Data-based modeling of the dynamics of a cellular signaling pathway

Jens Timmer

In cooperation with:
Klingmüller Lab, DKFZ

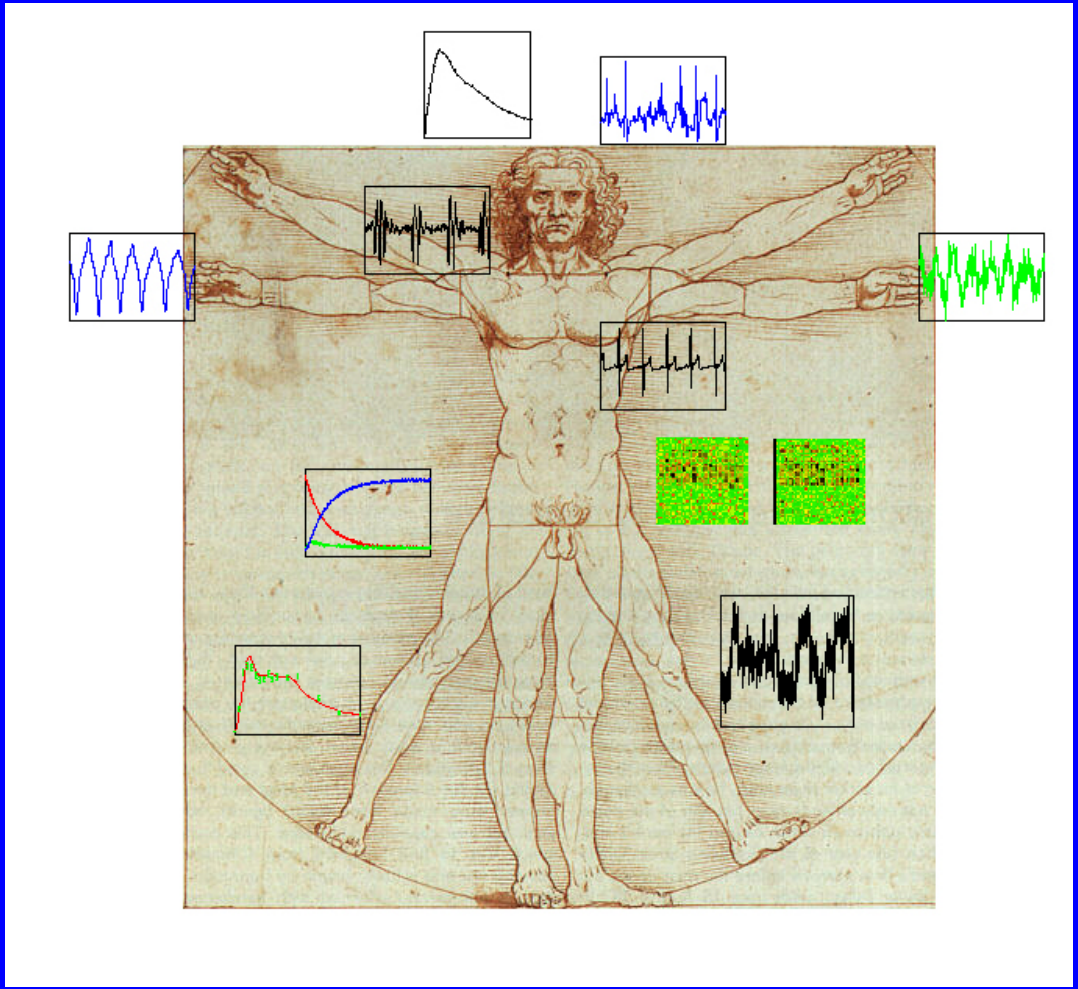
Center for Data Analysis and Modeling
Center for Systems Biology
Department for Mathematics and Physics
University of Freiburg

http://www.fdm.uni-freiburg.de/∼jeti/

Outline

- **Introduction**
- **JAK-STAT pathway of the Epo receptor**
- **Simulation and data based modeling**
- **A dynamical model for JAK-STAT pathway**
- **Observing the unobservable**
- *In silico* **biology: Predicting a new experiment**
- **Outlook**

Man : A Dynamical System



Diseases caused or expressed by malfunction of dynamical processes

General Goal

Understand biomedical systems by data-based analysis of their dynamical behavoir.

Time Series Analysis

Two Gaps In Time Series Analysis

Specific questions

Time series

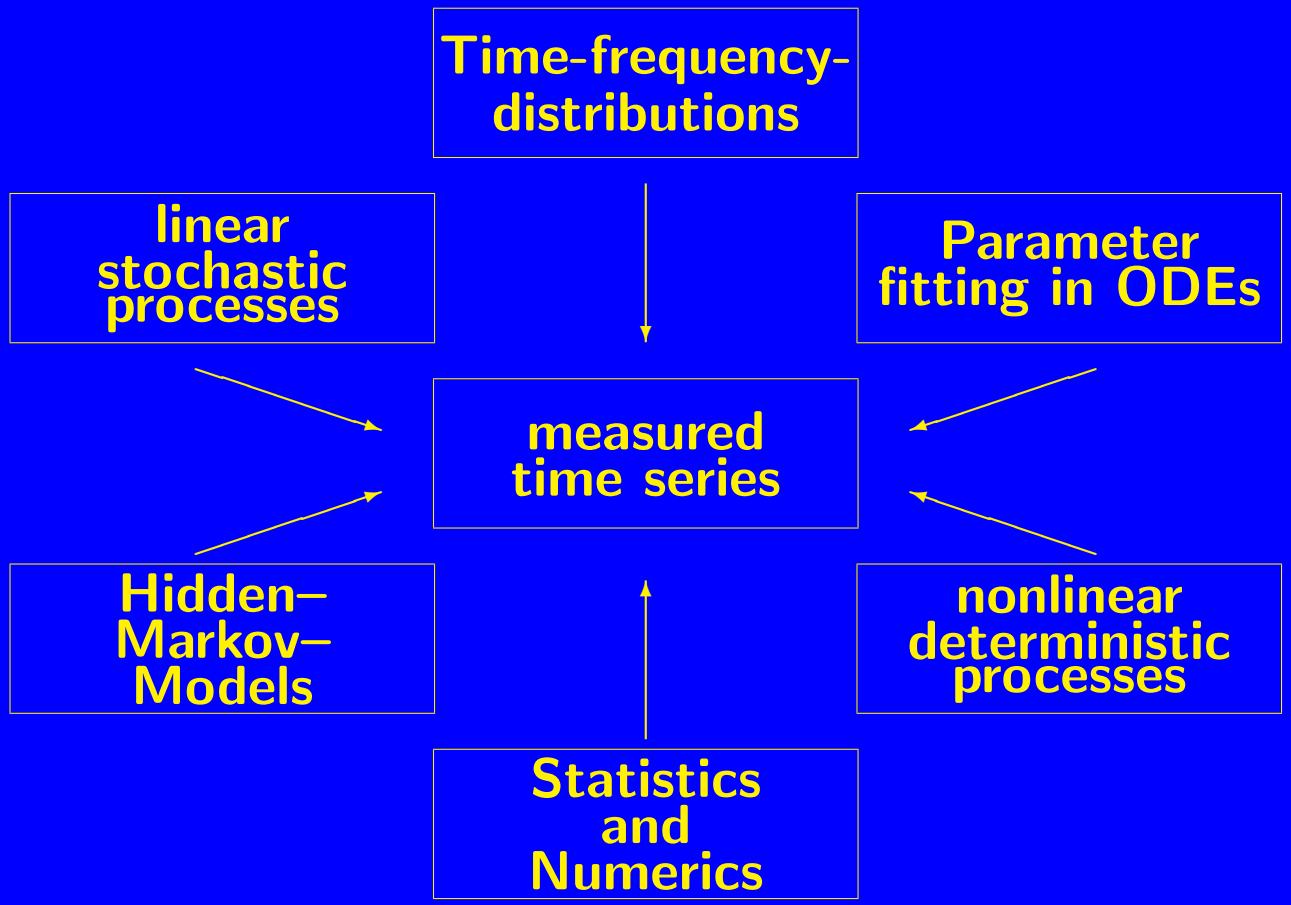
Time series analysis

Specific applications

Two Gaps In Time Series Analysis

Specific questions	Time series analysis	Time series
Time series analysis	Specific questions	Specific applications

Methods


Time-frequency-distributions
linear stochastic processes
Parameter fitting in ODEs
measured time series
Hidden–Markov–Models
nonlinear deterministic processes
Statistics and Numerics

Goals of Time Series Analysis

- Prediction
- Characterization / Classification
  - Improvement of diagnosis and therapy
- Modeling
  - Hypotheses testing
  - Understanding
  - Control

