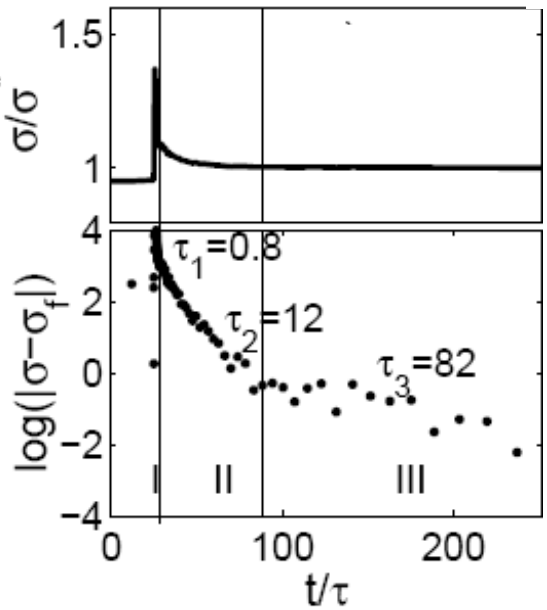




Diffusion is small

D~0.003-0.011 $\mu m^2 s^{-1}$

w~30-40 nm



O.Radulescu et al. Europhys.Lett. 2003

Application2: Diffusion dependent patterning of Drosophila

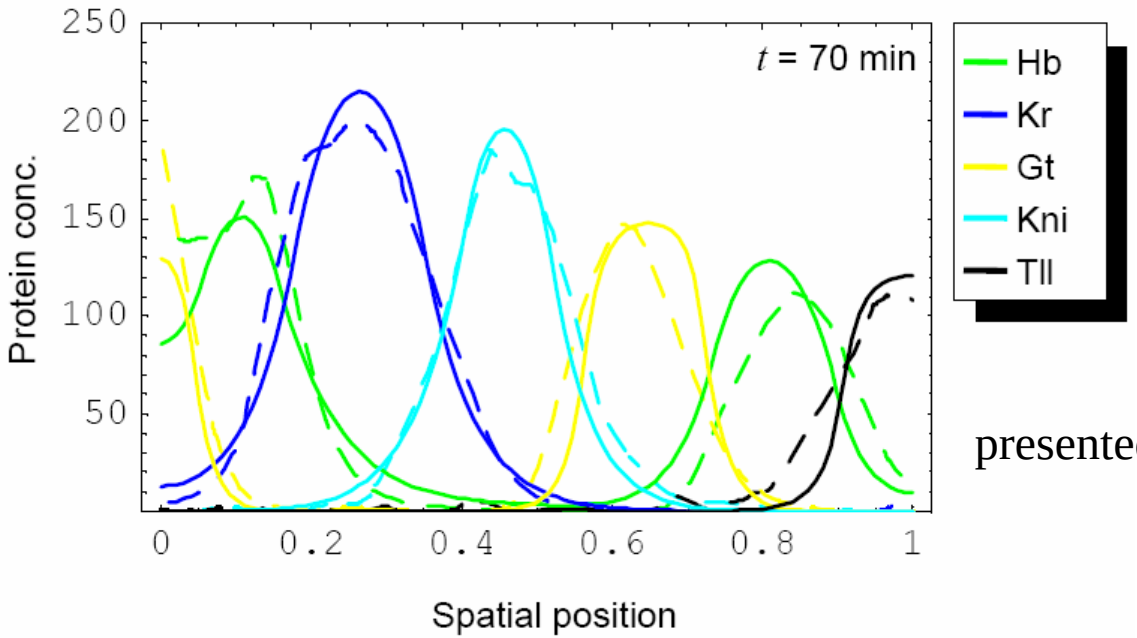
1) parameter fit of Reinitz model from time dependent data by simulated annealing

2) compute attractors of shorted equation

Result: patterning is diffusion dependent with Gursky, Manu, Vakulenko, unpublished

Improvement of model fit: Rapid method of parameter identification using interface kinetics

Resulted solution (solid curves) in comparison to data (dashed curves):



presented at Nanobio'06, St.Petersburg

Comment on the impact in biology

Compared to Turing models, the gene circuit model is realistic:

- the pattern is not a periodic modulation of a homogeneous state
- the pattern results from the interaction of development genes, is guided by maternal gradients and has aperiodic transients

Treating the set of segmentation genes as a dynamical system allows to understand:

- The logic of interactions (open problem) and transformations
- The stability of the result (open problem)
- The possible errors in mutants (open problem)

Conclusion and future projects

Start of a long term project: *produce powerful mathematical tools for analysis of complex systems*.