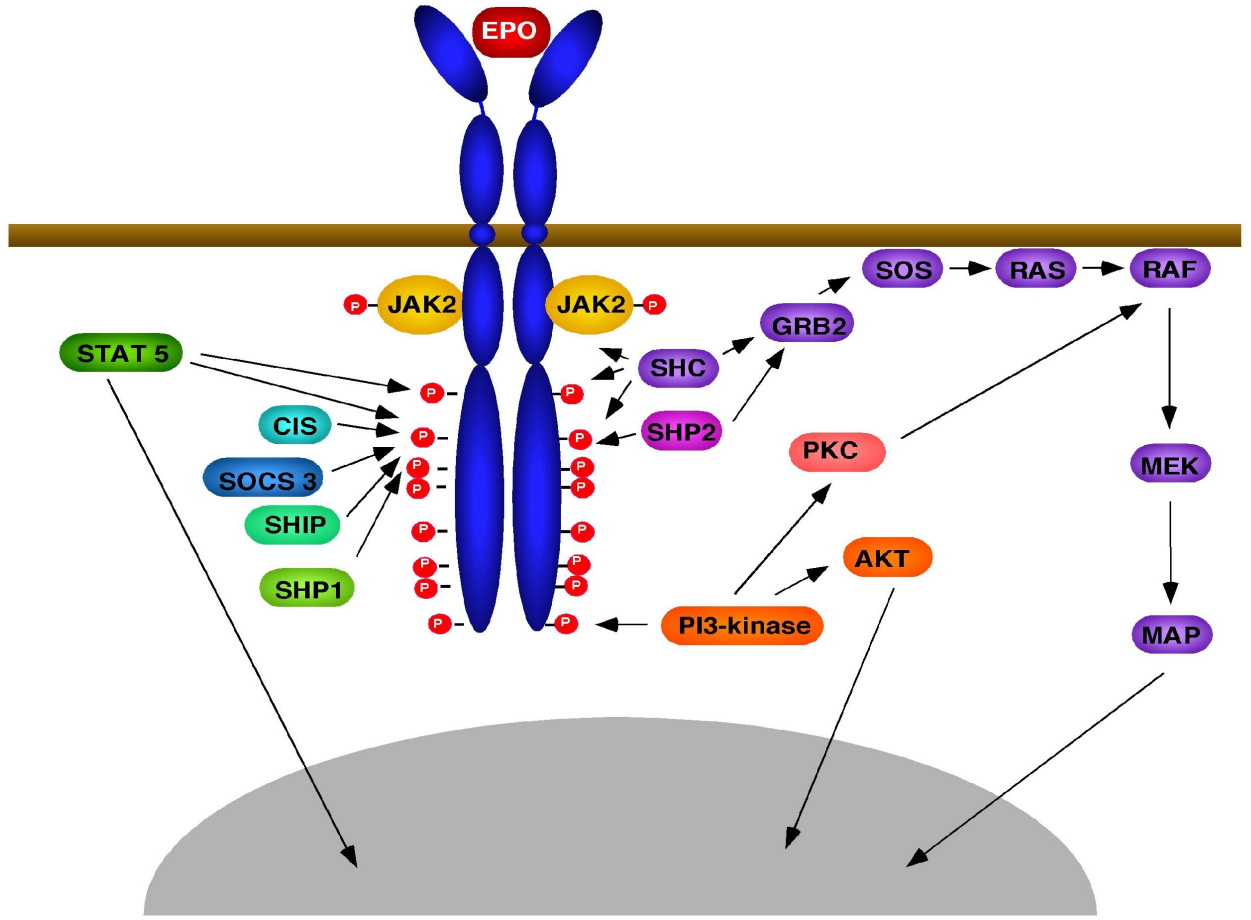
Why Modelling in BioMed?

- Make assumptions explicit
- Understand essential properties, failing models
- Condense information, handle complexity
- Understand role of dynamical processes, e.g. feed-back
- Prediction and control
- Discover general principles
- "You don't understand it until you can model it"

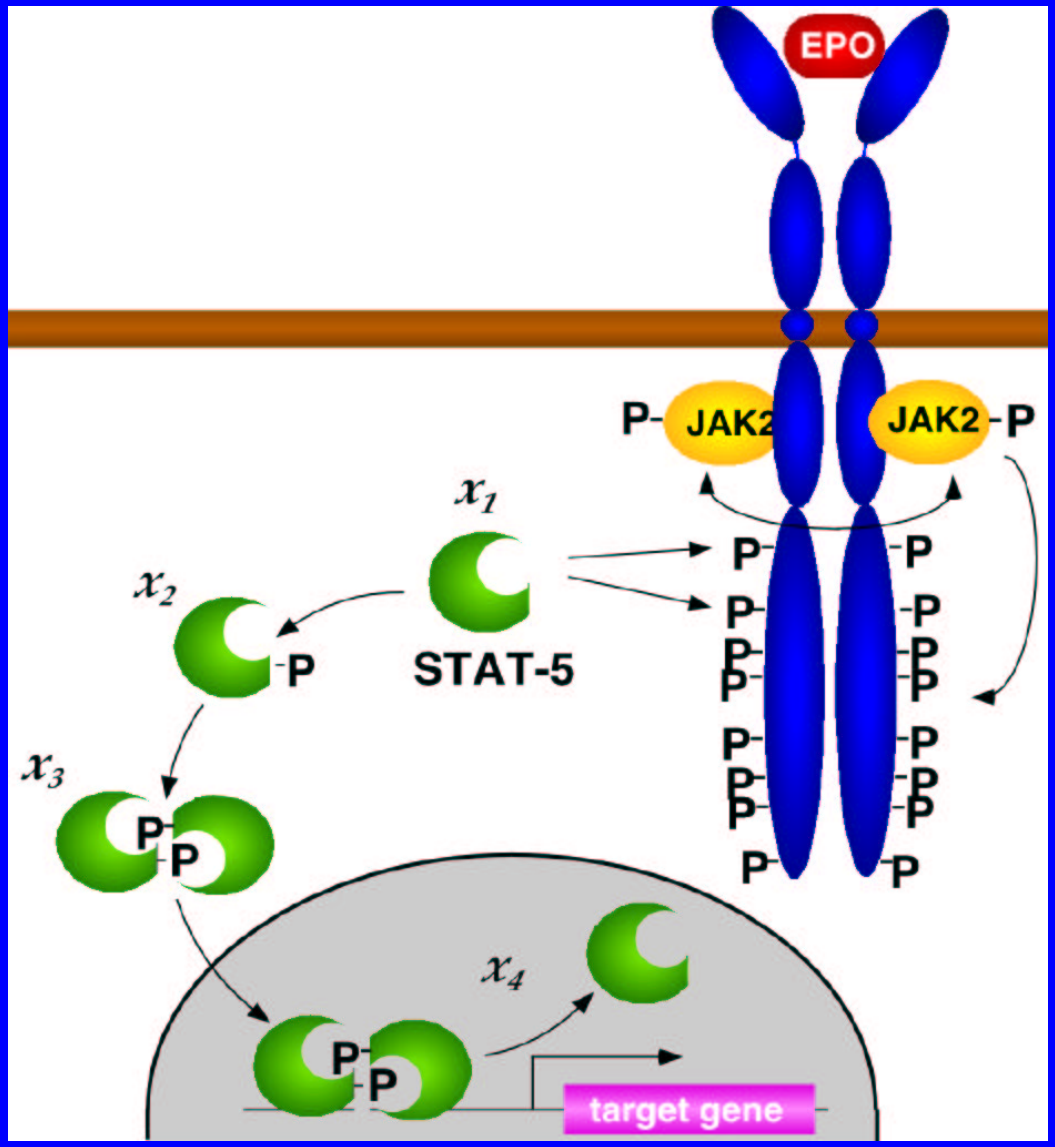
Why Modelling in Cell Biology?

- ...omics does not necessarily lead to understanding of function
- Function determined by regulation
- Regulation = Dynamics
- Function: Property of dynamic network
- "Systems Biology"

Signal transduction through the Erythropoietin receptor (EpoR)



JAK – STAT Pathway



Mass Action Yields :

$$
\begin{aligned}
\dot{x}_1 &= -p_1 x_1 \mathrm{EpoR_A} \\
\dot{x}_2 &= p_1 x_1 \mathrm{EpoR_A} - p_2 x_2^2 \\
\dot{x}_3 &= \frac{1}{2} p_2 x_2^2 - p_3 x_3 \\
\dot{x}_4 &= p_3 x_3
\end{aligned}
$$

Measurements

- $\mathrm{y_1(t)}$: **Phosphorylated STAT-5 in the cytoplasm**

$$\mathrm{y_1(t) = p_5(x_2(t) + 2\,x_3(t))}$$

- $\mathrm{y_2(t)}$: **All STAT-5 in the cytoplasm**

$$\mathrm{y_2(t) = p_6(x_1(t) + x_2(t) + 2\,x_3(t))}$$

- $\mathrm{y_3(t)}$: **Activation of the epo receptor**

$$\mathrm{y_3(t) = p_7\,EpoR_A(t)}$$

Simulation vs. Data Based Modeling I

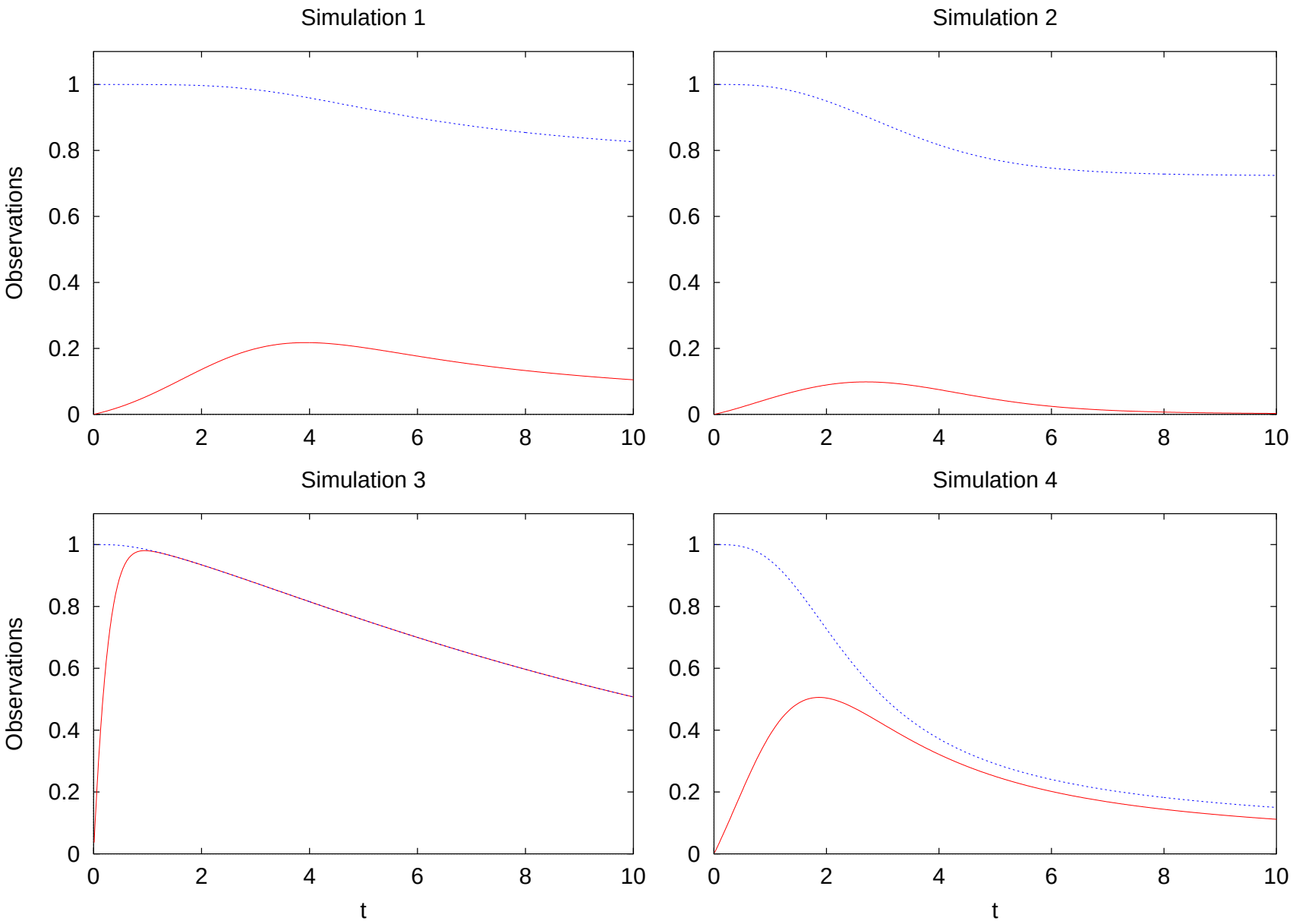
Model comprises:

- **Structure of the equations (the cartoon)**
- **Values of the parameters**

Simulation:

- **Structure from pathway cartoon**
- **Parameters from**
  - **Independent measurements**
  - **Literature**
  - **Educated guesses**

Simulations



Simulation vs. Data Based Modeling II

Simulation dilemma:

If discrepancies between experiment and model

- **Wrong structure or wrong parameters ?**

Data based modeling:

- **Structure from pathway cartoon**
- **Parameters estimated from data**

If discrepancies:

Think about the cartoon ! Learn biology !

Parameter Estimation

Dynamics:

$$\dot{\vec{x}} = \vec{f}(\vec{x}, \vec{p})$$

Observation:

$$\vec{y}(t_i) = \vec{g}(\vec{x}(t_i), \vec{p}) + \vec{\epsilon}(t_i) \quad \vec{\epsilon}(t_i) \sim N(0, \Sigma_i)$$

Log-Likelihood:

$$E = \chi^2(\vec{p}, \vec{x}(t_0)) = \sum_{i=1}^{N} \sum_{j=1}^{M} \left(\frac{(y_j^D(t_i) - g_j(\vec{x}(t_i; \vec{p}, \vec{x}(t_0))}{\sigma_{i\,j}} \right)^2$$

Initial Value Approach, Multiple Shooting, GO

Statistics I

- **Confidence regions for parameters**
  - **Asymptotically :**

$$\frac{\partial^2}{\partial p_i \partial p_j} \chi^2(\hat{\vec{p}}, \hat{\vec{x}}(t_0))$$

  - **Finite:**
    - **Log-Likelihood contours**
    - **Bootstrap**

Statistics II

- Model selection
  - Likelihood ratio test
  - Non-standard test situations :
    - Parameter on the boundary
    - Non-identifiability under the null
  - Non-nested models, Bootstrap

Really Good Data

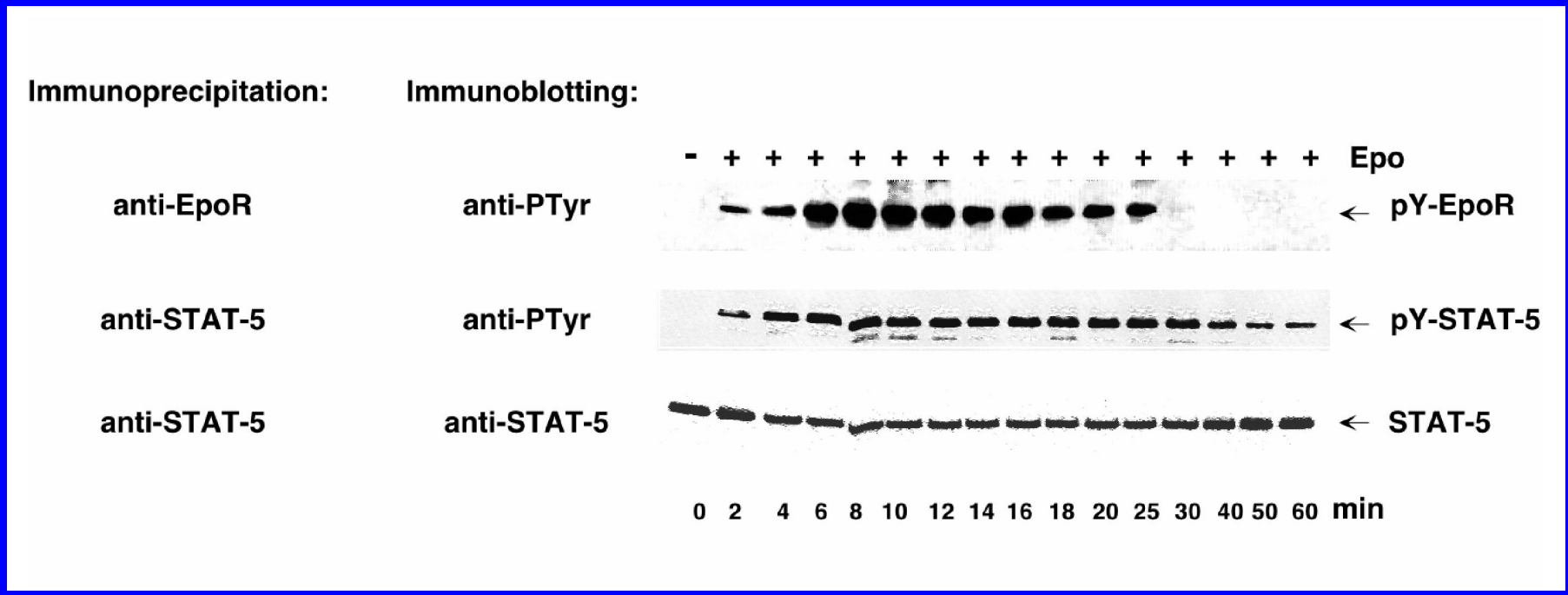
"What makes you feel good ?"

"Good data."