Strategy:

Model simplification

***The invariant manifold technique of Carr-Pego**
***piecewise deterministic approach for Markov processes**
***graph theory methods for chemical kinetics models**

An intrinsic relation exists between model reduction, stochasticity and robustness: concentration phenomena!

Physical chemistry for diffusion and transport in physiology.

Collaboration

Upi Bhalla NCBS Bangalore, planned co-tutored phD.
A.Gorban (Leicester) Egide/Alliance sponsorship
J.Reinitz (Stony Brook) and Samsonova (St.Petersburg)
ASC project with INRA on modeling lipid metabolism
project ANR SITCON with Curie
Symbiose team IRISA

Project 3: Robustness of biological systems

$$M = f(K_1, K_2, \ldots, K_n)$$

$$Var(logM) << Var(logK_i)$$

Distributed robustness

Cube concentration

$$M = (K_1 + K_2 + \ldots + K_n)/n$$

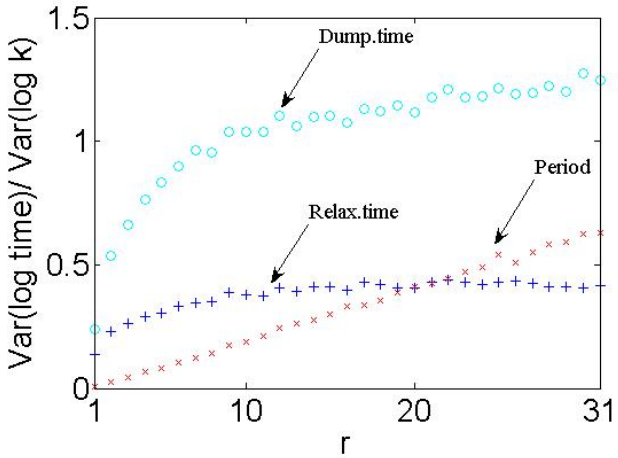
$$K_i = \begin{cases} K_i^0 s_i & i \in I_r \\ K_i^0 & i \notin I_r \end{cases}$$

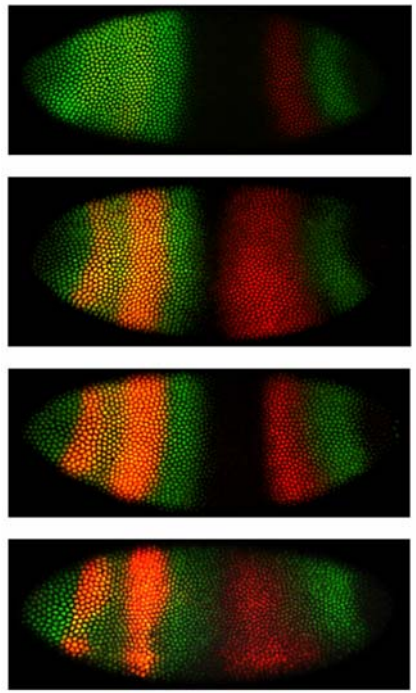
r-robustness

$$Var(logM) << Var(logs_i), 1 \le i \le r$$

Simplex concentration

$$M = K_{(r)}, \quad K_{(1)} \le K_{(2)} \le \ldots \le K_{(n)}$$





Stony Brook (Reinitz), Leicester (Gorban), St.Petersburg(Samsonova, Vakulenko)

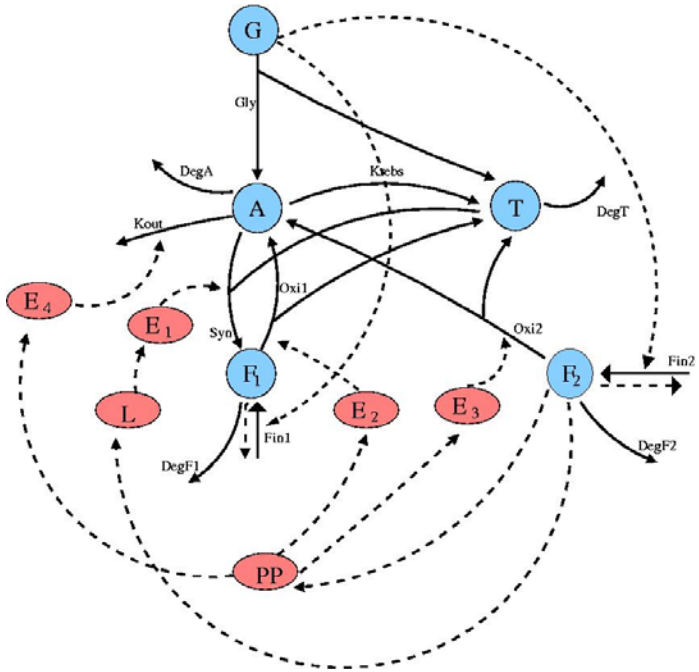
Project 4: LIPID METABOLISM

Hierachical modeling:

1)Extended model

2) Abstract model

Multiorgans, multispecies



Heterogeneous data

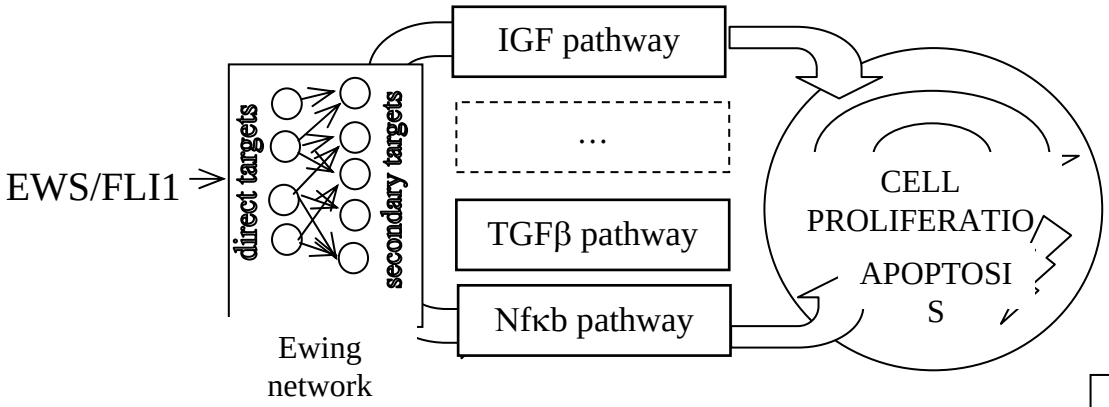
Microarrays

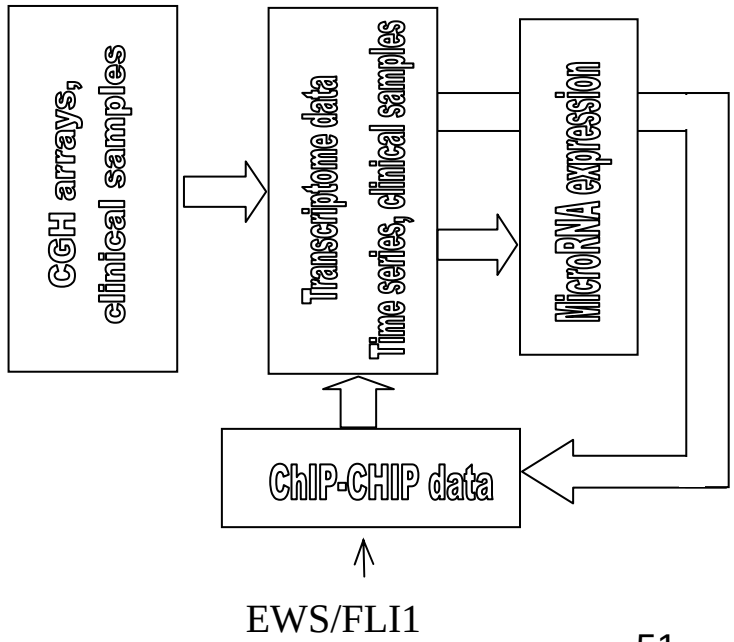
Biochemical dosages

Symbiose, INRA Rennes and Toulouse

Project 5: SITCON

Modeling signal transduction induced by a chimeric oncogene