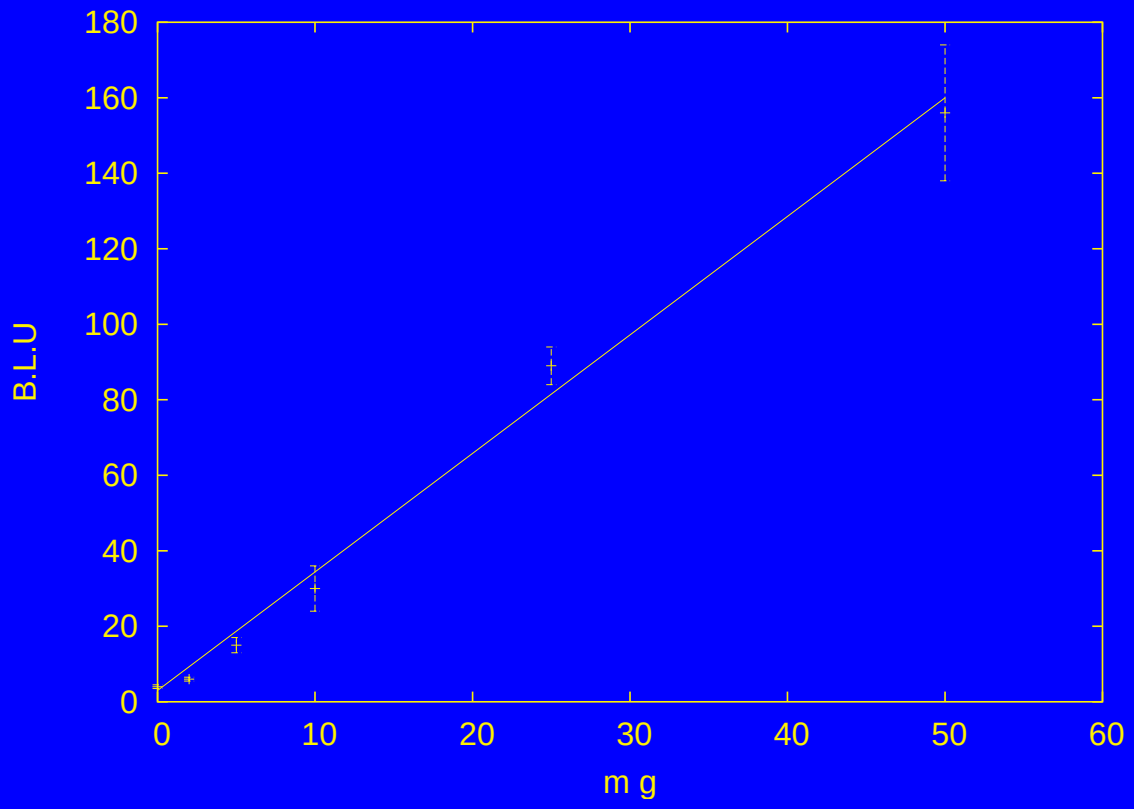
"What makes you feel really good ?"

"Really good data !"

Quantitative Immunoblotting



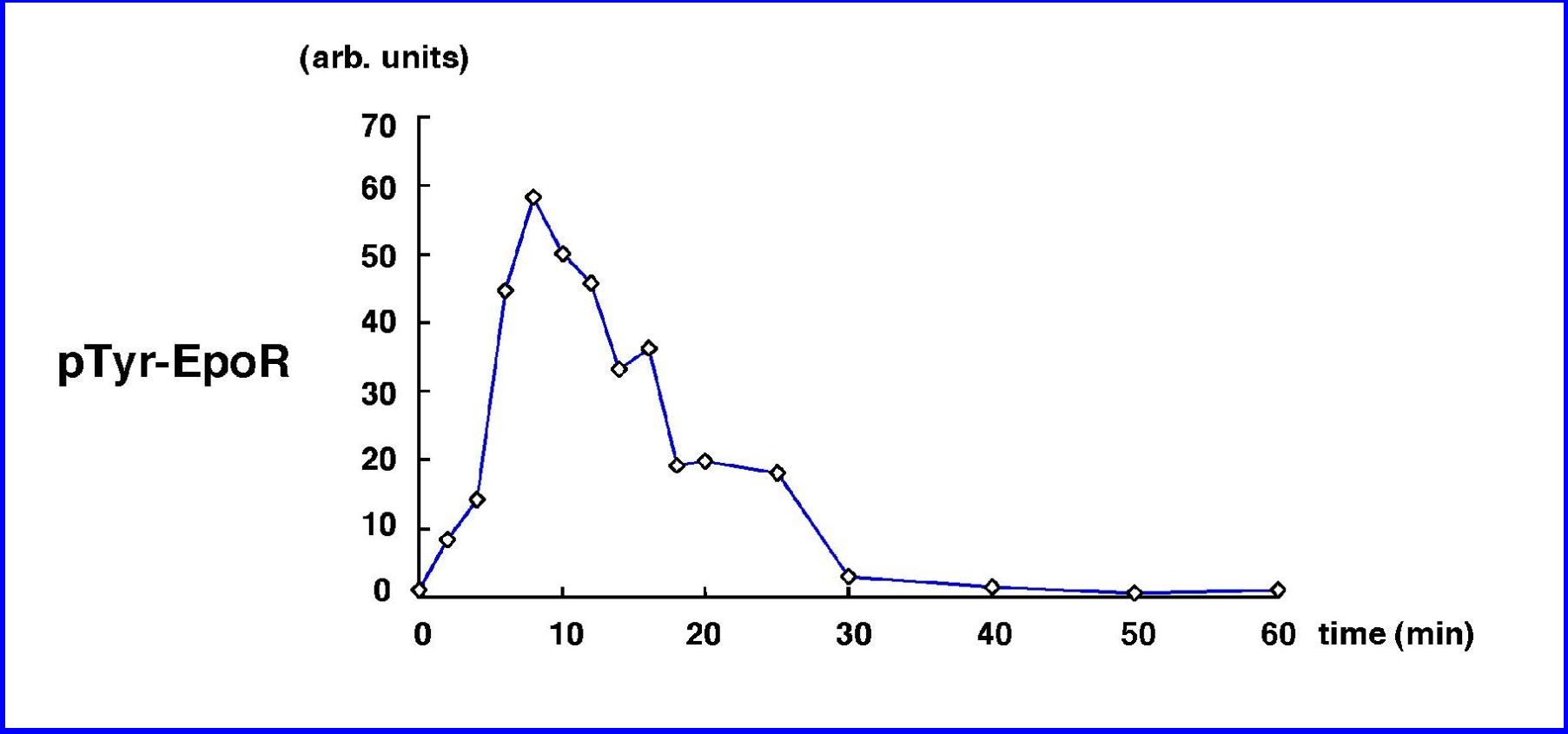
Really Good Data



$g(x)$ is linear

The data

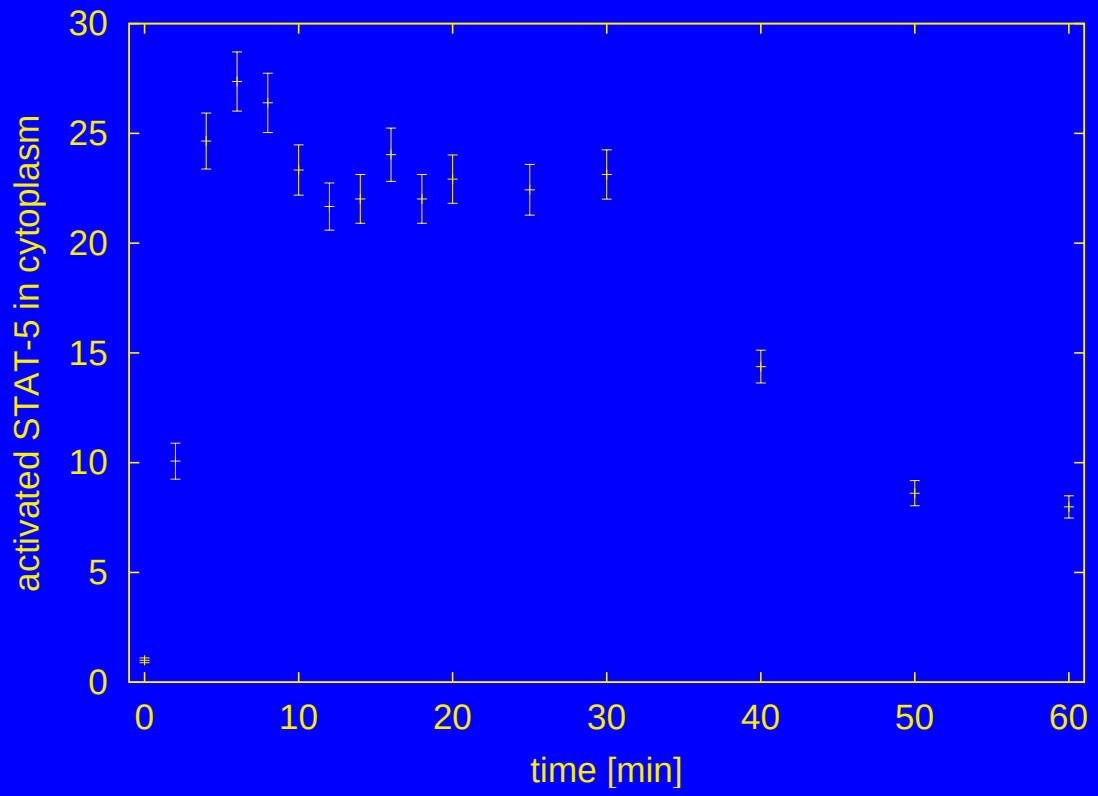
Activation of the epo receptor :



Maximum at 8 min

The data

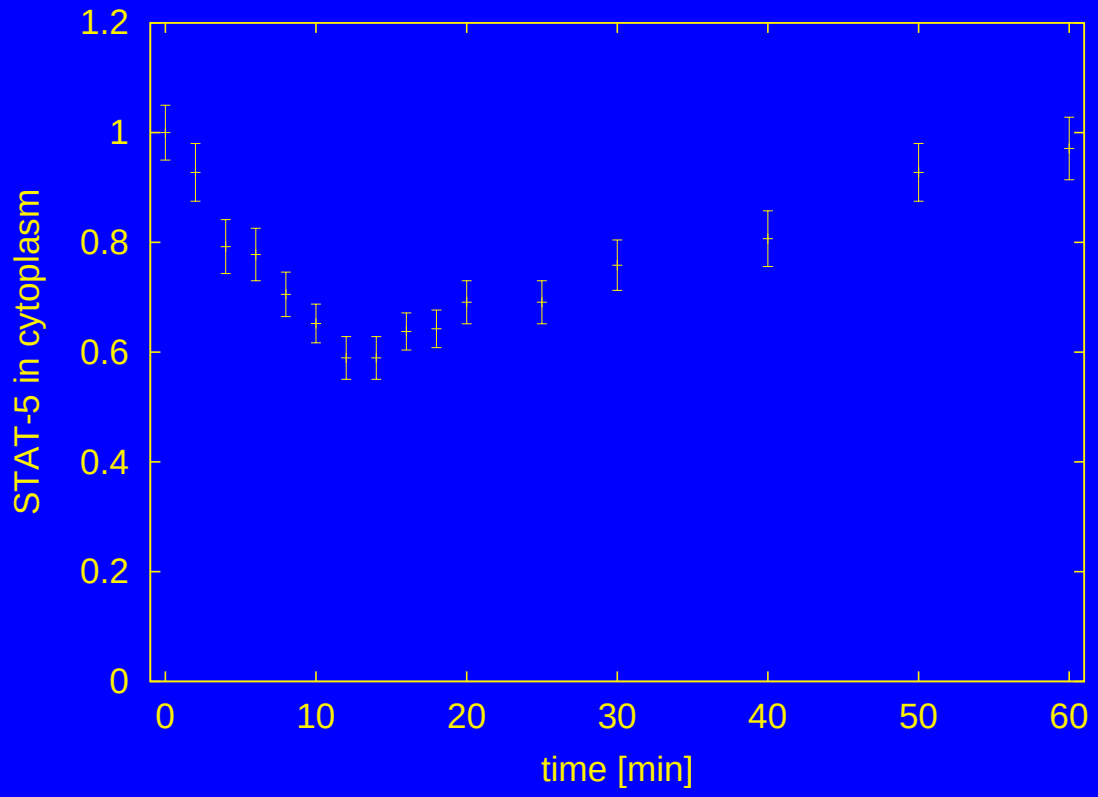
Phosphorylated STAT-5 in cytoplasm :



Plateau from 10 to 30 min

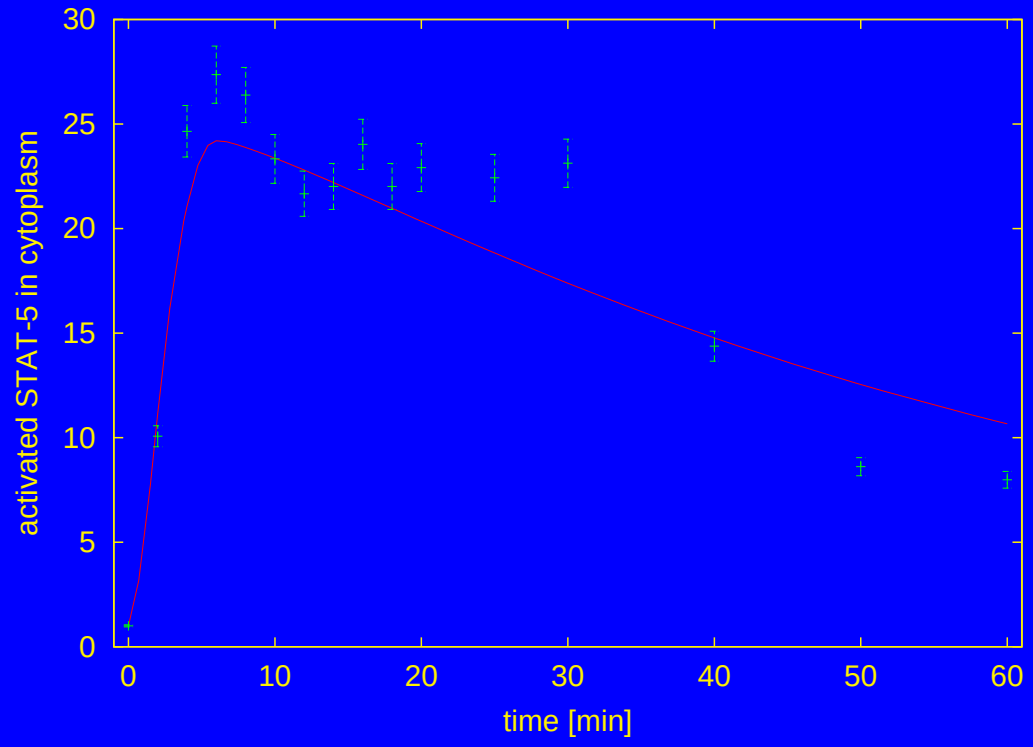
The data

All STAT-5 in cytoplasm :



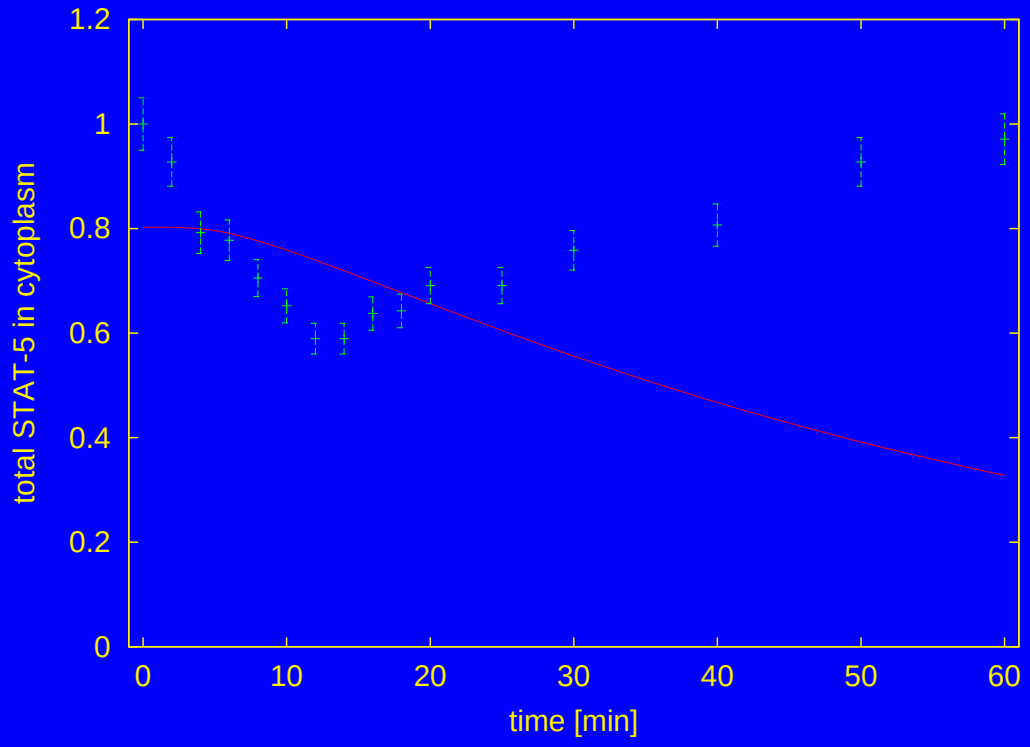
First results

Phosphorylated STAT-5 in cytoplasm :