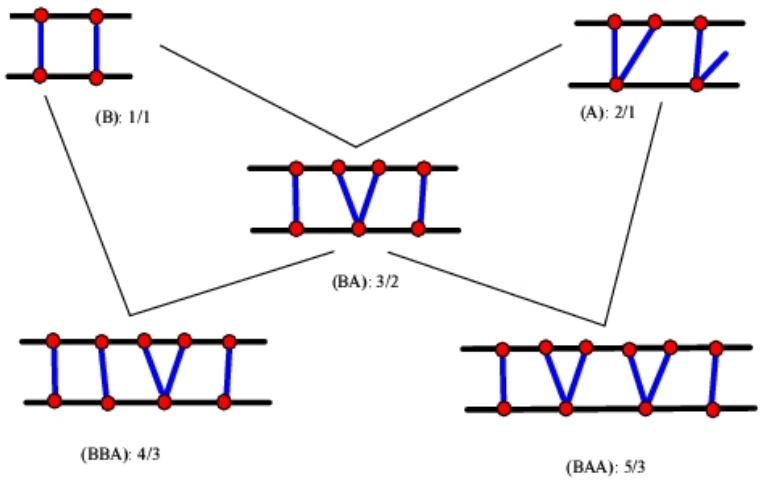


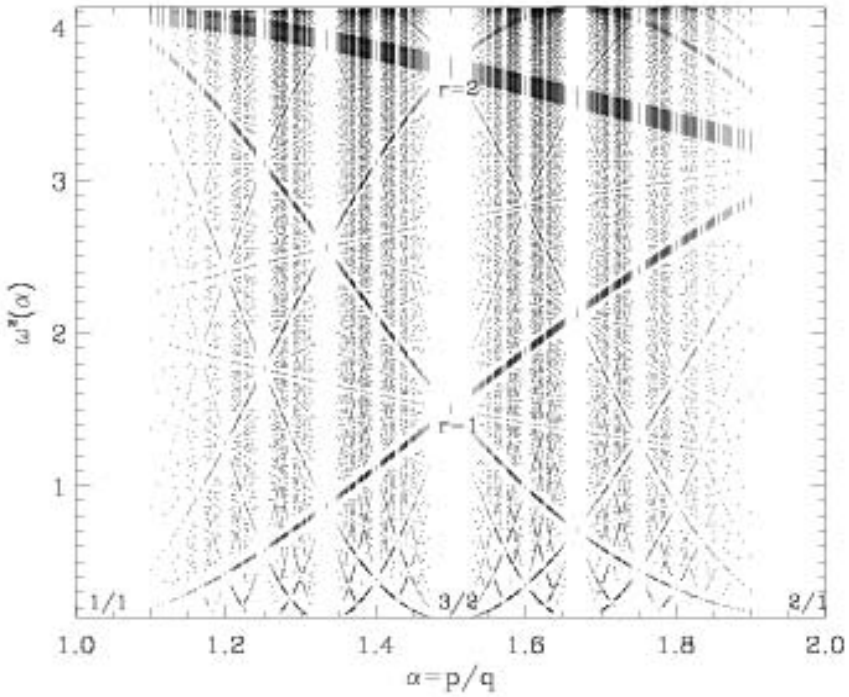


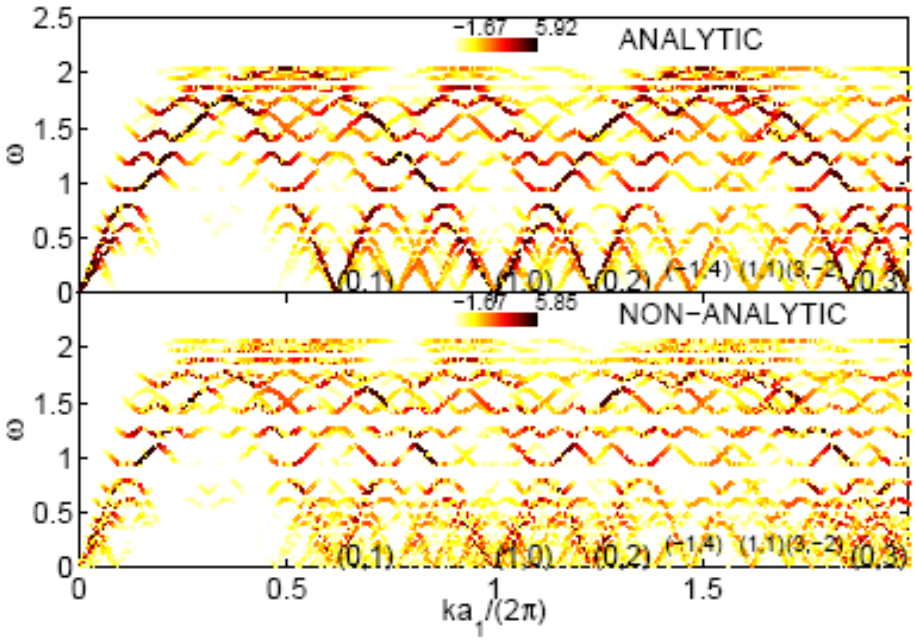
Institut Curie, Symbiose

Farey sets and spectra of incommensurate structures

Number theory and Incommensurate compounds







Related problems:
Hofstadter Butterfly
Bellissard's gap labelling
Arnold tongues