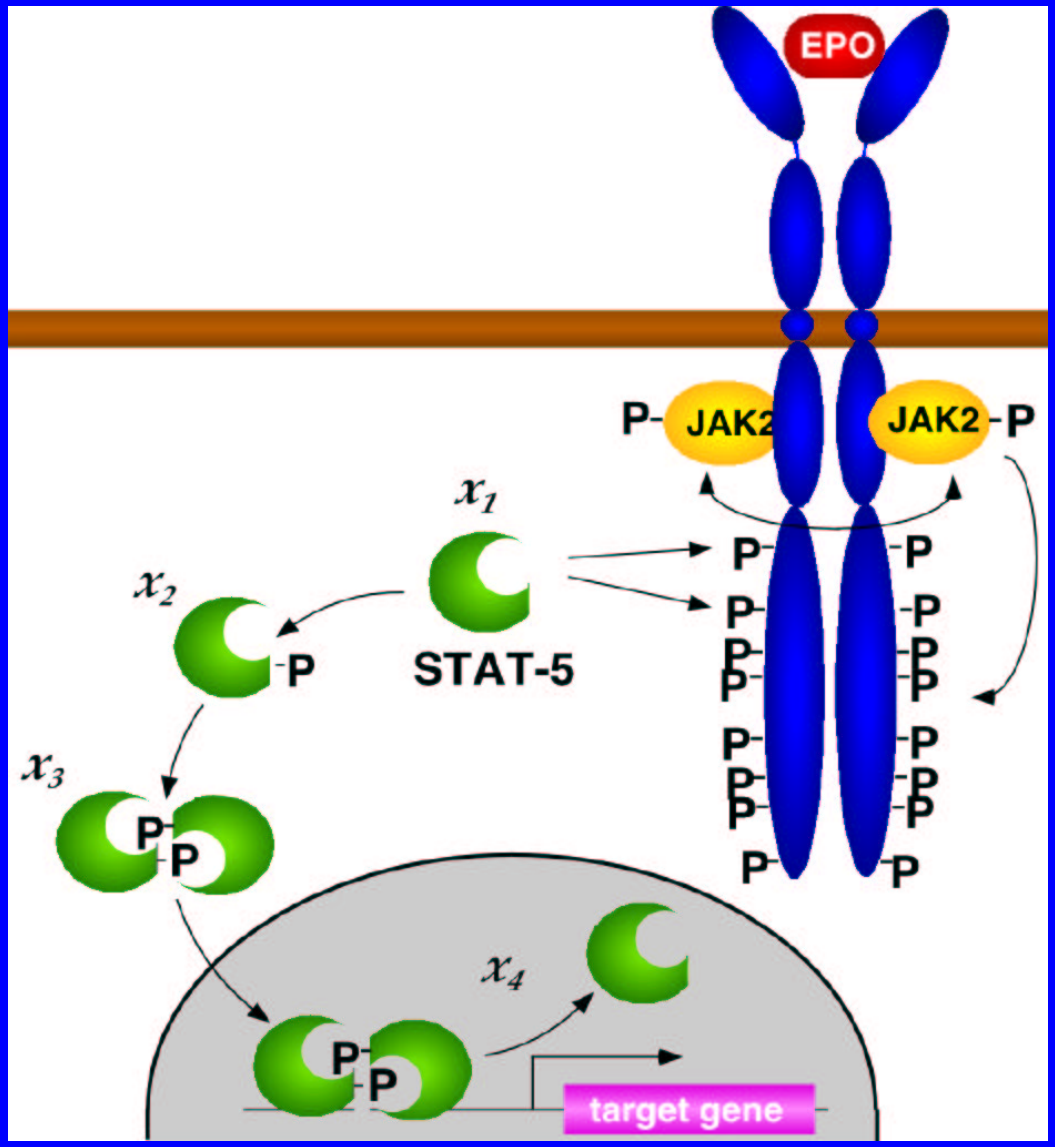


First results

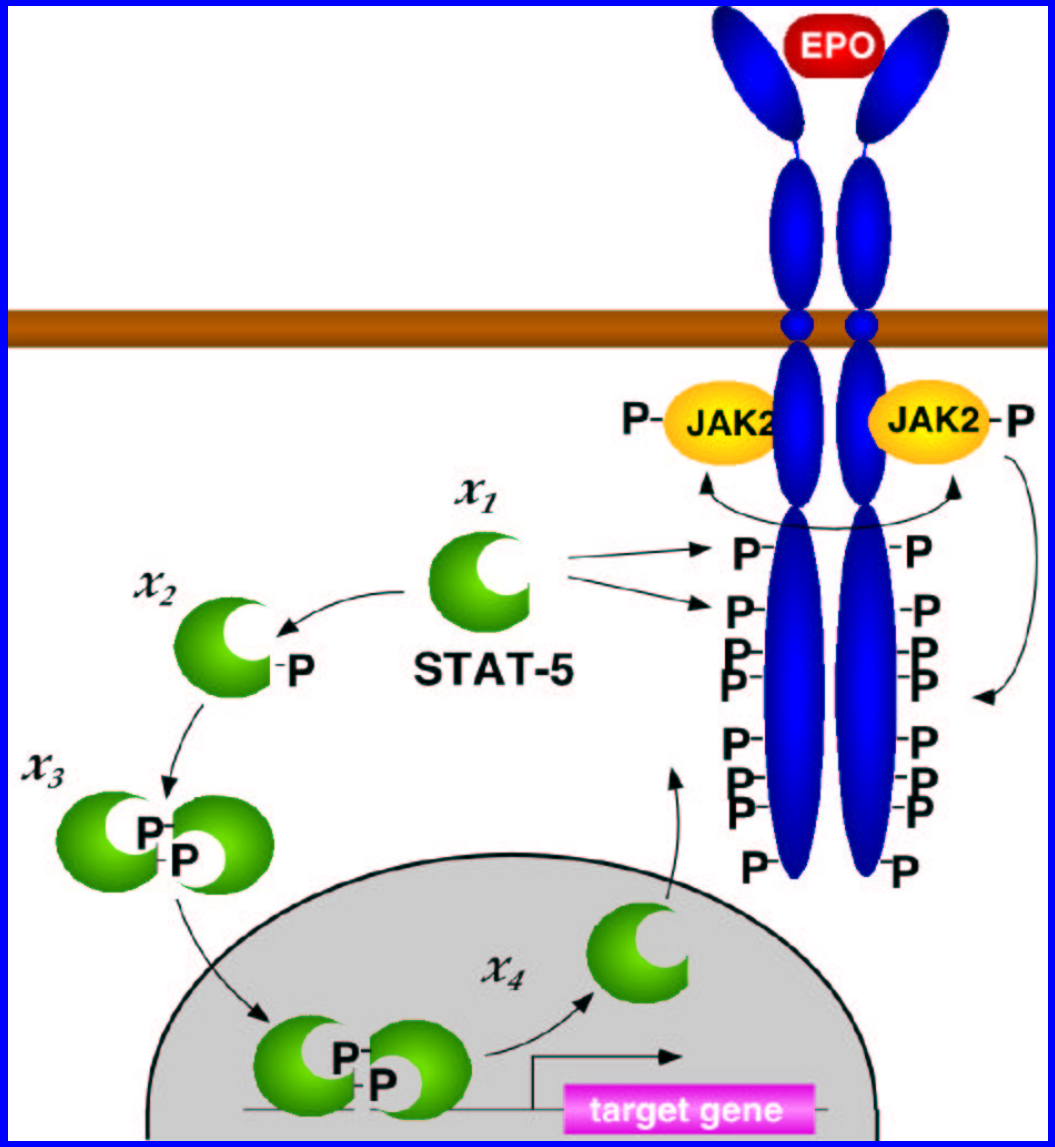
All STAT-5 in cytoplasm :



JAK – STAT Pathway



Model Extension



Modeling Nuclear Export

- One compartment:

$$\dot{x}_4 = p_3 x_3 - p_4 x_4$$

- Series of compartments $\approx$ delay

$$\dot{x}_4 = p_3 x_3 - p_4 x_3^{\tau}, \quad x_3^{\tau} = x_3(t - \tau)$$

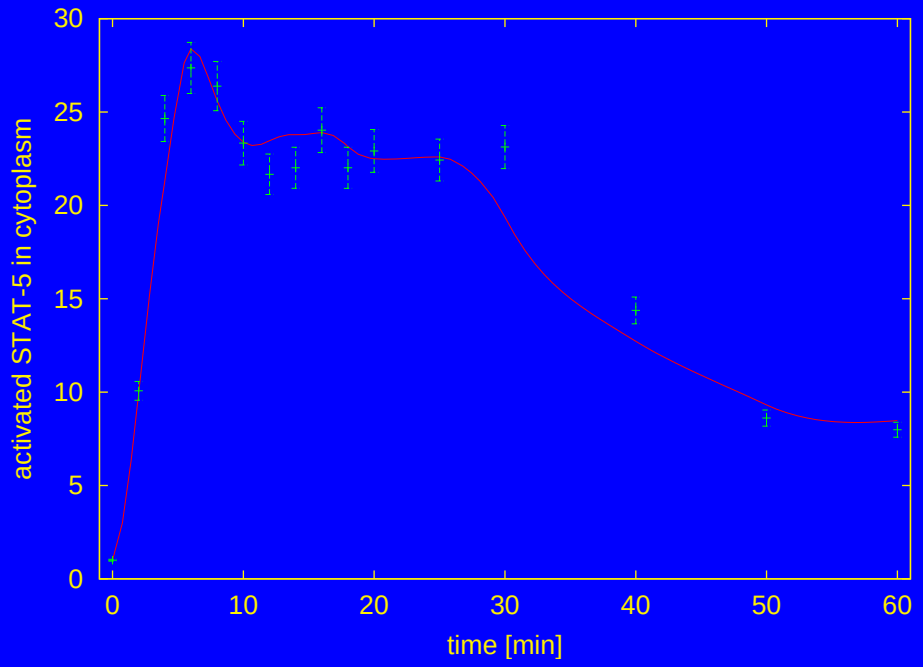
- Non-nested models

Second try

$$
\begin{aligned}
\dot{x}_1 &= 2p_4 x_3^{\tau} - p_1 x_1 \mathrm{EpoR_A} \\
\dot{x}_2 &= p_1 x_1 \mathrm{EpoR_A} - p_2 x_2^2 \\
\dot{x}_3 &= \frac{1}{2} p_2 x_2^2 - p_3 x_3 \\
\dot{x}_4 &= p_3 x_3 - p_4 x_3^{\tau}
\end{aligned}
$$

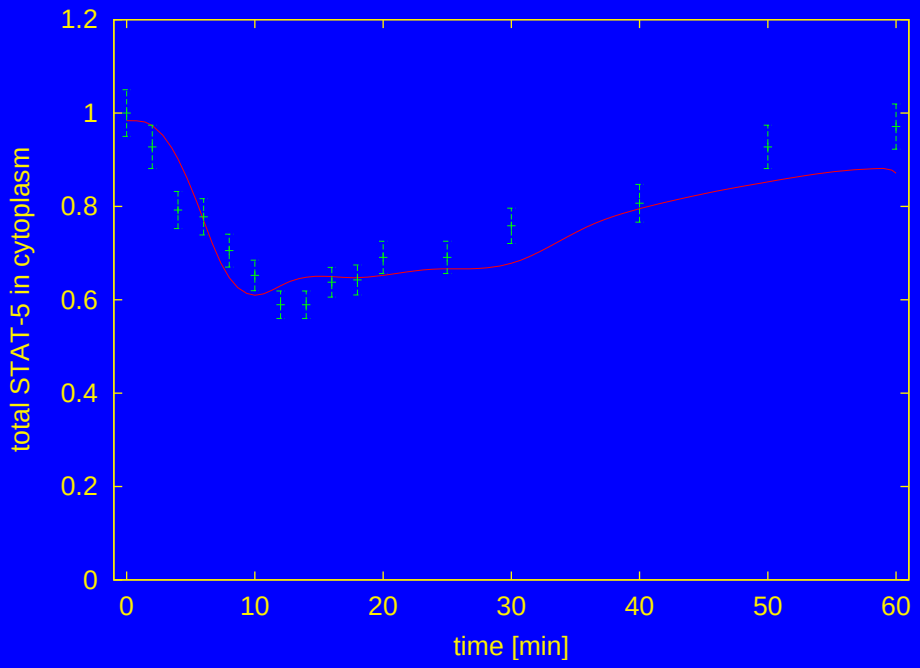
Results

Phosphorylated STAT-5 in cytoplasm :



$\tau \approx$ **6 min**

All STAT-5 in cytoplasm :



Observing the unobservable

Simulating the fitted model :

Access to dynamic variables x_i

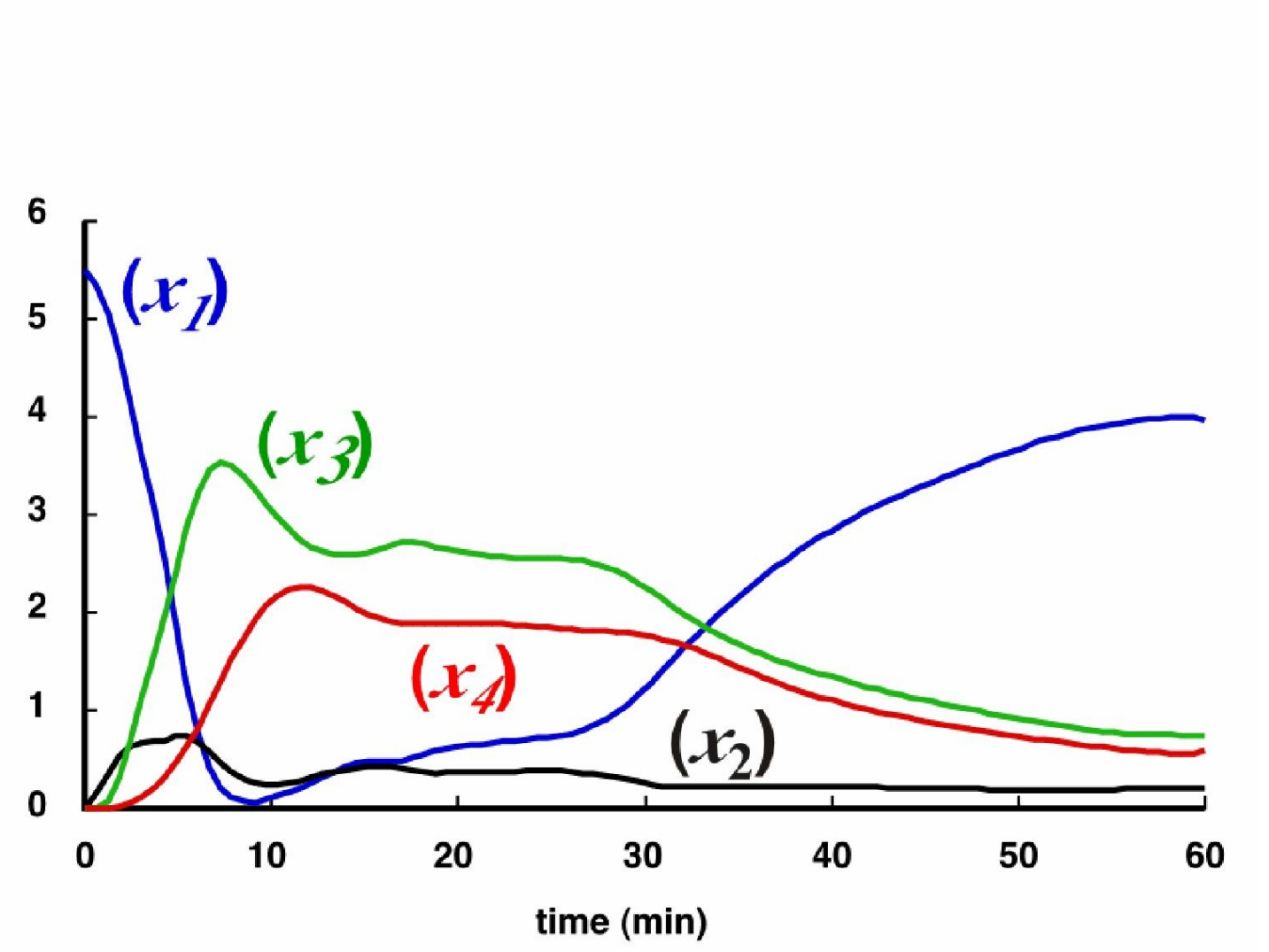
- **Unphophorylated STAT-5 is limiting factor**

- **Experimental fact:**

  Phosphorylated monomeric STAT-5 is hard to measure

Explanation by the model:

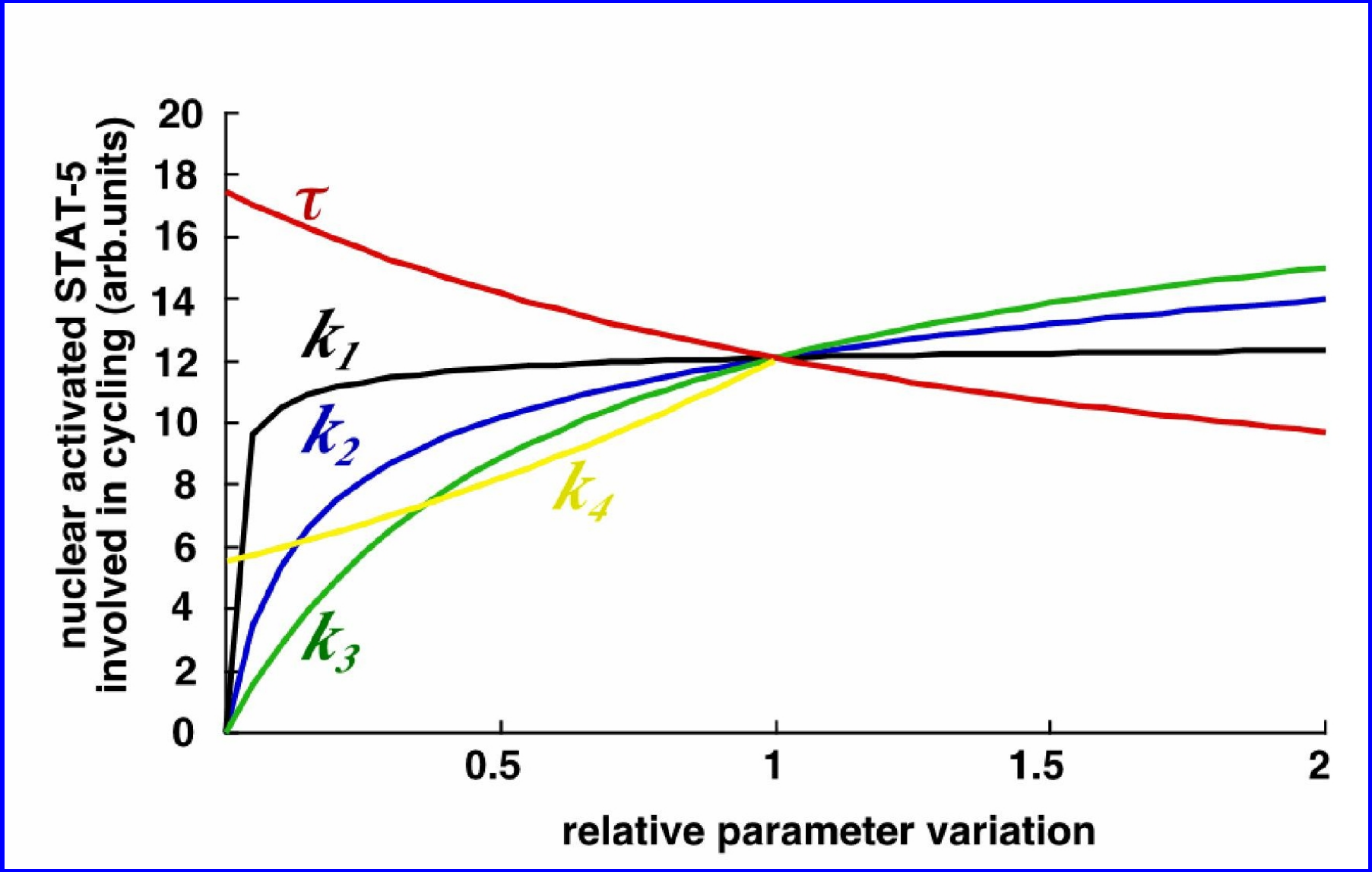
It is rapidly processed into dimeric STAT-5


6
5
4
3
2
1
0
(x_1)
(x_3)
(x_4)
(x_2)
0
10
20
30
40
50
60
time (min)

In silico biology

What happens if ... ?

- **Sensitivity analysis**
- **Change parameters in the model**
- **Calculate the transscriptional yield**


nuclear activated STAT-5
involved in cycling (arb.units)
20
18
16
14
12
10
8
6
4
2
0
τ
k1
k2
k4
k3
0.5
1
1.5
2
relative parameter variation

Prediction of New Experiment

- **Result of sensitivity analysis:**

  Transscriptional yield is most sensitive to nuclear shuttling parameters.

- **Setting $\mathrm{p}_4 = 0$ or $\tau = \infty$**

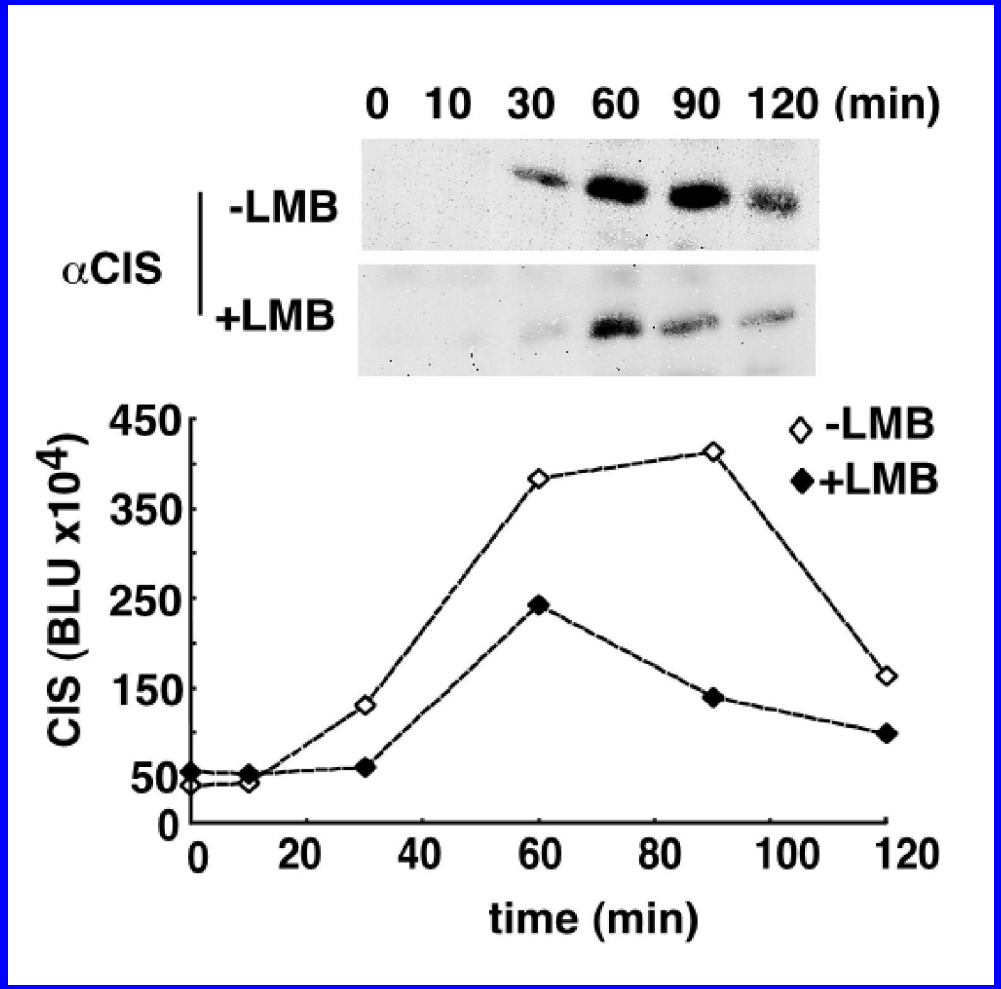
  Only one cycle : Only 45 % efficiency

- **Blocking nuclear export by leptomycin B confirms prediction.**

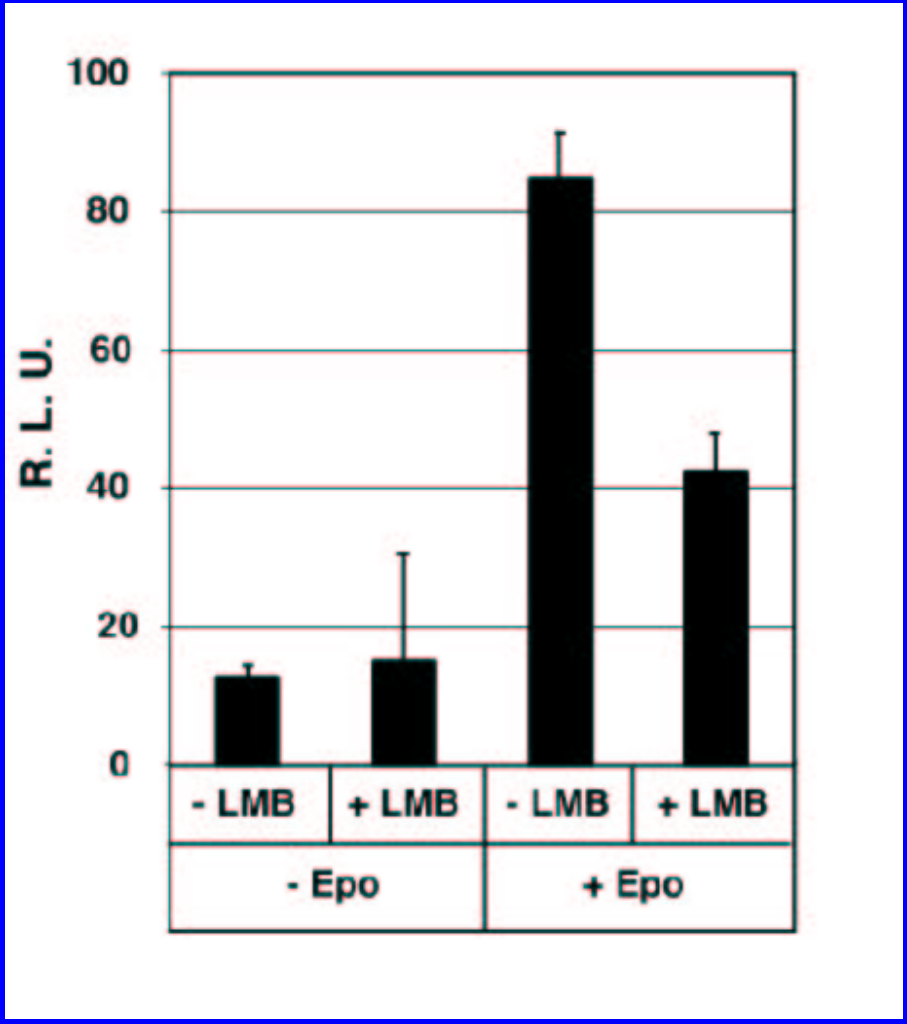
Perspective:

Identification of targets for medical intervention

Experimental Confirmation of Prediction



Experimental Confirmation of Prediction



Why Cycling ?

- Optimal use of limited pool of STAT-5
- Continuous monitoring of receptor activity :

Systems' property: "Remote Sensor"

Proc. Natl. Acad. Sci. 100, 2003, 1028-1033

"All models are wrong ..."

- No scaffolding for receptor–STAT-5 interaction, 200 eqs.
- Spatial effects, ODE vs. PDE
- Stochastic effects

"... but some are useful"

- Captures the main effect
- Makes testable prediction

Successful modelling: Making controlled "errors"

Summary

Given time-resolved "really good" data:

It is possible to turn qualitative cartoons into quantitative dynamical models

- **Testing the cartoon**
- **Calculating unobservable components**
- **Manipulating the system** $in\ silico$
- **Prediction of experiments**
- **Infering systems' properties**

Outlook: Scale It Up

BMBF Systems Biology of (Regenerating) Hepatocytes

- **SMAD, IGF, Wnt/β-catenin, NF-κB, ... pathways**
- **Crosstalk**
- **Interaction Transcription factors – DNA**
- **Genetic Networks**

The Mission of Systems Biology

Turn the life sciences

from a

static, qualitative, descriptive

into a

dynamic, quantitative, predictive

science